Research Report

WHEN MORE PAIN IS PREFERRED TO LESS: Adding a Better End

Daniel Kahneman,[1] Barbara L. Fredrickson,[2] Charles A. Schreiber,[1] and Donald A. Redelmeier[3]

[1]*University of California,* [2]*Duke University, and* [3]*University of Toronto*

Abstract—*Subjects were exposed to two aversive experiences: in the short trial, they immersed one hand in water at 14 °C for 60 s; in the long trial, they immersed the other hand at 14 °C for 60 s, then kept the hand in the water 30 s longer as the temperature of the water was gradually raised to 15 °C, still painful but distinctly less so for most subjects. Subjects were later given a choice of which trial to repeat. A significant majority chose to repeat the long trial, apparently preferring more pain over less. The results add to other evidence suggesting that duration plays a small role in retrospective evaluations of aversive experiences; such evaluations are often dominated by the discomfort at the worst and at the final moments of episodes.*

Decisions are often controlled by hedonic predictions. We choose the option that will cause most pleasure, or least pain—in Jeremy Bentham's terms, the option that will yield the greatest utility. Hedonic prediction usually relies on memories of previous experiences: We expect to like what we remember as pleasant and to dislike what we remember as unpleasant. How accurate are these evaluations of past experiences? Do they provide good guides for future decisions?

When we ask a friend who has recently returned from the Bahamas, or from the dentist, "How was it?" or "Was it better than last time?" we assume that the friend knows the answer. Retrospective assessments of the utility of past experiences are accepted in everyday interaction with almost as much confidence as the answers to questions about the affect of the moment: "Are you enjoying this?" or "Does it hurt?" This confidence could be unwarranted because two fallible mental processes separate retrospective assessments from the sequence of experiences that constituted the original episode: an operation of memory and an act of evaluation. Some recent research has called into question the accuracy of people's memories for their hedonic and affective experiences (Kent, 1985; Rachman & Eyrl, 1989; Thomas & Diener, 1990). This article focuses on the process of evaluating past episodes of pain.

Some rules for the evaluation of episodes have the appeal of logical principles. The most compelling is a rule of *temporal monotonicity*, which requires that adding moments of pain to the end of an episode can only make the episode worse, and that adding moments of pleasure must make it better. As we shall see, however, the psychology of evaluation does not obey this rule.

In one investigation (Varey & Kahneman, 1992), subjects made global evaluations of episodes of discomfort suffered by another person. The subjects were shown a series of "discomfort ratings" on a scale from 0 to 10; these ratings were purportedly made by an individual at 5-min intervals during an unpleasant experience (e.g., exposure to loud drilling noise). The episodes to be evaluated varied in duration, in average intensity, and in the temporal trend of the discomfort. Global evaluations were highly sensitive to intensity and to trend: An unweighted combination of peak discomfort and of the discomfort at the end of the episode accounted for 94% of the variance. The effect of duration, though statistically significant, was remarkably small: Adding this factor raised R^2 by only a further 3%. The neglect of duration and the emphasis on endings led to predictable violations of 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 judged much less aversive than the series 2-5-8, even though the only difference between the two episodes was the 5 extra min of discomfort in the former.

More recently, we have extended this research to the retrospective evaluation of episodes of pleasure or discomfort that subjects experience themselves. In the first of these studies (Fredrickson & Kahneman, 1993), subjects viewed a series of short, plotless films, varying in content from pleasant (penguins at play) to highly aversive (an amputation). There were two versions of each film, one three times longer than the other. Each subject saw the long version of some films and the short version of others. Subjects provided continuous ratings of affect while watching each film and assessments of overall pleasure or discomfort at its end. The results of this analysis were strikingly similar to those of the earlier (Varey & Kahneman, 1992) study: Retrospective evaluations were well predicted by a weighted average of the peak affect rating and the final rating recorded for each film; the duration of the film did not emerge as an independent predictor of the overall evaluation.

A subsequent study extended these findings to the retrospective evaluation of a painful medical procedure (Redelmeier & Kahneman, 1993). Patients undergoing diagnostic colonoscopy indicated their current discomfort every 60 s during the procedure. They also provided retrospective evaluations of the procedure, both immediately and 1 month later. Again, a combination of the ratings of the worst and the final moments of the colonoscopy predicted subsequent evaluations with substantial accuracy. The duration of the procedure, which varied between 4 min and 69 min for different patients, did not significantly affect any of the retrospective judgments. The attending physician and nurse also provided independent retrospective evaluations of each patient's

Address correspondence to Daniel Kahneman, Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08544-1013.

overall discomfort, without access to the patient's ratings. As might be expected from the earlier results (Varey & Kahneman, 1992), the judgments of these observers were also dominated by the worst and the final moments of the procedure, and unrelated to its duration.

The common finding of these studies is the relative neglect of duration in retrospective evaluations and the successful prediction of the global disutility of an extended aversive episode by the disutility of two singular moments—the peak and the end of the experience. The results suggest an averaging model for global evaluations (Anderson, 1991), which implies violations of temporal monotonicity. As illustrated by the fictitious sequences 2-5-8 and 2-5-8-4 mentioned earlier (Varey & Kahneman, 1992), the evaluation of an aversive episode can be improved by adding to it a period of diminishing discomfort. In the present study, we tested whether this process can lead subjects to prefer more pain over less pain in a direct choice.

Subjects had two separate unpleasant experiences in the course of an experimental session: a short trial in which they immersed one hand in water at the moderately painful temperature of 14 °C for 60 s and a long trial in which they immersed the other hand in water at 14 °C for 60 s, then kept the hand immersed 30 s longer while the temperature of the water was raised slightly, still within the uncomfortable range. Thus, the long trial included all the discomfort of the short trial, plus an extra period of slowly diminishing discomfort. The subjects expected to have a third unpleasant experience during the session, and they were given a choice of whether to repeat the first or the second trial. Our hypothesis was that subjects would retain a more favorable memory of the long episode because it ended at a lower level of discomfort, and that they would consequently choose to repeat that episode.

METHOD

Subjects

Thirty-two male University of California students, age 19 to 39 (median age = 22.5), were paid $10 for a 1-hr session. Participants were screened for health problems and use of drugs, including tobacco. Three of the subjects replaced others whose data were discarded, 1 because of technical difficulties with temperature control, and 2 because they did not indicate a consistent preference between the two trials.

Apparatus

A plastic tub in which subjects immersed their hand was filled to a depth of 11 cm with 7 L of water cooled to 14.1 °C (±0.3 °C). To maintain a constant temperature and a slight agitation of the water, an external pump circulated water from the tub through an aluminum coil submerged in ice water. Water temperature was controlled also by using another pump to circulate water through a coil submerged in room-temperature water (21 °C ± 1.1 °C). By simultaneously turning off the first pump and turning on the second, the water temperature in the subject's tub could be increased by 1.1 °C (±0.3 °C) in 30 s. The switching of pumps was not audible and produced no noticeable change in tub circulation. Pumps, coils, and switches were not visible to the subject.

An on-line measure of discomfort was obtained using a "discomfort meter," which consisted of a potentiometer and a linear array of 15 light-emitting diodes (LEDs). A single green LED at one end of the display remained lit at all times. By adjusting the potentiometer, subjects could control the number of red LEDs that were lit, thereby indicating the level of discomfort. The potentiometer was sampled five times per second, and the 1-s means were recorded by a computer, which also recorded water temperature. Discomfort values could range between 0 and 14.

Procedure

Subjects were tested individually by a female experimenter. They were told that the experiment concerned judgments of discomfort and that they would be asked to place a hand in a tub of cold water on three separate occasions. The cover story was that the study dealt with lateral differences in the experience of discomfort. As part of the consent procedure, subjects were asked to immerse both hands in the cold-water tub for 5 s before agreeing to participate. They were given no indication that the trials would differ, except that they were to use one hand in the first experience and the other hand in the second. The order of the long and short trials and their assignment to the dominant or nondominant hand were counterbalanced across subjects.

Immediately before each trial, subjects immersed both hands in room-temperature water for a 2-min baseline period. After each trial, they spent 7 min in a waiting area working on a personality inventory. Before the expected third trial, they were told that we needed their impressions of the first two trials because they would choose one of them to be repeated. They were then given a questionnaire titled "Impressions of Cold-Water Trials." The first question was, "Suppose we paid you to come back tomorrow to repeat just one of the two cold-water trials that you've experienced today. Which one would you choose?" The choice referred to the first and the second trials. The next question was, "For today's third trial, you can pick which of the previous cold-water trials you will repeat. Which one do you choose?" (Two subjects who gave inconsistent responses to these two questions were replaced.) Subjects then compared their two experiences using four Likert scales (ranging from −5 to +5). They were asked "Which trial caused the greater overall discomfort?" "Which trial lasted longer?" "At its most extreme moment, which trial was colder?" and "Which trial was tougher for you to cope with?" Finally, subjects depicted the discomfort they felt "moment-by-moment during each trial" by drawing a continuous line across a Discomfort × Time chart provided by the experimenter. Subjects were then informed that there would be no third trial and were fully debriefed.

RESULTS

Real-time measures of discomfort were essentially identical for the short trial and for the first 60 s of the long trial: The mean responses recorded at 60 s were 8.44 for the short trial and 8.34 for the long trial. The gradual increase of

water temperature (from a mean of 14.1 °C to 15.2 °C) during the final 30 s of the long trial caused a pronounced drop of the discomfort measure (M = 2.65 for the change score, $t[31] = 6.80$, $p < .01$). The high sensitivity of pain to minor changes of temperature in this range confirms prior results (Cabanac, 1981). However, not all subjects responded alike to the temperature change: Eleven of the 32 subjects indicated a decrement of discomfort of 1 point or less, and 2 of these subjects even reported an increase of discomfort during the last 30 s of the long trial. Figure 1 shows the time course of the discomfort measure in the long trial for these 11 subjects and for the majority who showed a larger decrement. Note that, even for the latter, the experience at the end of the long trial was still distinctly unpleasant.

The main dependent variable was the subject's choice for the third trial. As predicted, most subjects (22 of 32, or 69%) preferred to repeat the long trial ($z = 2.15$, $p < .05$ by sign test). Note that this proportion would be zero if subjects acted to minimize their exposure to pain. Additional tests showed that choices did not depend on whether the long trial was experienced first or second or with the dominant or nondominant hand. As might be expected, a preference for the long trial was correlated with the decrement of discomfort indicated during the last 30 s of that trial ($r_{bis} = .38$, $p < .05$). Among the 21 subjects who showed a decrement of 2 or more points, 17 (or 81%) preferred the long trial; the 11 subjects who showed little or no decrement of discomfort split 6:5 in favor of the short trial.

The comparative ratings that subjects provided after stating their choice were usually consistent with their decisions, but not always with the facts. Thus, most subjects indicated that the long trial had caused less overall discomfort ($M = -0.91$, $t[31] = 2.12$, $p < .05$), was less cold at its most extreme moment ($M = -0.91$, $t[31] = 1.90$), and was less tough to cope with ($M = -1.12$, $t[31] = 2.90$, $p < .01$). Since the long trial contained all the pain of the short trial and then some, these postchoice judgments are simply wrong. The bias in favor of the long trial may have affected some judgments of duration: six subjects reported that the long trial was actually shorter, and 9 did not report any difference. On average, however, the relative duration of the two trials was judged correctly ($M = 1.09$, $t[31] = 3.27$, $p < .01$). The duration difference evidently did not loom large in subjects' choices, although one was heard to mutter after comparing the duration of the two trials, "The choice I made doesn't seem to make much sense."

The correlations between subjects' choices and their postchoice comparisons of the two trials support two conclusions: First, subjects almost always chose to repeat the trial that they remembered as being easier; the biserial correlation was .80 between choice and the comparison of overall discomfort. Second, the neglect of duration is confirmed by a correlation of only .16 between choice and the comparison of durations. The intercorrelations among the comparative judgments tell the same story. The (mostly erroneous) judgments of which trial included the coldest temperature correlated .69 with ratings of overall discomfort and .62 with ratings of "tough coping." In contrast, the (mostly veridical) judgments of duration correlated only .08 with rated discomfort and .18 with rated difficulties of coping. Subjects evidently felt little pressure to distort their judgments of duration to fit their global impression of the trials.

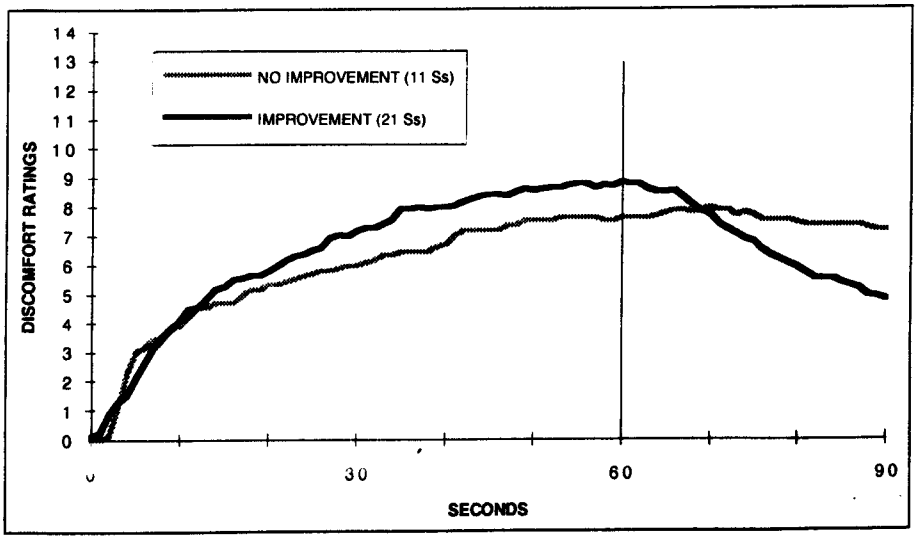

Fig. 1. Mean of real-time discomfort measure on the long trial, for 11 subjects who indicated little or no decrement of discomfort when temperature changed and for 21 subjects who indicated decreased discomfort.

GENERAL DISCUSSION

The present results are compatible with the peak-and-end pattern we have observed before, in which an average of the real-time responses to the worst and to the final moment predicts the retrospective evaluation of an aversive episode with fair accuracy, whereas duration is relatively neglected. This pattern entails different results for those subjects who experienced diminishing pain in the long trial and for those who did not. For the typical subject, the worst moments of the short and of the long trial were about equally bad, but the final moment was better in the long trial. A weighted average of these momentary utilities would therefore yield a more favorable evaluation of the long trial, as was found. A minority of subjects indicated no lessening of discomfort in the long trial; their worst discomfort and their final discomfort were therefore approximately the same within each trial, and similar across the two trials. The peak-and-end pattern predicts that the long and the short trials should be about equally aversive for these individuals, as was found. Thus, the peak-and-end pattern explains both the cases in which our initial prediction was confirmed and those in which it appeared to fail.

We suspected that the requirement to report affect in real time could enhance the salience of the worst and the final moments, but the results do not depend

on this feature of the design. The strong preference for the long trial was confirmed in a replication of the present experiment in which the real-time measure of discomfort was eliminated: Of 37 participants in that replication, 24 chose the long trial. We have also found in two other studies (Fredrickson & Kahneman, 1993; Redelmeier & Kahneman, 1993) that the peak-and-end pattern and the neglect of duration were maintained even when subjects did not provide explicit evaluations of ongoing experience.

We do not propose that duration is always neglected: It seems likely that duration could play a role in the evaluation of affective episodes that are either very much longer or very much shorter than expected. Nor do we propose that peak affect and end affect are the only relevant predictors of retrospective evaluation. For example, the velocity of an improving or deteriorating trend has been shown to be a factor in evaluations (Hsee & Abelson, 1991; Hsee, Abelson, & Salovey, 1991), and there are surely others. We suspect that there are cases in which memories of an episode are dominated by its initial moments. There are also situations in which the peak may be discounted and only the end matters: The positive affect associated with hope for a good outcome may not be counted in retrospective evaluations of an episode that eventually ends in disappointment (Carmon & Kahneman, 1993; see also Fredrickson, 1991).

We view the peak-and-end pattern as an instance of the broader proposition that people tend to use selected moments as proxies in evaluating temporally extended states or episodes. This proposition applies to both prospective and retrospective evaluations. For example, evidence from studies of decisions about monetary gambles suggests that the effective carriers of utility are changes of wealth (gains and losses), not states of wealth (Kahneman & Tversky, 1979). The behavior of subjects in simple exchanges also indicates that choices are governed by the affect associated with obtaining an attractive object or giving it up, not by the long-term utility of owning the object or retaining a sum of money (Kahneman, Knetsch, & Thaler, 1991). A general principle of mental representation may be involved: Just as the visual system appears to describe objects in terms of boundaries and singular points, the cognitive system may represent extended experiences in terms of transitions and singular moments. Note that we refer here to representations that are formed for the purpose of evaluation. We do not claim that other information is necessarily lost, only that it is often not used. Indeed, most subjects in the present experiment could correctly retrieve the relative durations of the two trials, but did not use that knowledge in making their choices. Thus, the neglect of duration that we have observed is an attentional phenomenon; it does not represent an inability to use duration as a cue to decision or a general policy to ignore this attribute.

In the present experiment, the neglect of duration led most subjects to expose themselves to more pain rather than less. Nothing in the subjects' comments referred to a particular advantage in the long trial that made the extra pain worthwhile. Furthermore, we do not believe that subjects who choose the long trial would actually prefer to keep their hand in slowly warming cold water, if after 60 s at 14 °C they were offered the alternative of a dry towel. In the absence of any valid reason for the choice, the preference for the long trial must be viewed as a violation of temporal monotonicity—and as a mistake.

We conclude that subjects chose the long trial simply because they liked the memory of it better than the alternative (or disliked it less), not because they were willing to suffer for the sake of obtaining a more favorable memory. As they normally do in other choices, we suppose, our subjects trusted their retrospective evaluations of the two episodes as a basis for a decision: What could be wrong with repeating the experience one now likes best? Indeed, evaluated memories are the only available guide for many decisions, but our experiment has shown this guide to be fallible. It is part of the human condition that people prefer to repeat the experiences that have left them with the most favorable memories—not necessarily the experiences that actually gave the most pleasure and the least pain.

A better understanding of the rules of retrospective evaluation could yield some valuable applications. For example, the peak-and-end rule suggests that the memory of a painful medical treatment is likely to be less aversive if relief from the pain is gradual than if relief is abrupt. A related hypothesis is that the provision of relief in the context in which pain has been experienced will yield a more favorable memory than immediate transition to a new context as the pain ends. These are meaningful issues in medical care, given the general availability of analgesics that vary in onset, duration, and strength. Furthermore, memories of treatment can affect medical outcomes if they influence patients' morale and compliance with treatment recommendations (Redelmeier & Kahneman, 1993).

While it offers new opportunities, the dissociation of retrospective evaluations from immediate experience also raises intricate dilemmas of informed consent. Consider, for example, a direct extension of the present study to a medical context: Will a physician be allowed to add an interval of diminishing pain to the end of a medical procedure if the sole benefit of the added pain is to cause patients to retain a more favorable memory of it? The answer is likely to depend on how the patient is informed. On the one hand, it is safe to assume that few patients will agree to expose themselves to pain for the sole purpose of improving a future memory. Thus, informed consent would probably be denied. On the other hand, the present results imply that patients who have actually experienced both versions of the procedure—a form of knowledge that is generally considered superior to a mere description—will generally prefer to repeat the longer one. What weight should be given to a choice that is informed by personal experience if this choice can be traced to a faulty evaluation process? The ethical question of which of these conflicting preferences should be considered authoritative may not have a straightforward answer.

Acknowledgments—This research was supported by grants from the Sloan Foundation and the National Science Foundation. We thank George Loewenstein and Amos Tversky for helpful suggestions and comments. We are particularly grateful to Kathy L. Knopoff, who was involved in the study from the early stages of pretesting and was the sole experimenter.

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.

Cabanac, M. (1981). Physiological signals for thermal comfort. In K. Cena & J. Clark (Eds.), *Bioengineering, thermal physiology and comfort* (pp. 181–192). Amsterdam: Elsevier.

Carmon, Z., & Kahneman, D. (1993). *Real-time and retrospective accounts of simulated waiting lines*. Manuscript in preparation, University of California, Berkeley.

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., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology, 65*, 45–55.

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.

Kahneman, D., Knetsch, J., & Thaler, R. (1991). The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives, 5*, 193–206.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47*, 313–327.

Kent, G. (1985). Memory of dental pain. *Pain, 21*, 187–194.

Rachman, S., & Eyrl, K. (1989). Predicting and remembering recurrent pain. *Behaviour Research and Therapy, 27*, 621–635.

Redelmeier, D.A., & Kahneman, D. (1993). *On the discrepancy between experiences and memories: Patients' real-time and retrospective evaluations of the pain of colonoscopy*. Manuscript submitted for publication.

Thomas, D.L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology, 59*, 291–297.

Varey, C., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making, 5*, 169–195.

(RECEIVED 12/20/92; ACCEPTED 3/5/93)