The Journey from Episodes to Evaluations:

How Travelers Arrive at Summary Evaluations

Abstract

Understanding how travelers evaluate their overall trip experience is important to travel research. Psychologists suggested that these retrospective evaluations are often made by temporally integrating multiple episodes following simple heuristics that draw upon key episodes only, typically the peak and end episodes, rather than considering every episodic evaluation, weighted by its respective duration. To test these aggregation rules, a survey adapting the Day Reconstruction Method was conducted in 2017 among 691 travelers to Macau. Our findings reveal that summary evaluations are better predicted using an arithmetic average of all episodic evaluations, instead of the peak-end rule. This may be explained by the lengthier and more complex nature of travel, compared with other extended experiences that psychologists have investigated. The immediate theoretical implications are that 1) aggregate trip evaluations are influenced by most episodes, and 2) the relative duration of individual episodes is disregarded. Theoretical, methodological and practical implications are discussed.

Keywords: peak-end rule; retrospective evaluation of experience; temporal integration; Day Reconstruction Method; simple average

The average vacation is one-tenth playing—nine-tenths paying. –Arnold H. Glasow

Introduction

When answering questions such as "How was your recent trip?", the traveler tend to evaluate the overall trip, perhaps assuming that this involves an integration of the multiple episodes over an extended traveling period. Such evaluations are central to much tourism research and have been shown to be strong mediators of loyalty (Baker and Crompton 2000; del Bosque and San Martín 2008), and behavioral intentions such as willingness to pay more (Bigné and Andreu 2004), willingness to recommend (Bigné et al. 2008; Hosany and Prayag 2013) and revisitation intention (Ross 1993). Respondents to these questions are evaluating destinations through retrospective recollections of their travels. A few fundamental questions that underpin many investigations are: "[W]hat were people feeling and thinking during the episode, how will it be remembered upon future recall, how will it contribute to overall satisfaction of the total activity or trip?" (Mannell and Iso-Ahola 1987, 315). The present paper focuses on the latter question and investigates how travelers draw upon individual episodes and integrate them into an overall satisfaction. The ensuing insights offer the prospect of designing more precisely customized studies to answer relevant questions and can assist researchers to interpret findings when relying on retrospective evaluations. Understanding the psychology of the aggregation rules of how travelers evaluate their experiences is also of potential theoretical importance. To the authors' knowledge, no previous tourism researchers have examined how episodic experiences are constructed through an aggregation of such evaluations.

The idea of aggregation rule has a long history in economics (Edgeworth 1881). Economist have been interested in measuring the *total utility* by summing the *moment utilities*. More recently, a number of contributors to the psychology literature have examined how multi-episodes are aggregated (e.g., Fredrickson and Kahneman 1993). Various aggregation rules have been proposed to explain how we retrieve our memories of diverse episodes to construct summary evaluations. Normatively, it might be anticipated that we would recall all episodes through the extended period, as well as our related perceptions and feelings. We construct evaluations when recalling these events. Next, episodic evaluations are integrated temporally in a weighted-average manner, with each episode purportedly carrying a weight proportional to its duration.

While this proposal is theoretically plausible, it depends on a perfect recall of the past events on the part of respondents and is thus unrealistic. It has been well-established that people usually retrieve only abstract representations of a few key episodes, rather than recalling every event (e.g., Fredrickson and Kahneman 1993; Kahneman et al. 1993). Following this logic, summary evaluations are disproportionately dependent on these few episodes. This idea is echoed by Arnold Glasow. In the epigraph, the renowned writer and humorist proposes that only a small fraction of a vacation is really worth remembering. The perspective leaves a related question unanswered, namely which episodes? Of the various possibilities, there are robust findings that the most extreme (or peak) episode and the closing (or end) episode are most critical to the formation of summary evaluations (Fredrickson and Kahneman 1993).

While the peak-end rule has support from numerous empirical studies, these have not included evaluations of traveling. The current researchers examine this hypothesis in the context of travel, noting its potential importance for tourism research. They investigate whether traveling

is simply "a subset of leisure experiences, subject to the same explanatory and descriptive models" (p.315, Mannell and Iso-Ahola 1987) or if traveling is unique and distinct from other extended activities. The closest parallel to the present study are the works of Geng et al. (2013) and Kemp, Burt and Furneaux (2008), which asked undergraduate students the experiences during their school breaks. In the former study, it was found that the performance of peak-end rule declines over time. In the latter, the evaluations of peak-end episodes were found highly correlated with the overall evaluations. However, the studies suffered from a small sample, and it reported only summary statistics such as correlation coefficients, which are inadequate to test the validity of the peak-end rule and its comparison with other aggregation rules. More importantly, there is a fundamental difference between the activities reported by the student respondents and traveling. Students engage in multiple activities during their school breaks and travel is only one of the many options. Even during their travels, the respondents visited a variety of destinations, thereby introducing too much heterogeneity to allow for direct comparisons.

Given the enduing question about whether travel is unique relative to other leisure activities (Mannell and Iso-Ahola 1987), it is unclear whether travelers rely heavily on peak and end episodes when forming their overall trip evaluations, thereby paralleling how consumers' temporal integration process in other domains. Geng et al. and Kemp et al.'s studies leave as many questions as answers. The gap that is evident in the literature may be addressed by using a larger sample and more homogeneous activities that are undertaken at the designated destination. Along with the peak and end episodes, the current study examines a few possible inputs to the respondent evaluations, such as the simple average. As will be presented in this paper, a destination which is particularly popular amongst Asian travelers—Macau SAR, China—offers an instructive example to test the relevant hypotheses.

Compared with previous studies, the present investigation adapts a more sophisticated means of measuring participant satisfaction with individual events during their trips—Kahneman et al.'s (2004a, 2004b) Day Reconstruction Method (DRM). To the researchers' knowledge, this is the first application of the approach in the tourism context. Established online assessment methods demand that participants repeatedly provide responses throughout the course of their traveling (e.g., Chhetri, Arrowsmith and Jackson 2004). The DRM has the merit of being less intrusive for respondents. Though it is still reliant on retrospective assessments of prior episodes, the intervention occurs immediately after the trip, when memories remain vivid.

There are a few objectives. First, we would like to examine how travelers summarize the evaluation of their trips using different aggregation rules. These rules vary in terms of their cognitive effort requirements from simply relying on inputs from the peak and end episodes to a comprehensive consideration of every episodes and their respective durations. Second, we examine how the order of episodic and overall evaluations may influence the construction of these evaluations. This could have important methodological implications. Third, we would like to examine the downstream consequences of the inputs of these aggregation rules through their effects on the summary evaluation. In particular, we examine the overall evaluation's effect on travelers' revisit intention and their willingness of recommend.

Literature Review

Tourist Experience

In contemplating a travel experience, one might think of feeling the wind and/or the sun and sand at the beach, or sensing the smell of a city, whether Vienna, New York, Sydney or Beijing. These micro-experiences are 'assembled' into a meaningful whole for the traveller, as a result of cognitive and emotion-based processes. The trip consists of many micro-experiences, perhaps running into millions or even billions, involving a series of instantaneous responses "to some event having just occurred" (Kim and Fesenmaier 2017, 19). The multi-faceted nature of experience is especially important for tourism researchers, since many evaluations or creations of meaning occur after traveling to a place.

Research have adopted various lines of investigation when evaluating travel experiences and tourism researchers have placed a particular focus on satisfaction in vacation experiences. For example, tourist outcomes are commonly assessed on the basis of level of satisfaction. This refers to "an emotional state of mind after exposure to the opportunity" (Baker and Crompton 2000, 785). The adoption of satisfaction as the medium is also appropriate when studying the aggregation of summary evaluations. It may be stated that for economists, "utility means satisfaction" (Viner 1925, 638). According to Baker and Crompton, satisfaction is influenced by travelers' social-psychological wellbeing, along with extraneous events beyond the control of suppliers, such as weather and social interactions. Researchers have commonly examined tourist satisfaction using a cognitive-affective approach (del Bosque and San Martín 2008). For instance, it has been argued that "specific emotions may intervene or act as a mediator, between performance and satisfaction" (Otto and Ritchie 1995, 39).

Theoretically, overall evaluations of the travel experience are determined independently or jointly by various cognitive and emotional processes (e.g., Taylor and Baker 1994). Unsurprisingly, it has been found that summary evaluations involve significant contributions from both cognitive and emotional inputs. More importantly, it is consumer satisfaction (a cognitive process) that triggers the positive emotions that in turn lead to a satisfying trip (del Bosque and San Martín 2008). These results support cognitive appraisal theory, which hypothesizes that a cognitive process precedes an emotional outcome (Hosany 2012) and is determined by an interpretation, rather than by the event itself (Scuttari and Pechlaner 2017).

Another line of research has investigated the role of prior expectations on overall satisfaction in the travel experience by applying expectancy-disconfirmation theory. This predicts that experiences are only deemed to be satisfactory when they meet or exceed expectations (e.g., Baker and Crompton 2000). High expectation often leads to low satisfaction and vice versa in the case of a given level of performance. We may predict a contrast between expectation and satisfaction. Alternatively, assimilation theory (Sherif and Hovland 1961) posits that travelers distort their perceptions in the direction of their prior expectations in order to resolve conflicts arising from disconfirmation. Thus we may predict an assimilation between expectation and satisfaction. Interestingly, tourism researchers found that prior expectations have a direct effect on overall evaluations with a leaning towards assimilation rather than contrast, thereby supporting assimilation theory (del Bosque and San Martín 2008).

Third, evaluations that are reliant on hindsight are memory based. This has prompted psychologists to study the bias associated with retrospective recall. It has been found that people tend to exaggerate the intensity of emotions that were experienced previously (Robinson and Clore 2002a). It has also been observed that retrospective reports about emotions are often biased

by respondents' current state of mind (Ross 1989). Respondents can only assess how they feel currently and if they are to reconstruct how they felt in the past, they need to draw upon their beliefs to make inferences. The incidence of memory-based distortion is also well-known in tourism. Since there is a prevailing belief that travel experiences are pleasant, retrospective reports tend to be relatively rosier than online reports (e.g., Mitchell et al. 1997; Wirtz et al. 2003). Another explanation is that any negative affects which are associated with an experience fade more rapidly over time than is applicable to positive affects. Given that traveler revisitation intentions and willingness to recommend are often made after the visit, more reliance may be attached to rosier recollections. Since most research evaluations are elicited ex post hoc, scholars have not taken the trouble to compute any differences between real-time experience and how they are remembered. Based on the preceding discussion, it is relevant and timely to investigate the psychological mechanism which drives how travelers arrive at summary evaluations.

Evaluations of Tourism Experience

In noting the relationship and prospective differences between retrospective and online evaluations, scholars have observed that travelers do not base their summary evaluations on perceived feelings of these experiences, instead constructing a recollection that is heavily dependent on conditions prevalent when the evaluation is prompted (e.g., Fredrickson 1991). Moreover, this construction depends on a rule-based process that integrates the various episodic evaluations which change through the course of the trip (Nawijn et al. 2013). Marketing scholars have examined the dynamics of episodic evaluations in a hospital context as the in-process consumption satisfaction (Dubé and Morgan 1998). It has been theorized that attribute-specific satisfaction is a major input to overall satisfaction (Spreng, Mackenzie and Olshavsky 1996).

The current paper makes particular reference to the psychology literature, which identifies rules that can be categorized under two broad headings: *Movie Models* and *Snapshot Models*.

Movie Models. This line of thinking derives originally from Edgeworth (1881), who proposed the hedonometer as an "ideally perfect instrument, a psychophysical machine, continually registering the height of pleasure experienced by an individual" (Edgeworth 1881, 101). The idea was that it is possible to measure moment-to-moment utility, and that overall evaluations can be formed by integrating the various measures of utility over time. When applied in the contemporary tourism context, researchers have deployed physiological measures such as electrodermal activity in an attempt to proxy continuous real-time experience (e.g., Kim and Fesenmaier 2015). Neuroscientists have made continuous use of magnetic resonance imaging to measure the striatal activity for predicting subjective well-being (Rutledge et al. 2014).

The normative model is called the *Duration-Weighted Average*. Ideally, when this movie-like memory of instant utility is available, the summary evaluation may be obtained by an averaging rule whereby the overall evaluation simply averages the utilities across time. Kahneman, Wakker and Sarin (1997) proposed a normative treatment of total utility and identified a set of axioms that yields a rule of temporal integration. Total utility may be derived by integrating the complete profile of moment-to-moment utility intertemporally. Total utility could also be represented as the area under the curve on a plot of instant utility over time. Due to technological limitations, it is impossible to obtain a continuous profile of instant utility. Instead, the process generates discrete measures from a variety of occasions during the experience. The most achievable outcome is an approximation of total utility by averaging the impressions of the series of events, with each event carrying a weight that is proportional to its duration. This leads to our first hypothesis:

H1a: The overall evaluation is best predicted by a duration-weighted average of the

episodic evaluation.

The less cognitive demanding version of the Movie Model is the *Simple Average* rule. A

direct implication of the averaging rule is that the overall evaluation depends heavily on the

duration of individual experiences. However, it has been found that retrospective evaluations of

experience are often insensitive to the duration of the episode (e.g., Fredrickson and Kahneman

1993). More generally, researchers have shown that the simple average outperforms other

models that assign weights to different factors (Dawes 1979) since simple averaging is less

susceptible to sampling errors (Kahneman 2011). Thus it may be possible to average the episodic

evaluation to reach a summary evaluation, instead of relying on a duration-weighted average. In

this case, summary evaluations are insensitive to the duration of the episodes. This prompts us to

hypothesize that:

H1b: The overall evaluation is best predicted by a simple average of the episodic

evaluation.

Snapshot Models. It has been commonly observed that the duration-weighted averaging

rule fails to describe the relationship between moment-to-moment utility and the overall

evaluation (e.g., Kahneman et al. 1993). Empirical studies have shown that this failure goes

deeper than simple neglect of the duration of individual episodes. There is a more fundamental

misspecification of how we encode and retrieve from our memories. It has been proposed that we

are *cognitive misers* (Fiske and Taylor 1984). In order to economize our time and effort, we often

rely on simpler *heuristics* than more sophisticated and complex processing. This is because our

memory does "not take film, it takes photographs" (Kundera 1991). Our memory only retains the

experiences of key moments, akin to a few discrete snapshots (Fredrickson and Kahneman 1993). It has been argued that less salient events may not even be encoded into memory (Tulving and Craik 2000). Thus, instead of accounting for the whole when summarizing an experience, travelers usually consider only a few of the most readily accessible snapshots. Summary evaluations are constructed by drawing on the experiences at these particular moments, taking a form that follows *snapshot models* (Fredrickson and Kahneman 1993). It has thus been concluded that summary evaluations involve a disproportionate reflection on these prototypical moments (Fredrickson 2000).

Psychologists have identified conditions under which Movie Models or Snapshot Models perform better. According to the accessibility model, our memory is organized into experiential (or episodic) and semantic knowledge (Robinson and Clore 2002a, 2002b). Episodic knowledge is experiential and concerns "conscious recollection of previous experiences of events, happenings, and situations" (Tulving and Markowitsch 1998, 202). Semantic knowledge is conceptual, decontextualized and crucial for the acquisition of factual knowledge. In the shorter term, we rely on experiential knowledge to construct evaluations. Thus the episodic information plays a greater role and Snapshot Models are better for capturing summary evaluations (Chajut et al. 2014). Since episodic memory becomes quickly inaccessible over time (Tulving 1984), we are more reliant on semantic information over the longer term when constructing evaluations. Other factors such as previous knowledge and beliefs also affect the summary evaluations and are better captured by Movie Models. A few studies have supported this temporal difference in the predictive power of aggregation rules (Chajut et al. 2014; Geng et al. 2013; Strijbosch et al. 2019). Geng et al. provided notable process evidence of the temporal shift in the reliance of experiential to semantic knowledge.

The most prominent of the various snapshot models is the peak-end rule (Fredrickson and Kahneman 1993). This posits that impressions of the overall experience may be predicted by formulating the average of two things. The first is the peak experience, namely the best or worst moment when the overall experience tilts toward either positive or negative. The second is the end and final impression. It has been found that the peak-end rule accounts for how people aggregate their experience of a series of events across different domains to form an overall evaluation. Examples include dining experiences (Rode, Rozin and Durlach 2007), perceptions of pain (e.g., Kahneman et al. 1993), consumer reactions to advertisements (Baumgartner, Sujan and Padgett 1997) and gambling experiences (Yu, Lagnado and Charter 2008). In one study, patients undergoing a colonoscopy were asked to rate their momentary pain during the procedure and to provide a retrospective summary evaluation (Redelmeier and Kahneman 1996). It was found that the summary evaluation followed the peak-end rule.

Fredrickson (2000) explained why summary evaluations rely disproportionately on these two snapshots. The peak experience represents "the personal capacity necessary for achieving, enduring, or coping with that episode" (p. 590), whereas the closing experience provides a turning point that facilitates evaluation. As suggested by Fredrickson, closure means that episodes "can be assessed with relative certainty. Endings signal that episodes are completed, safely in the past, and therefore knowable. While experiences are still ongoing, people may hold their global impressions somewhat tentatively" (p.591). Thus respondents keep track of these key episodes when proceeding through an experience. This allows the episodes to become memorable and available for subsequent retrieval. Though tourism researchers have proposed that the peak experience influences the overall evaluation (Mannell and Iso-Ahola 1987), to the authors' best knowledge, no researchers have conducted empirical testing of this hypothesis:

H1c: The overall evaluation is best predicted by the average evaluation of the peak and end episodes.

Significance of the Tourism Experience. It has been found that various behavioral intentions are positively affected by evaluations of the experience. For instance, it has been noted that satisfaction with the tourism experience leads to increased loyalty and enhanced reputation (Baker and Crompton 2000), willingness to pay more (Bigné and Andreu 2004), to recommend (Bigné et al. 2008; Hosany and Prayag 2013) and intentions to revisit (Ross 1993). Since downstream consequences are an important goal of tourism design, it is important and timely to conduct a focused investigation. The proposed investigation can also potentially validate the summary evaluation that is constructed from memories of episodic experiences. The authors are particularly interested in the intentions of travelers to revisit and their willingness to recommend. This leads to the following hypothesis:

H2: The episodic evaluations are integrated temporally to form an overall evaluation which in turn influences traveler revisitation intentions and their willingness to recommend.

Methods

Sampling and Data Collection Procedures

The researchers surveyed visitors at the two ferry terminals in Macau SAR, China and at the Macau International Airport. The study was conducted on 25 days over the period from August to November, 2017. Human research ethics approval was received from the university of the first named author and each participant provided written informed consent. Respondents departing Macau were intercepted randomly and interviewed by trained survey administrators. To ensure that the respondents were surveyed at the end their trip, the first screening question asked whether respondents were leaving Macau at the time of interception. More than 95% of the potential respondents accepted our invitation.

A total of 691 visitors (437 female) were surveyed (response rate of 95%). Three incomplete responses were excluded from the analysis. Among the 633 respondents who reported their age, the average was 32.3 ($\pm$11.3 S.D.). Among the 683 respondents who reported their residence of origin, 138 (20%) were international travelers while the rest (80%) were national (ie Chinese) travelers. On average, respondents stayed 1.72 ($\pm$1.55 S.D.) nights and recalled 3.04 ($\pm$.47 S.D.) events. The relatively short average length of stay is consistent with the small scale of Macau.

Respondents were first asked about the nature of their preceding trip. The questions included whether they stayed overnight and if so for how many nights. Respondents were also asked whether they were about to leave Macau. The researchers confined coverage to tourists aged 18 and above who stayed in Macau for at least one night and were about to leave. Respondents were then asked about the nature of their trip, particularly their purpose of travel,

whether they were Fully Independent Travelers (FITs) or Group Inclusive Travelers (GITs), and who planned the trip. The following section describes their summary trip evaluations and corresponding episodic evaluations. The interviewers collected demographic information towards the end of the survey, such as gender, age, place of residence, marital status, job status, income level and education level.

Order Manipulation. The researchers counterbalanced the order of the two sets of questions to avoid the risk that memory reconstructions might affect the evaluation of overall experiences and vice versa. Participants received one of the two versions randomly, with half of them responding to the DRM first and then answering the questions about their overall experiences. The other half of the participants were asked about their overall experiences and then responded to the DRM.

Day Reconstruction Method

A variety of methods are available to gauge people's momentary utility, either real-time (e.g., Redelmeier and Kahneman 1996) or memory based (e.g., Kahneman et al. 2004a). The current study deployed the latter approach. Specifically, the assessment of respondents' episodic experiences was inspired by the DRM (Kahneman et al. 2004a, 2004b). The method was originally designed to reconstruct events based on respondents' daily activities. Participants are first asked to reconstruct the activities of a given time period and then to evaluate the relevant events. Thus the DRM is retrospective. Accuracy of recall is achieved by requiring respondents to retrieve specific events from their memory of the complete experience. Information about the duration of the events is also gathered.

While yielding similar results to the real-time approach (Kahneman et al. 2004a, 2004b), it has been found that the DRM approach outperforms real-time measurement because of its ease of administration and lesser intrusiveness (Kahneman et al. 2004a). The DRM also enables a more detailed and contiguous record compared with real-time sampling methods, in which participants are asked to respond to the same set of questions repeatedly during the extended travel experience (Chhetri et al. 2004) or to proxy with real-time physiological measures such as electrodermal activity (e.g., Kim and Fesenmaier 2015). Such repeated reminders about making evaluations of the experience during the experience itself may cause the Hawthorne effect (where participants modify their behaviors in response to being aware that they are being observed; Shoval and Ahas 2016).

Survey Instrument

DRM. In the original DRM, respondents were asked to reconstruct their experiences by recalling the events of the preceding day. They completed a diary that listed the nature, timing, durations and venues of the various activities in chronological order. The current authors have adapted the DRM by surveying at the end of the trip instead of daily. Respondents were asked to recall as many episodes as possible and to provide a simultaneous rating of their satisfaction for each activity from 1 (not at all satisfied) to 7 (very satisfied). They were also asked the day, time and duration of each episode.

Explanatory Variables. Respondents were asked to use DRM to reconstruct their memories of the whole trip, either prior to or after making their summary evaluations. They were also asked to evaluate each episode that they were able to recall. These episodic evaluations allowed the researchers to construct several explanatory variables. First, since the average

traveling was reported as being positive ($M = 5.57$ out of 7, S.D. = 1.03), the peak experience is equated with the most positive episode. The researchers arranged the episodes that were recalled in chronological order with the most recent being treated as the end episode. To test the peak-end rule which hypothesizes that the overall evaluation is best predicted using the average of the most extreme and most recent experiences, the researchers took an arithmetic average of the most positive episode and the end episode. Second, the researchers also considered the worst episode during the whole experience. The overall experience was negative in many of the previously reported summary evaluations, e.g., pain in a colonoscopy (Redelmeier and Kahneman 1996). In such cases, the most extreme episode was the trough. Since it was unclear whether the trough is predictive in such cases because a) it is the extreme experience or b) simply because it is the worst experience, it was included in the present study, where it was anticipated that the overall tourism experience would be positive. Third, participants reported the duration of each episode in hours. This facilitated the calculation of the duration-weighted average using the following equation:

$$\frac{\sum_{i=1}^{n}(duration_i \times evaluation_i)}{\sum_{i=1}^{n}(duration_i)},$$

where there are n events recalled by a particular respondent, with each $event_i$ receiving an $evaluation_i$ and lasting for $duration_i$. Fourth, the last explanatory variable is obtained by taking an arithmetic average of the episodic evaluations as reported by a particular respondent.

Dependent Variable. The summary evaluation of the whole Macau trip was measured by asking travelers to answer the question: "What is your satisfaction level for this trip in Macau?" on a scale of 1 (very dissatisfied) to 7 (very satisfied). The two behavioral intentions — revisitation intention and willingness to recommend — were measured by the questions "How

likely is it that you will return to Macau in the next year?" and "How likely will you be to recommend Macau to others as a tourist destination?" on a scale of 1 (definitely will not) to 5 (definitely will).

Covariates. A number of covariates were included in the analyses. These may have a direct or indirect influence on the overall evaluation by interacting with the explanatory variables. For instance, visitors who plan their own travel arrangements may be more sensitive and pay greater attention to individual events during their trips, compared with those who are not involved in planning the trip. The latter may pay less attention and are thus more reliant on key events to reconstruct their overall evaluations. There is a prospect that simple heuristics such as the peak-end rule could be better predictors of the overall evaluation of the experiences amongst the latter types of traveler. Similarly, FITs may differ from GIT travelers who join package tours. Whether the purpose of travel is leisure or business may also play a role. Lastly, the length of the trip proxy by the number of nights spent may also influence how travelers make overall evaluations.

Analysis

A series of t-tests was conducted to check whether the covariates confound with the demographic variables (i.e., age and gender), and to determine whether age differs amongst the subsamples grouped around the three covariates, i.e., travel type, trip planning and purpose of the trip. None of these were found to be significant. Similarly, the authors ran three logit regressions with gender as the dependent variable and with the same three covariates as independent variables separately. Again, none were found to be significant. In addition, a t-test was conducted

to determine whether the summary evaluation is influenced by the order of its elicitation and/or by the DRM. T-test analysis showed that no significant difference was found between the conditions of summary evaluation first ($M = 5.52$) and the DRM first ($M = 5.62$, $t = 1.19$, $p > .2$). This prompted a pooling of the two order conditions for the subsequent analyses.

In order to test the different aggregation rules, we followed the previous practice of testing the peak-end rule using a few statistical analyses to identify the independent variable that best predicts the summary evaluation (Chajut et al. 2014). These include comparing the Pearson correlation between the independent variables and the summary evaluation using the Steiger's z test, a hierarchical regression and a stepwise regression. In particular, Pearson correlations of the summary evaluation with the explanatory variables were conducted at $p < .001$. The authors then used a hierarchical regression to test the relationships between the summary evaluation and the explanatory variables in an attempt to replicate the finding by including the different explanatory variables in a single analysis and with a few control variables. The predictive power of the few explanatory variables was further examined using stepwise regression. In this analysis, the summary evaluation was predicted by the *simple average, peak-end*, *trough* and *duration-weighted average*. The analysis was implemented using the stepwise function in the linear regression model of SPSS. This function was based on the *p*-value of the *F* of the regression. Both the forward and backward functions are used in the stepwise regression analysis.

Three different methods were used in the analyses since each has advantages and disadvantages which are mutually complementary. First, the comparison of Pearson correlation is independent of any model specification, though important variables may be neglected. Second, the hierarchical regression complements the Pearson correlation by including covariates in the analysis. It thus allows us to check the robustness of the predictive power of any particular

explanatory variables against different combinations of covariates. This is especially crucial when there is a chance that important covariates may be omitted (e.g., the order of elicitation) leading to biased estimation. Third, the stepwise regression model provides an automatic procedure to select the best predictor from a set of potential explanatory variables using predefined criteria. However, examining the explanatory variables in a stepwise regression is often not theoretical driven. This is a lesser problem in the present case since all explanatory entering the analysis are motivated theoretically. Another problem with stepwise regression is that there is no guarantee that the best predictor is unique. It may end up with multiple explanatory variables, thereby impeding the formulation of a definitive conclusion.

To examine the downstream consequences of the overall evaluation, the authors further ran a mediation model (Preacher, Rucker and Hayes 2007) with 10,000 bootstrapped samples using the SPSS PROCESS macro (Hayes 2013; Model 4). The simple average was used as the independent variable, with behavioral intentions (revisitation intention and willingness to recommend in two separate models) as the dependent variable, and the summary evaluation as the mediator.

Results

The respondents expressed general satisfaction with their visit, as their summary evaluations were quite high (M = 5.57 out of 7; Table 1). This favorable attitude is also reflected in the explanatory variables such as peak-end, weighted average and simple average. As expected, the average of peak and end (M = 5.90) is higher than both weighted average (M = 5.57) and simple average (M = 5.47). Similarly, the intention to revisit and willingness to recommend average 3.89 and 4.15 (out of 5), respectively. These suggest that the favorable attitude of the trip successfully translated into positive behavioral intentions. With the exception of the business travelers, all the variables and demographics are invariable with the few subsample categories. Although it did not reach statistical significance, the business travelers tended to be older and stayed longer. They also tended to have a higher revisitation intention.

--Insert Table 1 here--

Pearson Correlations

The results of Pearson tests indicated that all variables were significantly correlated (Table 2). Among the explanatory variables, the *simple average* had the highest correlation with summary evaluation (r = .3457). These correlations were then compared using the Steiger's z test. It was found that the *simple average* yields significantly higher correlations with summary evaluation than all of the other explanatory variables at $p < .05$ (i.e., Steiger's z's ≥ 1.70). Thus H1b is supported over H1a and H1c. Since *simple average* was found to be the best predictor of the summary evaluation, the authors further tested whether this correlation depends on the order

of elicitation. When the DRM was elicited first, the *simple average*–summary evaluation correlation was found to be .3374; whereas the correlation was .3506 when the summary evaluation was elicited first. A Steiger's z test showed no significant difference between the two correlations ($z = .195$, $p > .4$).

--Insert Table 2 here--

Hierarchical Regression

The correlation analyses indicate that the *simple average* is the best predictor of the summary evaluation. A hierarchical regression is conducted with the inclusion of a few control variables. In particular, the *peak-end* and *trough* were included as explanatory variables in Step 2, in addition to *simple average*. All of the three variables are mean-centered. Since the *duration-weighted average* was highly significant with *simple average* ($r = .91$), it was excluded from the analysis. Five covariates also entered the model in Step 3, including the *number of nights.* The *order* of elicitation was incorporated as a binary variable (0 = summary evaluation first, 1 = DRM first), whether traveling in a *group* (1 = GIT, 0 = FIT), whether the respondent *planned* the trip (= 1) or not (= 0) and the trip *purpose* (leisure = 0, business = 1). Lastly, two demographic variables, namely *gender* and *age*, entered in Step 4.

Results of the hierarchical regression confirm that *simple average* is the best indicator, and the elicitation order does not have a statistically significant effect on the relationship (Table 3). Again, H1b is supported. None of the other explanatory, control and demographic variables

reached significance at the .10 level. Variance inflation factors (VIFs) for our explanatory variables suggested that multicollinearity was unlikely (VIFs $\leq$ 10.86; Myers, 1986).

--Insert Table 3 here--

Stepwise Regression

The results of the stepwise regression confirm *simple average* as the best indicator. Analyses using both the forward and backward functions yielded the same result. Only the *simple average* entered the model. Other variables including *simple average, peak-end*, *trough* and *duration-weighted average* accounted for no additional variance in *overall evaluation*. The current analysis gives further converging support to H1b.

Behavioral Intentions

Next, the authors examined the behavioral consequences of traveler satisfaction. It was found that summary evaluation was highly significantly correlated with both revisitation intention ($r = .20, p < .001$) and willingness to recommend ($r = .38, p < .001$). Similarly, *simple average* was highly significantly correlated with both revisitation intention ($r = .12, p < .002$) and willingness to recommend ($r = .25, p < .001$). Conceptually, the summary evaluation is constructed using *simple average,* and the behavioral intentions are downstream consequences. In this section, the authors undertook further testing of the mediation relationship.

Mediation analyses are conducted to test this conceptual model (see Figure 1). The direct effect of the simple average was not significant on revisitation intention (.066, S.E. = .042; 95% CIs, -.017 to .148) and was significant on willingness to recommend (.129, S.E. = .036; 95% CIs, .060 to .199). The summary evaluation had a significant effect on revisitation intention (.170, S.E. = .039; 95% CIs, .094 to .247) and willingness to recommend (.292, S.E. = .0331; 95% CIs, .227 to .357). The indirect effect of the simple average through summary evaluation was significant on revisitation intention (.063, S.E. = .020; 99% CIs, .018 to .12) and willingness to recommend (.109, S.E. = .022; 90% CIs, .059 to .174). H2 is supported.

--insert Figure 1 here--

Discussion

In this paper the authors have explored different aggregation rules about how travelers construct summary evaluations, with downstream consequences for their behavioral intentions. While various previous studies have found that the peak-end rule outperformed alternatives such as the duration-weighted average rule, the current study uncovered a surprising finding — that travelers do not follow a peak-end rule. Instead, an arithmetic average of all episodic evaluations was found to provide a better prediction of the summary evaluation. Alternative characterizations of the experience, such as peak and end episodes, do not add significant explanatory power to the summary evaluation. A number of possible factors may explain the finding about the superiority of the simple-average rule in the current context.

First, the literature suggests two possibilities: 1) the accessibility model (Robin and Clore 2002a, 2002b), and 2) the complexity of the experience (Strijbosch et al. 2019). As was discussed previously, the former refers to the observation that summary evaluations are better captured by Snapshot Models such as Peak-End rule in the beginning and gradually shift to Movie Models such as Averaging rule over subsequent weeks. Since the relevant summary evaluation was elicited immediately after the trip, it is unlikely that the reliance of semantic information is responsible for the performance of the average rule. The latter explanation is more plausible. It has been argued that most previous findings supporting the peak-end rule came from relatively simple, short and one-dimensional experiences (Strijbosch et al. 2019). The authors found that in more complex and heterogenous experiences delivered using virtual reality, people's overall evaluation was better predicted by the average rule than the peak-end rule. With reference to the current case, travel differs from most other experiences that have been investigated previously, being typically more prolonged and complex, as well as relatively rare

and infrequent (Mannell and Iso-Ahola 1987). Travelers often visit multiple attractions that offer diverse experiences, thereby prompting the overall evaluation to draw upon most episodes. The simple-average rule appears to work best in capturing this comprehensiveness in their overall evaluations.

Second, the presence of duration neglect is reasonable and has the support of many previous findings. Indeed, the duration-weighted average rule according to which only relative and not absolute duration matters, is not as reasonable as was predicted. Individual travelers visit attractions for differing durations which are often an endogenous variable in the sense that those who are most engaged by the attraction will "decide" to linger. Thus when constructing an overall evaluation in a survey or making recommendations to friends, they understand that it could be misleading to consider the applicable durations in forming their evaluations. Third, it has been observed that little predictive power arises from models (termed *proper linear models*) with weightings assigned to different factors. Meanwhile simpler models (termed *improper linear models*) that weight each factor equally, such as the simple-average rule, are often superior (Dawes 1979). Fourth, the present study has adopted a methodology that differs from what prevails in the literature. While the current investigation has used DRM to reconstruct traveler memories and to evaluate the experience in each episode, most previous findings that support the peak-end rule undertook online assessments of participant experiences (e.g., Redelmeier and Kahneman 1996). In support of our argument, the duration-weighted average is favored by two reports using DRM (Miron-Shatz 2009) and the simple-average rule (Strijbosch et al. 2019) over the peak-end rule. The former study did not test the simple-average rule.

In support of the simple-average rule, it may be that the DRM helps travelers reconstruct their memories of each individual episode. This is potentially an artefact of the choice of

elicitation method, rather than a reflection of any difference in psychological process through the construction of summary evaluations. To address this potential issue, the authors randomly assigned half of the participants to respond to the summary evaluation prior to the DRM questions in the study design, while the other half of the respondents first went through the DRM. There were no differences between the two groups in the performance of the simple-average rule. Hence the methodological artefact hypothesis is not supported.

It is suggested that future researchers should address whether more trust can be assigned to online or retrospective evaluations. The somewhat clichéd answer is that "it depends". Researchers have found that remembered experiences differ from online evaluations of the same incident/s (Mitchell et al. 1997; Wirtz et al. 2003). Specifically, remembered experiences are often rosier. On this basis, retrospective evaluations are commonly more positive. Thus by design, the two types of evaluation will deliver different results. The suitability of using online or retrospective evaluations will depend on the intended purpose. For instance, when destination marketers develop and maintain attractions, the online experience is often the primary focus. In practice, travelers rely on remembered experiences when constructing their evaluations and making decisions about future events such as revisiting the destination. The findings of the simple-average rule appear to be most applicable.

Contributions to the Literature

The superiority of the simple-average rule, and the inferior predictive power of the peak-end rule may come as a surprise to psychologists working on the summary evaluation of extended experiences. The current paper is the first to identify the use of the simple-average rule

when individuals are summarizing their travel experiences. This highlights two important aspects of their evaluation process. First, the simple-average rule is a special case of the duration-weighted average rule in which every episode is considered to have an identical duration. The direct implication is that travelers do not consider the relative durations of the various episodes. This is consistent with most findings about duration neglect (Fredrickson and Kahneman 1993).

Second, previous researchers have observed that decisions and judgments frequently rely on mental shortcuts or heuristics. If the underlying psychological processes are to be understood and explained, it is important to understand the nature of such shortcuts, including when travel is evaluated. The simple-average rule averages over all episodic evaluations, unlike other rules that account for only a small subset of the whole experience. This is noteworthy in that travelers make maximum use of all information about individual episodes to arrive at retrospective summary evaluations. In this sense, using the simple-average rule is more effortful than the peak-end rule. Interestingly, this more deliberate approach to processing was found to be commonplace amongst those who undertook trip planning, as well as those who relied on others to plan.

Methodological Implications

Some methodological implications arise from the study. First, the authors have questioned how respondents arrived at their evaluations in the case of previous studies that involved retrospective assessments. The present findings suggest that the reported outcomes were probably influenced by all episodes of respondent travel experiences, neglecting the respective durations. Second, tourism scholars should elicit summary evaluations carefully when

designing future studies. For example, episodic evaluations provide researchers with detailed breakdowns and also allow respondents to reconstruct their memories, thereby reaching more accurate overall evaluations. Since switching the order of the diary and global evaluations has yielded a null result, it may be concluded that responses were unaffected by the use of diary reconstructions.

Third, researchers should consider timing. As previous researchers have suggested, retrospective evaluations are often rosier than their real-time equivalents (Wirtz et al. 2003). More importantly, when the current findings are compared with those of previous studies that relied on online evaluations, it may be observed that when both episodic and summary evaluations are elicited retrospectively, the relationship between the two tends to follow averaging rules. In contrast, the relationship between retrospective summary evaluations and online episodic evaluations tends to follow the peak-end rule (e.g., Kahneman et al. 1993). This distinction is particularly important now, because researchers can analyze real-time traveler experiences using GPS devices to geotag the real-time evaluations (e.g., Pettersson and Zillinger 2011), and physiological measures such as skin conductance responses (e.g., Shoval, Schvimer and Tamir 2018). As the available methods grow in sophistication, it is important for researchers to understand the relationship between these measures of episodic experience and overall evaluations of the extended experience.

Practical Implications

The current findings also shed light on tourist behaviors and may inform destination and attraction managers about the design of tourism facilities. If it were the case that travelers are

preoccupied with peak and end experiences, practitioners could emphasize their targeting at these moments. This approach might generate a disproportionate return on investment at departure ports by improving the traveler experiences when exiting the destination. Singapore, for example, has been recognized for its airport reinvestment that has enhanced the visitor experience by featuring the world's tallest indoor waterfall. This is especially useful since travelers often arrive early when they are departing the destination and spend significant time at the airport. Another response is the potential staging of more special events to ensure that visitors experience at least one such signature event during their stay and thereby enhance their peak experience. However, the current findings suggest that this strategy might be less fruitful than anticipated. The simple-average rule implies that there is no such shortcut. Tourism planners should therefore note that improving the peak and end experiences of their visits might help, though only to the extent that it enhances the overall experience when averaged during the whole visit over many episodes.

The current findings also suggest that the duration of various episodes has been neglected. Destination marketers should not devote disproportionate energies to prolonging the time spent by travelers at particular sites. In an era when overtourism is increasingly prevalent and where some sites are seriously overcrowded at peak times, it may be worth encouraging visits that are long enough to enjoy the site, but short enough to make space for others. In particular, tourism planners might reformulate the provision of supporting facilities in different attractions by considering the duration of the travelers' tendency to linger.

Neglecting duration may be reflected in the downstream behavioral consequences of satisfaction. Specifically, the authors have observed that the simple average has an indirect effect on both willingness to recommend and intentions to revisit. Interestingly, these relationships

were stronger in the case of willingness to recommend than for revisitation intentions. Making a recommendation to a friend might be more straightforward than a revisitation decision, which involves many contributing factors. It has been argued that taking account of duration in a recommendation may indeed be misleading since the duration may vary substantially among travelers (Ariely and Loewenstein 2000). This is especially important since most of these behavioral intentions happen after time has elapsed post-trip. As suggested by the accessibility model, the retrospective evaluation is better captured by the average rule after the time lapse. Thus this indirect effect of the simple average on behavioral intentions is expected to grow over time.

Limitations and Future Research

The study has a number of noteworthy limitations. The paper has examined how travelers arrive at an overall evaluation of their trip, rather than comparing online and retrospective evaluations. The decision to deploy DRM was a methodological choice. Our conclusions were drawn from retrospective assessment, rather than online assessments of individual episodes. As has been previously documented, this reliance on the reconstruction of activities can be problematic (e.g., Robinson and Clore 2002b), especially if it is not done daily (Mitas et al. 2012) since the experience can often be "vividly recalled but difficult to describe" (Arnould and Price 1993, 42). As noted by Mitas and Bastiaansen (2018), there is always a tradeoff between recall error and disrupting the experience. Thus a daily diary strategy may be more appropriate (Gao et al. 2019; Fredrickson 2000; Nawijn et al. 2013). An interesting question for future researchers will be whether this overall assessment has a different relationship with evaluations of individual events during the trip when such events are assessed online, versus when they are

assessed retrospectively. Furthermore, it may be worthwhile measuring tourist experiences using similar methods such as ESM, diary methods (Mitas and Bastiaansen 2018) or even cognitive neuroscience (Bastiaansen et al. 2019). There are potentially important methodological and theoretical implications if these methods are found to yield different results.

The study was conducted in a relatively small and confined, though highly visited destination. It is possible that the simple average may be favored because travelers did not engage in many activities. Future researchers may examine how the aggregation rules perform as the length of the trip increases. There may also be a cultural difference in the process of retrospective assessment. Our findings of the superior performance of the simple average over the peak-end rule may be unique to the sample. Amongst the 691 respondents in the current sample, only 60 were from countries outside Asia, whereas previous findings have mostly been obtained from North America. It will be interesting to study this possible cultural difference in the process of summary evaluation of extended experience. Another characteristic of our sample is that 63.2% are female. Thus we cannot write off the possibility that gender drives the superior performance of the simple average rule. As has been reported though, gender does not covary with any of the four subsample categories and there is no reason to believe that gender has interacted with other covariates to influence our results. Lastly, two behavioral intentions were measured instead of actual travel behaviors as a consequence of the travelers' satisfaction with their overall trip.

Conclusions

In this paper, the authors have reached a surprising result. Unlike most previous studies which have suggested that the peak-end rule guides summary evaluations of extended experience, the current authors conclude that the global evaluations by travelers are best predicted by a simple averaging of the episodic evaluations. Other characterizations of the travel experience, such as the peak and end episodes, do not add significant explanatory power to the summary evaluation. This may be a result of the more extended and complex nature of travel compared with other experiences that have previously been studied. The immediate implications are that: 1) most of the episodes in one's travel experience affect the summary evaluation, and 2) relative duration of individual episodes is not considered. These findings have important methodological implications and can potentially guide researchers to design better studies and to re-interpret previous findings. They have practical implications when destination marketers rely on traveler feedback when engaging in tourism design and engineering travel experiences.

References

Arnould, E. J., & Price, L. L., 1993. "River magic: Extraordinary experience and the extended service encounter." *Journal of consumer Research*, *20* (1): 24-45.

Baker, D.A., and Crompton, J.L., 2000. "Quality, satisfaction and behavioral intentions." *Annals of tourism research*, 27 (3): 785-804.

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

Bastiaansen, M., Lub, X. D., Mitas, O., Jung, T. H., Ascenção, M. P., Han, D. I., ... & Strijbosch, W., 2019. "Emotions as core building blocks of an experience." *International Journal of Contemporary Hospitality Management*, 31 (2): 651-668.

Bigné, J.E., and Andreu, L., 2004. "Emotions in segmentation: An empirical study." *Annals of Tourism Research*, 31 (3): 682-696.

Bigné, J.E., Mattila, A.S., and Andreu, L., 2008. "The impact of experiential consumption cognitions and emotions on behavioral intentions." *Journal of Services Marketing*, 22 (4): 303-315.

Chajut, E., Caspi, A., Chen, R., Hod, M. and Ariely, D., 2014. "In pain thou shalt bring forth children: the peak-and-end rule in recall of labor pain." *Psychological science*, 25 (12): 2266-2271.

Chhetri, P., Arrowsmith, C., and Jackson, M., 2004. "Determining hiking experiences in nature-based tourist destinations." *Tourism management*, 25 (1): 31-43.

Dawes, R.M., 1979. "The robust beauty of improper linear models in decision making." *American psychologist*, 34 (7): 571.

del Bosque, I.R., and San Martín, H., 2008. "Tourist satisfaction a cognitive-affective model." *Annals of tourism research*, 35 (2): 551-573.

Dubé, L., & Morgan, M. S., 1998. "Capturing the dynamics of in-process consumption emotions and satisfaction in extended service transactions." *International Journal of Research in Marketing*, *15* (4): 309-320.

Edgeworth, F.Y. 1881. *Mathematical psychics: An essay on the application of mathematics to the moral sciences* (Vol. 10). Kegan Paul.

Fiske, S. T. and Taylor, S. E. 1984. *Social cognition*. 1st ed.. Reading, MA: Addison-Wesley.

Fredrickson, B. L. 1991. "Anticipated endings: An explanation for selective social interaction." *Dissertation Abstracts International, 51*(12-B, Pt 1): 6150.

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

Fredrickson, B.L. and Kahneman, D., 1993. "Duration neglect in retrospective evaluations of affective episodes." *Journal of personality and social psychology*, 65 (1): 45.

Gao, J., Zhang, Y., Kerstetter, D. L., & Shields, S., 2019. "Understanding changes in tourists' use of emotion regulation strategies in a vacation context." *Journal of Travel Research*, *58* (7): 1088-1104.

Geng, X., Chen, Z., Lam, W. and Zheng, Q., 2013. "Hedonic evaluation over short and long retention intervals: The mechanism of the peak–end rule." *Journal of Behavioral Decision Making*, *26*(3): 225-236.

Hayes, A.F., 2013. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Hosany, S., 2012. "Appraisal determinants of tourist emotional responses." *Journal of travel Research*, 51 (3): 303-314.

Hosany, S. and Prayag, G., 2013. "Patterns of tourists' emotional responses, satisfaction, and intention to recommend." *Journal of Business Research*, 66 (6): 730-737.

Kahneman, D. 2011. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., Fredrickson, B.L., Schreiber, C.A., and Redelmeier, D.A., 1993. "When more pain is preferred to less: Adding a better end." *Psychological science*, 4 (6): 401-405.

Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., and Stone, A.A., 2004. "A survey method for characterizing daily life experience: The day reconstruction method." *Science*, 306 (5702): 1776-1780.

Kahneman, D., Krueger, A.B., Schkade, D., Schwarz, N., and Stone, A., 2004. "Toward national well-being accounts." *American Economic Review*, 94 (2): 429-434.

Kahneman, D., Wakker, P.P. and Sarin, R., 1997. "Back to Bentham? Explorations of experienced utility." *The Quarterly Journal of Economics*, 112 (2): 375-406.

Kemp, S., Burt, C.D., and Furneaux, L., 2008. "A test of the peak-end rule with extended autobiographical events." *Memory & Cognition*, 36 (1): 132-138.

Kim, J., and Fesenmaier, D.R., 2015. "Measuring emotions in real time: Implications for tourism experience design." *Journal of Travel Research*, 54 (4): 419-429.

Kim, J.J., and Fesenmaier, D.R., 2017. "Tourism experience and tourism design." In *Design science in tourism* edited by D. R Fesenmaier and X. Zheng, 17-29. Springer, Cham.

Kundera, M. 1991. *Immortality*. New York: Grove Press.

Mannell, R.C., and Iso-Ahola, S.E., 1987. "Psychological nature of leisure and tourism experience." *Annals of tourism research*, 14 (3): 314-331.

Miron-Shatz, T., 2009. "Evaluating multiepisode events: Boundary conditions for the peak-end rule." *Emotion*, 9 (2): 206.

Mitas, O., & Bastiaansen, M., 2018. "Novelty: A mechanism of tourists' enjoyment." *Annals of Tourism Research*, *72*: 98-108.

Mitas, O., Yarnal, C., Adams, R., & Ram, N., 2012. "Taking a "peak" at leisure travelers' positive emotions." *Leisure Sciences*, *34* (2): 115-135.

Mitchell, T. R., Thompson, L., Peterson, E., & Cronk, R., 1997. "Temporal adjustments in the evaluation of events: The "rosy view"." *Journal of experimental social psychology*, *33*(4): 421-448.

Nawijn, J., Mitas, O., Lin, Y., & Kerstetter, D., 2013. "How do we feel on vacation? A closer look at how emotions change over the course of a trip." *Journal of Travel Research*, *52* (2): 265-274.

Otto, J. E., and Ritchie, J. B., 1995. "Exploring the quality of the service experience: A theoretical and empirical analysis." *Advances in services marketing and management*, *4* (1): 37-61.

Pettersson, R., and Zillinger, M., 2011. "Time and space in event behaviour: Tracking visitors by GPS." *Tourism Geographies*, 13 (1): 1-20.

Preacher, K.J., and Kelley, K., 2011. "Effect size measures for mediation models: quantitative strategies for communicating indirect effects." *Psychological methods*, 16 (2): 93.

Preacher, K.J., Rucker, D.D., and Hayes, A.F., 2007. "Addressing moderated mediation hypotheses: Theory, methods, and prescriptions." *Multivariate behavioral research*, 42 (1): 185-227.

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

Robinson, M. D., and Clore, G. L., 2002a. "Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes." *Journal of personality and social psychology*, *83* (1): 198.

Robinson, M. D., and Clore, G. L., 2002b. "Belief and feeling: evidence for an accessibility model of emotional self-report." *Psychological bulletin*, *128* (6): 934.

Rode, E., Rozin, P., and Durlach, P., 2007. "Experienced and remembered pleasure for meals: Duration neglect but minimal peak, end (recency) or primacy effects." *Appetite*, 49 (1): 18-29.

Ross, G.F., 1993. "Destination evaluation and vacation preferences." *Annals of Tourism Research*, 20 (3): 477-489.

Ross, M., 1989. "Relation of implicit theories to the construction of personal histories." *Psychological review*, *96*(2): 341.

Rutledge, R.B., Skandali, N., Dayan, P., and Dolan, R.J., 2014. "A computational and neural model of momentary subjective well-being." *Proceedings of the National Academy of Sciences*, 111 (33): 12252-12257.

Scuttari, A., and Pechlaner, H. 2017. "Emotions in tourism: From consumer behavior to destination management." In *Design Science in Tourism*, edited by D. R Fesenmaier and X. Zheng, 41-53. Springer, Cham.

Sherif, M., and Hovland, C. I., 1961. *Social judgment: Assimilation and contrast effects in communication and attitude change*. Oxford, England: Yale University Press.

Shoval, N., and Ahas, R., 2016. "The use of tracking technologies in tourism research: the first decade." *Tourism Geographies*, 18 (5): 587-606.

Shoval, N., Schvimer, Y., and Tamir, M., 2018. "Real-time measurement of tourists' objective and subjective emotions in time and space." *Journal of Travel Research*, 57 (1): 3-16.

Spreng, R. A., MacKenzie, S. B., and Olshavsky, R. W., 1996. "A reexamination of the determinants of consumer satisfaction." *Journal of marketing*, *60* (3): 15-32.

Tulving, E., 1984. "Precis of elements of episodic memory." *Behavioral and Brain Sciences*, *7*(2): 223-238.

Tulving, E., and Craik, F. I. (Eds.), 2000. *The Oxford handbook of memory*. Oxford University Press.

Tulving, E., and Markowitsch, H. J., 1998. "Episodic and declarative memory: role of the hippocampus." *Hippocampus*, *8* (3): 198-204.

Viner, J., 1925. "The utility concept in value theory and its critics." *Journal of Political Economy*, *33* (6): 638-659.

Wirtz, D., Kruger, J., Scollon, C.N., and Diener, E., 2003. "What to do on spring break? The role of predicted, on-line, and remembered experience in future choice." *Psychological Science*, 14 (5): 520-524.

Yu, E.C., Lagnado, D.A., and Charter, N., 2008. "Retrospective evaluations of gambling wins: evidence for a 'peak-end' rule." *Proceedings of the Cognitive Science Society*, 30 (30): 77-82.

Table 1

	Evaluation order		Travel type		Trip planning		Purpose of the trip		Overall
	Overall first	Diary first	FIT	GIT	Own	Others	Leisure	Business	
N =	363	325	423	265	324	364	653	35	688
Overall evaluation	5.53	5.62	5.57	5.57	5.63	5.52	5.56	5.71	5.57
	(1.06)	(.98)	(.95)	(1.14)	(1.05)	(1.00)	(1.04)	(.86)	(1.03)
Explanatory variables									
Peak-End	5.92	5.88	5.90	5.90	5.90	5.89	5.90	5.90	5.90
	(.92)	(.77)	(.80)	(.94)	(.91)	(.80)	(.85)	(.92)	(.85)
Trough	4.59	4.63	4.65	4.53	4.70	4.52	4.60	4.77	4.61
	(1.54)	(1.50)	(1.42)	(1.68)	(1.50)	(1.54)	(1.52)	(1.63)	(1.52)
Weighted average	5.54	5.60	5.58	5.54	5.57	5.57	5.56	5.68	5.57
	(.98)	(.91)	(.86)	(1.07)	(1.03)	(.87)	(.95)	(.98)	(.95)
Simple average	5.45	5.50	5.50	5.43	5.51	5.43	5.46	5.62	5.47
	(1.00)	(.90)	(.87)	(1.08)	(1.01)	(.91)	(.96)	(.91)	(.96)
Behavioral intention									
Revisitation	3.90	3.88	3.83	3.98	3.88	3.90	3.87	4.26	3.89
	(1.00)	(1.01)	(.98)	(1.03)	(1.03)	(.98)	(1.00)	(138)	(1.00)
Recommend	4.12	4.19	4.09	4.27	4.23	4.09	4.15	4.17	4.15
	(.94)	(.87)	(.86)	(.97)	(.88)	(.93)	(.90)	(1.01)	(.91)
Demographics									
% female	65.6%	60.9%	64.3%	61.9%	63.0%	63.7%	63.4%	62.9%	63.4%
Age^	32.87	31.60	31.52	33.49	32.09	32.44	31.81	41.83	32.28
	(11.11)	(11.49)	(10.70)	(12.12)	(12.02)	(10.63)	(11.03)	(12.56)	(11.3)
Number of nights	1.82	1.62	1.70	1.75	1.89	1.57	1.67	2.71	1.72
	(1.79)	(1.24)	(1.53)	(1.59)	(1.85)	(1.22)	(1.54)	(1.51)	(1.55)

Table 1. Summary statistics. The standard deviations are in brackets. ^ Only based on the 633 respondents who reported their age.

Table 2

	1	2	3	4	5	6
1. Overall evaluation	-					
2. Peak-End	.2583***					
3. Trough	.2942***	.5198***				
4. Simple average	.3457***	.7731***	.8765***			
5. Weighted average	.3201***	.7557***	.7393***	.9127***		
6. Revisitation	.1957***	.1052*	.1082**	.1225**	.1438***	
7. Recommendation	.3774***	.1865***	.1961***	.2506***	.2504***	.3039***

Table 2. Correlation matrix. (*$p < .01$, **$p < .005$, ***$p < .001$)

Table 3

	Step 1		Step 2		Step 3		Step 4	
	B	SE	B	SE	B	SE	B	SE
Simple average	.372***	.039	.482***	.126	.464***	.127	.460***	.127
Other explanatory variables								
Peak-end			-.061	.079	-.048	.080	-.048	.080
Trough			-.049	.059	-.044	.059	-.034	.060
Control variables								
Number of nights					.021	.024	.021	.024
Order					.071	.074	.074	.075
Group					.003	.080	-.002	.081
Plan					-.072	.079	-.076	.079
Purpose					.036	.173	.026	.174
Demographics								
Age							.009	.013
Gender							-.024	.077
Intercept	5.568***	.037	5.568***	.037	5.534***	.092	5.496***	.133
R^2	.119		.121		.124		.125	
ΔR^2			.001		.004		.001	
F	93.101***		31.263***		12.043***		9.678***	

Table 3. Hierarchical regression analyses of predictors of summary evaluation. ($*p < .05$, $**p < .005$, $***p < .001$)
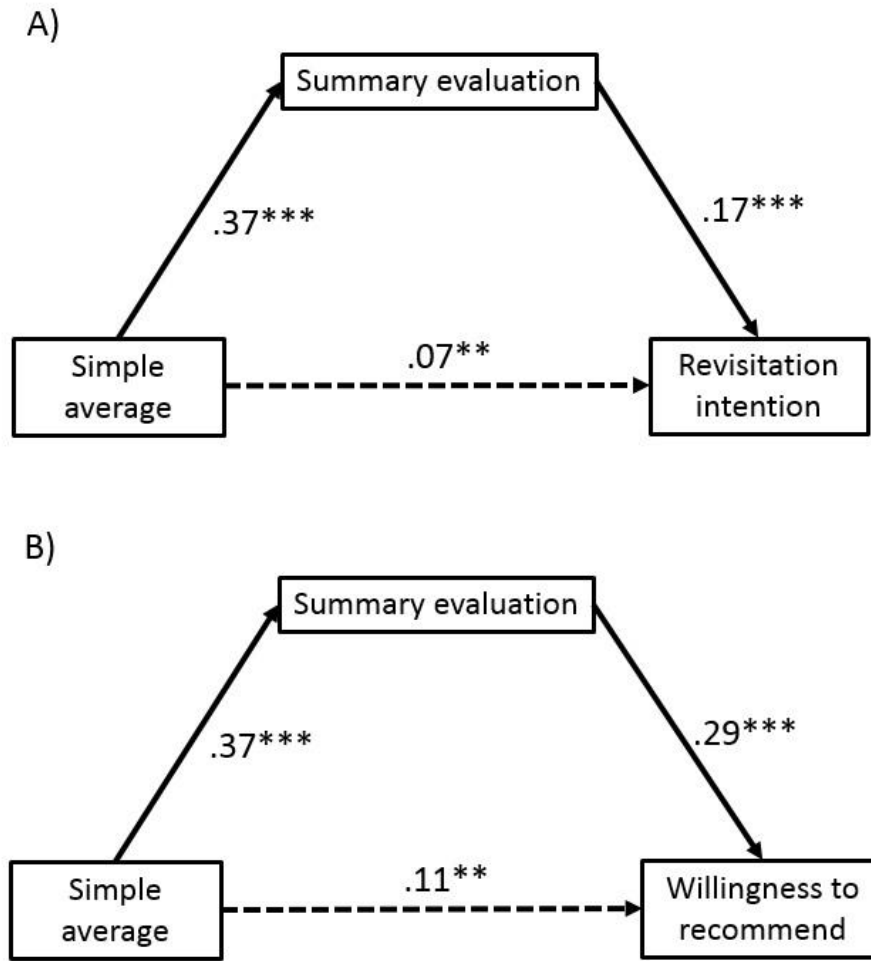
Figure 1

A)

B)

Figure 1.   The effects of simple average on behavioral intentions (A. revisitation intention;

B. willingness to recommend) is mediated by summary evaluation. (Notes:

Numbers are beta values; *p < .05, **p < .01, ***p < .001)