Journal of Applied Research in Memory and Cognition 8 (2019) 337–346

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

No Peak-End Rule for Simple Positive Experiences Observed in Children and Adults

Eric Y. Mah[1] and Daniel M. Bernstein*

Kwantlen Polytechnic University, Canada

We investigated the tendency of children and adults to rely on the most intense and final moments when judging positive experiences, a heuristic known as the peak-end rule. This rule allows us to judge experiences quickly, but it can bias judgments. In three experiments involving various age groups ($N = 988$, ages 2–97), we attempted to replicate prior findings of a peak-end rule for small and simple positive experiences (e.g., receiving small gifts; Do, Rupert, & Wolford, 2008). Based on the original study and peak-end rule predictions, we hypothesized that individuals of all ages would be less satisfied with a highly desirable gift followed by a less desirable gift than with a highly desirable gift alone. We failed to observe the peak-end rule in preschoolers, school-aged children, younger adults, or older adults in any of the contexts we investigated. Our results show little support for positive peak-end rule effects and mark boundary conditions for the rule.

General Audience Summary

This research examined peoples' tendency to judge events mainly by the best/worst moment (the peak) and the last moment (the end)—a tendency known as the *peak-end rule*. The peak-end rule is generally useful for judging events quickly, but it can result in irrational judgments. Many studies have examined the peak-end rule for negative (e.g., painful or unpleasant) events, but there is relatively little research on how the peak-end rule affects positive experiences. A previous study found evidence of peak-end rule effects in children who experienced a simple positive event: receiving candy (Do, Rupert, & Wolford, 2008). Children who received a highly desirable candy (high peak) were more satisfied than children who received that same highly desirable candy followed by a less desirable one (high peak and low end). This is irrational, because two candies should be more satisfying than one. We attempted to replicate and extend this work by testing the peak-end rule in various age groups in three experiments—we examined how the peak-end rule affects judgments of a small gift (e.g., a toy, candy) in individuals of various ages. People either received a highly desirable gift alone, a less desirable gift alone, a highly desirable gift followed by a less desirable one, or a less desirable gift followed by a highly desirable one, and then rated their satisfaction with the gift. Unlike the original study, we found little evidence for the peak-end rule in children or adults. This finding suggests that the peak-end rule likely either does not apply to or has little effect on our judgments of small, simple positive experiences.

Keywords: Peak-end rule, Children, Adults, Cognition, Heuristic

Author Note.

Eric Y. Mah and Daniel M. Bernstein, Kwantlen Polytechnic University, Surrey, B.C., Canada.

* Correspondence concerning this article should be addressed to Daniel M. Bernstein, Kwantlen Polytechnic University, 12666 72nd Avenue, Surrey, B.C. V3W 2M8, Canada. Contact: dbernste@kpu.ca.

[1] Present address: University of Victoria, 3800 Finnerty Road, Victoria, B.C. V8P 5C2, Canada.

Imagine you've just finished a meal consisting of an excellent grilled steak followed by a decent side dish. The highlight of the meal was the steak, followed by the slightly less satisfying side. After the meal, you reflect on your experience. How happy were you with your meal? Might your answer differ if you had finished the meal after the steak?

When we retrospectively judge affective experiences, research suggests that we may focus on the most affectively intense moment (the peak) and most recent moment (the end)—an effect known as the *peak-end rule* (Kahneman, 2000, 2011). In our hypothetical example, the peak-end rule suggests that you might be happier with your meal if you finished with the most pleasurable part (i.e., the steak) rather than with the slightly less pleasurable part (i.e., the side). This rule was first demonstrated for unpleasant experiences in an experiment where participants preferred a longer painful ice-water experience to a shorter one because the longer experience was manipulated to end on a less painful note (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993).

The tendency to judge negative experiences by a retrospective snapshot of specific memorable moments instead of the length of the event and overall amount of unpleasantness experienced is counterintuitive (Kahneman, 2000). The peak-end rule affects judgments in different negative contexts, including videos, medical procedures, noises, advertisements, and learning experiences (Finn, 2010; Kahneman, 2011; Hoogerheide & Paas, 2012). Although the peak-end rule often biases judgments, it can be a useful heuristic. Accurately reconstructing an affective event and performing a hedonic calculation based on total duration and the quality of each individual moment is slow and inefficient, if not impossible. Judging an event based on the quality of the highly memorable peak and end moments is fast and efficient (Kahneman, 2000). However, these savings come with costs: the heuristic occasionally results in judgments that might seem irrational. A central aspect of the way we remember events is duration neglect: the duration of affective events is often heavily discounted such that longer painful episodes can be judged as more pleasant than shorter ones if the final moments of the longer episode are less painful (Kahneman, 2000).

Although the peak-end rule for negative experiences has been studied extensively, the peak-end rule for positive experiences has been relatively neglected. Some studies have found evidence of a positive peak-end rule. For example, consider the *James Dean effect*: people rated a fictitious life that ended sooner but on a highly positive note as more pleasurable and desirable than the same life that ended after five additional pleasurable, but less happy years (Diener, Wirtz, & Oishi, 2001). The peak-end rule has also been observed for pleasant learning experiences (Hoogerheide & Paas, 2012), positive peer assessments (Hoogerheide, Vink, Finn, Raes, & Paas, 2017), and memories of musical pieces (Rozin, Rozin, & Goldberg, 2004). Conversely, other work has found little to no peak-end rule effects on vacation memories or judgments of pleasurable meals (Kemp, Burt, & Furneaux, 2008; Rode, Rozin, & Durlach, 2007)—eating a so-so side dish after a fantastic steak might not spoil your memories of the meal after all.

One study examined how the peak-end rule affects judgments of gifts. In one experiment, undergraduates received one of various combinations of highly rated and lower rated DVDs (of their choice); in another experiment, child trick-or-treaters received one of various combinations of a highly desirable candy (Hershey's chocolate bar) or a less desirable candy (bubblegum). Consistent with the peak-end rule, young adults who received a great DVD followed by an average DVD of their choice were less satisfied than those who received just one great DVD. Similarly, children who received a Hershey's bar followed immediately by bubblegum were less satisfied than children who received the Hershey's bar alone (Do, Rupert, & Wolford, 2008).

These experiments are of interest because they appear to show that the peak-end rule affects a class of events qualitatively different from peak-end events that have typically been studied. Do et al.'s (2008) experiments involved discretely segmented events that are shorter and simpler than typical peak-end experiences. Most peak-end experiments involve comparatively longer events that are either continuous (e.g., cold-water task in Kahneman et al., 1993) or that consist of a series of discrete events (e.g., sequences of positive or negative peer ratings in Hoogerheide et al., 2017). The DVD and candy events of Do et al. only involve a series of events in the loosest sense and arguably lack a meaningful duration (to be neglected). A further disconnect from prior peak-end work is that the DVD and candy experiments did not involve "direct" experiences of the stimuli (like in the original ice-water experiment). Finally, at least in the candy experiment, there appears to have been no delay between the receipt and rating of the candy. Most peak-end experiments involve a delay to make the evaluation retrospective and set the stage for heuristic remembering.

Nonetheless, Do et al. (2008) observed what resembles a peak-end rule in immediate evaluations of short, simple and discrete events—exactly where one would *not* expect to find a memory heuristic. If the effects that Do et al. (2008) observed are robust and replicable, there are important implications for how we judge simple positive experiences. Firstly, work by Do et al. (2008) suggests that the peak-end rule can distort evaluations of very simple and short discrete events that are not "directly" experienced. On a more theoretical level, by utilizing a simple event with negligible duration, their findings show the power of salient moments, largely separate from the duration neglect aspect of the peak-end rule. In Do et al.'s (2008) peak-end experiment, some participants experienced a longer event with a less pleasurable end. However, Do and colleagues' focus was more on the manipulated quality of the peak and end moments than the added duration. Despite the potential implications, their experiments were underpowered. Assuming a medium-sized effect ($d = .5$) and a one-tailed test of the critical peak-end difference, Do and colleagues' DVD experiment attained a power of only around.53, and their candy experiment reached a power of only around .3. If we assume a large-sized effect ($d = .8$), the DVD experiment has acceptable power ($\sim$.87), but the candy experiment remains woefully underpowered (∼.55).[2] Thus, more powerful tests of their predictions are necessary.

Our primary aim was to provide a more comprehensive and adequately-powered test of the peak-end rule predictions examined by Do et al. (2008). In doing so, we extended the original paradigm to include different age groups, event timing, and stimuli. Based on findings by Do et al. (2008) and other prior research, we predicted that people of all ages would demonstrate the peak-end rule for small, simple positive experiences (i.e., people who receive a single, highly desirable gift would be more satisfied than those that received a highly desirable gift followed by a less desirable one). We tested this prediction in three experiments.

Experiment 1

Method

Participants. We determined sample sizes for our experiments using *a priori* power analyses via G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). In the original experiment, Do et al. (2008) observed a very large peak-end effect ($d = 1.35$, $f = .68$). In light of the uncertainty surrounding that effect and possible effect size inflation due to low power, we assumed a more conservative medium effect size for power analyses in the current study ($f = .25$).

In Experiment 1, we tested 608 individuals ages 2 to 97 drawn from a community and university population. After excluding participants with missing data on any peak-end variables except preference ($n = 41$), and those who indicated a preference for both or neither of the candies ($n = 7$), our final sample included 560 individuals. We divided this sample into four age groups: 2–7 year olds ($n = 119$, $M_{age} = 5.63$, $SD = 1.39$, 50% female), 8–15 year olds ($n = 128$, $M_{age} = 11.51$, $SD = 2.29$, 44% female), 16–50 year olds ($n = 156$, $M_{age} = 33.20$, $SD = 11.33$, 74% female), and 51+ year olds ($n = 157$, $M_{age} = 67.19$, $SD = 9.19$, 71% female). We chose these age groups to permit rough comparisons to age groups tested in Experiments 2 and 3. The total sample size exceeded the required N of 260 to achieve a power of .8 to detect all potential main and interaction effects in a 4 (Candy condition: high peak, low peak, high peak-low end, high peak-high end) × 4 (Age group: 2–7, 8–15, 16–50, 50+) between-subjects ANOVA. We pre-registered Experiment 1 midway through data collection to reflect design changes and all subsequent experiments via the Open Science Framework. The pre-registration and data for our pilot test and Experiments 1–2 can be viewed here: https://osf.io/avpvx/. The pre-registration and data for Experiment 3 can be viewed here: https://osf.io/smp7d/. See Supplementary Material A for more details about our pre-registrations.

Materials and procedure. Experiment 1 examined the peak-end rule (as it applies to receiving candy) in children and adults. Our materials and procedure for the main experiment closely followed those of the original experiment (Do et al., 2008: Experiment 2) with minor changes to the stimuli (i.e., different candies). We used Mars bars instead of Hershey's to avoid potential nut allergies, and substituted lollipops for bubblegum because some people (e.g., younger children, older adults) may have had trouble chewing the gum. We made these stimulus changes for all subjects in the interests of safety (e.g., researcher error or participant failure to report allergies). We validated these stimuli (i.e., that they were viewed as highly and less desirable, respectively) via pre-registered pilot testing (see Supplementary Material B for more details). For the main experiment, participants were assigned to one of four Candy combinations: the high peak (a highly desirable Mars bar, abbreviated *A*), the low peak (a less desirable lollipop, abbreviated *B*), the high peak-low end (a highly desirable Mars bar followed by a less desirable lollipop, abbreviated *AB*), or the high peak-high end (lollipop followed by Mars bar, abbreviated *BA*).

For all candy conditions, we told participants that they would receive a small prize for participating in a longer experiment that contained our peak-end task, and then gave them the Mars bar (or lollipop). For conditions AB and BA, we remarked that we had an additional candy for the participant ("I have another candy for you"), and we then gave participants the lollipop (or Mars bar). After receiving the candy, participants rated how "they felt about the candy [candies] we gave them." Participants indicated their rating on the same 7-point smiley face scale used by Do et al. (2008), which ranged from *Neither happy nor unhappy* to *Very happy*. The satisfaction question was worded in a way that encouraged participants to rate the single candy on its own or the two candies together (i.e., in the single-candy conditions, participants rated their satisfaction with the candy we gave them; in the dual-candy conditions, participants rated their satisfaction with both candies we gave them). Though we did not include a preference question initially in the experiment, we introduced such a question partway through data collection. For this question, after rating the gift, participants indicated which of the two candies they preferred (participants given a single candy at this point received the candy they did not originally rate). This question was aimed at supplementing our pilot data on candy preference rates and was used to identify participants who indicated a preference for the lollipop over the Mars bar (or a preference for both or neither candy). Because we introduced this preference question partway through the study, we only obtained preference data for roughly 40% ($n = 223$) of our final sample. Participants completed our candy peak-end task at the end of a longer 60–90 minute cognitive psychology experiment.

Note that in both the original experiment and our replication, participants did not eat the candy as part of the experiment or rating procedure. As such, the positive event we examined was the experience of receiving a gift, not eating candy. However, prior studies have shown that directly experiencing an event may not be necessary for the peak-end rule to occur (e.g., hypothetical meals, Rode et al., 2007; imagined lives, Diener et al., 2001)

Results

We predicted that satisfaction ratings of the highly-desirable Mars bar (condition A) would be higher than ratings of the

[2] Power analyses conducted via G*Power 3 (Faul et al., 2007), as if they were *a priori*.

less-desirable lollipop (B) *and* higher than ratings of the Mars bar followed by the lollipop (AB). Recall that the peak-end rule predicts that people mainly focus on the peak (best moment) and the end (final moment) when making overall retrospective judgments about affective experiences. Accordingly, ending with the lollipop should reduce satisfaction ratings. However, beginning with the lollipop should not lower ratings, because the highly-desirable Mars bar serves as both the peak and the end. Thus, we also predicted that those who received the Mars bar after the lollipop would be more satisfied than those who received the lollipop after the Mars bar. Finally, we predicted that the peak-end rule would operate similarly across all age groups (i.e., no interaction between age group and Candy condition).

We first examined candy preference data. Though we only had preference data for about half of our participants, those who answered the preference question preferred the Mars bar to the lollipop (83% of 8–15 year olds, 79% of 16–50 year olds, and 93% of 51+ year olds). The exception was 2–7 year olds, who preferred the Mars bar (56%) only slightly more than the lollipop. In total, 49 participants indicated a preference for the lollipop. Though the original pre-registered plan was to perform separate analyses on participants who indicated a preference for the lollipop, we did not collect enough data from lollipop-preferers to warrant a separate analysis. Instead, we reversed their conditions (e.g., we treated the lollipop as highly-desirable and the Mars bar as less-desirable for these participants) and combined them with participants who indicated a preference for the Mars bar. A cursory manipulation check in our lollipop-preferring subsample suggested that a simple condition reversal was valid. Those in the highly-desirable A condition (lollipop for these participants) gave significantly higher satisfaction ratings than those in the less-desirable B condition (Mars bar for these participants), $t(22.97) = 2.21$, $p = .04$. Though a corresponding Wilcoxon rank-sum test (in light of dubious normality and variance homogeneity) was *not* statistically significant ($p = .07$), the difference was in the appropriate direction.

We analyzed candy condition satisfaction ratings by age group (4 × 4 between-subjects ANOVA), and found a significant main effect of age group, $F(3, 544) = 11.90$, $p < .001$, partial $\eta^2 = .05$. Due to evidence of assumption violations (heterogeneity of variance and non-normality in satisfaction ratings across Candy combinations and age groups), we opted to use non-parametric Kruskal–Wallis tests to follow up on significant main effects/interactions. In all cases, parametric and non-parametric results agreed. We also used Wilcoxon rank-sum tests in place of planned and unplanned pairwise comparisons (with Bonferroni-corrected alpha levels). Following up on our age group main effect: 2–7 year olds were more satisfied with candy than 16+ year olds ($p \leq .001$), but there were no other age group differences. There was also a main effect of candy condition, $F(3, 544) = 9.81$, $p < .001$, partial $\eta^2 = .05$. We successfully manipulated candy desirability (A > B), $p < .001$, Cohen's $d = .71$, 95% CI [.39, 1.03], but observed no peak-end differences. There was no difference between A and AB, $p = .99$, Cohen's $d = -0.04$,95% CI [−.23, .15] or between AB and BA, $p = .43$, Cohen's $d = -.19$, 95% CI [−.52, .12]. Finally, there was no significant interaction between age group and candy combination, $F(9, 544) = 1.24$, $p = .27$. We also conducted the above analyses with only participants who indicated a preference for the Mars bar and obtained identical results. Table 1 lists overall candy condition means collapsing across age groups (with the critical conditions bolded).

Table 1
Mean Satisfaction Ratings by Candy Combination and Age Group

Candy condition	Mean	SD	*n*
A	**5.72**	**1.54**	**220**
B	4.53	2.02	64
AB	**5.76**	**1.41**	**217**
BA	5.59	1.55	59

Note. A = Highly desirable candy; B = Less desirable candy; AB = Highly desirable candy followed by less desirable candy; BA = Less desirable candy followed by highly desirable candy.

Because we lacked preference data for most of our sample, it is likely that there were undetected lollipop-preferers for whom we couldn't appropriately recode conditions. Based on our cursory analysis of the Mars bar versus lollipop difference in these participants, it is likely that lollipop preferers' response pattern was opposite that of the full sample. This could have plausibly led to the attenuation of group differences. However, recall that the vast majority of participants who *were* asked about their preference indicated that they preferred the Mars bar (≥83% in all but the youngest age group). We also found a consistent A > B difference in the overall sample, and our pilot testing suggests that our stimulus manipulation worked as intended. Given all this, we think it unlikely that undetected lollipop-preferers in our full sample are the culprits behind our null results.

The standard null hypothesis significance testing (NHST) approach does not allow one to draw substantive conclusions when null results emerge—*p* values only quantify evidence *against* the null and cannot be used to corroborate it (Loftus, 1996). Bayesian hypothesis testing is a compelling alternative to the standard NHST approach and offers a way of quantifying relative evidentiary support for competing hypotheses—for example, H_0 versus H_1. Aside from allowing one to directly compare the plausibility of competing hypotheses, Bayesian inference offers several advantages over NHST (for a comprehensive review, see Wagenmakers et al., 2018).

To better understand our null results, we supplemented our NHST analyses with Bayesian tests. For our first Bayesian analysis, we tested the following hypotheses: H_0 (that there is no peak-end rule for small positive experiences) and H_1 (that there is a peak-end rule for small positive experiences). This kind of Bayesian analysis results in a *Bayes factor* (BF), which quantifies the relative support for one hypothesis over another. We obtained a BF_{01} (subscript indicating support in favor of H_0 relative to H_1) of 12, indicating that given our data, we should update our belief in no peak-end rule relative to our belief in a peak-end rule by a factor of 12:1. This constitutes strong evidence for a null effect (Wagenmakers, Wetzels, Borsboom,

& Van Der Maas, 2011). Details about this analysis (e.g., prior choice) and Bayesian analyses in general (e.g., implications of priors) appear in our Supplementary Material (1A).

In addition to a standard Bayesian analysis, we examined how our results differed across a wide range of priors (i.e., pre-data assumptions about the peak-end rule effect size). For most reasonable prior choices, this "robust" Bayesian analysis resulted in moderate-to-strong evidence for H_0 (see Supplementary Material 1B for more details about this analysis). Finally, we conducted two additional analyses using non-standard priors that represented more specific predictions one might have about the peak-end rule, based specifically on the effect size observed in Do and colleagues (2008). The resulting BF_{01}s for these analyses were 7.69 and 5—in both cases, moderate evidence for H_0. The details for these analyses, including the priors used and background for non-standard prior analyses, can be found in the Supplementary Material (1C).

Though the results of our omnibus Candy condition × Age group test pertain less directly to the peak-end rule, our decision to collapse across age groups relies on the assumption of no interaction. Using a Bayesian ANOVA with default specifications (see Rouder, Morey, Verhagen, Swagman, & Wagenmakers, 2016 for an overview), we obtained a BF_{01} of 19.94. Thus, a model of the peak-end rule *without* a Candy condition × Age interaction is 19.94 times as likely as a model with such an interaction. This analysis provides strong evidence that the peak-end rule (lack thereof) does not differ across our age groups

Discussion

The peak-end rule that Do and colleagues (2008) observed failed to materialize in our considerably larger, all-ages sample. Our NHST analyses suggest the absence of, or a trivially small, peak-end rule effect. A variety of Bayesian analyses provided moderate-to-strong evidence *against* a peak-end rule across all age groups. Even though we successfully manipulated candy desirability (as evidenced by our A > B difference and our pilot testing), the addition of the less desirable candy to the highly desirable one did not produce a less satisfying experience.

Experiment 2

We did not observe the peak-end rule for small positive experiences in any of the age groups we tested. However, in Experiment 1, the peak-end task followed a 60–90 minute study. Remembering events is prerequisite to judging them. Consequently, memory for events depends on the ability to segment experiences into discrete events (Sargent et al., 2013). Thus, the candy event could have been judged as part of the larger psychology study experience. Experiment 2 addressed this possibility by comparing Candy combinations either at the beginning or end of a longer experience. For Experiment 2, our predictions were the same as Experiment 1 (i.e., A > B, A > AB, A = BA), but only when the peak-end task occurred before a longer experience. When the peak-end task occurred after a long experience we predicted that people would not show peak-end effects—only the non-peak-end differences we observed in Experiment 1.

Method

Participants. Participants in Experiment 2 included 277 undergraduate students. After excluding 8 participants who indicated a preference for both or neither candy, our sample included 269 participants. Of these participants, 66 (24.5%) indicated a preference for the lollipop. Like Experiment 1, our original analysis plan was to analyze lollipop-preferers as a separate sample. Due to the low *n* in this subsample, we instead considered reversing their conditions (i.e., the lollipop is treated as the highly-desirable candy for these participants, and the Mars bar the less-desirable candy). However, an exploratory manipulation check in this subsample was not successful, $t(24.63) = .06$, $p = .95$ (the corresponding Wilcoxon rank-sum test was also non-significant, $p = .98$). On this basis, we could not justify a simple condition reversal for lollipop-preferers in our sample. Thus, we chose to exclude these participants from subsequent analyses. Our sample size of 203 ($M_{age} = 21.29$, $SD = 4.25$, 81% female) exceeded the recommended *N* of 179 to achieve a power of .8 to detect all potential main and interaction effects for a 4 (Candy condition) × 2 (Timing) between-subjects ANOVA.

Materials and procedure. Student participants completed the same peak-end task used in Experiment 1 as part of 1 of 4 longer psychology experiments that they completed for course credit (average experiment length approximately 40 minutes). Experiment 2's procedure was identical to Experiment 1's except that participants received one of the four candy combinations either at the beginning or end of one of the psychology experiments. As participants received the candy or candies, the experimenter mentioned that the candy was a thank-you gift for participating in the experiment.

Results

For our main hypothesis, we predicted that the peak-end rule is reduced or eliminated if the peak-end experience occurs after another longer experience. To test this, we used a 4 (Candy condition) × 2 (Timing) between-subjects ANOVA. We found a main effect of candy condition, $F(3, 195) = 4.82$, $p = .003$, partial $\eta^2 = .07$, but no interaction between candy condition and timing, $F = .57$, $p = .64$. A Shapiro–Wilk test suggested non-normality in our data, but because cell sizes were essentially equal and homogeneity of variance was maintained, we used only parametric tests for Experiment 2 analyses. Planned follow-up *t*-tests revealed a successful candy desirability manipulation (A > B), $t(100) = 3.55$, $p < .001$, Cohen's $d = .70$, 95% CI [.30, 1.11]. However, we did not observe a peak-end rule; there was no significant difference between A and AB, $t(100) = 0.48$, $p = .63$, Cohen's $d = .09$, 95% CI [−.30, .49]; or between AB and BA, $t(99) = -.80$, $p = .43$, Cohen's $d = .16$, 95% CI [−.24, .55]. We did find a significant main effect of timing, $F(1, 195) = 7.81$, $p = .006$, partial $\eta^2 = .04$, such that satisfaction with the candy gift was higher at the end of a longer experience than at the beginning. Table 2 lists candy condition mean satisfaction ratings, overall and by timing (with the critical conditions bolded).

Table 2
Mean Satisfaction Ratings by Candy Condition and Timing

Candy condition	Overall (SD)	Beginning of study (SD)	End of study (SD)
A	**6.02 (1.20), $n = 52$**	5.93 (1.36), $n = 30$	6.14 (.94), $n = 22$
B	5.08 (1.47), $n = 50$	4.88 (1.48), $n = 24$	5.27 (1.46), $n = 26$
AB	**5.90 (1.33), $n = 50$**	5.50 (1.42), $n = 26$	6.33 (1.09), $n = 24$
BA	5.68 (1.48), $n = 51$	5.32 (1.57), $n = 25$	6.02 (1.32), $n = 26$
Overall	*5.67 (1.41), n = 203*	*5.44 (1.49), n = 105*	*5.92 (1.28), n = 98*

Note. A = Mars Bar; B = Lollipop; AB = Mars then Lollipop; BA = Lollipop then Mars.

As with Experiment 1, we supplemented our NHST analysis with Bayesian analyses. First, it was crucial to determine whether we could rule out the possibility of timing effects on the peak-end rule (i.e., a Candy condition × Timing interaction). A Bayesian ANOVA analogous to the one conducted in Experiment 1 returned a BF_{01} of 8.03—moderate evidence against such an interaction. Next we examined the lack of observed differences between our critical peak-end conditions (A vs. AB). Using the same default prior as Experiment 1, we obtained a BF_{01} of 3.21, indicating moderate evidence for H_0 relative to H_1. A robustness analysis revealed that under more conservative prior widths the evidence in favor of H_0 was anecdotal—less strong evidence than that observed in Experiment 1 (see Supplementary Material 2A for the robustness plot).

Across a wide range of reasonable prior widths, we failed to observe more-than-anecdotal evidence against the peak-end rule—weaker than the evidence observed in Experiment 1. Observing compelling evidence for null or trivially small effects with Bayesian analyses typically requires larger sample sizes to detect small deviations from the null (e.g., in excess of $N = 50{,}000$ to obtain compelling evidence for effects of $d = .02$ or smaller; Rouder, Speckman, Sun, Morey, & Iverson, 2009). Thus, it is likely that our results are due to Experiment 2's smaller sample size (rather than any substantive differences in peak-end rule effects across experiments).

To see how our Experiment 2 results should affect our post-Experiment 1 beliefs in the peak-end rule, we conducted an informed Bayesian analysis. For this analysis, the prior distribution was the distribution of likely effect sizes given the Experiment 1 results.[3] Here the BF represents the change in our beliefs in a null effect before and after the Experiment 2 data. We obtained a BF_{01} of 1.01, suggesting that the Experiment 2 results tell us little beyond what we learned from Experiment 1. Again, this is unsurprising given the smaller sample size of Experiment 2. Though it might seem intuitive to characterize our current belief in the peak-end rule as a simple product of the BF_{01}s observed in both experiments (e.g., Experiment 1 belief in H_0:H_1 of 12:1 multiplied by Experiment 2 updating factor of 1.01), doing so is not valid (Rouder & Morey, 2011). Individual experiment BFs respect sample size—with smaller samples, small effects are more likely to be considered as evidence for the null (2011).

[3] Specifically, a *t* distribution centered on the A versus AB raw scale difference observed in Experiment 1, with a *SD* equal to the standard error of that difference and *df* equal to the Experiment 1 analysis *df*.

Discussion

Our Experiment 2 analyses failed to provide evidence for the peak-end rule regardless of event timing. This suggests that the results of Experiment 1 were not substantially tainted by the experimental context and thus reflect a meaningful null effect.

Experiment 3

Experiment 1 provided strong evidence *against* a peak-end rule for small positive experiences, and Experiment 2 provided no support for the peak-end rule. However, one could argue that there is something about candy in general that precludes peak-end effects. Therefore, we endeavored one final test of the peak-end rule, this time by giving small, non-candy gifts (toys) to children and youth. Our hypotheses mirrored those of Experiments 1 and 2: We predicted that children (3–7 year olds) and youth (8–15 year olds) would be less satisfied with a highly desirable gift followed by a less desirable gift (AB) than a highly desirable gift alone (A). We included age group as a factor but did not predict any age group differences in the peak-end rule.

Method

Participants. Our sample included 117 children ages 3–7 and 121 children ages 8–15. Examining toy preference data, we found that 82% of 3–7s who received both toys preferred A to B (i.e., a glass rock to a brown paper bag), while 94% of 8–15s preferred A to B (i.e., magnetic rocks to a wooden craft stick). Because we lacked enough participants to evaluate our manipulation in the subgroups with reversed preference, we excluded from further analysis all participants who indicated a preference for B (3–7s, $n = 10$; 8–15s, $n = 3$). Thus, our final sample included 107 children ages 3–7 ($M_{age} = 5.08$, $SD = 1.41$, 41.3% female) and 118 children ages 8–15 ($M_{age} = 9.57$, $SD = 1.68$, 44.7% female). Our sample exceeded the recommended N of 179 to attain a power of .8 to detect all potential main and interaction effects in a 4 (Toy condition) × 2 (Age group) between-subjects ANOVA. The study was run at a local science center, where we approached parents entering the center and asked if their children would be interested in participating in a short psychology study on how children and youth rate small toys. Specifically, we told children that we were going to give them a small gift and that we would ask them to tell us how they felt about it (children in the dual toy condition were not initially told they would receive two toys). After explaining the study purpose and obtaining consent, children received one of the toy combinations and were asked to tell us how they felt about the toy/toys. Though we had no control over what kind of experience children had before entering the science center, it seems unlikely that participants viewed our peak-end study as a continuation of a longer, prior experience—at least not in the same way participants may have viewed the peak-end experience in Experiments 1 and 2. Additionally, real-world affective experiences necessarily occur before and after other affective experiences.

Materials and procedure. We conducted pilot testing to select toys and found that the age groups differed in terms of what toys they preferred. Our final highly desirable toys were a glass rock for 3–7 year olds and three magnetic rocks for 8–15 year olds; our less desirable toys were a brown paper bag for 3–7 year olds and a wooden craft stick for 8–15 year olds. Other than the stimulus change, the experimental procedure was identical to Experiment 1, except that we only collected preference data from participants who received two toys (due to costs associated with having to give all children both toys). Again, it is worth noting that children were rating their experience of receiving the gifts (i.e., as opposed to playing with and then rating gifts).

Results

To test whether children and youth showed peak-end effects for a small positive experience *other* than candy, we conducted a 4 (Toy condition) × 2 (Age group) between-subjects ANOVA. We found a main effect of age group, $F(1, 217) = 9.34$, $p = .003$, partial $\eta^2 = .03$. Overall, 3–7 year olds were happier with the prizes than 8–15 year olds. We also found a main effect of toy condition, $F(3, 217) = 8.84$, $p < .001$, partial $\eta^2 = .07$.

However, both our main effects were qualified by a significant interaction, $F(3, 217) = 6.95$, $p < .001$, partial $\eta^2 = .09$. A follow-up one-way ANOVA comparing toy condition ratings in 2–7 year olds revealed that ratings for all Toy combinations did not significantly differ, $F(3, 103) = .355$, $p = .79$. Youth ages 8–15 drove our toy coondition main effect, with a significant omnibus test in this age group, $F(3, 114) = 16.06$, $p < .002$. Planned follow-up tests (Bonferroni-corrected *t*-tests) revealed that we successfully manipulated toy desirability, $t(68) = 5.33$, $p < .002$, Cohen's $d = 1.36$, 95% CI [.80, 1.91]. However, despite successfully manipulating toy desirability in 8–15 year olds, we observed no peak-end rule in this age group (no difference between A and AB), $t(43) = 1.75$, $p = .09$, Cohen's $d = .52$, 95% CI [−.09, 1.13] or between AB and BA, $t(46) = .11$, $p = .91$, Cohen's $d = .03$, 95% CI [−.55, .62]. Table 3 lists mean toy combination satisfaction ratings by age group (with the critical conditions bolded).

Though we couldn't fully investigate the peak-end rule in younger children due to unsuccessful manipulation of toy desirability, we successfully manipulated toy desirability in older children. They rated the magnetic rocks (A) as highly-desirable ($M = 6.04$ out of 7), and the wooden craft stick (B) as less-desirable ($M = 3.70$ out of 7). This difference is not only significant, but more in line with the differences observed in Do et al.'s (2008) original study. Despite this, we found no peak-end rule: Older children did not rate A alone as better than AB, as the peak-end rule would predict.

Table 3
Mean Satisfaction Ratings by Toy Condition and age Group

Toy condition	3–7s (SD)	8–15s (SD)
A	5.52 (1.94), $n = 33$	**6.04 (1.02, $n = 23$**
B	5.76 (1.83), $n = 29$	3.70 (1.98), $n = 47$
AB	6.00 (1.60), $n = 23$	**5.50 (1.06), $n = 22$**
BA	5.64 (1.65), $n = 22$	5.46 (1.30), $n = 26$

Note. A = Highly desirable toy (a glass rock for 3–7s, three magnetic rocks for 8–15s); B = Less desirable toy (a brown paper bag for 3–7s, a wooden craft stick for 8–15s).

Despite Experiment 3's smaller sample size, we conducted a final Bayesian test of the critical peak-end conditions (A vs. AB). Using a default prior, we obtained a BF_{01} of 0.53, indicating anecdotal support for a peak-end rule. However, a corresponding robust version of this analysis failed to produce more-than-anecdotal evidence in favor of a peak-end rule for the entire range of prior widths we examined (See Supplementary Material 3A for the robustness plot). Given our small sample size (relative to Experiment 1) and mounting evidence for a likely null and possibly small true effect size, these inconclusive results are not surprising.

How should these results affect our belief in the peak-end rule? An informed Bayesian analysis using the combined results of Experiments 1 & 2 as the prior resulted in a BF_{01} of .64. In other words, our belief in H_0 is slightly reduced in light of Experiment 3's results. As previously mentioned, one cannot simply combine BFs across experiments (Rouder & Morey, 2011). To better quantify our Bayesian evidence for and against the peak-end rule across all three experiments, we conducted a final Bayesian meta-analysis. Under the default prior width, the combined BF_{01} was 7.77—moderate evidence against a peak-end rule for small positive experiences (See Supplementary Material 3B for the robustness plot).

Discussion

Again, our results failed to support a peak-end rule for toys in children. Though the Experiment 3 peak-end difference (A > AB) was the largest numerically out of all three experiments, it was not significant. Additionally, there is an unexpected difference we observed that runs counter to the peak-end rule: the anomalous A > BA (numerical) difference. Recall that the peak-end rule predicts no difference between A and BA (because the peak and end in both cases are the same). Given that this anti-peak-end rule numerical difference is of similar magnitude to the peak-end rule difference, and Experiment 3's smaller sample size, we are cautious to conclude that Experiment 3 provides more support for the peak-end rule than the other experiments. Finally, though the individual BF test of Experiment 3 was uninformative (likely due to small sample size), the sum of our Bayesian evidence weighed against a peak-end rule for small positive experiences.

General Discussion

We attempted to replicate and extend findings of a peak-end rule for small positive experiences and failed to find any effects in children or adults (or any of the various age groups examined in Experiment 1) with two different types of small gifts (candies and toys). People of all ages who received a highly desirable gift were no more or less satisfied than people who received the same gift followed by a less desirable one. Our tests of candy (Experiments 1 & 2) resulted in either evidence against, or a lack of evidence for, the peak-end rule. Similarly, our toy experiment failed to provide convincing evidence for the peak-end rule. The Bayesian evidence that we obtained in Experiment 1 was unambiguous, while it was ambiguous in Experiments 2 and 3. Taken together, our experiments provided moderate evidence that the peak-end rule does not affect short and simple positive experiences.

Ultimately, the most likely explanation for our failure to replicate Do et al.'s (2008) original results is that the peak-end rule either does not apply to these kinds of experiences or has negligible effects on them. The experiences under examination were simple and brief. Other experiments have found evidence for positive peak-end effects using longer events (e.g., pleasant study sessions, DVD gifts separated by a delay; Do et al., 2008; Hoogerheide & Paas, 2012, Experiment 1). Thus, it is possible that short and simple events such as receiving gifts do not invoke the *evaluation by moments* heuristic. The qualitative difference between the gift event we studied and the affective events typically studied in the peak-end literature may explain our lack of peak-end findings. People may be able to accurately judge the affective quality of simple and discretely segmented events without relying on the peak-end rule.

One possible explanation for our null results lies with our manipulation. Because retrospective peak-end judgments are a combination of peak and end affect, ratings of an event with a similar peak and end will be like ratings of the peak or end alone (i.e., if A = B, then A = AB). Our manipulation of affective quality was weaker than the manipulation in the original study (Do et al., 2008). However, we still observed a significant difference in ratings between the stimuli, and the peak-end differences we observed fell far short of the magnitude one might predict (i.e., the average of A & B ratings). Additionally, the peak-end rule was originally observed with a small manipulation of stimulus quality (e.g., a gradual one-degree shift in ice water temperature; Kahneman et al., 1993). Thus, we do not think that our weaker manipulation explains our findings.

Finally, one may argue there are other paradigms more relevant to the evaluation of short, simple experiences. As we have discussed, the candy/toy events differ from typical events in the peak-end paradigm in terms of duration, continuity, complexity, and delay between event and evaluation. Regardless of whether Do et al.'s (2008) original results are replicable, it may be more appropriate to frame the experiences that they (and we) examined using a different theoretical perspective. For instance, people were less satisfied with a DVD or candy gift when given a less desirable gift after a highly desirable one. This finding is compatible with a general preference for "happy endings" (Ross & Simonson, 1991)—given sequences consisting of discrete, essentially duration-less positive and negative experiences, people prefer sequences that end with a positive experience. Yet another possibility is that aversion to affectively decreasing sequences can be explained in terms of adaptation, where affect is tuned to an initial set point by the first stimulus and even small changes in stimulus quality result in substantial changes in affect (Haisley & Loewenstein, 2011). From an adaptation perspective, people in Do et al.'s (2008) experiment may have been initially happy with the great Hershey's bar; however, participants' affect dropped when the stimulus quality of the additional lollipop departed from the higher set point. Similarly, the peak-end pattern observed by Do and colleagues may have been due to expectation violation. When people receive a gift and learn that another one is forthcoming, expectations about the gift-to-be are based on the gift that was already received (Haisley & Loewenstein, 2011). Thus, when expectations for another Hershey's-quality gift are subverted, satisfaction with the overall gift declines.

These alternative ideas present attractive and perhaps more appropriate ways to conceptualize evaluations of short, simple, and discrete affective experiences. However, we chose to remain within the peak-end perspective adopted by Do et al. (2008). The primary goal of our three experiments was to investigate a surprising (and important) finding of a peak-end rule for experiences far different than those typically found in the peak-end literature. While it may seem evident *a priori* that the peak-end rule should not apply to simple positive experiences, Do and colleagues' findings challenge that reasoning. Without further comprehensive tests supporting or opposing their results, there is a tension between *a priori* notions about the peak-end rule and empirical results. Our research serves as a stronger test of Do and colleagues' findings (i.e., larger samples, lack of rating ceiling effects, more variability in ratings), and provides evidence that the peak-end rule does not affect short positive experiences.

That is not to say that these other theories are irrelevant to the peak-end rule or our experiments. Though we did not test these other theories, our results show that alternate biases or heuristics were not at play in the experiences we examined. Across our three experiments, participants generally viewed the highly desirable gift as more satisfying than the less desirable gift, but did not appear to show order effects, preference for a better end, or aversion to a worse end. However, regardless of the status of these theories relative to our results, we believe that our adequately powered failure to replicate prior findings provides good evidence that the peak-end rule does not apply to such experiences.

Some open questions remain. It is possible that short, simple positive experiences elicit a peak-end rule, but only when a delay separates experience and evaluation. Given that the peak-end rule is based on *retrospective* evaluations, it is possible that the lack of delay in our study eliminated a peak-end rule that would otherwise exist. Because our main aim was to replicate Do et al. (2008), our lack of a delay manipulation leaves us unable to eliminate that possibility. Outside of direct replication, the fact that Do and colleagues observed a peak-end rule with no delay or negligible delay makes their finding even more surprising, and worthy of further tests. However, addressing the possibility that small positive experiences elicit a peak-end rule with longer delays is certainly a worthwhile next step in further tests of peak-end rule boundary conditions. Finally, we did not directly examine evaluations of small positive experiences through the lenses of the different paradigms that we described previously. Though our results suggest that small positive experiences are relatively resilient to evaluation biases, more direct tests are necessary to make conclusive claims.

Our study provides evidence that the peak-end rule does not substantially affect our judgments of short, simple positive experiences. In failing to replicate the surprising findings of a previous peak-end study (Do et al., 2008), we highlight potential boundary conditions for the peak-end rule. Though some important questions remain, it seems likely that our in-the-moment and soon-after-the-fact judgments of these experiences are unbiased. Further research along these lines will allow us to better understand how we think about the positive experiences that shape our lives. Such research will help us decide whether we really should save the best—whether it's steak, a good movie, or a big gift—for last!

Author Contributions

Daniel Bernstein conceived of and designed the study, collected the data and assisted with the writing of the manuscript. Eric Mah assisted with the conception and design of Experiments 2 & 3, collected, analyzed and interpreted the data, and wrote the manuscript.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Acknowledgements

We acknowledge Maryanne Garry for her helpful suggestions regarding the writing and data analysis, and André Aßfalg and Mike Masson for their helpful suggestions regarding our statistical analyses and interpretation of results. We would also like to thank TELUS World of Science for their cooperation in running Experiment 3. This work was supported by the Canada Research Chairs Program (950-228407) and the Social Sciences and Humanities Research Council of Canada (435-2015-0721).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jarmac.2019.05.002.

References

Diener, E., Wirtz, D., & Oishi, S. (2001). End effects of rated life quality: The James Dean effect. *Psychological Science*, *12*, 124–128. http://dx.doi.org/10.1111/1467-9280.00321

Do, A. M., Rupert, A. V., & Wolford, F. (2008). Evaluations of pleasurable experiences: The peak-end rule. *Psychonomic Bulletin & Review*, *15*, 96–98. http://dx.doi.org/10.3758/PBR.15.1.96

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. http://dx.doi.org/10.3758/BF03193146

Finn, B. (2010). Ending on a high note: Adding a better end to effortful study. *Journal of Experiment Psychology: Learning, Memory, and Cognition*, *36*, 1548–1553. http://dx.doi.org/10.1037/a0020605

Haisley, E., & Loewenstein, G. (2011). It's not what you get but when you get it: The effect of gift sequence on deposit balances and customer sentiment in a commercial bank. *Journal of Marketing Research*, *48*, 103–115. http://dx.doi.org/10.1509/jmkr.48.1.103

Hoogerheide, V., & Paas, F. (2012). Remembered utility of unpleasant and pleasant learning experiences: Is all well that ends well? *Applied Cognitive Psychology*, *26*, 887–894. http://dx.doi.org/10.1002/acp.2890

Hoogerheide, V., Vink, M., Finn, B., Raes, A. K., & Paas, F. (2017). How to bring the news ... peak-end effects in children's affective responses to peer assessments of their social behavior. *Cognition & Emotion*, *32*, 1114–1121. http://dx.doi.org/10.1080/02699931.2017.1362375

Kahneman, D. (2000). Evaluation by moments: Past and future. In D. Kahneman, & A. Tversky (Eds.), *Choices, values and frames* (pp. 693–708). New York: Cambridge University Press and the Russell Sage Foundation.

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

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. http://dx.doi.org/10.1111/j.1467-9280.1993.tb00589.x

Kemp, S., Burt, C. D. B., & Furneaux, L. (2008). A test of the peak-end rule with extended autobiographical events. *Memory & Cognition*, *36*, 132–138. http://dx.doi.org/10.3758/MC.36.1.132

Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science*, *5*, 161–171. http://dx.doi.org/10.1111/1467-8721.ep11512376

Rode, E., Rozin, P., & Durlach, P. (2007). Experienced and remembered pleasure for meals: Duration neglect but minimal peak, end (recency) or primacy effects. *Appetite*, *49*, 18–29. http://dx.doi.org/10.1016/j.appet.2006.09.006

Ross, W. T., & Simonson, I. (1991). Evaluations of pairs of experiences: A preference for happy endings. *Journal of Behavioral Decision Making*, *4*, 273–282. http://dx.doi.org/10.1002/bdm.3960040405

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review*, *18*, 682–689. http://dx.doi.org/10.3758/s13423-011-0088-7

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2016). Bayesian analysis of factorial designs. *Psychological Methods*, *22*, 304–321. http://dx.doi.org/10.1037/met0000057

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

Rozin, A., Rozin, P., & Goldberg, E. (2004). The feelings of music past: How listeners remember musical affect. *Music Perception: An Interdisciplinary Journal*, *22*, 15–39. http://dx.doi.org/10.1525/mp.2004.22.1.15

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailtey, H. R., . . ., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*, 241–255. http://dx.doi.org/10.1016/j.cognition.2013.07.002

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . ., & Rouder, J. N. (2018). Bayesian inference for psychology: Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*, 35–57. http://dx.doi.org/10.3758/s13423-017-1343-3

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*, 426–432. http://dx.doi.org/10.1037/a0022790

Received 16 November 2018;
received in revised form 17 May 2019;
accepted 17 May 2019
Available online 6 July 2019