Service Encounters as a Sequence of Events

The Importance of Peak Experiences

Peter C. Verhoef
Erasmus University

Gerrit Antonides
Wageningen University

Arnoud N. de Hoog
MeesPierson Private Wealth Management (Curaçao) N.V.

A service encounter can be considered as a sequence of events. In the early service literature, it was assumed that firms should deliver a consistent performance during a service encounter. However, research in psychology states that this is not necessarily true. In addition to the average performance, the peaks in the performance are important. Likewise, some service researchers have stressed the importance of a happy ending. The authors test a model on how events contribute to the overall evaluation of a sequence of events. They show that the average performance during the encounter is important. However, their results also stress the importance of peak experiences for satisfaction formation. Thus, managers of service encounters should not only manage the overall performance of a service encounter. To further elevate satisfaction, they could also provide some positive peak experiences.

Keywords: *sequence evaluation; satisfaction; call center; services*

A service delivery process often concerns a sequence of related events occurring at different points in time. An example is the visit of an amusement park in which the visitor experiences a number of attractions during the day. Generally, it is expected that the value of these experiences adds up to the total utility of the service independent of the time of occurrence of each outcome (Loewenstein and Prelec 1993). For the amusement park, this implies that an attraction experienced at the beginning of the day adds to the evaluation in the same way as an attraction visited at the end of the day. Stated differently, the overall evaluation of the amusement park does not depend on the order in which the attractions were visited. On the basis of this no-

The authors acknowledge the support of Fortis Bank Netherlands and their call center director Jan Doornenbal in their data collection. They also thank Daniel Read, the editor Roland Rust, and three anonymous reviewers for their helpful comments on a prior version of this article. Correspondence regarding this article should be addressed to Peter C. Verhoef, Erasmus University, Rotterdam School of Economics, Department of Marketing and Organization, Office H15-17, P.O. Box 1738, NL-3000 DR Rotterdam, the Netherlands; phone: +31 (10) 408-2809; fax: +31 (10) 408-9169; e-mail: Verhoef@few.eur.nl.

Journal of Service Research, Volume 7, No. 1, August 2004 53-64
DOI: 10.1177/1094670504266137

tion, service marketers are advised to aim for customer satisfaction in every service encounter (e.g., Zeithaml and Bitner 1996). The notion that services concern a sequence of related events holds for many other contexts, such as store visits (i.e., entrance store, choosing products, service delivery by store employees, and paying at cashier), restaurants (i.e., meal conversations with servant, consumption of meals, and payment of meals), and airline flights (i.e., entrance of terminal, check-in, boarding, flight, landing at destination, baggage claiming, and leaving terminal).

Recently, the assumption of time independence of outcomes has been questioned in the economic and psychological literature (e.g., Kahneman 1994; Loewenstein and Prelec 1993). Mechanisms, such as negative time preference, imply that sequences of outcomes in which the most preferred outcome is served at the end are preferred to sequences that start with the most preferred outcome. At the same time, other researchers have questioned the summation of outcomes in general. They argue that the utility provided by a sequence of outcomes is mainly determined by the average utility of both the most extreme event and the final event of the sequence (Kahneman, Wakker, and Sarin 1997). This is referred to as the peak-end rule.

The evaluation of service delivery processes has been of interest to marketing and service researchers already for years. A number of studies focused on how each separate element of the service delivery process contributes to the overall evaluation of the service. For instance, Szymanski and Hise (2000) studied the effect of different aspects of e-tailing on the overall customer satisfaction with an e-tailing service. Heskett, Sasser, and Hart (1990) advised a consistent performance within the service encounter implying that the sequence of events is not that important. In contrast, Bolton and Drew (1992) emphasized a strong start of a service encounter. Recently, service researchers incorporated some of the principles from the referred psychological and economic literature on sequences of events (Cook et al. 2002). Chase and Dasu (2001, p. 81) proposed that service encounters should finish strongly, and bad experiences should be overcome as early as possible. Hansen and Danaher (1999) examined the impact of inconsistent performance during the service encounter. Their general finding was that an improvement in performance during the encounter resulted in higher evaluations than a decline in performance.

Despite the interest in the service literature in considering services as sequences of events, there is still a lack of knowledge on this topic. Chase and Dasu (2001) and Cook et al. (2002) based their opinions on results from Daniel Kahneman and George Loewenstein and did not provide any empirical evidence of whether these findings also hold in service contexts. Using a qualitative pilot study and an experiment, Hansen and Danaher (1999) focused on the performance trend and the role of the final event, but they did not consider the role of peaks in the service process.

The objective of this article is to provide an empirical test on the theories proposed in the psychological and economic literature on the evaluation of sequences of events in a service context. In line with these theories, we specifically consider the effect of (1) the average performance, (2) performance peaks, (3) the final event, and (4) performance increases (decreases) during the sequence of events. The service context of our study is an inbound service call between a call center employee and a customer of a financial service provider, which is considered as a sequence of events or episodes.

The structure of this article is as follows. We first discuss theories on the evaluation of a sequence of events. Subsequently, we discuss our theoretical model. We continue with a description of our research methodology and empirical results. We end with a discussion and managerial implications.

THEORETICAL BACKGROUND

Sequences of Events

According to Loewenstein and Prelec (1993), it is often difficult to consider a particular series of events as a sequence. For example, when the time between events is rather long, one could dispute whether this should be called a sequence. A service example might be the occurrence of different events in a customer's relationship with an insurance provider, such as claiming and paying a bill. Such events generally are separated by a relatively long time delay. Loewenstein and Prelec (1993, p. 93) argued that when outcomes are commensurable and tightly spaced, the logic for treating these events as a sequence will be more compelling. Commensurability of events refers to the fact that the events should have approximately the same characteristics or attributes. Thus, even when outcomes occur quickly after one another, they may not be considered a sequence when they have different attributes. For example, in the case of a visit to a shopping mall, shopping for groceries, fitting clothes, and visiting the bank should be considered as events with different attributes. As a result, they cannot be treated as a sequence of events.

In the literature on sequences of outcomes, two evaluation modes can be distinguished. Some researchers focused on the ex ante preference formation of the sequence (e.g., Loewenstein and Prelec 1993). In this case, the utility of a sequence of outcomes was considered in advance. Other researchers investigated how a sequence of events is evaluated ex post (i.e., after the occurrence of events) (e.g.,

Kahneman, Wakker, and Sarin 1997). The latter focuses on the experienced utility of the outcome sequence. Our research deals with experienced utility because we study ex post evaluations of service encounters.

Preferences for Sequences of Events

Research on preferences for multiple outcomes has focused on two important issues. First, researchers investigated whether consumers prefer separated or combined outcomes. Prospect theory suggests that humans prefer separation of positive outcomes, whereas they prefer the combination of negative outcomes (Thaler 1980). However, Thaler and Johnson (1990) and Thaler (1999) showed that people also prefer separation of negative outcomes, whereas Linville and Fischer (1991) also could not find evidence for a preference for combining negative outcomes. Loewenstein and Prelec (1993) provided more conclusive evidence. They showed that people generally prefer separation of both positive and negative outcomes, which implies spreading of the outcomes in a sequence of events.

A second important issue in research on preference formation for sequences of events concerns the preferred flow of outcomes. According to standard discounting models, people should prefer the best outcomes occurring at the beginning of a sequence, because people aim to maximize the net present value of future outcomes. However, research has shown that people rather prefer sequences of events that show a positive development. For example, Loewenstein and Sicherman (1991) reported that employees prefer an increasing over a decreasing wage profile, the latter having the highest net present value. This phenomenon is also referred to as negative time preference, which is considered irrational in an economic sense. Two mechanisms are believed to cause this type of preference: (1) savoring and dread, and (2) adaptation and loss aversion. Savoring (dread) refers to the fact that people derive positive (negative) utility from anticipating a future event. For example, knowing that they will go on a summer holiday, people can already derive pleasure in advance. Adaptation means that people get used to a certain stimulus level. New stimuli will be considered as either positive or negative deviations from the adaptation level. As people generally prefer gains with respect to the current adaptation level (Tversky and Kahneman 1991), the new adaptation level tends to be higher than the previous one. Hence, people generally prefer improvements over time. Read and Powell (2002) provided some additional reasons for the preferences of a positive development in the sequence: appropriateness, people's expectations, and convenience. Although people have a negative time preference for the development of outcomes, research also shows that people prefer a fast improvement over a slow improvement of outcomes (Hsee and Abelson 1991).

Evaluation of Sequences of Events

Kahneman, Wakker, and Sarin (1997) distinguished two types of experienced utility: instant utility and remembered utility. Instant utility is the pleasure or distress felt at a particular moment. Remembered or retrospective utility is defined as the retrospective evaluation of a temporally extended outcome (Kahneman 1994). Each event in a sequence provides an instant utility to a consumer. If one evaluates the total sequence of events, one focuses on remembered utility. Remembered utilities have an adaptive function, as they determine whether a situation experienced in the past should be approached or avoided (Kahneman, Wakker, and Sarin 1997). Usually, it is assumed that the remembered utilities of a sequence of events equal the sum of the instant utilities of all events. This implies that the remembered utility can be predicted best with the average value of these instant utilities. However, recent research from Nobel Prize winner Daniel Kahneman and his coauthors showed that this is often not the case.

The remembered utility of a sequence of events is accurately predicted by averaging the peak (most intense value) of instant utility recorded during an episode and the instant utility recorded near the end of a sequence of events. This rule is referred to as the peak-end rule (Kahneman, Wakker, and Sarin 1997). Evidence for this rule has been gathered using experiments in which participants were exposed to unpleasant experiences, such as watching aversive movies, undergoing a colonoscopy in a hospital, and immersing a hand to the wrist in cold water (e,g., Fredrickson and Kahneman 1993; Kahneman et al. 1993; Redelmeier and Kahneman 1996).

The peak-end rule has two consequences. First, the remembered disutility of a bad episode can be reduced with the addition of an extra period of somewhat less discomfort that reduces the peak-end average. For example, Kahneman et al. (1993) showed that individuals exposed to an unpleasant episode, which was followed by a shorter less unpleasant episode, had higher remembered utilities than individuals exclusively exposed to the unpleasant event. A second consequence of the peak-end rule is duration neglect, which implies that the duration of experiences has little or no independent effect on the individual's remembered utility.

In line with the peak-end rule, the well-known recency effect also predicts that the last outcome of a sequence of events should be the most prevalent in the evaluation process (Anderson 2000). This effect assumes that the last event should be the most salient. In practice, this effect is

often used to instruct service employees to create a happy ending of a service experience.

Combining Theories

Ariely and Carmon (2000) combined the insights from the literature on the preferences for sequences of outcomes and literature on the evaluation of sequences of outcomes. In a hospital study, patients rated their pain every hour. These pain ratings were related to evaluations of overall pain. In their model, they included the final pain rating, the slope, the peak, and the average of the pain ratings. Their results showed that both the end and the slope had a significant positive effect on the overall evaluations, whereas the peak and the end had no effect. In our model, we will follow the approach of Ariely and Carmon (2000) and apply it in a service setting. Note furthermore that we also extend the model with the inclusion of the lowest peak in the sequence.

Service Research on Sequences of Events

In the foregoing we focused on the psychological and economic theories underlying evaluations of sequences of events. Service researchers sometimes have considered services as prolonged experiences. Some studies have focused on how satisfaction is formed during the service encounter. Oliver and Burke (1999) showed that prior expectations affected experiences early in the service encounter, the effect declining later on. Hansen and Danaher (1999) focused on customer evaluation of inconsistent service delivery. In a qualitative pilot study, they showed that inconsistent beginnings and endings were more extremely judged than consistent service performance. Their study also showed that judgments were relatively favorable when performance improved (positive trend). In summary, there is some evidence that some of the principles arising from the psychological and economic literature on sequences of events indeed also hold for service encounters. However, other principles (i.e., the effect of peaks) have not yet been tested within the service area. According to Rust et al. (1999, p. 89), peaks may be important in service evaluations. They suggest that consumers are not only sensitive to the average performance of a service but also to its variability around it. However, from a methodological point of view, the evidence for service evaluation based on sequence characteristics is limited. We therefore propose a model that incorporates all of the principles discussed so far and test this model in a service context.

EVALUATION MODEL FOR SEQUENCES OF SERVICE EVENTS

On the basis of the above theories, we develop models that explain customer *i*'s experienced utility of a sequence of events of a particular service (u_i). As noted, standard thinking in services research and economics assumes that μ_i can best be predicted by the average utility of the outcomes (μ) during the service process by customer *i*. We define the average utility of service events, μ_i, for individual *i* as follows:

$$\mu_i = \frac{1}{N_i} \sum_{n=1}^{N_i} U_{n,i}, \tag{1}$$

where N_i is the number of events occurring in the sequence. Our first model only includes the averaged utilities as an explanatory variable and is formulated as follows:

$$u_i = \beta_0 + \beta_1 \times \mu_i + \xi_i, \tag{2}$$

with β the quantified effects and ξ a normally distributed error term.

Subsequently, we account for the peak-end rule, that is, the implication that the experienced utility of a sequence of events is accurately predicted by the utilities of the peaks and the instant utility at the end of the sequence (EndU_i). The peaks in the sequence include both the minimum (MinU_i) and maximum values (MaxU_i) of the utilities of the outcomes in the sequence. Including the above terms, equation (2) becomes:

$$u_i = \beta_0 + \beta_1 \times \mu_i + \beta_2 \times \text{Max}U_i + \beta_3 \times \text{Min}U_i + \beta_4 \times \xi_i. \tag{3}$$

Our final model also accounts for preferences regarding the trend in the sequence of events. In this respect, we focus on the finding that people prefer a positive development of events over time. This positive development is measured by estimating a separate regression of utility of each outcome on its order in the sequence. The resulting regression coefficient (γ_i) is used subsequently as a measure of the trend in the instant utilities of the outcomes in the sequence. Including the trend parameter in (3), our final model becomes:

$$u_i = \beta_0 + \beta_1 \times \mu_i + \beta_2 \times \text{Max}U_i + \beta_3 \times \text{Min}U_i + \beta_4 \times \text{End}U_i + \beta_5 \times \gamma_i + \xi_i. \tag{4}$$

Based on the above theories, we expect that the estimated coefficients β each have a positive sign.

In line with prior research in the satisfaction and service literature, we use satisfaction with the service as a measure of experienced utility (Anderson and Sullivan 1993; Bolton 1998).

RESEARCH METHOD

Context

The context of the study is a service call center in a financial service market. Customer service call centers have been the most important medium for customers to communicate with companies in the past 10 years (Anton 2000). Hence, call centers are an important part of a firm's customer relationship management (CRM) (Winer 2001). Many firms use call centers as focus of their customer satisfaction strategy (Feinberg et al. 2000). As a consequence, the customer's evaluation of a service call is rather important. This becomes even more prevalent from a CRM perspective, as empirical evidence shows that satisfied customers are more loyal (Bolton 1998) and thus generate more profits during their lifetime.

An important question with respect to our context is whether a service call can be considered as a sequence of events. First, in comparison with, for example, cruises, amusement park visits, and restaurant visits, service calls have a short duration. Considering their short duration, it might be argued that service calls do not consist of separate events but rather should be considered as one event. However, the practitioner-oriented literature concerning call center operations shows clear guidelines and scripts for call center employees on how they should handle calls that are in line with the proposed theories. For example, in practice, it is suggested that call center employees should take care of a positive ending of the service call (Bencin and Jonovic 1989). Second, economic and psychological research on sequences of events considered both continuous streams of instant utilities (e.g., pain) and series of discrete events (e.g., film clips, series of numbers, life descriptions). Although a service call may consist of more discrete stimuli, possibly breaking the sequence, the stimuli in a service call are often closely related. Service call events all are part of the same customer service process that frequently occurs with the same service agent and in the same customer setting. Hence, we consider them as distinct, though related events that are considered as a sequence by customers.

Data Collection

During 2 months in the beginning of 2001, inbound service calls of a large European financial service provider were selected. We employed the following criteria to select service calls that were used in our empirical study:

1. Service calls were limited to calls with existing customers of the financial service provider (not prospects).
2. The topics of the service calls were representative for the normal incoming service calls.
3. Service calls that were very emotional (e.g., angry) were not selected, because the investigation might have further harmed the customer's relationship with the company.
4. Each customer was selected only once.
5. A maximum of five service calls per agent was allowed.

The selection of service calls occurred after each working day, the main focus being that the calls were representative for the usual incoming service calls. During our study, it appeared that only a few service calls (2-5) were excluded, because they were very emotional. The customers of the selected service calls were approached with a short telephone questionnaire during the evening of the same day. The minimal time interval between the service call and the questionnaire was 1 hour. The questionnaire measured the customer's evaluation of the service call. This resulted in 97 usable service calls with its accompanying questionnaires. For privacy reasons, the tape containing the calls was destroyed after coding of the calls. In addition to the 97 calls, 7 calls were saved in order to study the coding reliability.

Sample Description

In line with the age of the customer population of the financial service provider, the average age of the respondents was approximately 40 years, with a minimum of 21 and a maximum of 84 years. The majority of the respondents had completed at least high school education (82%), and 34% at least obtained a bachelor's degree. An explanation for the relatively high percentage of highly educated people is that this financial institution focuses on wealthy people in its marketing strategy. The average duration of the service calls was 208 seconds with a standard deviation of 145.27. In line with our selection criteria, the topics varied among the service calls. In total, the service calls considered 31 different and sometimes related topics, the most frequent of which were the arrangement of money

TABLE 1
Topics of Service Calls

Topic	*Frequency*
Phone accessibility	27
Payment methods	7
Bankcard	8
Payment advice	3
Account information	30
Collection payment	6
Money transfer	30
Other services	16

transfers and the provision of balance information. An overview of these topics clustered by general theme (e.g., bankcard) is given in Table 1.

Measurement

The evaluation of the service call was measured using two 5-point scales with the following adjectives: *very unpleasant—very pleasant*; *very dissatisfied—very satisfied*. These questions were based on Oliver and Swan (1989) and Crosby and Stephens (1987). The reliability coefficient alpha of these two items was .82 ($r = .71$, $p < .01$).

To measure the event utilities in the service calls, we distinguish between episodes and events, the latter including the former. Using this method, we followed the approach used by Doucet (1998), who considered each time period the same person was speaking as a separate episode. Subsequently, for each episode, utilities were measured as follows. We used a qualitative judgment to assess the emotional content of each episode, which was characterized by one single word, for example, *reassurance*. Next, the judge picked out the most representative words with respect to their emotional loading in each episode. These words of each customer used in the conversation with the service representative were scored with respect to pleasantness using the *Dictionary of Affect in Language* (Sweeney and Whissel 1984).[1] Positive words, such as *friendly*, obtained a high score, whereas negative words, such as *worried*, obtained a low score. The minimum score was 1, whereas the maximum score was 3. The scores were averaged for each episode. This way, an indication of episode utilities was obtained that was independent of the customer's experienced utility of the conversation.

In our view, episodes can hardly be considered as psychologically meaningful events constituting a series. Although all changes of voice occurred within the same call, the episodes can hardly be remembered separately and frequently are superficial (e.g., episodes containing only a few words, such as *OK*, *yes*, *I don't remember*). For this reason, we divided each call into psychologically meaningful entities, which we call events. We therefore instructed the judge to divide the call into a number of events based on the topics covered in the call. Specifically, if, for example, a customer both asked for account information and reported problems with his or her bankcard, this was considered as two separate events, each containing a number of episodes, defined by changes of voice. In most cases, the introduction and the ending of a call were also considered as separate events. In some instances, the beginning of the call coincided with the topic of the call, for example, "Hi, I am Mr. . . . and I lost my credit card." Sometimes the ending of a call was very short and went hardly unnoticed, for example, "OK, I will go to the bank office again, bye." We have chosen our methodology because the division into events is more in line with how customers would categorize different elements of a call. It is also in line with prior research in service marketing, where different elements of the service process (i.e., ordering, payment and delivery of products by an e-tailer) are considered as separate parts of the service.[2] The episode utilities within each event were simply averaged across the number of episodes, resulting in the event utilities used in our estimations.

Measurement Reliability

Reliability of qualitative judgments is essential (Rust and Cooil 1994). To assess the reliability of the judgments for a sample of seven calls (not included in the main analysis),[3] we used the following approach. First, the three judges summarized the emotional content of each episode

1. The *Dictionary of Affect in Language* contains 8,742 words that were rated regarding pleasantness, activation, and imagery. On average, pleasantness was rated on average by eight different judges. The average ratings correlated more than 80% with subsamples of ratings. We consider the *Dictionary of Affect in Language* as an independent and reliable instrument for assessing the emotional content of conversational text. This dictionary has been used in several instances (i.e., the emotional analysis of novels). For more information on this dictionary, we refer to publications by Cynthia Whissel (e.g., Sweeney and Whissel 1984; Whissel 1994). The only subjective part of our utility assessment of the episodes in the telephone conversations that we studied is the selection of words. The scoring of the words is based on objective criteria provided by the *Dictionary of Affect in Language*.

2. Despite the arguments in favor of our method, we also divided the service calls into episodes using Doucet's (1998) approach. Using this method, we did not find any significant effects of the variables considered. Considering each time period the same person was speaking as a separate episode appeared to be meaningless in explaining customer satisfaction with the call.

3. The assessment of the reliability was executed as a response on the comments of the reviewers. Unfortunately, the 97 calls in this article were no longer at our disposal on tape. Company policy did not allow us to keep these tapes, and they destroyed the content of these tapes. Thus, we could not do the check for the total sample. However, fortunately, we were able to collect seven additional calls, where we at least could get some idea about the reliability of the scores of the judge. Hence, we report this reliability check. Nonetheless, we note that ideally, the reliability check should have been performed for the total sample of calls.

independently and then picked out the most representative words with respect to their emotional loading in each episode. Subsequently. the words were scored using the *Dictionary of Affect in Language*. The resulting event utilities based on the scores of the judges were compared. The correlation coefficients between the three judges, based on a total of 196 episodes in the seven calls, were .69, .64, and .64, respectively ($p < .01$). The high correlation coefficients indicate that the resulting utility assessments of the judges were relatively comparable.

Next, the judges decided about the different topics of the calls to be considered as events in our analysis. Topics included, for example, greeting, explaining a problem, treatment of a problem, referring the problem to someone else (usually the bank office), providing information, and ending the call. The judges independently indicated the episodes belonging to each event for each of the seven calls. A change of topic usually was quite clear and differed by two changes of voice at most between the judges. The number of episodes distinguished differed between the judges by one at most. The correlation coefficients for the number of events between the three judges were .88, .88, and .94, respectively ($p < .05$). A further analysis revealed that the distinguished events by the judges were almost always the same. Thus, we conclude that the measurement of both the utilities and the indication of episodes belonging to an event was reliable. On the basis of this exercise, we have substantial confidence in the reliability of the data of our 97 calls. The utility score per event was calculated by averaging the pleasant scores of the underlying episodes.

EMPIRICAL RESULTS

Descriptive Analysis

Table 2 shows the averages and standard deviations for the variables concerned. The average event utilities of the service calls (μ) equaled 2.39. The event utilities varied between 1.77 and 2.69. As expected, the mean minimum instant utility (MinU) was well below the average utility, whereas the mean maximum instant utility (MaxU) exceeded the average utility. The average utility of the last event (EndU) was 2.51, somewhat higher than the average utility. The trend coefficient γ indicated that the instant utilities generally increased during the service call. Note, however, that this was only a slight increase and that the average value was 0.045 with a standard deviation of 0.10. Finally, most customers were rather satisfied with the service call, given the average value of 4.27 on a 5-point scale with a standard deviation of 0.66 (minimum = 2, maximum = 5). Thus, even though we excluded a few very emotional service calls, the variation in our satisfaction scores was substantial. The skewness of the satisfaction scores was 0.07, indicating a nonskewed distribution.

Table 2 shows the correlation coefficients between the variables. The correlation coefficients were rather high, ranging from .09 to .68. These high correlations can be explained by the fact that event utilities interacted with one another, while also some independent variables were to some extent related to each other. For example, the average event utility (μ) included MaxU, MinU, and EndU. Furthermore, there were significant positive correlations between satisfaction and μ, and between satisfaction and MinU, and there was a significant negative correlation between the trend parameter (γ) and satisfaction. The correlation analysis also reveals that satisfaction is significantly positively correlated with the average event utility and MinU ($p < .05$), whereas, as suggested, the trend parameter has a negative significant effect (γ) ($p < .05$).

Regression Analysis

After deleting two outliers, we estimated equations (2), (3), and (4) with with ordinary least squares (OLS) regression analysis. The estimation results are shown in Table 3. To assess whether the model fit increased significantly, we used Wald tests, in which we compared the F values of the restricted models with the less restricted (or extended) model in each case (Pindyck and Rubinfeld 1998). The high correlations between the independent variables might suggest possible multicollinearity problems in our regression analysis. We computed variance inflation factor (VIF) scores to assess the presence of multicollinearity.[4] In our most extended model, the maximum VIF score was 4.3. As this score was below 6, multicollinearity should not severely affect our regression results according to Hair et al. (1998). Furthermore, μ remained significant in equations (2) and (3) despite the inclusion of other correlated variables. This is another indication that multicollinearity was not problematic.

Equation (1) only included a constant term and the average event utility (μ). This model explained approximately 5% of the variance and was statistically significant ($p = .02$). The estimated coefficient of μ was positive and significant ($p = .03$). The addition of MaxU, MinU, and EndU in equation (2) significantly improved the model fit ($p = .02$) according to a Wald test. The R^2 of the extended model was .147. The coefficient for MaxU was positive and significant ($p = .04$); however, no significant coeffi-

4. An indicator of the effect the other predictor variables have on the variance of a regression coefficient is that it is directly related to the tolerance value ($VIF_i = 1/R_i^2$). Large variance inflation factor (VIF) values also indicate a high degree of multicollinearity among the independent variables (Hair et al. 1998).

TABLE 2
Averages, Standard Deviations, and Correlations of Event Utilities and Satisfaction

			Pearson Correlation Coefficients					
	Average	SD	μ	*MaxU*	*MinU*	*EndU*	γ	*Satisfaction*
μ	2.39	0.20	1.00					
MaxU	2.66	0.18	.18**	1.00				
MinU	1.83	0.48	.68**	.06	1.00			
EndU	2.51	0.11	.52**	.28**	.15	1.00		
Trend (γ)	0.04	0.10	–.49**	.09	–.51**	.24**	1.00	
Satisfaction	4.21	0.55	.24**	.15	.23**	–.081	–.22**	1.00

NOTE: μ = average event utility; MaxU = man maximum instant utility; MinU = mean minimum instant utility; EndU = average utility of the last event.
**$p < .05$.

TABLE 3
Estimated Regression Results for Equations (2), (3), and (4)

		Equation (2)		*Equation (3)*		*Equation (4)*	
		Coefficient	t-*Value*	*Coefficient*	t-*Value*	*Coefficient*	t-*Value*
Constant	(β_0)	2.64	3.86**	4.32	3.12**	4.58	3.13**
μ	(β_1)	0.68	2.38**	0.90	1.93**	1.11	1.90**
MaxU	(β_2)			0.63	1.80**	0.62	1.74**
MinU	(β_3)			0.08	0.54	0.09	0.57
EndU	(β_4)			–1.61	2.55**	–1.91	–2.36**
γ	(β_5)					0.48	0.59
F-value (*df*)		5.68** (1, 93)		3.88** (3, 91)		3.15** (5, 89)	
R^2 (adjusted R^2)		.058 (.048)		.147 (.109)		.150 (.103)	

NOTE: A two-sided *p* value is reported only for β_4 because the coefficient sign contrasts our expectations. μ = average event utility; MaxU = man maximum instant utility; MinU = mean minimum instant utility; EndU = average utility of the last event; γ = trend.
**$p \leq .05$.

cient was found for MinU ($p = .30$). Surprisingly, a significant negative coefficient ($p = .01$) was found for EndU. Note that the coefficient for the average utility (μ) remained significant ($p = .03$). Thus, both the average utility of the service call and the positive peak explained the satisfaction with the service call. In the final model, the trend parameter (γ) was added. The model fit did not increase significantly ($p = .56$). This was reflected in the very small increase of R^2, valued at .150.[5] The regression coefficient for γ was positive but not significant ($p = .28$). The significance of the other explanatory variables was almost the same as in (3).[6]

Despite satisfactory VIF scores (e.g., Hair et al. 1998), the relatively small sample size, combined with the relatively high correlation between the explanatory variables, might still raise suspicion about whether or not multicollinearity affected our results (Mason and Perrault 1991). To assess this issue, we used ridge regression (Hair et al. 1998; Judge et al. 1988; Mahajan, Jain, and Bergier 1977). In ridge regression, biased estimators for the regression coefficients are estimated in the hope that their smaller variances offset their bias so that the estimator mean square error is reduced below that of OLS.[7] The ridge regression results for the full model are shown in Table 4. The $R^2$of the ridge regression was a bit lower than the OLS regression model, due to the fact that this model accounted for multicollinearity. However, the results were highly comparable with the OLS estimation results. Again, the

5. Prior research in this area usually used experiments with simple hypothesis testing techniques (e.g., *t* tests). Thus, no R^2's are available from prior research. If we compare our results with usual satisfaction research in which respondents evaluate processes/attributes using statements and subsequently provide an overall satisfaction score, our R^2's are relatively lower. However, these R^2's are inflated due to common-method variance. In our method, the event utilities are not provided by the respondent using a questionnaire. As a result, there is no common-method variance. Hence, it is logical that we find a lower R^2.

6. In our analysis, we do not control for the purpose of the call as our sample size is rather small and there are a number of different topics. If we would include all these topics, this would reduce the degrees of freedom in our model substantially. It also makes our model less parsimonious.

7. In ridge regression, the regression estimator is defined as follows: $b(k) = (X'X + kI)^{-1}X'y$, where $k > 0$ is a constant. An important decision in ridge regression is the choice of *k*. We estimated this model for several values of *k*. This resulted in approximately the same results. We report the results for $k = 0.5$. This value has been chosen, because the estimated regression coefficients became more stable after a *k* value of 0.5. For an extensive discussion on ridge regression, we refer to Judge et al. (1988, chap. 21). We estimated ridge regression in SPSS 10.0.

TABLE 4
Ridge Regression Results for Equation (4)

		Coefficient	SE	*Coefficient*/SE
Constant	(β_0)	3.75	1.02	3.65**
μ	(β_1)	0.36	0.16	2.22**
MaxU	(β_2)	0.39	0.23	1.74**
MinU	(β_3)	0.11	0.07	1.49*
EndU	(β_4)	–0.62	0.32	–1.95**
γ	(β_5)	–0.30	0.33	-0.90
R^2 = .12 (adjusted R^2 = .07)				
F-value = 2.42** (*df* = 5, 89)				

NOTE: A two-sided *p* value is reported only for β_s because the coefficient sign contradicts our expectations. μ = average event utility; MaxU = man maximum instant utility; MinU = mean minimum instant utility; EndU = average utility of the last event; γ = trend.
**$p \leq .05$.

average utility (μ) and the positive peak (MaxU) had a significant positive effect. The utility of the final event (EndU) also had a significant negative effect, as before. However, there is one important difference. In the ridge regression, the minimum peak also showed the expected positive effect. However, it is only marginally significant ($p < .10$). This effect possibly did not show up in OLS, because MinU has relative high correlations with μ, EndU, and γ. Because ridge regression to some extent adjusts the OLS estimations for multicollinearity, this effect could now show up. Note furthermore that the found effect is in line with the significant positive correlation between MinU and satisfaction (see Table 2). Finally, the effect of the trend coefficient (γ) remained insignificant.

DISCUSSION

We considered a service process as a sequence of events. On the basis of both economic and psychological theories about the evaluation of sequences of events, we developed a model that was tested in the context of service calls in the financial service market. Our main objective was to test the theories proposed in the psychological and economic literature on the evaluation of sequences of events in a real-life service context, thus extending the available literature on sequences of events in services (e.g., Hansen and Danaher 1999). Furthermore, we did not only consider the performance trend and the end of the service sequence but also the peaks in the service performance. However, because our study used only a relatively small sample of customers in only one particular service process (service calls), our results should be considered as exploratory. As such, we discuss some preliminary findings that we believe should stimulate more comprehensive research in other service settings, which finally could lead to some generalizations.

According to the peak-end rule, one would predict that both the peaks and the end of the service call would mainly affect its evaluation. Our results were not fully in line with this prediction. However, our results do support that specific service processes can be considered as a sequences of events. In contrast with the peak-end rule, our results showed that the average utility of the service call was a significant predictor of the experienced utility (customer satisfaction). In addition, the positive peak (MaxU) of the sequence had a positive effect on the experienced utility. This effect is in line with prior research on sequences of events (e.g., Ariely and Carmon 2000). After accounting for possible multicollinearity effects, we also found some evidence for a negative peak effect (MinU). Our research on services confirms findings of Rust et al. (1999), who explicitly argued that both the average performance of the service and the deviations from the average (peaks) are important in shaping a customers' service evaluation.

Surprisingly, we found a negative effect of the end utility of the service call on customer satisfaction.[8] This result is difficult to explain and contradicts prior research (Ariely and Carmon 2000; Hansen and Danaher 1999). We have several reasons for this finding. First, it might be due to the fact that service calls usually have a happy ending, because representatives are instructed to act this way. Consequently, customers may discount the happy ending, leading to less pleasant experiences. Second, customers may feel obliged to end the call in a nice way. Then, when they think of the call in retrospect, they may become relatively dissatisfied. This reasoning implies that a routinely applied happy ending strategy might have negative consequences for customer satisfaction. Third, when we further investigated the endings of several calls, these endings were rather meaningless and phrased in the same fashion. Many calls more or less ended with words, such as "goodbye," "thanks for the service," and so forth. Still, these words obtained a relatively high pleasantness score compared to other events in the call. What might have occurred is that a negative call, which was evaluated negatively, still had a relatively positive ending. This could result in a negative effect of the end of the call on the total evaluation. Fourth, from a methodological standpoint, the relatively short duration of the calls resulting in few events and thus high interrelationship between different event scores (μ, MaxU, EndU, and MinU) has led to relatively high correlations in the independent variables, which might have led to multicollinearity resulting in wrong parameters. Note, however, that as we corrected for multicollinearity using

8. The finding of a negative end effect is indeed surprising and counterintuitive. The categorization of episodes into events might probably have led to this result. We assessed this issue by also estimating a model in which we used the original episodes, that is, each conversation of the customer or agent as a separate observation. This model also resulted in a negative end effect.

ridge regression, we still found the negative end effect. Moreover, we also found a negative correlation between EndU and satisfaction. Together, these explanations force us to emphasize that the negative end effect we found should be interpreted with caution.

Although we found a negative significant correlation between the trend and satisfaction, we did not find evidence for a sequence trend effect on satisfaction in our model. This is not in line with prior research (Hansen and Danaher 1999). We explain this as follows. First, the time period of the sequence of events was rather short, compared with time periods in earlier research. In this short time period, people may not have noticed the development of the event utilities. Second, the positive trend hypothesis has been tested mainly with preferred sequences of utilities instead of experienced utilities. In this research, we focused on the experienced utilities.

Summarizing, our results are not completely in line with prior research. We already provided some explanations for these deviations. Methodological issues may further explain our findings. First, in contrast with prior studies, we studied consumers in a real-life setting, in which a number of additional factors may shape customer satisfaction. We do not control for these factors, as is usually possible in experimental settings. Second, whereas studies of, for example, Frederickson and Kahneman (1993) were based on continuous measures concerning one topic (e.g., the subjective evaluation of pain—putting hands in cold water, undergoing colonoscopy), our study reports findings from sometimes-related topics. Although the service call topics were all related to banking, none of the calls had exactly the same content. As a result, it appeared difficult to replicate experimental findings in real-life research. Third, as already noted, a service call has a relatively short duration. Prior research mainly considered sequences of events with longer durations (e.g., Hansen and Danaher 1999). This might have made it particularly difficult to find a trend effect. Moreover, in a real-life service call setting, trends are not easily found, as the performance may go up and down over time, like a random process. In experimental research, however, performance can easily be manipulated (e.g., Hansen and Danaher 1999). One might also argue that the short average duration of the studied service calls might also make it less plausible that respondents experiencing these calls perceive a kind of peak-end rule. Therefore, we recommend that this study should be extended to service calls with longer durations (i.e., for getting information on more complex financial services, such as house loans). From a data viewpoint, this should lead to more events and as a result lower correlations between the different sequence characteristics.

MANAGEMENT IMPLICATIONS

Our research suggests that service managers should think of service processes as sequences of events. This especially holds for services that cover a relatively short time period and consist of rather related events. In the management of these processes, managers should be aware that satisfaction is not created solely by the average quality of the events in the service process. Satisfaction can be further enhanced with the provision of a positive peak experience. Thus, managers should probably not solely focus on a consistent performance during the service encounter. In addition to this, they should aim to provide a positive peak experience. This peak experience will enhance satisfaction.

Despite these guidelines, managers should be aware that the experiences in the service call are still rather subjective and customer specific. These experiences may be affected by the mood of the customer when calling. Managers can only affect the sequence by ordinary quality control methods. However, the use of these methods does not guarantee positive changes of customers' mood.

RESEARCH LIMITATIONS AND FURTHER RESEARCH

This research has the following limitations. First, we only considered service calls as an example of a service process that can be considered as a sequence. The application of the peak-end theory to the sequence of experiences in a telephone call can be considered as a stretch. Moreover, a service call consisted of related events. However, in other services (i.e., airline flights), the events might be less interrelated. This might have consequences for the appearance of the peak-and-end rule. Future research might therefore consider other service contexts. Second, our sample was rather small. However, the effort involved in measuring the utilities of the service calls did not allow us to include more cases. Future research could be extended to other industries and other service providers. Third, the financial service provider did not allow us to collect survey data on a few very emotional calls. This might have caused a lower variance in both the dependent and independent variables, which subsequently might have affected the significance levels in the regression analysis. Fourth, our research concerned customers of only one financial service provider. In this study, we also neglected the consumer's expertise in our model. However, research shows that satisfaction formation may differ between experienced and nonexperienced customers (e.g., Rust et al. 1999). In future research, the effect of experience may be incorporated. Fifth, the conducted study measured satisfaction

afterward, whereas the utilities were measured in an objective way during the sequence of events. However, we did not consider expectations and also did not consider dynamic updating of these expectations during the sequence (e.g., Oliver and Burke 1999). Future research might incorporate these issues in a more extended model.

Besides the future research topics that arise from our research limitations, we also propose some additional avenues for future research. First, future research may focus on the conditions in which a negative end effect in service sequences occurs. Second, in this research, we focused exclusively on customer experiences. Future research might also consider the agent's experience and interactions between agents and customers. It would be especially interesting to study how customers react to different agents and how agents react to different customers. Third, in this research, we used theory from economics and psychology to explain customer satisfaction in service calls. However, customer satisfaction might also be explained by other factors (i.e., waiting time and quality of the phone connection). Future research might develop and test models that incorporate these other factors. Finally, future research might consider how the evaluation of event sequences affects customer loyalty. This issue is especially interesting in light of the increasing attention for CRM (Hogan, Lemon, and Rust 2002).

Finally, we express the hope that our exploratory study stimulates further research on the interesting topic of evaluation of service event sequences.

REFERENCES

Anderson, Eugene W. and Mary W. Sullivan (1993), "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science*, 12 (2), 125-43.

Anderson, John R. (2000), *Learning and Memory: An Integrated Approach*. New York: John Wiley.

Anton, Jon (2000), "The Past, Present and Future of Customer Access Centers," *International Journal of Service Industry Management*, 11 (2), 120-30.

Ariely, Dan and Ziv Carmon (2000), "Gestalt Characteristics of Experiences: The Defining Features of Summarized Events," *Journal of Behavioral Decision Making*, 13, 191-201.

Bencin, Richard L. and Donald J. Jonovic (1989), *Encyclopedia of Telemarketing*. Englewood Cliffs, NJ: Prentice Hall.

Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17 (1), 45-65.

——— and James H. Drew (1992), "Mitigating the Effect of Service Encounters," *Marketing Letters*, 3 (1), 57-70.

Chase, Richard B. and Sriram Dasu (2001), "Want to Perfect your Company's Service? Use Behavioral Science," *Harvard Business Review*, 79 (June), 79-84.

Cook, Lori S., David E. Bowen, Richard B. Chase, Sriram Dasu, Doug M. Stewart, and David A. Tansik (2002), "Human Issues in Service Design," *Journal of Operations Management*, 20 (2), 159-74.

Crosby, Lawrence A. and Nancy Stephens (1987), "Effects of Relationship Marketing on Satisfaction, Retention, and Prices in the Life Insurance Industry," *Journal of Marketing Research*, 24 (4), 404-11.

Doucet, Lorna (1998), "Responsiveness: Emotion and Information Dynamics in Service Interactions," working paper 98-15, the Wharton School, University of Pennsylvania, Philadelphia.

Feinberg, Richard A., Ik-Suk Kim, Leigh Hokama, Ko de Ruyter, and Cherie Keen (2000), "Operational Determinants of Caller Satisfaction in the Call Center," *International Journal of Service Industry Management*, 12 (2), 131-41.

Fredrickson, Barbara L. and Daniel Kahneman (1993), "Duration Neglect in Retrospective Evaluations of Affective Episodes," *Journal of Personality and Social Psychology*, 65 (1), 45-55.

Hair, Joseph F., Rolph E. Anderson, Ronald L. Tatham, and William C. Black (1998), *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall.

Hansen, David E. and Peter J. Danaher (1999), "Inconsistent Performance during the Service Encounter," *Journal of Service Research*, 1 (3), 227-35.

Heskett, James L., W. Earl Sasser Jr., and Christopher W. Hart (1990), *Service Breakthrough: Changing the Rules of the Game*. New York: Free Press.

Hogan, John E., Katherine N. Lemon, and Roland T. Rust (2002), "Customer Equity Management: Charting New Directions for the Future of Marketing," *Journal of Service Research*, 5 (1), 4-12.

Hsee, Christopher K. and Robert P. Abelson (1991), "Velocity Relation: Satisfaction as Function of the First Derivative of Outcome over Time," *Journal of Personality and Social Psychology*, 60 (3), 341-47.

Judge, George E., R. Carter Hill, William E. Griffits, Helmut Lütkepohl, and Tsoung-Chao Lee (1988), *Introduction to the Theory and Practice of Econometrics*. New York: John Wiley.

Kahneman, Daniel (1994), "New Challenges to the Rationality Assumption," *Journal of Institutional and Theoretical Economics*, 150 (1), 18-36.

———, Barbara L. Fredrickson, Charles A. Schreiber, and Donald A. Redelmeier (1993), "When More Pain Is Preferred to Less: Adding a Better End," *Psychological Science*, 4 (6), 401-5.

———, Peter P. Wakker, and Rakesh Sarin (1997), "Back to Bentham? Explorations of Experienced Utility," *Quarterly Journal of Economics*, 112 (2), 376-405.

Linville, Patricia and Gregory W. Fischer (1991), "Preferences for Separating or Combining Events," *Journal of Personality and Social Psychology*, 60, 5-23.

Loewenstein, George and Drazan Prelec (1993), "Preferences for Sequences of Outcomes," *Psychological Review*, 100 (1), 91-108.

——— and Nachum Sicherman (1991), "Do Workers Prefer Increasing Wage Profiles?" *Journal of Labor Economics*, 9 (1), 67-84.

Mahajan, Vijay, Arun K. Jain, and Michel Bergier (1977), "Parameter Estimation in Marketing Models in the Presence of Multicollinearity: An Application of Ridge Regression," *Journal of Marketing Research*, 14 (4), 586-91,

Mason, Charlotte H. and William D. Perrault, Jr. (1991), "Collinearity, Power and Interpretation of Multiple Regression Analysis," *Journal of Marketing Research*, 23 (3), 268-80.

Oliver, Richard L. and Raymond R. Burke (1999), "Expectation Processes in Satisfaction Formation: A Field Study," *Journal of Service Research*, 1 (February), 196-214.

——— and John E. Swan (1989), "Consumer Perceptions of Interpersonal Equity and Satisfaction in Transactions," *Journal of Marketing*, 53 (2), 21-35.

Pindyck, Robert S. and Daniel L. Rubinfeld (1998), *Econometric Models and Economic Forecasts*. Boston: McGraw-Hill.

Read, Daniel and Melanie Powell (2002), "Reasons for Sequences of Preferences," *Journal of Behavioral Decision Making*, 15 (5), 433-60.

Redelmeier, Donald A. and Daniel Kahneman (1996), "Patients' Memories of Painful Medical Treatments: Real-Time and Retrospective Evaluations of Two Minimally Invasive Procedures," *Pain*, 66 (1), 3-8.

Rust, Roland T. and Bruce Cooil (1994), "Reliability Measures for Qualitative Data: Theory and Implications," *Journal of Marketing Research*, 31 (1), 1-14.

———, Jeffrey Inman, Jianmin Jia, and Anthony Zahorik (1999), "What You Don't Know about Customer Perceived Quality: The Role of Customer Expectation Distributions," *Marketing Science*, 18 (1), 77-92.

Sweeney, Kevin and Cynthia M. Whissel (1984), "A Dictionary of Affect in Language: I Establishment and Preliminary Validation," *Perceptual and Motor Skills*, 59, 695-98.

Szymanski, David M. and Richard T. Hise (2000), "e-Satisfaction: An Initial Examination," *Journal of Retailing*, 76 (3), 309-22.

Thaler, Richard H. (1980), "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organisation*, 1, 39-60.

———(1999), "Mental Accounting Matters," *Journal of Behavioral Decision Making*, 12 (3), 183-206.

——— and Eric J. Johnson (1990), "Gambling with the House Money and Trying to Break Even," *Management Science*, 36 (6), 643-60.

Tversky, Amos and Daniel Kahneman (1991), "Loss Aversion in Riskless Choice—A Reference Dependent Model," *Quarterly Journal of Economics*, 106 (4), 1039-1061.

Whissel, Cynthia (1994), "A Computer Program for the Objective Analysis of Style and Emotional Connotations of Prose: Hemingway, Galsworthy, and Faulkner Compared," *Perceptual and Motor Skills*, 79, 815-24.

Winer, Russell S. (2001), "A Framework for Customer Relationship Management," *California Management Review*, 42 (4), 89-105.

Zeithaml, Valarie A. and Mary Jo Bitner (1996), *Service Marketing*. New York: McGraw-Hill.

Peter C. Verhoef is an assistant professor of marketing in the School of Economics at Erasmus University in Rotterdam, the Netherlands. His main research interest is customer asset management. He has also conducted research on other topics, such as waiting times, private labels, and out-of-stocks. He has been a visiting research scholar at Tuck School of Business, Dartmouth College in fall 2003. He has published a wide variety of articles in journals, such as the *Journal of Marketing*, the *Journal of Marketing Research*, the *Journal of Retailing*, *Marketing Letters*, the *Journal of the Academy of Marketing Science*, and the *Journal of Consumer Psychology.*

Gerrit Antonides is a professor of economics of consumers and households at Wageningen University and senior fellow of the Mansholt Institute in the Netherlands. He has published research articles and textbooks in the areas of economic psychology, consumer behavior, and behavioral economics and teaches courses in these fields.

Arnoud N. de Hoog is a former student of the School of Economics at Erasmus University in Rotterdam, the Netherlands. The topic of his master's thesis concerned the evaluation of sequences of events in a financial service setting. He currently works as an investment adviser at MeesPierson Private Wealth Management in Curaçao, Netherlands Antilles.