

Duration Neglect in Retrospective Evaluations of Affective Episodes

Barbara L. Fredrickson and Daniel Kahneman

Two experiments documented a phenomenon of duration neglect in people's global evaluations of past affective experiences. In Study 1, 32 Ss viewed aversive film clips and pleasant film clips that varied in duration and intensity. Ss provided real-time ratings of affect during each clip and global evaluations of each clip when it was over. In Study 2, 96 Ss viewed these same clips and later ranked them by their contribution to an overall experience of pleasantness (or unpleasantness). Experimental Ss ranked the films from memory; control Ss were informed of the ranking task in advance and encouraged to make evaluations on-line. Effects of film duration on retrospective evaluations were small, entirely explained by changes in real-time affect and further reduced when made from memory. Retrospective evaluations appear to be determined by a weighted average of "snapshots" of the actual affective experience, as if duration did not matter.

Given their choice, most reasonable people would prefer aversive experiences to be short rather than long and would opt to reduce the duration of their exposure to disgusting objects, horrifying images, or painful procedures. Yet choices between experiences that vary in duration need not be expressed explicitly in terms of time. Suppose instead that individuals are asked which of two aversive experiences in their past they would prefer to have again. Presumably, the experience they choose will be the one that evokes the more favorable (or less aversive) global evaluation at the moment of choice. Can we be certain that decision makers relying on retrospective evaluation will always choose the shorter of the two episodes? The answer depends in part on how the relevant experiences are represented in memory. It also depends on the processes by which people make global evaluations of extended affective episodes.

The question of how time influences people's thoughts about affective events and experiences has captivated quite a few researchers in recent years (Coombs & Avrunin, 1977; Hsee & Abelson, 1991; Hsee, Abelson, & Salovey, 1991; Johnson & Sherman, 1990; Linville & Fischer, 1991; Loewenstein & Pre-

lec, 1993; McFarland, Ross, & DeCourville, 1989). A puzzle concerning time also motivated the present research, where we identify a circumstance in which time should matter but does not. We argue that episode duration plays a surprisingly small role in people's evaluations of prior affective experiences. Two related theoretical contrasts help us develop this argument. First, we contrast two metaphors for the representation of emotional episodes in memory: one likens this representation to a film, the other to a collection of snapshots. Second, we contrast two models for the process of global evaluation: temporal integration of momentary affect or weighted averaging of selected moments.

Introspecting about how people recall love affairs, Milan Kundera (in his novel *Immortality*) concluded that "memory does not make films, it makes photographs," and at best only a few photographs (1991, p. 314). As metaphors for memory, films and photographs differ in many respects. One important difference is that a representation of time is inherent in a film, but not in a set of photographs. A secondary aspect of the metaphor is that a film suggests comprehensive coverage, whereas a collection of snapshots represents experience by selected salient moments. Running a film in memory could provide a cumulative impression of the episode, in which constituent incidents are weighted by their duration. Consulting a collection of remembered stills is more likely to yield a weighted average of the impressions of separate moments, because time is not directly represented.

Whether adding or averaging best represents how people form global evaluations is hardly a new question (for recent discussions of information integration, see Anderson, 1991; Massaro & Friedman, 1990). A special feature of the context in which we raise this question is that time is cumulative: Extending the duration of an aversive episode adds to the experience of negative affect. Thus, a rule of *temporal monotonicity* has the force of a logical principle: Adding moments of misery should make the overall experience worse, and adding moments of pleasure should make it better. As we found, there is a tension between such normative principles and the descriptive rules that actually govern the processes of global evaluation.

An appealingly simple model for global evaluation—one that

Barbara L. Fredrickson, Department of Psychology: Social and Health Sciences and Center for the Study of Aging and Human Development, Duke University; Daniel Kahneman, Department of Psychology, University of California, Berkeley.

This research was facilitated by a postdoctoral fellowship from the National Institute of Mental Health awarded to Barbara L. Fredrickson and by grants from the Sloan Foundation, the National Science Foundation, and the John D. and Catherine T. MacArthur Foundation Research Program on Determinants and Consequences of Health-Promoting and Health-Damaging Behavior awarded to Daniel Kahneman. We thank Paul Grant, Amy Hayes, Kathy Knopoff, Christy Monahan, and Rachel Wolfe for their valuable assistance at various stages of this research. We also thank George Loewenstein and Erika Rosenberg for helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Barbara L. Fredrickson, Department of Psychology: Social and Health Sciences, Duke University, Box 90085, Durham, North Carolina 27708-0085, or to Daniel Kahneman, Department of Psychology, 3210 Tolman Hall, University of California, Berkeley, California 94720. Electronic mail may be sent to blf@acpub.duke.edu.

obeys temporal monotonicity—is summation over time, or *temporal integration*. This is most easily explained by an example (Kahneman & Snell, 1990; Varey & Kahneman, 1992). Imagine that you undergo a series of painful daily injections. How do you measure the overall suffering associated with a given number of injections? If you agree that the difference between suffering nine injections and suffering seven injections is the pain associated with the eighth and ninth injections, you have committed yourself to an additive measure of overall suffering. The same reasoning extends readily to continuous episodes of discomfort: If affect intensity is plotted as a function of time, the integration model represents the overall affect of an episode by the area under the curve. In the simplest version of this model, equal units of time are assigned equal weight, regardless of their content or of the duration of the episode. More elaborate versions of temporal integration could also be defined, for example, by allowing the sum of weights to be a subadditive function of duration and by allowing the weights of individual moments to vary with their content.

Weighted averaging is a competing model of how people might form global impressions of affective episodes. Note that an averaging model is likely to violate the logic of temporal monotonicity: Extra moments of misery could make the overall experience less aversive if the added moments are less miserable than others and are given substantial weight. In an averaging model, the weights assigned to individual moments are constrained to add to unity for any episode, regardless of its duration. In its simplest form, all moments would be weighted equally. We favor a special case of the averaging model, in which most moments of an episode are assigned zero weight in the evaluation, and a few select “snapshots” receive larger weights.

The key difference between temporal integration and weighted averaging is that the two models imply different predictions about the importance of duration as an attribute of affective experience. The core of any temporal integration model is that people’s global evaluations of episodes should be determined jointly by the intensity of the experienced affect and by the duration of the episode. Weighted averaging, in contrast, entails duration neglect, which we define as little or no independent effect of duration on retrospective evaluations of affective episodes.

In earlier work that we conducted independently, we offered some initial evidence for duration neglect and for the idea that selected moments of an extended episode play a privileged role in its global evaluation. Varey and Kahneman (1992, Experiment 2) examined subjects’ global evaluations of another person’s experience, based on schematic reports of momentary discomfort. A target person was said to have made a series of “discomfort ratings” using a 0–10 scale, at 5-min intervals during an unpleasant episode (e.g., standing in an uncomfortable position and exposure to a loud drilling noise). Recovery from the discomfort was said to be immediate. The episodes ranged from 15 to 35 min long and varied in the intensity and trend of reported discomfort. Subjects’ global evaluations of these episodes were highly sensitive to the trend of discomfort and largely insensitive to duration. A recent reanalysis of the published data indicates that a simple rule that we call the *peak & end rule* predicted evaluations quite accurately: An unweighted combination of the peak discomfort in the series and of the *terminal discomfort* accounted for 94% of the variance in

global evaluations. The factor of duration, although significant, accounted for only 3% of additional variance. The neglect of duration and the emphasis on endings were in violation of temporal monotonicity. For example, the series of discomfort ratings 2-5-8-4 (indicating a 20-min episode ending with a discomfort rating of 4) was reliably judged as less aversive than the series 2-5-8, although the only difference between the two episodes is that the latter includes 5 min fewer of discomfort.

In a quite different context, Fredrickson (1991) tested the disproportional weighting of endings in people’s retrospective evaluations of their own social interactions. Subjects in this study conversed with a new acquaintance for several hours; some believed the interaction would end that day, others believed it would continue on a subsequent day. At three points in the experiment, subjects viewed a videotaped portion of their conversation to provide on-line ratings of how they were feeling during the actual interaction. After three 30-min sessions of conversation, subjects indicated their global impressions of the relationship. Among those who believed that the interaction had ended, a strong correspondence emerged between subjects’ evaluations of the entire social relationship and the affect they had experienced during the final conversation. This result was not merely an effect of recency: The correspondence was completely absent for those who perceived the final conversation as a midstream portion of an ongoing relationship. Fredrickson’s results show that affect experienced during the perceived ending of a social episode is weighted heavily in retrospective evaluations of that episode.

When people evaluate their own past experiences, we know from Fredrickson’s (1991) work that some units of time can matter more than others. We also suspect from Varey and Kahneman’s (1992) work that the duration of experiences may contribute little to global evaluations, but this has been demonstrated only for judgments of other people’s experiences. The aim of the two studies reported here was to examine the role duration plays in people’s assessments of their own past experiences. Consistent with weighted averaging of salient moments as a rule of global evaluation, we hypothesized that retrospective evaluations of affective episodes would not depend on duration.

Our strategy in testing the hypothesis of duration neglect was to use brief emotionally evocative film clips that varied in valence (aversive or pleasant¹), intensity (mild to moderately intense), and duration (short or three times as long). To maximize experimental control, we selected film clips for which we could manipulate duration without drastically altering the quality or meaning of the affective experience. These were essentially plotless clips, in which the basic content and affective impact of the images did not vary over the course of the clip. Using a within-subjects design, we showed individuals short and long versions of both pleasant and aversive clips and had them report on the affective experiences that these clips generated.

¹ We realize that affective experiences are more complex than these two broad categories suggest. For instance, aversive experiences could evoke disgust, fear, or some other negative affect, and pleasant experiences could evoke amusement, contentment, and so forth. Yet, because the approach–avoidance dimension is particularly important to choice situations, we concentrated our initial research efforts on valence alone.

Although we attempted to keep the content of films constant, we suspected that people's ability to endure an unchanging aversive experience or to enjoy an unchanging pleasant experience might in fact change over time, perhaps according to Coombs and Avrunin's maxim that "good things satiate and bad things escalate" (1977, p. 224; see also the renewable resources model proposed by Linville & Fischer, 1991). The hypotheses of weighted averaging and of duration neglect do not preclude such systematic changes in affect over time. They only entail that duration should make little or no independent contribution to global evaluations above and beyond its impact on momentary affect. In statistical terms, we propose that the partial correlation between the global evaluation of an episode and its duration, controlling for real-time affect, should be close to zero.

Study 1 was designed to permit statistical separation of the effects of time on immediate affective experiences and on subsequent global evaluations. To this end, we required subjects to provide continuous reports of affect using a real-time rating device (Levenson, 1988; Levenson & Gottman, 1983). The device we used is similar to instruments that are routinely used in applied marketing research to assess the affective impact of different segments of advertisements or political speeches (see "Manhattan Project," 1992). Using this device, subjects readily indicate their affective state of the moment, and they are quick to respond to changes. We have observed in other work using this rating device that subjects report large changes in affect within seconds of changes in relevant stimuli (Fredrickson & Levenson, 1992; Kahneman, Fredrickson, Schreiber, & Redelmeier, in press). The high temporal resolution of the real-time measure provides the needed operational distinction between a continuous recording of momentary affect and a cumulative evaluation of the experience.

In sum, the objective of Study 1 was to examine the effects of duration on real-time affect and on global evaluations for a set of affective experiences. Here, global evaluations were made immediately after each film, and subjects knew they would make them. Study 2 examined the effects of delaying global evaluations.

Study 1

Method

Subjects

Subjects were 32 students (16 women and 16 men) at the University of California, Berkeley, who received \$10 for their participation in a 90-min study of emotion. They ranged in age from 16–35 years, with a median age of 22. Male and female subjects were randomly assigned to view one of four stimulus videotapes, as described below.

Apparatus

We used a positive–negative *affect meter* (Levenson & Gottman, 1983) to obtain moment-by-moment reports of affective experience. Subjects using the affect meter manipulate a sliding knob attached to a potentiometer that controls an array of 15 colored lights positioned above the video monitor. The 7 rightmost lights in this array are green, and subjects are told that these represent degrees of any positive feelings. The 7 leftmost lights are red and represent degrees of any negative feelings. The center light is yellow and represents neutral feelings.

When the sliding knob is centered, only the yellow light is illuminated. As the knob is slid to the right, the green lights come on one by one, so that all 7 green lights are illuminated when the sliding knob is set to the extreme right. Negative ratings are indicated similarly by the number of red lights illuminated. A computer records the mean position of the sliding knob every second. Subjects are asked to adjust the sliding scale as often as necessary so that it always reflects how positive or negative they are presently feeling. The light display allows subjects to maintain their visual focus on the video monitor while receiving continuous, synchronized feedback on the affect they are currently reporting. (Validity data for this type of continuous, real-time measure of affect can be found in Gottman & Levenson, 1985.)

Visual Materials

Selection. To select film clips for the main experiments, we conducted a pilot test of 20 clips ranging in length from 55 to 138 s. Some of these clips had sound, others were silent. As mentioned previously, all clips were essentially plotless: The affect-eliciting images were introduced in the initial moments of each clip (e.g., playful penguins or dying people) and remained present without substantive change until the end of the clip. Ten subjects individually viewed these 20 film clips while rating their moment-by-moment affective responses using the affect meter. On the basis of these pilot data, we selected 12 test clips (6 pleasant and 6 aversive) by the following criteria: (a) Each clip was consistently rated as either positive or negative by all subjects, and (b) the range of second-by-second affect ratings across each clip was small. We made a short version of each test clip by taking approximately the first third of it. The leftmost columns of Tables 1 and 2 describe the contents of the pleasant and aversive test clips, respectively.²

Stimulus presentation. We formed two sets of test clips matched roughly on affective intensity (labeled *Set A* and *Set B* in Tables 1 and 2). Half the subjects saw the short versions of films in *Set A* and the long versions of those in *Set B*; the other subjects saw the complements of both sets. Subjects also viewed four additional film clips, each shown at the same duration to all subjects: one pleasant and one aversive clip and two clips that induced varied or mixed affective experiences. We only report the data for the 12 test films that were shown in two durations.

We made four stimulus videotapes, each consisting of 16 clips, as described above. Pleasant and aversive clips were mixed in viewing order on each tape. In Video 1, *Set A* clips were short and *Set B* clips were long. In Video 2, the clips were in the same order as in Video 1, but *Set A* clips were now long, and *Set B* clips were short. In Videos 3 and 4, the clips were presented in orders that were the reverse of Videos 1 and 2.

Written Materials

We obtained global evaluations in Study 1 by asking subjects "Overall, how much pleasure [displeasure or discomfort] did you experience during the film you just saw?" Subjects indicated their rating by marking a single vertical slash through a 100-mm line anchored at the left with *no pleasure [displeasure or discomfort] at all* and on the right with *a great deal of pleasure [displeasure or discomfort]*. These global evaluations were obtained immediately after each clip ended.

Procedure

Subjects participated individually. A female experimenter described the session as a study on the subjective experience of emotion. Subjects

² Similar versions of two of these clips (the puppy and amputations) were first used by Ekman and Friesen (1974). We thank Paul Ekman for making these films, along with the one of waves, available to us.

Table 1
Description of Pleasant Clips and Mean Affective Responses From Studies 1 and 2

Description of film clip	Duration (s)	Study 1			Study 2	
		Peak affect ^a	End affect ^a	Global evaluation ^b	On-line ^c	Delayed ^c
Puppy playing with flower (B)						
Short	33	1.87	1.72	33.50	3.71	5.00
Long	100	2.81	-0.18	30.38	3.67	4.54
Waves breaking on beach (A)						
Short	29	3.13	2.75	50.94	4.58	3.42
Long	100	3.33	1.31	45.19	3.79	3.83
Gregory Hines tap dancing (B)						
Short	38	2.53	2.06	44.81	3.50	3.42
Long	113	3.38	2.27	50.19	4.08	4.67
Ski jumping (A)						
Short	37	3.44	2.97	52.00	3.50	4.54
Long	106	3.53	2.89	56.31	4.50	4.08
Flying over African landscape (B)						
Short	41	3.00	2.51	57.50	6.08	4.96
Long	119	4.44	3.89	67.62	6.12	5.42
Penguins diving off a glacier (A)						
Short	41	4.25	3.58	59.56	5.50	5.21
Long	125	5.00	3.69	67.50	5.88	6.83
Average of pleasant clips						
Short						
<i>M</i>	36.5	3.05	2.61	49.72	4.48	4.42
<i>SD</i>		1.47	1.42	23.07	0.92	0.98
Long						
<i>M</i>	110.5	3.74	2.30	52.86	4.67	4.90
<i>SD</i>		1.20	1.39	18.98	1.11	1.17
<i>t</i> tests		2.79*	1.05	0.87	0.77	1.68
Mean biserial <i>r</i>		.25	-.06	.10	.06	.11

Note. The two sets of films are identified by the labels *A* and *B*.

^a *n* = 31. ^b *n* = 32. ^c *n* = 48.

* *p* < .01.

were told that they would view a series of film clips that would range from very pleasant to very unpleasant. The experimenter showed subjects how to use the sliding affect meter and told them that their task was to adjust the light array as often as necessary so that it always indicated how they were feeling moment-by-moment as they watched each film clip. The experimenter also described the global evaluation task and told subjects that instructions on the video monitor at the end of each clip would prompt them to indicate their evaluations on the written materials. After answering any questions, the experimenter started the videotape, dimmed the lights, and left the room. After viewing all 16 clips, subjects repeated their global evaluations of the films in the order of viewing. Finally, subjects estimated the duration of each clip that they had seen.

Results and Discussion

Preliminary Analyses

Tables 1 and 2 present descriptive statistics for each of the pleasant test films and aversive test films, respectively, in both short and long versions.³ To facilitate later comparison between Studies 1 and 2, the tables show dependent variables for both studies. For Study 1, these include two measures derived from the real-time ratings of affect obtained during each film (peak and end affect), along with the global evaluation made after each film ended. Note that mean values associated with each

version of individual clips are based on the 16 subjects who viewed that version of the clip.⁴

Tables 1 and 2 also include, for each dependent measure, a within-subjects test of the significance of the duration effect followed by a correlational measure of the size of this effect. For each subject, we computed mean responses across the three films that the subject had seen in short versions and compared these with the corresponding means computed across the three films seen in long versions. For half of the subjects, the long clips were those of Set A and the short ones of Set B. The relation of duration to set was reversed for the remaining subjects. (Differences between the two sets of films increase error variance but do not confound the effect of duration.) The results of these within-subject *t* tests are reported in the next-to-last row of Tables 1 and 2. To estimate the size of the duration effects, we computed, within the data of each subject, a biserial correlation between duration (short vs. long) and each of the dependent measures, over the six films that the subject saw. The means of

³ All tests of significance reported in this article are two-tailed, with an alpha level of .05.

⁴ Real-time affect ratings for 1 subject were unusable because of technical problems. Consequently, half the mean values for peak and end affect for individual clips are based on 15 subjects, and statistical tests using these measures are based on 31 subjects.

Table 2
Description of Aversive Clips and Mean Affective Responses From Studies 1 and 2

Description of film clip	Duration (s)	Study 1			Study 2	
		Peak affect ^a	End affect ^a	Global evaluation ^b	On-line ^c	Delayed ^c
Aftermath of Hiroshima (A)						
Short	35	-2.81	-2.16	50.13	4.04	4.46
Long	109	-3.27	-2.69	52.94	5.50	5.04
Dying people in Singapore (B)						
Short	36	-3.07	-2.63	56.38	4.25	4.33
Long	84	-4.44	-3.64	64.94	5.46	4.00
Killing wild pigs with clubs (A)						
Short	21	-3.31	-3.20	49.63	3.42	4.08
Long	64	-3.73	-3.51	62.50	4.17	4.46
Man being forcibly drowned (B)						
Short	27	-2.33	-2.07	41.75	3.42	3.42
Long	83	-4.19	-3.16	48.38	4.21	3.83
Victims of the Holocaust (A)						
Short	29	-3.94	-3.69	62.06	4.92	5.21
Long	90	-5.53	-5.10	81.06	5.96	5.96
Medical film of amputation (B)						
Short	25	-4.47	-4.03	71.13	5.71	5.92
Long	75	-6.44	-5.82	82.19	6.54	7.08
Average of aversive clips						
Short						
<i>M</i>	28.8	-3.33	-2.96	55.18	4.29	4.57
<i>SD</i>		1.49	1.54	23.25	0.68	0.71
Long						
<i>M</i>	84.2	-4.60	-4.00	65.33	5.30	5.06
<i>SD</i>		1.60	1.64	21.88	0.68	0.81
<i>t</i> tests		5.98***	4.42***	3.08**	5.63***	2.47*
Mean biserial <i>r</i>		.41	.32	.26	.24	.12

Note. The two sets of films are identified by the labels *A* and *B*.

^a *n* = 31. ^b *n* = 32. ^c *n* = 48.

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

these individual correlations are shown in the last row of Tables 1 and 2. These results are discussed in detail in subsequent sections.

We examined several other variables that yielded no results of interest. First, the repeated global evaluations that subjects in Study 1 made at the end of the session were highly correlated with the ratings they made immediately after each film and provided little new information. Second, the estimates of film duration indicated that subjects clearly discriminated the long from the short films: On average, subjects estimated the long versions to be twice the length of the short versions. Third, all initial analyses of mean levels included the between-subjects factors of subject sex, viewing order, and clip set. Because no reliable main effects or interactions emerged for these control factors, we collapsed across them in all subsequent analyses. Fourth, in addition to abstracting peak and end affect from the real-time ratings, we also calculated mean affect over the course of each film (excluding the first 10 s). Because analyses with this measure offered no new information, we judged the two snapshot measures to be more efficient indexes of momentary affect.

Real-Time Affect

Second-by-second means of subjects' affect meter responses to the six pleasant and six aversive films are illustrated in Fig-

ures 1 and 2, respectively. The ratings indicate that the first several seconds of each film were marked by sharp displacements from neutral affect, which is expected given that the affect meter was reset to the neutral zero point before each clip. The substantial leveling off beyond this point indicates that we were successful in avoiding films that produced large, sudden changes in affect. However, gradual changes were common.

The real-time responses to pleasant films (Figure 1) were quite variable. Final affect ratings were distinctly more positive for the long version of one film (Clip 5), approximately equal for three others (Clips 3, 4, and 6), and distinctly less positive for the remaining two films (Clips 1 and 2). Indeed, mean ratings of affect dipped into the negative at the end of the long puppy film. We suspect that subjects' responses to each of the pleasant clips would eventually become negative if shown long enough. Perhaps all good things satiate, although not all at the same rate.

The real-time responses to aversive films were much more consistent. The longer versions reliably produced some escalation of negative affect: As can be seen in Figure 2, mean affect ratings were more negative at the ends of the long versions of each of the six films. The increasing aversiveness of continued exposure to unpleasant images fits Coombs and Avrunin's (1977) dictum that "bad things escalate."

For statistical analyses of the measure of real-time affect, we extracted two variables from each subject's responses to each

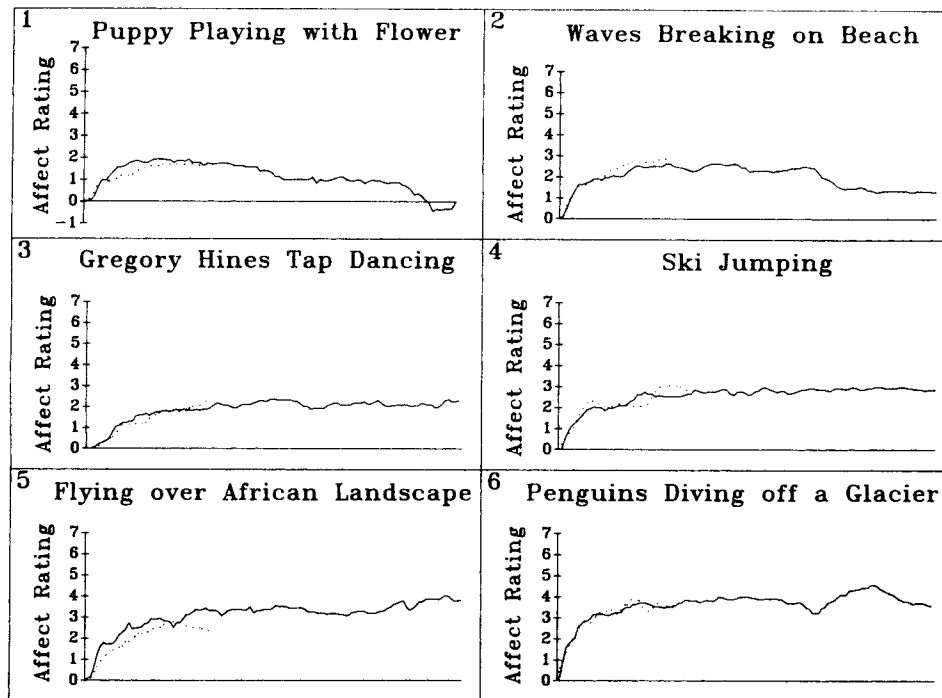


Figure 1. Mean second-by-second real-time ratings of affect for six pleasant test clips. (Means were calculated separately across the 16 subjects who viewed long and short versions of each clip. Means for short clips are represented by dashed lines.)

film: (a) the most extreme rating given at any point during the clip (peak affect) and (b) the mean rating across the final 10 s of the clip (end affect). These two variables were indexes of particular moments, or snapshots, of the momentary affective experience that we predicted would greatly influence global evaluations.

Likely reflecting the mix of trends evident across the pleasant clips (see Figure 1), no reliable duration effect emerged for the end affect measure in Table 1. In contrast, the significant duration effect for end affect in Table 2 confirms the escalation of negative affect during the aversive films. The peak affect measure also differed significantly between short versus long clips for both pleasant and aversive clips (see Tables 1 and 2). This result must be interpreted with caution, however, because the correlation between these variables is constrained to be nonnegative: Prolonging a film necessarily affords increasing opportunities for extreme affective responses to occur.

Our interpretation of the trends in Figures 1 and 2 in terms of escalation and satiation assumes that the changes in subjects' emotional responses are not simply tracking changes in the content of the films. As we have mentioned, our aim was to select plotless clips to ensure stability of content, but were we successful? To answer this question, we showed the long versions of our test films to 20 additional subjects. Immediately after viewing each clip, these subjects were asked whether, over the course of the clip (a) their "emotional responses got worse, got better, or did not change in any particular direction" and (b) "the basic content of the film got worse, got better, or did not change in any particular direction." The responses to the first question were in accord with the trends observed in Figures 1

and 2. For aversive films, a considerable majority of responses (72%) indicated escalation of negative affect, most others indicated no change (24%), and a small minority indicated diminishing negativity (4%). As in Figure 1, the responses to positive films were less consistent: The percentages of responses indicating improvement, no change, and deterioration were 54%, 25%, and 21%, respectively. Although we might expect subjects to attribute their changed feelings to changes in film content, only a minority did so; the percentage indicating that the basic content of the films had not changed was 83% for aversive films and 77% for pleasant films (almost all other responses indicated change for the worse in aversive films and for the better in pleasant films). These results bolster the conclusion that the changes in real-time affect illustrated in Figures 1 and 2 were not caused by changes in film content. They also make the point that people can discriminate endogenous processes of affective escalation and satiation from the exogenous effects of stimulus changes.

Global Evaluations

The results for the global evaluations, summarized in Tables 1 and 2, again suggest different patterns for pleasant and aversive films. For pleasant films, increasing clip duration did not reliably increase global evaluations (see Table 1). Indeed, the effects of increasing duration varied across the different pleasant films: Long versions were evaluated more favorably for four films but less favorably for the puppy film and the waves film. These were the same two films for which end affect was clearly less positive for the long versions (see Figure 1). Note, however,

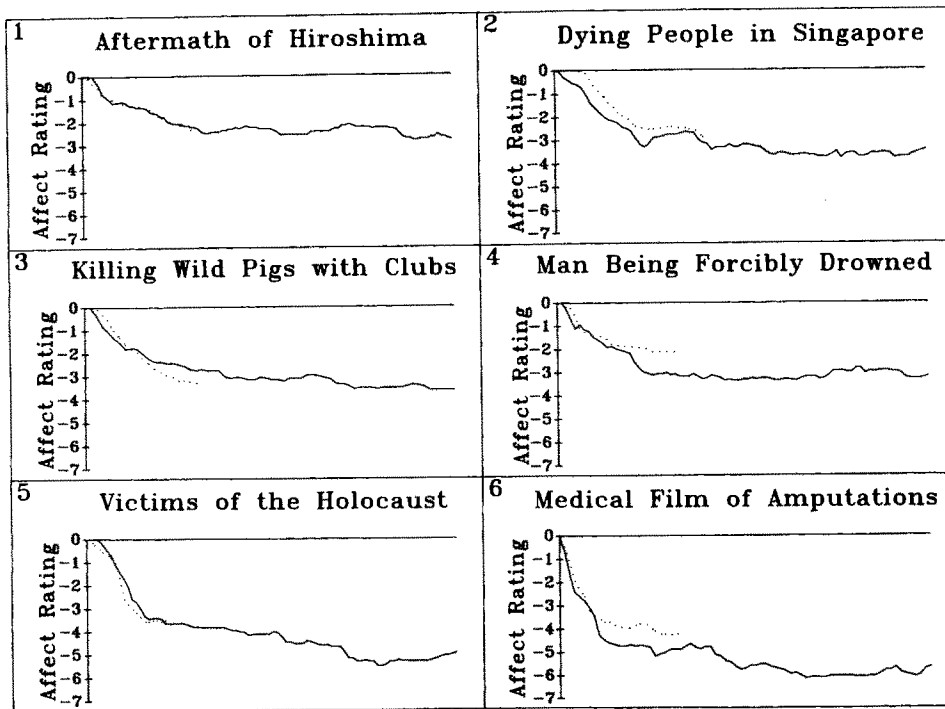


Figure 2. Mean second-by-second real-time ratings of affect for six aversive test clips. (Means were calculated separately across the 16 subjects who viewed long and short versions of each clip. Means for short clips are represented by dashed lines.)

that the mean real-time ratings remained positive throughout the waves film and almost to the end of the puppy film. The global evaluations of these two films therefore run counter to the logic of temporal monotonicity: Adding positively valued segments to a pleasant episode did not improve its global evaluation and could even make it worse. Thus, the results for pleasant films are compatible with a model in which global evaluations are derived by weighted averaging of selected moments and incompatible with a model of temporal integration.

As observed in the real-time measures (see Figure 2), results for the global evaluations of aversive films were quite consistent. Increasing the duration of these films always made global evaluations more negative (see Table 2). This effect was rather small, however, as indicated by the modest mean value (.26) of the within-subject biserial correlation between duration (short or long) and global evaluations. Looking across the mean biserial correlations in Table 2, it is notable that duration appears to have less effect on the global evaluations of aversive films than on real-time ratings of affect. Temporal integration models suggest the opposite pattern.

As we noted earlier, the observation that global evaluations covary with duration does not necessarily contradict the hypothesis of duration neglect. An averaging model for global evaluations would also yield a correlation with duration, if the real-time affective experience escalates over the course of the films, which it did for many of the films we used. Therefore, to more closely examine the possibility of duration neglect, we turned to an analysis of the within-subject correlations between retrospective global evaluations and various attributes of the real-time affective experience of viewing each film (i.e., peak

affect, end affect, and clip duration). We acknowledge, of course, that correlations and partial correlations based on six observations are not impressive statistical objects, but the average of 32 such measures can be quite informative. Table 3 presents the results of these analyses. For each coefficient, the null hypothesis that the mean of the individual correlations is zero was examined by *t* test.

Not surprisingly, the correlations indicate that global evaluations are predicted much more accurately by measures of real-time affect than by clip duration. The two measures of real-time affect, peak affect and end affect, were very highly correlated for the aversive films ($r = .91$) and more moderately so for

Table 3
Mean Within-Subject Correlations Between Global Evaluations and Aspects of the Real-Time Affective Experience

Correlation type and predictor	Variable controlled	Pleasant clips		Aversive clips	
		Mean <i>r</i>	<i>t</i> (30)	Mean <i>r</i>	<i>t</i> (30)
Zero-order					
Peak		.77	19.88**	.76	13.80**
End		.70	12.51**	.66	8.75**
Peak + end		.78	20.02**	.69	9.21**
Duration		.13	1.61	.25	3.44*
Partial					
Peak	End	.44	6.29**	.28	3.57*
End	Peak	.25	2.82*	.04	0.49
Duration	Peak + end	.06	0.72	-.02	0.18

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

the pleasant films ($r = .69$). The significant partial correlations suggest that both "moments" make separate contributions to the global evaluations of the pleasant films, but not of the aversive films. The fact that peak affect contributed to predictions independently of end affect also permits us to reject the possibility that subjects used the real-time ratings to express their cumulative global evaluation of each film, rather than their affect of the moment.

An unweighted combination of peak affect and end affect has been useful in predicting global evaluations in other studies (Redelmeier & Kahneman, 1993; Varey & Kahneman, 1992). Here again, this sum is an accurate predictor of evaluations for pleasant films, although for aversive films it is somewhat less accurate than peak affect alone (see Table 3). A particularly suggestive finding is that the mean of partial correlations of global evaluations with clip duration is near zero when the sum of peak affect and end affect is controlled. This is precisely the outcome that would be predicted for this test of duration neglect. Although gratifying, this apparently precise verification of the hypothesis should be treated with due caution, in view of the well-known vagaries of partial correlations, and of the fact noted earlier that the correlation of peak affect with duration is necessarily nonnegative. The best evidence for duration neglect, in our view, comes from the overall pattern of the data: the small and inconsistent effects of duration on the global evaluations for pleasant films and the finding that duration appears to be more strongly related to real-time negative affect than to global evaluations for aversive films. Note also that the conditions of this experiment could very well have made the attribute of time unusually salient: The distribution of film durations was notably bimodal, and evaluations were made immediately after the termination of each film, when feelings of impatience for it to end (or perhaps frustration that it ended too soon) were presumably still salient. The finding that duration had so little impact on global evaluations under these circumstances is remarkable.

Study 2

We conducted Study 2 both to extend the results of Study 1 and to test the speculation that the zero-order correlation between global evaluations and duration might be further reduced in global evaluations elicited after a delay. At first glance, the notion that a neglected attribute might become even more neglected with delay may seem odd, but it draws on plausible intuitions about escalating negative affect and how people might recall such affect. We suppose that escalation of negative affect during an aversive episode in part reflects the gradual depletion of people's resources for coping with negative events (Linville & Fischer, 1991), perhaps compounded by the uncertainty of not knowing how much longer the episode will last. In subsequent recall of the experience, however, it seems likely that such time-related feelings of exhaustion or impatience will be less salient than the retrieved images of the original episode and the affect that these images may evoke. Differential retrieval of the content of an episode and of these time-related aspects of the original experience could cause the representations of long and short versions of the same film to become increasingly similar as evaluation is delayed. Thus, we see no contradiction between the following two propositions: (a) Dura-

tion is neglected in immediate global evaluations, in the sense of a small or vanishing partial correlation between duration and evaluations when real-time affect is controlled; (b) duration is neglected even more by delaying global evaluations, in the sense that the zero-order correlation between duration and evaluations diminishes with delay.

We tested this idea in Study 2 by having subjects rank the films at the end of the experimental session. Subjects in the delayed condition were not informed in advance about the ranking task and presumably evaluated the films solely on the basis of their memories of them. In contrast, before viewing the films, subjects in the on-line condition reviewed the instructions for the ranking task they would later perform. Advance instructions should cause people to make evaluations as episodes unfold and encourage them to draw relevant information from current experience rather than from memory (Hastie & Park, 1986; Mackie & Asuncion, 1990).

Besides the timing of the global evaluations, Studies 1 and 2 differed in three other ways. First, subjects in Study 1 provided real-time reports of their affect, whereas those in Study 2 did not. This allowed us to explore whether providing real-time ratings contaminates later evaluations. Second, the main dependent variable in Study 1 was a rating of each film; in Study 2, it was a ranked choice. Third, subjects in Study 1 made global ratings of their own emotional experiences, whereas those in Study 2 considered how the films would affect most people. This measure provides a better index of subjects' retrospective evaluations of these films than would a question about their willingness to see them again (people often recommend to friends a movie that they have already seen, without being willing to join them to see it).

Method

Subjects

The subjects were 96 students (48 women and 48 men) at the University of California, Berkeley, paid \$10 for participating in a 1-hr study of emotion. They ranged in age from 18 to 57, with a median age of 21. Male and female subjects were randomly assigned to view one of the four stimulus videotapes under one of two evaluation conditions (on-line or delayed).

Visual Materials

Stimulus tapes were the same four used in Study 1.

Written Materials

Subjects provided their global evaluations of the films by ranking them at the end of the experimental session. They were told that researchers sought their advice in constructing one pleasant and one unpleasant videotape that other subjects would view in future research. The task was to rank the films according to the contribution each made to the overall affective experience of viewing the videotape. Pleasant and aversive clips were ranked separately. The ranking of the aversive clips was explained by the following instructions:

The unpleasant videotape will consist of several of the unrelated unpleasant clips that you just saw. Of course, people who will watch this unpleasant videotape will experience some unpleasantness. Yet our goal is to MINIMIZE the overall unpleasantness of the experience of viewing the unpleasant videotape

that we make. Your task is to help us select the unpleasant clips we should include in order to MINIMIZE the overall experience of unpleasantness for most people.

Instructions for ranking the pleasant film clips were similar, with the stated goal "to MAXIMIZE the overall pleasantness of the experience of viewing the pleasant videotape that we make." Note that both ranking tasks invoke a benevolent attitude and create no moral conflict for the subjects. Subjects were asked to rank eight clips in each task (the six test clips from the 12 listed in Tables 1 and 2, along with two other clips of the same valence). They assigned a rank of 1 to the clip that was their first choice for inclusion and 8 to the clip that they most wanted to keep out of the videotape. (Rankings for pleasant clips are reversed in all analyses so that higher values indicate greater overall pleasantness.)

Procedure

Subjects viewed film clips and made their judgments in small-group sessions (ranging from 1 to 9 individuals) run by a male experimenter. At the beginning of the session, the experimenter distributed an overview of the experiment and asked subjects to read it carefully. The first paragraph of this overview was identical for all subjects and simply described the scope and general procedures of the study ("This is a preliminary study of people's emotional responses to films. We will be showing you 16 short film clips that range from very pleasant to very unpleasant"). The text that followed this general introduction provided the experimental manipulation. In the on-line condition, this text included a full preview of the instructions for the ranking task that would be performed at the end of the session. It concluded with the request to

Please watch the clips with an eye toward helping us choose which of the pleasant clips would MAXIMIZE the overall experience of pleasantness, and which of the unpleasant clips would MINIMIZE the overall experience of unpleasantness for most people.

In the delayed condition, subjects were only given the following information:

For purposes of research, we are interested in your own emotional responses and also in what you think the responses of most people would be. Your task later in the hour will be to answer some questions about the experience of viewing these clips. Please watch the clips with an eye toward what you think most people's emotional responses would be.

Whenever possible, both the on-line and the delayed conditions were represented in each group screening. Subjects viewed one of the four stimulus videotapes in a darkened room and completed the two ranking tasks after viewing the entire sequence. The order of the ranking tasks was counterbalanced.

Results and Discussion

Tables 1 and 2 included descriptive statistics of the rankings for the on-line and delayed conditions of Study 2 and separate summary statistics for the effects of duration in the two conditions.

We first explored whether the procedural differences between Studies 1 and 2 influenced subjects' average global assessments of the films. Recall that in Study 1, global evaluations were obtained using scale-free ratings made immediately after each film; in Study 2, they were obtained by having subjects rank the films at the end of the session. To compare responses across the two studies, we simply computed correlations be-

tween the means of these various global evaluations (reported in Tables 1 and 2), separately for the 12 pleasant clips (i.e., short and long versions of six test films) and for the 12 aversive clips. These correlations describe the degrees to which the relative orderings of the films were similar across different evaluative circumstances.

Despite the differences in the ways global evaluations were obtained, the mean rankings made by subjects in the on-line condition in Study 2 were quite similar to the mean ratings made by subjects in Study 1. The correlations between the entries in the relevant columns of Tables 1 and 2 were .81 for the pleasant clips and .86 for the aversive clips. This correspondence may not be unduly surprising, however, because subjects in both Study 1 and in the on-line condition of Study 2 were aware as they watched the films that they would later provide global evaluations of them.

The mean rankings of the aversive clips made by subjects in the delayed condition in Study 2 were also quite similar to the evaluations made under other evaluative circumstances: The correlations with the mean evaluations in Study 1 and in the on-line condition were .86 and .85, respectively. This suggests that even when evaluations were made from memory, the average evaluations of these aversive film clips remained largely unchanged.

Surprisingly, however, the delay procedure made a larger difference on the evaluations of the pleasant clips: The correlations of mean rankings made in the delayed condition of Study 2 with the mean evaluations made in Study 1 and in the on-line condition were .48 and .58, respectively. This finding suggests that the average evaluations of these pleasant clips changed, depending on when evaluations were made. Related evidence for this idea comes from the finding that the mean rankings for the on-line and delayed groups differed beyond the .05 level for 5 of the 12 pleasant clips listed in Table 1 (for aversive clips, only one difference reached significance).

It seems that when evaluated from more distant memory, some of the pleasant films became more pleasant (e.g., the films of the puppy and of penguins), whereas others became relatively less pleasant (e.g., the films of waves and of flying). We speculate that such differences might arise if subjects in the delayed condition did not necessarily recall the emotion they experienced during the films, but rather reconstructed their affective experience on the basis of the mere idea of each film, perhaps along with some of its images. It is plausible that the idea of a puppy playing with a flower is more appealing than the idea of waves breaking on a beach, even if the latter film was actually more pleasant to watch. This reasoning suggests that delaying evaluations may have had a smaller effect on the evaluations of aversive films, because the affect associated with the ideas of each of these films and with the actual experience of viewing them may have been more comparable than for the pleasant films. Although we did expect that subjects making on-line versus memory-based evaluations would rely on different sources of information (Hastie & Park, 1986), we had not anticipated that the probes used to elicit retrospective evaluations (the written descriptions of the films as shown in Tables 1 and 2) might operate differently for pleasant and aversive clips.

The main hypothesis to be tested in Study 2 was that any correlation between global evaluations and film duration could be reduced by delaying evaluations. In Study 1, we observed

significant relationships between clip duration and global evaluations for aversive films, but not for pleasant films; thus, we could only expect duration effects to become weaker for the aversive films.

In general, we expected the duration effects obtained in the on-line condition of Study 2 to be comparable to those obtained in Study 1, again because in both cases subjects were aware of the subsequent evaluation task as they watched the films. This turned out to be the case: Table 1 shows that for pleasant films, the mean within-subject biserial correlations with duration for the on-line rankings in Study 2 and the global ratings in Study 1 were each low and nonsignificant (.06 and .10, respectively); whereas Table 2 shows that for aversive films these mean biserial correlations were slightly higher and significant (.24 and .26, respectively). The similarities between Studies 1 and 2 illustrate that duration neglect generalizes across different evaluative circumstances and is not caused by the requirement of reporting momentary affective experience.

We expected and found duration effects to be quite small in the delayed condition: The mean biserial correlations were .11 for pleasant films and .12 for aversive films. The prediction that the correlation between global evaluations and duration for aversive films would be decreased in the delayed condition relative to the on-line condition was confirmed (.12 vs. .24), although this difference only reached marginal significance, $t(94) = 1.90, p < .10$.

In evaluating the hypothesis of duration neglect, it is important to note that the delayed condition in Study 2 may best represent the circumstances in which people typically make retrospective evaluations. After all, we do not normally go through episodes of life continually articulating our momentary experience, as did subjects in Study 1, or with the explicit intent of constructing a global evaluation of the episode later, as did subjects in the on-line condition in Study 2. We believe that the results of the delayed condition provide a realistic estimate of the restricted role that duration plays in people's everyday evaluations of their past experiences.

General Discussion

The central finding of the present research was that the duration of affective episodes had little effect on people's retrospective evaluations of their experiences. For pleasant experiences, we found no systematic effects of film duration. For aversive experiences, effects of episode duration were (a) small, (b) entirely explained by the observed escalation of negative affect during the films, and (c) reduced by delay in probing the memory of the episode.

Two recently completed studies have extended the observation of duration neglect from affective experiences to physical pain. Redelmeier and Kahneman (1992) asked patients undergoing colonoscopy to rate their discomfort at 1-min intervals from the insertion of the colonoscope to its removal. Across 101 patients, the duration of this procedure ranged from 4 to 69 min, and discomfort ratings varied substantially from minute to minute. Patients' retrospective evaluations of the procedure, obtained both immediately and 1 month later, were predicted with good accuracy by an unweighted combination of the peak intensity of discomfort and of the discomfort reported during the last few minutes of the procedure. The duration of the pro-

cedure was not a significant predictor of global evaluations made by the patients themselves or by the attending physician and nurse.

Subjects in another experiment (Kahneman et al., in press) endured two cold-pressor experiences in the course of an experimental session: a short trial in which they immersed one hand in cold water (maintained at 14 °C) for 60 s; a long trial in which they kept the other hand in cold water for a total of 90 s, the first 60 s at 14 °C as in the short trial, then 30 s more of diminishing discomfort, during which water temperature was gradually raised by 1 °C. When given a choice of which of the two trials to repeat, most subjects preferred the long trial, even though it contained all the discomfort of the shorter trial and then some. This preference was evidently determined by discomfort at the end of the trial, as if duration did not matter.

Duration neglect does not imply that people do not appreciate or remember how long their experiences last. The subjects in our experiments and the patients in the Redelmeier-Kahneman study were quite aware of the duration of their experience, yet they did not give this attribute much weight in their evaluations. Nor does the neglect of duration in evaluations reflect a general view that time is not an important attribute of experiences. To the contrary, the severities of childbirth, hospital stays, and jail sentences are commonly measured in units of duration. Rather, we view duration neglect as an attentional phenomenon: Although people may be aware of duration and consider it important in the abstract, we suggest that what comes most readily to mind in evaluating episodes are the salient moments of those episodes and the affect associated with those moments. Duration neglect might be overcome, we suppose, by drawing attention more explicitly to the attribute of time.

Although converging evidence for duration neglect in retrospective evaluations now appears quite strong, several qualifications are in order. First, it should be obvious that caution is required in extrapolating to the more interesting and important cases of retrospective evaluation, in which episodes are measured in days or months, rather than in seconds or minutes. For example, how might retrospective evaluations of prolonged aversive episodes change as individuals gradually adapt to adversity? If, for instance, convicts' adjustment to prison improves from year to year, do their global views of their own prison experiences also improve over time? How does this relate to the ex-convict's willingness to return to prison? Do longer sentences actually make prison life seem less bad on the whole, perhaps encouraging recidivism?

Another qualification concerns the range of factors that might predict retrospective assessments. We have found in several studies that global evaluations of affectively homogeneous episodes were reasonably well predicted by an unweighted combination of the most extreme affect experienced during the episode and by affect at the end moments. It is quite certain, however, that this peak and end rule does not exhaust the factors that govern global evaluations. Other variables, such as the velocity of changes for the better or for the worse may also be important (Hsee & Abelson, 1991; Hsee et al., 1991). Additional complexities arise in analyzing the role of anticipatory emotions, such as hope and fear, in retrospective evaluations made after uncertainty is resolved. For example, Carmon and Kahneman (1992) measured both real-time affect and retro-

spective evaluations in situations of waiting in simulated queues. They found that joining an unusually short queue elicits substantial positive affect, interpreted as hope for a quick ending of the episode. Affect remained positive for quite a while, even if the queue moved slowly. However, when the prospect of a short wait was eventually given up, momentary affect became sharply negative. The global evaluations reflected only the final disappointment, not the time spent in pleasant anticipation. In general, the relative impact of how people feel during endings is likely to depend on the meaning attached to the outcomes of episodes.

Finally, a comprehensive analysis of retrospective evaluations must recognize the subtleties of people's memory for emotion. Clearly, retrospective accounts of emotional experience depend greatly on the probes and cues offered at the time of retrieval. The results we present suggest that abstract global queries, which are common in both research and everyday conversations ("How was the movie?" ". . . your trip to the dentist?" or ". . . your trip to Portugal?"), may adequately capture the most intense or ending moments of extended affective episodes, or both, but neglect episode duration, and probably also gloss over the distinct emotions that may have been experienced over the course of the episode. There is growing consensus among emotion researchers that one solution is to present people with concrete, time-yoked cues about their past experiences, allowing them to relive these episodes to some degree and thus provide richer retrospective evaluations (e.g., cued-review method used by Rosenberg & Ekman, in press, and video-recall method used by Fredrickson, 1991, and originally by Levenson & Gottman, 1983).

As the preceding comments have shown, the study of the process by which people make retrospective evaluations has barely begun. Understanding this process is important because what people think about the past often determines what they do about the future. The phenomenon of duration neglect suggests the sobering conclusion that people's memories for their past experiences may not always be good guides for their choices among future actions.

References

- Anderson, N. H. (1991). Functional memory in person cognition. In N. H. Anderson (Ed.), *Contributions to information integration theory: Cognition*. (Vol. 1, pp. 1-55). Hillsdale, NJ: Erlbaum.
- Carmon, Z., & Kahneman, D. (1992). *Real-time and retrospective evaluations of queuing in simulated lines*. Unpublished manuscript, University of California, Berkeley.
- Coombs, C. H., & Avrunin, G. S. (1977). Single-peaked functions and the theory of preference. *Psychological Review*, *84*, 216-230.
- Ekman, P., & Friesen, W. V. (1974). Detecting deception from body or face. *Journal of Personality and Social Psychology*, *29*, 288-298.
- Fredrickson, B. L. (1991). Anticipated endings: An explanation for selective social interaction (doctoral dissertation, Stanford University, 1990). *Dissertation Abstracts International*, *3*, AAD91-00818.
- Fredrickson, B. L., & Levenson, R. W. (1992). *Positive emotions speed recovery from negative emotional arousal*. Manuscript submitted for publication.
- Gottman, J. M., & Levenson, R. W. (1985). A valid procedure for obtaining self-report of affect in marital interaction. *Journal of Consulting and Clinical Psychology*, *53*, 151-160.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, *93*, 258-268.
- Hsee, C. K., & Abelson, R. P. (1991). Velocity relation: Satisfaction as a function of the first derivative of outcome over time. *Journal of Personality and Social Psychology*, *60*, 341-347.
- Hsee, C. K., Abelson, R. P., & Salovey, P. (1991). The relative weighting of position and velocity in satisfaction. *Psychological Science*, *2*, 263-266.
- Johnson, M. K., & Sherman, S. J. (1990). Constructing and reconstructing the past and the future in the present. In E. T. Higgins & R. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 482-526). New York: Guilford Press.
- Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (in press). When more pain is preferred to less: Adding a better end. *Psychological Science*.
- Kahneman, D., & Snell, J. (1990). Predicting utility. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 295-310). Chicago: University of Chicago Press.
- Kundera, M. (1991). *Immortality*. New York: Grove Press.
- Levenson, R. W. (1988). Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity. In H. L. Wagner (Ed.), *Social psychophysiology and emotion: Theory and clinical applications* (pp. 17-42). London: Wiley.
- Levenson, R. W., & Gottman, J. M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology*, *45*, 587-597.
- Linville, P. W., & Fischer, G. W. (1991). Preferences for separating or combining events. *Journal of Personality and Social Psychology*, *60*, 5-23.
- Loewenstein, G., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review*, *100*, 91-108.
- Mackie, D. M., & Asuncion, A. G. (1990). On-line and memory-based modification of attitudes: Determinants of message recall-attitude change correspondence. *Journal of Personality and Social Psychology*, *59*, 5-16.
- "Manhattan Project," 1992. (1992, November/December). *Newsweek* (special election issue), pp. 40-42, 55-56.
- Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, *97*, 225-252.
- McFarland, C., Ross, M., & DeCourville, N. (1989). Women's theories of menstruation and biases in recall of menstrual symptoms. *Journal of Personality and Social Psychology*, *57*, 522-531.
- Redelmeier, D. A., & Kahneman, D. (1993). *On the discrepancy between experience and memory: Patient's real-time and retrospective evaluations of the pain of colonoscopy*. Manuscript submitted for publication.
- Rosenberg, E. L., & Ekman, P. (in press). Coherence between expressive and experiential systems in emotion. *Cognition and Emotion*.
- Varey, C., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making*, *5*, 169-185.

Received August 3, 1992

Revision received January 7, 1993

Accepted January 27, 1993 ■