

Evaluations of pleasurable experiences: The peak–end rule

AMY M. DO, ALEXANDER V. RUPERT, AND GEORGE WOLFORD
Dartmouth College, Hanover, New Hampshire

Prior research suggests that the addition of mild pain to an aversive event may lead people to prefer and directly choose more pain over less pain (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). Kahneman et al. suggest that pain ratings are based on a combination of peak pain and final pain. Similarly, people rate a happy life that ends suddenly as being better than one with additional years of mild happiness (Diener, Wirtz, & Oishi, 2001), even though the former objectively consists of less pleasure than the latter. Applying these concepts to material goods, we investigated the impact of positivity and timing on the retrospective evaluations of material goods. We found strong evidence that the peak–end rule applies to both material goods and pain.

Kahneman, Fredrickson, Schreiber, and Redelmeier (1993) found that participants preferred 60 sec of 14°C ice water followed by 30 sec of 15°C ice water to 60 sec of 14°C ice water alone. This result is counterintuitive, since water at both 14°C and 15°C is unpleasant. People seemed to prefer and directly choose more pain over less pain in this situation. The authors suggested that pain ratings are influenced by the peak pain experienced during the episode and the final level of pain—the peak–end rule. The peak–end rule for pain ratings has been confirmed in several other experiments. Redelmeier and Kahneman (1996) found that patients rated colonoscopies as less unpleasant if an interval of mild pain was added to the end of the procedure. Patients' memories of the pain experienced during a colonoscopy were correlated with predictions of the peak–end rule (Redelmeier & Kahneman, 1996). Schreiber and Kahneman (2000) showed that ratings of unpleasantly loud sounds showed clear peak–end effects. Fredrickson (2000) reviewed a large number of experiments supporting the peak–end rule. The peak–end rule yields the counterintuitive result that conditions can readily be arranged in which people prefer and choose a condition with objectively more pain as long as the episode ends on a relatively less painful note.

In the realm of positive experiences, Diener, Wirtz, and Oishi (2001) conducted studies in which participants rated a wonderful life that ended suddenly as better than one with the addition of mildly pleasant years. They termed this the *James Dean effect*. In addition, television commercials that induce positive feelings are rated more highly by consumers if the commercials have high peaks of intensity and strong positive endings (Baumgartner, Sujan, & Padgett, 1997). Participants in Fredrickson and Kahneman's (1993) study viewed a series of video clips that varied in pleasantness and duration. The data indicate

that people essentially disregard the duration of an affective episode when making retrospective assessments and that they heavily rely on just one or two moments in the episode (Fredrickson & Kahneman, 1993).

Simonson, Carmon, and O'Curry (1994) showed that in some cases, added features and offered promotions can lead participants to rate a brand lower, even when the features are seen as neutral or positive and the promotions are completely optional. Their research is not directly relevant to the peak–end rule in that the information about products, promotions, and/or features was available simultaneously rather than sequentially. Further, their research differed from both previous research on the peak–end rule and our research in that the presence of the features and promotions affords inferences about the primary product that could affect the ratings (e.g., “the product must be deficient if they have to offer promotions”). Our research does not afford such inferences.

By extrapolating the various peak–end findings to the domain of pleasurable goods, we predicted that people would retrospectively report lower levels of overall pleasure for a desirable gift if a positive but less desirable gift were added to it, even though the addition of this second gift objectively increases the total worth. In this way, people would irrationally rate experiences with objectively less overall pleasure higher than experiences with objectively greater overall pleasure. We conducted two experiments in order to investigate this counterintuitive hypothesis.

EXPERIMENT 1

Method

Over the course of several months, three charitable fundraisers were conducted at Dartmouth College during which undergraduate students made monetary donations to nonprofit organizations. All of the money collected went to the designated charities. Donors were

G. Wolford, wolford@dartmouth.edu

informed that they could enter a "raffle" for a chance to win free DVDs. All of the donors were later contacted via e-mail and told that they had won one or two DVDs from the raffle and that they would be able to choose their free DVDs from the list(s) attached to the e-mail. Participants were contacted through the e-mail account of a nonprofit organization in order to eliminate suspicion that the DVD raffle was part of a psychological experiment.

Participants were presented with the pleasure-inducing opportunity to select free DVDs from one or two predetermined lists. Two different DVD lists were used: List A consisted of 10 movies that were rated highly on www.rottentomatoes.com (ratings are based on users' opinions); List B consisted of 10 movies that were rated positively on www.rottentomatoes.com, but less so than those on List A. The "very pleasing" experience involved choosing a free DVD from List A. The "mildly pleasing" experience involved choosing a free DVD from List B. Participants voluntarily took part in the experiment by responding to the DVD offer via e-mail. All participants who were offered a free DVD—even one from List B—voluntarily chose to receive one, suggesting that the movies from List B were seen as positive for these participants.

The participants were randomly divided into one of five groups (see Table 1). Group A received a DVD from List A only. Group A+B received a DVD from List A and a second DVD from List B. The participants in Group A+B viewed List B via a second e-mail only after they had chosen their first DVD from List A. There were three additional control groups to aid in interpreting any observed differences between Groups A and A+B. The control groups were B, B+A, and A+A. After making all of their DVD selections, participants were asked to rate how pleased they were with the overall experience of receiving the DVD offer on a scale that ranged from 1 (*least pleased*) to 7 (*most pleased*). Participants were also asked to provide an explanation for their rating. We had approval from the Institutional Review Board (IRB) to carry out this experiment without prior informed consent. The DVDs were delivered to the participants on a later date, along with a letter that served to debrief them.

Results

The participants recruited from all three fundraisers were combined in the statistical analysis, for a total of 104 participants. On the basis of their written comments to the free DVD offer, 4 of the participants were eliminated. They made comments that indicated that our DVD lists were not viewed as intended in their cases (one in each of four different groups). For instance, 1 participant in Group A+B wrote that one of the movies from List B was his "all time favorite movie." A 2nd participant thought the List A movies were all "establishment." The analyses are based on the 100 remaining participants.

The results are displayed in Table 1. Because our data may not fully have met the assumptions for parametric tests, we carried out a Kruskal-Wallis H test on the five groups, followed by three Mann-Whitney U tests to examine three specific contrasts. The H test was significant [$\chi^2(4) = 41.8, p < .001$]. Using Mann-Whitney tests, we found that the mean rank pleasure rating of Group A was significantly greater than that of Group A+B ($z = 3.31, p = .001$). Group A was not significantly different from Group B+A ($z = 1.34, p = .179$). Group A+B was significantly different from Group B+A ($z = 2.08, p = .045$).

EXPERIMENT 2

Method and Procedure

A smaller-scale experiment was conducted in order to attempt to replicate the results of Experiment 1 in a different population and

context. Experiment 2 was conducted at a residential house that was routinely visited by elementary school and middle school children on Halloween night (see Santino, 1981, for a description of Halloween customs). Participants consisted of "trick-or-treaters" who were old enough (mean age = 10.03, $SD = 1.79$) to answer the survey question "How happy are you with the candy that I gave you?" by pointing to a "smiley face" on a smiley-face happiness scale. Seven out of 28 (25%) of the participants were female.

On the night of Halloween, trick-or-treaters were given different combinations of candy and asked to rate their level of happiness in relation to the candy that they were given. The "very pleasing" treat was a full-size Hershey's chocolate bar. The "mildly pleasing" treat was a piece of bubble gum. Permission was granted by the IRB to conduct the experiment without the prior consent of the participants.

Trick-or-treaters were assigned to one of four groups (A, A+B, A+A, and B). Because the comparison of Groups A and A+B was the critical assessment in the experiment, and because it was not possible to accurately predict how many participants would be recruited that night, group assignment was biased toward the first two groups in order to ensure that there would be enough participants for a statistical comparison of these two groups. Specifically, participants had a three-eighths probability of being assigned to Group A, a three-eighths probability of being assigned to Group A+B, and a one-eighth probability of being assigned to each of the other groups.

All participants who were offered the bubble gum voluntarily accepted it, which suggests that the experience of receiving the gum was positive. After participants were presented with their candy, they were shown a picture of seven smiley faces, each expressing a different level of happiness. The first smiley face appeared "neutral" (the mouth was drawn as a horizontal line), and the seventh smiley face appeared "most pleased" (the mouth was drawn as a wide, open-mouthed grin). The other smiley faces varied on a continuum between these two extremes. Participants were asked to indicate which smiley face reflected how happy they were with the candy that was given to them. Afterward, trick-or-treaters and their adult companions were given a letter that served to debrief them.

Results

The results are displayed in Table 2. Every child in Groups A and A+A responded with the top rating of 7. We carried out a Kruskal-Wallis H test on the four groups, followed by one planned comparison. The H test was significant [$\chi^2(3) = 18.04, p < .001$]. Using a Mann-Whitney U test, we found that the mean rank pleasure rating of Group A was significantly greater than that of Group A+B ($z = 2.96, p = .003$).¹ The results of Experiment 2 replicated the key finding of Experiment 1, demonstrating that two

Table 1
Mean Ratings As a Function of Group in Experiment 1

Group	<i>N</i>	<i>M</i>	Median	<i>SE</i>
A	29	5.21	5.00	0.17
A+B	21	4.14	4.00	0.26
B+A	17	4.82	5.00	0.23
A+A	19	5.50	5.50	0.22
B	14	2.57	2.50	0.29

Table 2
Mean Ratings As a Function of Group in Experiment 2

Group	<i>N</i>	<i>M</i>	Median	<i>SE</i>
A	10	7.00	7.00	0.00
A+B	11	5.91	6.00	0.39
A+A	4	7.00	7.00	0.00
B	3	3.00	3.00	0.00

positives are given a lower rating than a single positive if the second item is less positive than the first.

DISCUSSION

The results of both Experiment 1 and Experiment 2 support the hypothesis that people report lower levels of overall pleasure with the addition of mild pleasure at the end of an experience. In this way, people rate less pleasure higher than more pleasure. All of the participants in Group B rated the experience as positively pleasurable, with a mean rating of 2.57 in Experiment 1 and a mean rating of 3.00 in Experiment 2. We made it clear in both experiments that a rating of "1" was considered *neutral*. In addition, participants voluntarily made a selection from List B movies in Experiment 1 and happily accepted the gum in Experiment 2, which implies that these items were seen as positive. There were even spontaneous expressions of pleasure from a few of the participants in Group B in both Experiments 1 and 2. It is, of course, possible that a particular participant would find the List B movies or the "mildly pleasing" candy aversive, but we doubt that was the case for many of the participants.

In Experiment 1, Group A+A had a higher average rating than did Group A, although this difference is not statistically significant. In Experiment 2, the two groups were equal. The lack of significance could be due to a ceiling effect—especially in Experiment 2—but this result does eliminate the possibility that simply receiving two items (e.g., two DVDs or two pieces of candy) is worse than receiving a single item.

In Experiment 1, the mean rating of Group A was not significantly different from that of Group B+A, and the mean rating of Group B+A was significantly higher than that of Group A+B. These results suggest that order is crucial, and they confirm that the peak–end rule is applicable to pleasurable experiences, particularly to pleasure derived from material gains.

The peak–end rule presumably applies to episodes that are viewed as one joint event. In Experiment 1, participants were specifically asked to rate how pleased they were with the "overall" DVD offer. In Experiment 2, participants were asked to rate how pleased they were "with the candy that I gave you." This wording was intended to encourage the participants in Experiment 1 to perceive the overall experience of receiving two free DVDs as one cohesive event, and to encourage those in Experiment 2 to perceive the overall experience of receiving two pieces of candy as one joint experience. The strongly significant difference between Groups A+B and B in both experiments rules out

the possibility that the ratings were determined entirely by the last item received. However, an implementation of the peak–end rule in which the two items are combined in a weighted average—with the latter item receiving a higher weight—is consistent with our results.

The results of this study support the idea that the effects found in retrospective evaluations of pain are applicable to evaluations of pleasure. If you were planning to give multiple gifts for a future holiday, you might consider giving only the best one—or at least making sure that you give the best one last.

AUTHOR NOTE

We thank the Lincoln Filene Fund at Dartmouth College for supporting this research. Address correspondence to G. Wolford, 6207 Moore Hall, Dartmouth College, Hanover, NH 03755 (e-mail: wolford@dartmouth.edu).

REFERENCES

- BAUMGARTNER, H., SUJAN, M., & PADGETT, D. (1997). Patterns of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgments. *Journal of Marketing Research*, *32*, 219-232.
- DIENER, E., WIRTZ, D., & OISHI, S. (2001). End effects of rated life quality: The James Dean effect. *Psychological Science*, *12*, 124-128.
- FREDRICKSON, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion*, *14*, 577-606.
- FREDRICKSON, B. L., & KAHNEMAN, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality & Social Psychology*, *65*, 44-55.
- 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.
- 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.
- SANTINO, J. (1981). Halloween in America: Contemporary customs and performances. *Western Folklore*, *41*, 1-20.
- SCHREIBER, C. A., & KAHNEMAN, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology*, *129*, 27-42.
- SIMONSON, I., CARMON, Z., & O'CURRY, S. (1994). Experimental evidence on the negative effect of product features and sales promotions on brand choice. *Marketing Science*, *13*, 23-40.

NOTE

1. Because there were so many tied scores, we also carried out a χ^2 test of Group A versus Group A+B. We dichotomized the ratings into those of 7 versus all of the others [$\chi^2(1) = 9.545, p = .002$]. A two-sided Fisher exact test yielded a p value of .004.

(Manuscript received August 22, 2006;
revision accepted for publication May 15, 2007.)