

Evaluating Multiepisode Events: Boundary Conditions for the Peak-End Rule

Talya Miron-Shatz
Princeton University

This study advances our understanding of how people arrive at retrospective evaluations of multiepisode experiences. Large samples from the United States, France, and Denmark (810, 820, and 805 participants, respectively) reported their feelings during each episode of the previous day using the Day Reconstruction Method. The duration-weighted average of these feelings represented the normative approach to evaluation, and, contrary to the predictions of the peak-end rule, the average was the best predictor of retrospective evaluations of the day. To capture participants' heuristic evaluation, they also reported having a wonderful (peak) and/or awful (low) moment during the previous day. The results indicate that retrospective evaluations of multiepisode events rely on the averaged ratings of emotions, ignore ends, and also consider the presence of lows, and occasionally peaks, as subjectively defined by those experiencing them. Peaks and lows contribute more to comparative, rather than absolute evaluations. Future research should examine whether these findings extend to other multiepisode events that, unlike days, form cohesive units in terms of their content, goal, and emotionality.

Keywords: peak-end rule, judgment, well-being, heuristics, Day Reconstruction Method

Caroline had a long day: she woke up early, took a shower, got dressed, and drove to work. She spent 7 hr at the office, drove back, did some grocery shopping, exercised, cooked, had dinner, watched TV, and went to read in bed. When her best friend calls her the next morning asking “how was your day” what will Caroline say? Apart from being the naturally occurring unit of our lives, days are multiepisode events that constitute a relevant example of the complexity of our experiences. During a day, we flow in and out of activities, constantly shifting between locations, partners, and foci, feeling a wide array of emotions. It is unclear how we sum all this information into a retrospective evaluation, such as the one Caroline is asked to administer.

This study examined how people evaluate days: are these evaluations normative and based on the aggregated feelings during the day, which are a close representation of experience, or are they heuristic and based on peaks, lows, and the feelings when the day ended. Discrepancies between experience and evaluation often result in a failure to choose or predict options that will maximize experienced happiness (Hsee & Hastie, 2006; Wirtz, Kruger, Napa Scollon, & Diener, 2003).

Normative and Heuristic Evaluations

Normative models posit that people form moment-by-moment judgments of experiences. The summations or integrals of these momentary judgments lead to holistic judgments. Kahneman,

Wakker, and Sarin define instant utility as “the pleasure or distress of the moment” (Kahneman, Wakker, & Sarin, 1997, p. 379). Total utility is “constructed from temporal profiles of instant utility according to a set of normative rules” (p. 376). Another determinant of total utility is “the retrospective evaluation of a temporally extended outcome” (p. 379). If the normative assumptions are met, a duration-weighted measurement of how a person felt throughout a multiepisode experience would be the best predictor of his retrospective evaluation of the experience. Thus, Caroline would multiply the affective value of each of her daily episodes by its relative duration and aggregate the products to evaluate how her day went.

Heuristic, bounded-rationality models (Simon, 1957; Tversky & Kahneman, 1973) challenge the normative stand. They do not assume that computations determine judgments. Rather, people's judgments and evaluations rely on mental shortcuts, and are often based on segments that represent the whole experience, even if the representation is inaccurate. The peak-end rule demonstrates how people use experienced utility to make retrospective heuristic judgments. This descriptive rule states that judgment and recollection of a past event are based on one's feelings when the experience reached extreme intensity, either positive or negative, and how the event ended. These distinct feelings override a duration-weighted average of ongoing affective reports throughout the entire experience, as prescribed by the normative approach (see Bell, Raiffa, & Tversky, 1988, for a comparison of these models).

In manifestations of the peak-end rule (Redelmeier & Kahneman, 1996; Redelmeier, Katz, & Kahneman, 2003) some patients experienced a typical colonoscopy, whereas others underwent a modified, slightly longer procedure that ended less painfully and was associated with a significantly less painful memory of the procedure. Similar findings emerged for other instances of physical discomfort (Kahneman, Fredrickson, Schreiber, & Redelmeier,

Talya Miron-Shatz, Woodrow Wilson School for Public and International Affairs, Princeton University.

Correspondence concerning this article should be addressed to Talya Miron-Shatz, Woodrow Wilson School for Public and International Affairs, Princeton University, 327 Wallace Hall, Princeton, NJ 08544-1013. E-mail: tmiron@princeton.edu

1993; Schreiber & Kahneman, 2000), over extended periods of time (Jensen, Martin, & Cheung, 2005; Stone, Schwartz, Broderick, & Schiffman, 2005, respectively), and for pleasant stimuli (Fredrickson & Kahneman, 1993).

Thus, the normative assumption is that Caroline's evaluation of her day will not differ from the average of her ongoing experiences. The heuristic assumption is that when Caroline is asked about her day, she will base her judgment on moments that aroused extreme feelings—such as finally incorporating a handstand in her morning exercise routine, and on the end of the event. In this case, the joy of the handstand and the serenity of going to sleep would lead her to conclude that she had a good day, above what the average of how she rated her episodes throughout the day might indicate. Figure 1 in the Appendix provides an illustration of the episodes comprising Caroline's day and their emotional ratings.

Conceptual Considerations in Examining Peaks and Lows During Days

Previous studies of the peak-end rule examined instances of either positive or negative intense emotion, but not both. Entire days consist of diverse experiences, evoking a gamut of emotions, so a person could even report extreme emotionality of both valences, which would not necessarily cancel each other out. Research indicates that it is important to distinguish emotional valence. Positive affect is not the reverse of negative affect (Isen, 1999). Rather, they differ in their structure and in their impact on cognition, memory, and motivation and even lead to different brain activation (Ito, Larsen, Smith, & Caccioppo, 1998). Therefore, this study denotes moments of intensive positive and negative affect as *peaks* and *lows*, respectively.

Previous studies deduced each participant's peak and low experiences from the participants' moment-by-moment affective ratings (e.g., of pain). This study, however, focused on multiepisode events, that cannot be captured by a single emotion and also last too long to allow for a moment-by-moment measurement of emotion. Yet, above these technical considerations, this inquiry was premised on the assumption that an experiencing individual is best equipped to determine the peaks and lows in her own life. Therefore, participants were asked to note the presence of peak and low experiences, what these were, and when they occurred. To avoid reports of affectively neutral events (Langston, 1994), the phrasing was overstated, defining peaks as moments that were "unusually wonderful or thrilling," and lows as "unusually unpleasant or

awful." To avoid problems of subjective interpretation participants were not asked to rate the intensity of events (Almeida & Horn, 2004; Almeida, Neupert, Banks, & Serido, 2005; Britt, Davidson, Bliese, & Castro, 2004).

Previous inquiries applied a statistical approach to determine when the peak (that was of either positive or negative valence) occurred and the extent of emotion it was associated with. To emulate this in the present study, I denominated the highest episode affective rating as the *statistical peak* and the lowest as the *statistical low*. From a phenomenological perspective, the statistical peak would not necessarily correspond with the peak reported by the participant. This is because a reported peak was described as a "moment," and could be very brief. Indeed, a peak may involve a surprising and delightful phone call from a friend, embedded in an episode of work. In addition, while it was possible to identify for each participant the episodes that were highest and lowest in affective ratings, participants could choose whether or not to report having had a peak or a low moment. The reported peaks and lows that the participants singled out were at the core of this investigation.

The Methodology of Examining Peaks and Lows During Days

Participants in this study followed the Day Reconstruction Method (DRM; Kahneman, Krueger, Schkade, Schwarz, & Stone 2004a, 2004b). The DRM was developed to capture time use: to record participants' emotions during each episode of the previous day. In current versions of the DRM (Kahneman, Schkade, Fischler, Krueger, & Krilla, 2008), participants indicated whether they experienced a peak and/or a low, and what rendered these moments so exceptional. Extending the peak-end rule to multiepisode experiences, I hypothesized that the variables marking the presence or absence of peaks and lows would significantly add to the prediction of the previous day's evaluations based on the duration-weighted episode ratings.

Previous studies demonstrated the peak-end rule using both absolute and relative evaluations (Redelmeier et al., 2003). Similarly, in this study, participants made two retrospective evaluations of the previous day—an absolute one calling for a detailed emotional account of the day (e.g., "overall, how happy were you yesterday?"), and an evaluation comparing it with other days. The comparative evaluation might have more ecological validity as prospect theory suggests that preferences depend on the reference point, rather than remaining invariable (Kahneman & Tversky, 1979). Comparisons carry important emotional outcomes: the prospective pleasure and pain conceived by comparing various plausible results affects decisions and choices (Mellers, 2000). Thus, standards are actively constructed through comparison, rather than remaining constant regardless of the context.

To summarize, this study examined whether multiepisode experiences are evaluated, absolutely and comparatively, using normative or heuristic models, specifically the peak-end rule. The research questions were whether people apply normative or heuristic thinking when evaluating their day and whether objectively defined (statistical) peaks and lows, or subjectively defined (reported) peaks and lows would be better predictors of the retrospective evaluation. The research hypotheses was that heuristic evaluation will prevail in comparative judgments.

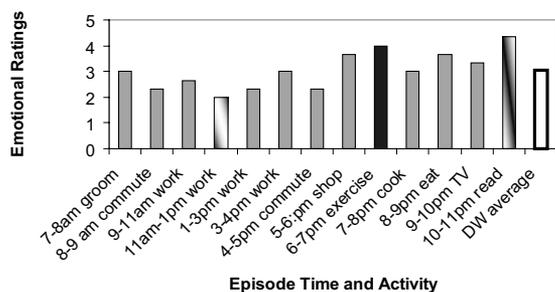


Figure 1. Caroline's day: activities and emotional ratings of episodes. DW = duration weighted.

Method

Participants

The participants were 810 women from Columbus, OH, 820 from Rennes, France, and 805 from Odense, Denmark. They were recruited by survey companies, using random-digit dialing. Their mean age was 42.30 in the United States ($SD = 10.94$), 38.61 in France ($SD = 11.23$), and 40.26 in Denmark ($SD = 11.60$). The samples were comparable concerning major demographic variables, such as the percentages of women who are married or cohabiting (69.8%, 60.9%, and 71.9%, respectively), who have a regular job (64.4, 60.8, and 71.9, respectively), or have a biological child at home (55.4%, 52.1%, and 48.6%, respectively). All participants spoke the dominant language at home. They filled out the questionnaires individually.

Materials and Procedure

The participants followed the DRM protocol (Kahneman et al., 2004a, 2004b). They reconstructed the episodes of the previous day, from when they woke up to when they went to sleep. For reasons of feasibility, episodes were specified to last between 20 min and 2 hr. Participants could decide how many episodes to define. For each episode they indicated the location, starting and finishing time, their actions and partners and the extent to which they experienced various feelings, from 0 (*not at all*) to 6 (*very much*). They then rated the previous day overall on the same emotional scales and noted whether the previous day was typical for that day of the week. Finally, they reported whether there was “a moment that was unusually wonderful or thrilling” that day (a peak) and what made it so great, and likewise “unusually awful or unpleasant” (a low) and what made it so bad.

Measures

Measuring the experience of episodes. Several variables were derived from the participants’ rating of how they felt during each episode measured experience. These variables all ranged from -6 , denoting highly negative affect, to 6 (highly positive affect).

Episode net affect was the average of positive emotions (happy, friendly, calm) minus the average of negative emotions (angry, tense, depressed) as rated for a specific episode. *Duration-weighted net affect* measured the participants’ reported emotions throughout the day. It was calculated by multiplying the net affect of each episode with the proportion of time it took out of ones waking hours and adding up these products. *Statistical peak and statistical low* were the highest and lowest episode net affect ratings for each participant, respectively.

Measuring the presence of a peak or a low in the participant’s day. Reported peak was a binary measure, with 1 noting that the person reported having an unusually wonderful or thrilling moment, and 0 noting that she did not report such a moment. Reported low was a binary similar measure: 1 noted that the person reported having an unusually awful or bad moment, and 0 noted the absence of such a report.

Measuring the retrospective evaluation. Two variables measured the retrospective evaluation of one’s day. *Yesterday net affect* was the average of positive emotions minus the average of negative emotions as rated for the day as a whole (e.g., “overall,

how tense were you yesterday?”). This formed the absolute evaluation of the day.

Typical measured how the day compared to what that day of the week usually is, using a scale of 1 (*much worse*) to 5 (*much better*). This formed the comparative evaluation of the day.

Results

The three samples yielded similar results. However, it is possible that participants in each country used the scales slightly differently, so the samples were analyzed separately.

This inquiry was based on the assumption that women would identify and report instances of intense emotionality in the experiences of their previous day. Reports of peaks and lows were prevalent despite being optional, which suggests that the concept of peaks and lows is integral to women’s experiences. Over half the participants in the United States, France, and Denmark reported a peak (52.9%, 70.7%, and 58%, respectively). A slightly smaller proportion (46.2%, 47.1%, and 31%) reported a low.

The method of reporting the peaks and lows required that the participant write when the peak or low happened, and what made it so great (or bad) but not when the peak or low ended. This made the overlap between peak (or low) moments and episodes hard to detect, if such existed. Every effort was made to detect the episode in which the peak or low occurred, using both the time in which they happened, and their content (e.g., taking care of a grandchild, or having a family quarrel). Still, one should bear in mind that the peak or low could have taken up a small portion of the episode. Table 1 lists the correlations between all the variables included in the analyses.

It indicates that the net affect of the episode in which the reported peak was embedded, and the highest episode net affect (statistical peak) correlate at a level of between .61 (United States) and .71 (Denmark), and the correlations for the net affect of the episode in which the reported low was embedded, and the lowest episode net affect are slightly higher (.72 in the United States and Denmark, .75 in France). This suggests a certain convergence of the reported peak and lows with their statistical equivalents, despite the conceptual differences.

Predicting the Retrospective Evaluation of Days

To compare the heuristic peak-end rule with the normative model (represented by the duration-weighted net affect), I included the duration-weighted net affect, the reported peak and low variables, and the net affect of the end episode as predictors of the day’s retrospective evaluations. As Table 2 indicates, the experienced duration-weighted net affect was most highly predictive of the retrospective net affect rating of the previous day ($\beta = .84$ in the United States and Denmark, $\beta = .82$ in France, $p < .001$). When considering this, one should take into account that duration-weighted and retrospective net affect ratings shared the same structure. Furthermore, the participants gave their net affect retrospective evaluations of the previous day immediately after rating the episodes on the same scales, which could account for some of the shared variance.

Table 2 also illustrates that the “reported low” variable consistently added to the predictions of the retrospective evaluations of the previous day. The presence of a reported low was associated

Table 1
Correlations Between Major Variables

Variable	1	2	3	4	5	6	7	8	9	10
1. Typical										
United States ^a	—									
France ^b	—									
Denmark ^c	—									
2. ODNA										
United States	.43	—								
France	.43	—								
Denmark	.42	—								
3. DWNA										
United States	.34	.87	—							
France	.33	.87	—							
Denmark	.35	.87	—							
4. Peak										
United States	.27	.15	.13	—						
France	.13	.14	.15	—						
Denmark	.19	.10**	.12**	—						
5. Low										
United States	-.26	-.38	-.34	.11**	—					
France	-.23	-.36	-.30	.17	—					
Denmark	-.30	-.38	-.31	.12	—					
6. EENA										
United States	.23	.60	.69	.07*	-.28	—				
France	.20	.58	.64	.09**	-.22	—				
Denmark	.13	.45	.54		-.18	—				
7. PENA										
United States	.17**	.44	.56	. ^d	-.12*	.39	—			
France	.14**	.47	.55	. ^d	-.17	.39	—			
Denmark	.17	.53	.63	. ^d	-.11*	.27	—			
8. LENA										
United States	.30	.58	.58	.12**	. ^d	.38	.16*	—		
France	.23	.52	.55		. ^d	.36	.20**	—		
Denmark		.60	.63		. ^d	.35	.29	—		
9. SPNA										
United States	.21	.59	.72	.15	-.15	.58	.61	.25	—	
France	.14	.60	.72	.25		.51	.64	.23	—	
Denmark	.17	.66	.75	.16	-.08**	.48	.71	.40	—	
10. SLNA										
United States	.34	.71	.74		-.43	.48	.31	.72	.31	—
France	.27	.69	.72		-.45	.45	.31	.75	.28	—
Denmark	.32	.70	.76		-.45	.42	.33	.77	.39	—

Note. All *ps* are < .001 unless marked otherwise. Nonsignificant correlations are not shown. Typical = yesterday typicality; ODNA = overall day net affect; DWNA = duration-weighted experienced net affect; Peak = reported a peak (dummy-coded); Low = reported a low (dummy-coded); EENA = end episode net affect; PENA = peak episode net affect; LENA = low episode net affect; SPNA = statistical peak net affect; SLNA = statistical low net affect.

^a *n* = 810. ^b *n* = 819. ^c *n* = 810. ^d No correlation could be computed because all cases with valid values for the continuous measure had the same value for the dummy variable.

* *p* < .05. ** *p* < .01.

with lower absolute evaluations compared with the absence of a reported low ($\beta = -.10$ in the United States, $\beta = -.12$ in France, and $\beta = -.13$ in Denmark, $p < .001$). The presence of a reported peak was only predictive of retrospective evaluations in the United States ($\beta = .006$, $p < .001$).

The duration-weighted net affect was also the best predictor of the retrospective comparative evaluation ("typical"), though with lower betas than for the overall evaluations ($\beta = .25$ in the United States, $\beta = .28$ in France, and $\beta = .29$ in Denmark, $p < .001$). Once again, reported low was a better predictor than reported peak, except in the United States.

The end episode did not add to the participants' overall evaluations of the previous day beyond its contribution to the duration-

weighted net affect. It was only a significant predictor of the comparative ("typical") evaluations in the United States ($\beta = -.08$, $p < .05$).

Previous studies of the peak-end rule used statistically determined peaks and lows. Likewise, Table 3 displays an analysis of the extent to which the net affect of the statistical peak and statistical low episodes added to the prediction of the day's evaluations.

As indicated by Table 3, the statistical peak and especially the statistical low were good sole predictors of typical and excellent predictors of yesterday net affect. Once duration-weighted net affect was entered in the regression as a predictor, however, the statistical peak only added to the prediction of "typical" for the

Table 2

Summary of Regression Analysis for Predicting Retrospective Evaluations of the Previous Day With Duration-Weighted (DW) Net Affect, Net Affect for the End Episode, and Peak Low

Variable	United States		France		Denmark	
	B	β	B	β	B	β
Predicted variable: Overall day net affect (absolute judgment)						
DW experienced net affect	1.11 (.03)	.84***	1.13 (.03)	.82***	1.05 (.03)	.84***
End episode net affect	-.01 (.02)	-.01	.02 (.02)	.02	-.02 (.02)	-.02
Reported a peak	.26 (.08)	.06***	.15 (.08)	.03	.06 (.07)	.02
Reported a low	-.45 (.08)	-.10***	-.49 (.08)	-.12***	-.51 (.07)	-.13***
Predicted variable: Yesterday typicality (comparative judgment)						
DW experienced net affect	.13 (.02)	.25***	.16 (.03)	.28***	.21 (.03)	.29***
End episode net affect	-.01 (.02)	-.02	-.01 (.02)	-.03	-.04 (.02)	-.08*
Reported a peak	.48 (.06)	.26***	.23 (.07)	.12***	.38 (.07)	.19***
Reported a low	-.39 (.06)	-.21***	-.31 (.06)	-.18***	-.54 (.07)	-.25***

Note. SEs for unstandardized betas are in parentheses.

* $p < .05$. *** $p < .001$.

French sample, though the beta value was negative. The statistical low added significantly to the predictions of yesterday net affect. However, it did not add to the prediction of "typical" in France. Thus, adding the presence of reported peaks and lows to duration-weighted net affect results in stronger predictions of previous day evaluations than adding the net affect values of the statistical peak and low episodes.

The Nature of Reported Peaks and Lows

Reported peaks and lows are pivotal in this investigation, so a brief discussion of their nature is warranted. These instances seldom revolved around extraordinary events, and yet were meaningful enough for the participants to single them out and report them, when they could have moved on to the next question.

When relating what made a peak exceptional, the first responses in the United States (sorted by participant number) were "rocking the baby [grandchild]," and "calculations were right for the first

time." French peaks started with "returning home to my husband," and "shopping." The Danes mentioned "two deer passed by us," and "relaxing in the yard, doing what I like to do."

Similarly, when recounting what made the lows so bad, American participants reported: "my husband and I have been fighting a lot and I was wondering if our marriage will last," and "computer crashed." Some French lows were: "had to pay bills," and "fast food dinner even though I was on a diet." The Danes mentioned "allergic reaction," and "there was a mess everywhere and it was dirty." This gives a taste of what people identified as their peaks and lows, which merit a more systematic investigation.

Discussion

This paper examined whether the retrospective evaluation of multi-episode events, focusing on days, the naturally occurring unit of our lives, was normative or heuristic. The normative approach involved aggregating and duration weighting the feelings associated with the

Table 3

Summary of Regression Analysis for Predicting Retrospective Evaluations of the Previous Day With Duration-Weighted (DW) Net Affect, Net Affect for the End Episode, and Net Affect for the Statistical Peak and Low Episode

Variable	United States		France		Denmark	
	B	β	B	β	B	β
Predicted variable: Overall day net affect						
DW experienced net affect	1.02 (.05)	.77***	1.06 (.06)	.76***	0.95 (.05)	.75***
End episode net affect	0.01 (.02)	.01	.04 (.05)	.04	-.02 (.05)	-.02
Statistical peak net affect	-.03 (.05)	-.01	0.02 (.05)	.01	0.11 (.05)	.07
Statistical low net affect	0.12 (.02)	.14***	0.10 (.03)	.11***	0.09 (.02)	.10**
Predicted variable: Yesterday typicality						
DW experienced net affect	0.10 (.04)	.18*	0.31 (.05)	.53***	0.32 (.06)	.45***
End episode net affect	0.00 (.02)	-.01	0.00 (.02)	.00	-.03 (.02)	-.07
Statistical peak net affect	0.02 (.04)	.02	-.017 (.04)	-.24***	-.014 (.05)	-.16**
Statistical low net affect	0.07 (-.02)	.20***	-.02 (.02)	-.05	0.03 (.03)	.07

Note. SEs for unstandardized betas are in parentheses.

* $p < .05$. ** $p < .01$. *** $p < .001$.

previous day's episodes, and the heuristic approach involved having the participants identify instances of extreme positive and negative emotionality during their day (peaks and lows). The normative variable, duration-weighted net affect, was composed of about a 100 observations per participants, whereas the presence of peaks and lows was marked by two binary variables.

Contradictory to the predictions of the peak-end rule, and most likely because of the method of current examination, the normative variable was the best predictor of retrospective evaluations of the previous day, both absolute and comparative. The presence of lows, and to a lesser degree, of peaks, added to the retrospective evaluations of the previous day that were generated using the normative model. The greater predictive power of lows over peak was consistent with literature regarding the greater effect of negative events and procedures relative to positive ones (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Kanouse, 1984; Van den Bos & Van Prooijen, 2001). The affective value of the end episodes rarely added to the prediction of retrospective evaluations, perhaps because the end of the day does not carry a distinct emotional meaning. Large samples from three countries yielded similar results, though future investigation could shed light on the intricacies of retrospective evaluations in each country.

A key corollary of the peak-end rule is duration neglect, an insensitivity to the duration of the event, as demonstrated by basing the rating on two time points and ignoring the rest. Our participants displayed this tendency, as the correlation between the duration-weighted average of their affective episode ratings was almost identical to the non-duration-weighted average. This could imply that arriving at the episode rating involved a local process of relying on segments of the event, possibly using the peak-end rule. Because the day does not comprise a cohesive emotional unit, the overall evaluation of the day, however, was arrived by using the normative aggregation of the episode ratings.

Absolute Versus Comparative Evaluations

The proportion of the contribution of heuristic and normative variables depended on the type of retrospective evaluation. Absolute evaluations, such as estimating how friendly, angry, or depressed you were yesterday call for a thorough scrutiny of the previous day and elicit the experiencing self, especially when retrospective and episode-based evaluations share the same structure. Therefore, absolute evaluations relied heavily on the normatively generated duration-weighted net affect variable. On the other hand, deciding whether yesterday was better or worse than a typical day emphasizes the perspective of the remembering self. This form of evaluation still relied on the normative evaluation, but relied to an almost similar extent on the heuristic notion of the lows in one's day and, to a lesser extent, on peaks. Previous research (Parkinson, Briner, Reynolds, & Totterdell, 1995) also found that peaks and lows contributed significantly to comparative evaluations of typicality. They attributed this to the fact that peaks and lows call for unusual events, which may also be atypical. Yet, participants in the present study rated the weekend days significantly higher on the "typical" scale than weekdays, hinting that typicality also provided a general evaluation of the day.

Why Ends Do Not Matter in This Study

In many instances, the end is one of an event's defining features (Ariely & Carmon, 2003), as are the rate and valence of changes and moments of extreme intensity. Most human experiences are goal directed, which explains why ends should be given extra weight (Spiegel, 1998). In cases such as queuing (Carmon & Kahneman, 1996), the end dominates the situation: there are no benefits for partial completion, and unless the end is successful, the experience needs to be repeated. In the past, the peak-end rule was mostly demonstrated for events that displayed a clear trend. This is predicted to increase the meaning of the end (Lowenstein & Prelec, 1993; Varey & Kahneman, 1992; Zauberman, Diehl, & Ariely, 2006). Partitioning events so they displayed an ascending or descending trend increased the weight of the end in overall evaluations (Ariely & Zauberman, 2003).

The participants in the present study evaluated days, which do not resemble this scenario: the end does not determine the outcome of the situation, if a day can be said to have an outcome at all, and the end is not finite in the sense that more days are to follow. Thus, days are incompatible to the situations in which the end was meaningful. Concordantly, the affective value of the participants' final episode added null predictive value to the retrospective evaluations of the previous day. Returning to Caroline, her final episode involved going to sleep, which carried scarce input for the evaluation of the day. This suggests a boundary condition for the nature of the experiences in which the end is meaningful and receives extra weight in the overall evaluation of the event.

Advantages of the Present Method for Examining the Peak-End Rule

Former studies of the peak-end rule were criticized as confronting two key limitations: using "affect-inducing stimuli. . . [that are] fairly uniform, likely to produce variations in valence and intensity, but not in specific emotions" and measuring continuous real-time ratings on one dimension, rather than "multiple discrete emotions" (Fredrickson, 2000, p. 594). The present study overcame the first limitation by examining entire days, which are anything but uniform. Open-ended reports of peaks conveyed a rich emotional array, including tranquility, achievement, love, pleasure, friendship, and faith. Likewise, lows conveyed loneliness, hostility, frustration, anger, fatigue, and anxiety. The second limitation was overcome by allowing the participants to rate 10 emotions for each episode, 6 of which were compiled into the net affect variable.

Another limitation of past investigations was their restricted context. Current advocacy calls for examining "personality in its natural habitat" and listening to people's daily experiences (Mehl, Gosling, & Pennebaker, 2006, p. 862), as well as letting people use their own words to delineate the troughs and summits in their life (Diener, Wirtz, & Oishi, 2001). Consistent with this, the present exploration was based on people's subjective definition of peaks and lows. The participants were at liberty to decide which elements were "unusually wonderful or thrilling" or "unusually awful or unpleasant," if any. Most of the peaks and lows revolved around events that may seem mundane, but were substantial enough to be singled out by the participants, as moments of unique emotionality. Consider, for example, "my husband won two trees in a silent auction" (a peak), and "going to Wal-Mart" (a low). Furthermore,

allowing the participants to highlight such instances of varying length helps alleviate a major caveat of the DRM, namely that episodes should last at least 20 min, so that events that are meaningful but brief, are not registered.

Limitations of the Present Study

One major caveat of this study involved the manner of data collection. Previous studies included a moment-by-moment measurement of affect, which could be conducted over an entire day. Instead, in this study participants used the DRM protocol (Kahneman et al., 2004a, 2004b) to record their affect during each episode of the previous day. Despite the retrospective nature of the DRM, the instructions encouraged participants to take the time to relive each episode in detail—their activities, who they were with, and their own feelings. This evokes the contextual experience, as opposed to the semantic and decontextualized remembering self that involves one's beliefs about emotions (Robinson & Clore, 2002). The DRM replicated affective patterns obtained with experience sampling, in which participants report their feelings as they are experiencing them during random moments throughout the day (Stone, Shiffman, & DeVries, 1999).

Another caveat of the DRM protocol was that episodes had to last at least 20 min, so briefer instances would be lost. Reporting peak and low moments, however, served as a partial compensation, as it allowed for noting events that were subjectively meaningful to the participant, regardless of their length.

Conclusion

The findings suggest boundary conditions to the peak-end rule and indicate that the gap between experience, as captured by the episodes that make up a day, and memory, as captured by how that day is evaluated, is in fact quite small. This was particularly true of overall evaluations, which, in the present study, followed the specific episode ratings. Parsing events of intense emotionality to peaks (highly positive) and lows (highly negative), indicated that the later were more predictive of summative evaluations than the former. Additionally, it appears that for ends to matter in the overall evaluation, they need to define a trend and carry a defining meaning for the event.

Thus, the memory-experience gap and duration neglect that underlined the peak-end rule, seem to be less characteristic of the evaluation of multi-episode events, though they may describe the way discrete episodes are evaluated. Future research should examine whether these findings extend to other multi-episode events that, unlike days, form cohesive units in terms of their content, goal, and emotionality.

References

- Almeida, D. M., & Horn, M. C. (2004). Is daily life more stressful during middle adulthood? In O. G. Brim, C. D. Ryff, & R. C. Kessler (Eds.), *How healthy are we?: A national study of well-being at midlife* (pp. 425–451). Chicago: University of Chicago Press.
- Almeida, D. M., Neupert, S. D., Banks, S. R., & Serido, J. (2005). Do daily stress processes account for socioeconomic health disparities? *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 60B(2, SpecIssue), 34–39.
- Ariely, D., & Carmon, Z. (2003). Summary assessment of experiences: The whole is different from the sum of its parts. In G. Lowenstein, D. Read, & R. Baumeister (Eds.), *Time and decision: Economic and psychological perspectives on intertemporal choice* (pp. 323–349). New York: Russell Sage Foundation.
- Ariely, D., & Zauberman, G. (2003). Differential partitioning of extended experiences. *Organizational Behavior and Human Decision Processes*, 91, 128–139.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5, 323–370.
- Bell, D. E., Raiffa, H., & Tversky, A. (1988). Descriptive, normative, and prescriptive interactions in decision-making. In D. Bell, H. Raiffa, & A. Tversky (Eds.), *Decision making* (pp. 9–30). New York: Cambridge University Press.
- Britt, T. W., Davison, J., Bliese, P. D., & Castro, C. A. (2004). How leaders can influence the impact that stressors have on soldiers. *Military Medicine*, 169, 541–545.
- Carmon, Z., & Kahneman, D. (1996). *The experienced utility of queuing: Real time affect and retrospective evaluations of simulated queuing*. Working paper, Duke University.
- Diener, E., Wirtz, W., & Oishi, S. (2001). End effects of rated life quality: The James Dean effect. *Psychological Science*, 12(2), 124–128.
- Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion*, 14, 577–606.
- Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, 65, 45–55.
- Hsee, C. K., & Hastie, R. (2006). Decision and experience: Why don't we choose what makes us happy? *Trends in Cognitive Sciences*, 10, 31–37.
- Isen, A. M. (1999). Positive affect. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 521–539). New York: Wiley Ltd.
- Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, 75, 887–900.
- Jensen, M. P., Martin, S. A., & Cheung, R. (2005). The meaning of pain relief in a clinical trial. *Journal of Pain*, 6, 400–406.
- Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2004a). A survey method for characterizing daily life experiences: The Day Reconstruction Method. *Science*, 306, 1776–1780.
- Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2004b). *The Day Reconstruction Method (DRM): Instrument documentation*. Retrieved April 3, 2005, from <http://www.sciencemag.org/cgi/content/full/306/5702/1776/DC1>
- Kahneman, D., Schkade, D. A., Fischler, C., Krueger, A. B., & Krilla, A. C. (2008). *The structure of well-being in two cities*. Working paper, Princeton University.
- Kahneman, D., & Tversky, A. (1979). Prospect theory. *Econometrica*, 47, 263–292.
- Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics*, 112, 375–405.
- Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, 4, 401–405.
- Kanouse, D. E. (1984). Explaining negativity bias in evaluative and choice behavior: Theory and research. In Kinnear, T. (Ed.), *Advances in Consumer Research* (Vol. 11, pp. 703–308). Provo, UT: Association for Consumer Research.
- Langston, C. A. (1994). Capitalizing on and coping with daily-life events: Expressive responses to positive events. *Journal of Personality and Social Psychology*, 67, 1112–1125.
- Loewenstein, G., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review*, 100, 91–108.
- Mehl, M. R., Gosling, S. D., & Penebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90, 862–877.

- Mellers, B. A. (2000). Choice and the relative pleasure of consequences. *Psychological Bulletin*, *126*, 910–924.
- Parkinson, B., Briner, R. B., Reynolds, S., & Totterdell, P. (1995). Time frames for mood: Relations between momentary and generalized ratings of affect. *Personality and Social Psychology Bulletin*, *21*, 331–339.
- Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*, 3–8.
- Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain*, *104*, 187–194.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*, 934–960.
- Schreiber, C. A., & Kahneman, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology: General*, *129*, 27–42.
- Simon, H. A. (1957). *Models of man*. New York: Wiley.
- Speigel, D. (1998). Getting there is half the fun: Relating happiness to health. *Psychological Inquiry*, *9*, 66–68.
- Stone, A. A., Schwartz, J. E., Broderick, J. E., & Shiffman, S. S. (2005). Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. *Personality and Social Psychology Bulletin*, *31*, 1340–1346.
- Stone, A. A., Shiffman, S. S., & DeVries, M. W. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61–84). New York: Russell-Sage Institute.
- Tversky, A., & Kahneman, D. (1973). *Judgment under uncertainty: Heuristics and Biases*. Oxford, England: Oregon Research Institute.
- Van den Bos, K., & Van Prooijen, J. W. (2001). Referent cognitions theory: The psychology of voice depends on closeness of reference points. *Journal of Personality and Social Psychology*, *81*, 616–626.
- Varey, C. A., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making*, *5*, 169–185.
- Wirtz, D., Kruger, J., Napa Scollon, C., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science*, *14*, 520–524.
- Zauberman, G., Diehl, K., & Ariely, D. (2006). Hedonic versus informational evaluations: Task dependent preference for sequences of outcomes. *Journal of Behavioral Decision Making*, *19*, 191–211.

Appendix

Examining Caroline's Day

This paper involves several conceptualizations of the peak (and low) notion. To illustrate what the components of the day are and clarify the nature of the variables, I demonstrated them using the hypothetical Caroline's day. Figure 1 follows Caroline throughout her day, from when she woke up (7 a.m.) till when she went to sleep (11 p.m.), listing all her episodes and their affective ratings.

This was the first time Caroline managed to do a hand stand as part of her exercise at home before dinner. The handstand occurred at 6:22 p.m. and the whole exercise episode took place between 6 and 7 p.m. Caroline gave the exercise episode ratings that resulted in a net affect score of 4 (the specific ratings were: calm = 6, happy = 6, friendly = 4 for the positive emotions, and angry = 2, tense = 1, depressed = 1 for the negative emotions). For the question, "Was there a moment yesterday that was unusually wonderful or thrilling?" (peak), Caroline answered yes and noted the handstand as her reported peak. The exercise episode, in which the reported peak was embedded is marked by a darker shade in the Figure. For the question, "Was there a moment yesterday that was unusually bad or awful?" (low), Caroline did not note anything.

The statistical peak episode (marked by dark diagonal shading in the Figure) was the one that received the highest net affect value of all episode ratings (4.33): reading in bed, 10 to 11 p.m. Note that the net affect rating for this episode was higher than that for the episode in which the reported peak was embedded. Similarly, the statistical low episode (marked by bright diagonal shading in the figure) was the one that received the lowest net affect value of all episode ratings (2): work, between 11 a.m. and 1 p.m. The end episode was Caroline's last episode during the day, which happened to converge with her statistical peak episode.

The duration-weighted (dw) average of the net affect of Caroline's day (2.91) appears at the rightmost column (marked by a dark bold border in the figure). To calculate the dw average, the net affect of each episode was multiplied by the episode's relative length in Caroline's 16 waking hours. For example, the net affect of the exercise episode, lasting 1 hr, was multiplied by 1/16 when entered in the dw score.

Received May 9, 2007

Revision received March 24, 2008

Accepted October 20, 2008 ■