COGNITION AND EMOTION
https://doi.org/10.1080/02699931.2019.1675599

Routledge
Taylor & Francis Group

Check for updates

How does the peak-end rule smell? Tracing hedonic experience with odours

Benjamin Scheibehenne[a] and Géraldine Coppin[b,c,d]

[a]Faculty of Economics and Management, University of Geneva, Geneva, Switzerland; [b]Swiss Center for Affective Sciences, Geneva, Switzerland; [c]Laboratory for the Study of Emotion Elicitation and Expression, University of Geneva, Geneva, Switzerland; [d]Swiss Distance University Institute (UniDistance), Centre d'Etudes Suisse Romand, Sierre, Switzerland

ABSTRACT

The peak-end rule predicts that retrospective evaluations of affective events heavily depend on their most intense and last moment and imply duration neglect. It was originally proposed for negative experiences such as painful medical procedures. It is unclear, however, to what degree it also applies to positive experiences. Previously, rigorous comparisons between the two domains were limited due to the use of qualitatively different stimuli. Hence, it is not clear if the peak-end rule holds for short positive and negative experiences alike. To address these questions in a genuinely emotional domain, we conducted two experiments ($n = 48$ each) in which we used odours as stimuli. Participants repeatedly evaluated continuous odour sequences delivered into their noses via an olfactometer. The sequences differed in valence (positive vs. negative), length (36 vs. 72 s), and trajectory (increasing, decreasing, U-shaped, and inverse U-shaped). Results provide evidence for the peak-end rule for both positive and negative experiences alike. Results further show an overweighting of intense negative experiences for sequences that contain both pleasant and unpleasant episodes but provide little evidence for an effect of the trajectory manipulation.

ARTICLE HISTORY
Received 19 December 2018
Revised 11 June 2019
Accepted 23 September 2019

KEYWORDS
Peak-end rule; odours; positive and negative affect; emotions

The peak-end rule predicts that retrospective evaluations of affective events heavily depend on the most intense and last moments. For example, when a person is asked: "How was the movie you watched last night?" the rule predicts that the judgment will not rely on an unbiased sum or average of the affective experiences (in our example, consider all the movie scenes equally) but rather that people will overweight the peak (i.e. the most intense) and the end (i.e. the most recent) experience (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). The peak-end rule further predicts duration neglect: In the example above, the overall assessment would not depend on the length of the movie. Instead, the peak-end rule predicts that retrospective evaluations can be improved by adding a positive ending to the event or by enhancing the peak moment, even if the average experience remains unchanged or worsens (Redelmeier & Kahneman, 1996).

Most seminal studies on the peak-end rule used painful experiences such as medical examinations, the cold pressor test, or giving birth (Chajut, Caspi, Chen, Hod, & Ariely, 2014; Kahneman et al., 1993; Redelmeier & Kahneman, 1996; Redelmeier, Katz, & Kahneman, 2003; Varey & Kahneman, 1992). In support of the peak-end rule, these past studies showed that people often prefer longer negative experiences to shorter ones, as long as the longer ones involve a less painful peak and/or end. Similar results were found for other types of unpleasant stimuli such as aversive sounds (Ariely & Loewenstein, 2000; Schreiber & Kahneman, 2000) or effortful learning experiences (Finn, 2010).

As retrospective evaluations influence subjective experiences and eventually choice, both duration

CONTACT Benjamin Scheibehenne benjamin.scheibehenne@unige.ch
Supplemental data for this article can be accessed at https://doi.org/10.1080/02699931.2019.1675599.

neglect and an overweighting of the peak and end moments have important real-world implications, for example, when helping people increase their long-term happiness (Wirtz, Kruger, Napa Scollon, & Diener, 2003) or when trying to improve consumer experiences (Cornil & Chandon, 2016; Dixon, Victorino, Kwortnik, & Verma, 2017). This raises the question of the extent to which the results obtained using aversive stimuli generalise to positive emotional experiences.

Does the peak-end rule hold for positive affective experiences?

An early literature review by Fredrickson (2000) concluded that duration neglect and peak-end are robust phenomena for positive and negative affective experiences alike. However, when considering more recent empirical work on the peak-end rule, a less consistent picture arises.

In support of the peak-end rule for positive experiences, Hsee and Abelson (1991) found that people prefer monetary payouts that increase rather than decrease over time and are also sensitive to the magnitude of the peak payout. Langer and Weber (2005) reported similar findings. The peak-end rule was also supported when consumers evaluated television advertisements (Baumgartner, Sujan, & Padgett, 1997), pleasant music clips (Schäfer, Zimmermann, & Sedlmeier, 2014), or candies and DVDs (Do, Rupert, & Wolford, 2008). Some studies also found evidence for duration neglect for positive experiences, for example, using pleasant videos (Fredrickson & Kahneman, 1993) or well-liked food (Rode, Rozin, & Durlach, 2007).

In contrast to these supporting findings, Fredrickson and Kahneman (1993) could not verify the predictions of the peak-end-rule for positive video clips. Rode et al. (2007) reported only minimal peak and end effects and did not find evidence of a difference between meal sequences with different trajectories (e.g. rising, falling, U-shaped, or inverse U-shaped). When people retrospectively evaluated different episodes during the day and then rated the day as a whole, Miron-Shatz (2009) did not find evidence of a difference for end effects and only little evidence for the overweighting of high or low peaks. Research on retrospective evaluations of vacations also found mixed results. Geng, Chen, Lam, and Zheng (2013) indicated that, at least for periods of less than 3–7 weeks, the peak-end rule predicted retrospective holiday evaluations. On the other hand, data produced by Kemp, Burt, and Furneaux (2008) suggested that the peak-end rule was not a good predictor for retrospective happiness ratings of holiday experiences. When analysing retrospective evaluations of average daily and weekly affect, Ganzach and Yaor (2018) found that positive emotions lend themselves to an end effect but not a peak effect. They further found a peak effect but not an end effect (i.e. the reverse pattern) for negative emotions. A series of experiments by Seta, Haire, and Seta (2008) where participants remembered positive live events failed to reveal significant order effects and hence evidence for the peak-end rule. Finally, Hui, Meyvis, and Assael's (2014) analysis of a large set of data containing people's moment-to-moment ratings while watching TV shows failed to find evidence for a peak effect. While Hui et al. found that the evaluation of the end quintile carries four times as much weight for the overall evaluation, they noted that for TV shows, the final moments may actually determine the overall quality of the show as a whole. They further pointed out that the end rating may already reflect an overall, aggregated evaluation. This would limit the generalisability of the results beyond TV shows (see Tully & Meyvis, 2016, for a similar argument).

In summary, these inconsistent findings indicate a need for a more systematic investigation of the peak-end rule for momentary positive experiences. Rigorous comparisons between negative and positive evaluations in the existing literature are difficult though because the stimuli used in previous experiments are hard to compare between domains, for example, painful medical experiences on the one hand (e.g. Kahneman et al., 1993) and funny movies on the other (e.g. Fredrickson & Kahneman, 1993).

Are negative experiences over weighted?

There is an ongoing debate regarding the theoretical explanations and the cognitive processes underlying the peak-end rule (e.g. Cojuharenco & Ryvkin, 2008; Geng et al., 2013), but many psychological explanations refer to memory-related processes. For example, in support of a peak effect, extreme values might be better remembered (Ludvig, Madan, & Spetch, 2014; Tsetsos, Chater, & Usher, 2012). Likewise, explanations for an end effect often refer to recency effects and the sequential weighting of information (Hogarth & Einhorn, 1992). From an evolutionary perspective, an ultimate reason for the relevance of the peak intensity could be that it signals if the experience can be handled or endured again in the future (Fredrickson, 2000). This would suggest that the peak experience is more relevant and hence carries more weight for aversive as compared to pleasant experiences. In line with this idea of a negativity effect, some researchers have pointed out that negative information may be more relevant than positive information and hence processed more intensely and better remembered (Vaish, Grossman, & Woodward, 2008). Likewise, as mentioned above, Ganzach and Yaor (2018) reported a peak effect for negative episodes but not for positive ones. However, the authors explicitly pointed out that their theorising applies to retrospective evaluations of extended experiences that last for days or even weeks and it is not clear if the results also generalise to short-term or momentary emotional experiences.

The idea that negative and positive stimuli are processed differently is also debated in the affective science literature where evidence for such a negativity effect has been mixed (Hilgard, Weinberg, Proudfit, & Bartholow, 2015). Furthermore, there is increasing evidence that relevant stimuli will be processed in an enhanced fashion (Pool, Brosch, Delplanque, & Sander, 2016; Sander, Grafman, & Zalla, 2003; Stussi, Pourtois, & Sander, 2018) and be better remembered (Montagrin, Brosch, & Sander, 2013; Montagrin et al., 2018) regardless of their valence. This raises the question of the extent to which negative information in general and negative peaks in particular are overweighted in retrospective affective evaluations.

Do trajectories matter?

Besides the peak and the end, past empirical research also found that people care for the sequential order of consumption experiences. In particular, for many domains people seem to prefer trajectories that improve over time, for example when receiving payments, choosing medical treatments, or planning vacations (Alba & Williams, 2013; Ariely & Loewenstein, 2000). A sequence that improves over time will deliver the best experience at the end, hence tallying with the peak-end rule. As mentioned above, some researchers (e.g. Rode et al., 2007) found no difference between rising and falling profiles though. This raises the question in how far hedonic trajectories influence retrospective affective evaluations.

Research questions

To summarise, here we address three closely connected research questions on how people integrate momentary emotional experiences into overall retrospective evaluations: Do the peak-end rule and duration neglect apply to both positive and negative emotional experiences alike? Is there evidence for an overweighting of strong negative experiences (i.e. negative peaks) as compared to strong positive experiences (i.e. positive peaks) when evaluating short, momentary experiences? Does the trajectory of the observed sequence matter?

Using odours as stimuli

To test these research questions empirically we conducted two experiments in which we used odours as stimuli. Using odours has several advantages in the context at hand: Perhaps most importantly, odours can be perceived as both pleasant and unpleasant, which allows for a comparison of positive and negative emotional experiences within the same domain. In addition, odours can elicit potent affective reactions (e.g. Herz, Eliassen, Beland, & Souza, 2004) and their hedonic dimension is salient (e.g. Mohanty & Gottfried, 2013), allowing the experimental manipulation of different trajectories (e.g. increasing or decreasing in valence). Also, odours can be easily and precisely applied in a laboratory setting when using appropriate equipment (e.g. an olfactometer), and they have been used in numerous studies, including research on perceptual and decision-making processes (Coppin et al., 2014; Coppin, Delplanque, Porcherot, Cayeux, & Sander, 2012; Oud & Coppin, 2012). Because different odours can be combined into uninterrupted sequences at will, they lend themselves to a continuous presentation and hence increase the chance of a holistic evaluation. Past research indicates that this facilitates the use of the peak-end rule (Ariely, Kahneman, & Loewenstein, 2000; Ariely & Loewenstein, 2000).

Experiment 1

Materials and method

We conducted an experiment in which participants experienced and evaluated eight different odour sequences. Each sequence consisted of six different discrete odours that were delivered into participants' noses. By combining different odours, we could experimentally manipulate the valence, trajectory, and length of each sequence in a within-subject design.

Valence

To test for the influence of valence, half of the sequences presented to each participant consisted of a combination of presumably unpleasant odours and the other half consisted of presumably pleasant odours. Odour (un)pleasantness was based on data from past research (e.g. Chrea et al., 2009; Delplanque et al., 2008; Ferdenzi et al., 2011; Von Helversen, Coppin, & Scheibehenne, 2019). To avoid repetition, six different odours were used within each sequence, drawn from a pool of 16 odour channels in total. The pleasant odours were lavender, strawberry, laundry, and lily of the valley. The unpleasant ones were sulphur and onion, cigarette, cheese, and civet. All odorants were complex compounds.

Length

For long sequences, each of the six odours was diffused into the participants' noses for 12 s, resulting in a total length of 72 s. For the short sequences, the diffusion time was half as long (i.e. 36 s in total). Given that a typical respiratory rate for a healthy adult at rest is 12–18 breaths per minute (Blows, 2001), these lengths were set to ensure that participants experienced each of the six different odours within each sequence while breathing normally.

Trajectory

We manipulated the trajectory of the odour sequences by combining odours with low (L) and high (H) intensity within each sequence. Concentrations for low intensity odorants were 20% for cigarette, 10% for lavender, strawberry, laundry, lilly of the valley, 5% for civet, and 1% for sulfur and onion and for cheese. Concentrations for high intensity odorants were of 100% for cigarette, 50% for lavender, strawberry, laundry, lilly of the valley, 10% for civet, and 5% for sulfur and onion and for cheese. These concentrations were chosen so that all low intensity odorants were perceived as isointense and relatively weak and all high intensity odorants as isointense and relatively potent. This manipulation was based on the assumption that diluting odour intensity diminishes both the pleasantness and unpleasantness of the odours. Thus, we constructed increasingly pleasant (+) trajectories by presenting four presumably pleasant odours of low intensity followed by two presumably pleasant odours of high intensity (i.e. LLLLHH+). Accordingly, the patterns for decreasing, U-shaped, and inverse U-shaped sequences were HHLLLL+, HLLLLH+, and LLHHLL+, respectively. Likewise, for unpleasant (−) sequences, the patterns were HHLLLL −, LLLLHH−, LLHHLL−, and HLLLLH−. This way, the expected mean of all sequences was constant a priori and the variance in participants' online ratings was increased, which creates a stronger test for the peak-end rule (e.g. Tully & Meyvis, 2016). Figure 1 provides a schematic visualisation of the trajectories. However, given that the perception and valuation of odours greatly varies between individuals, the observed online ratings may differ from the intended trajectories.

Experimental design

A fully factorial within-subject design would have required 2 (Valence) × 2 (Length) × 4 (Trajectory) = 16 sequences. To reduce the risk of olfactory habituation (i.e. decreased perceived odour intensity after prolonged exposure, see Cain, 1974), this design was reduced to eight sequences per participant by varying some combinations of sequence length and trajectory between participants rather than within participants. Hence, each participant encountered all eight possible combinations of trajectory and valence as well as all four possible combinations of

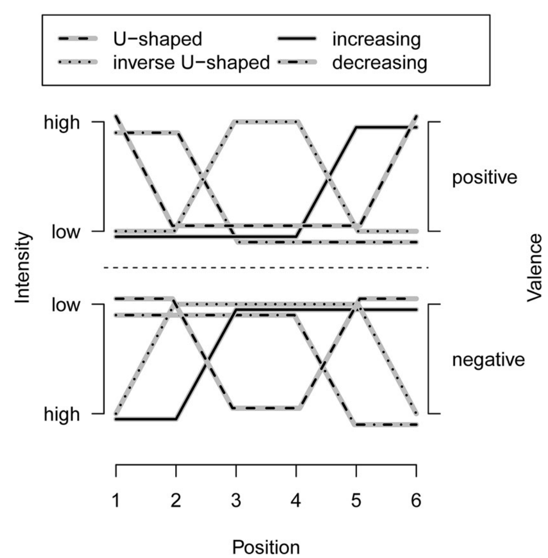

Figure 1. Trajectories in Experiment 1. The *x*-axis indicates the sequential order of the six odours within each sequence. The area below the horizontal dotted line indicates negative valences and the area above indicates positive valences. Jitter added on the *y*-axis for better visibility.

length and valence, but only half of all possible combinations of length and trajectory, counterbalanced between participants.

Half of participants first experienced all four unpleasant sequences followed by the remaining four sequences with pleasant odours. We chose this order to enhance comparisons within sequences of the same valence and hence circumvent possible dichotomous answer patterns that would be mainly driven by the valence factor (i.e. "good" vs. "bad" odours). Within the positive and negative sequences, 4! = 24 different permutations of sequences could be presented to a single participant in theory. To ensure a balanced experimental design, we presented each combination twice throughout the study, thus requiring 48 participants. The ethics committee of the Psychology Department of the University of Geneva approved the study protocol.

Participants

Participants were students at the University of Geneva. The mean age was 22 years and 33 of the 48 participants were female.

Odour delivery

The odours were delivered with an olfactometer, a mechanical device consisting of a non-metallic array of up to 28 odour-containing glass tubes. Each glass tube was pressure fed by a corresponding computer-controlled air valve. The odours were connected via a mixing chamber and diffused into the participants' noses through a flexible silicone tube that was loosely strapped around their necks (see Ischer et al., 2014 for additional information). The whole system, connected to the medical air supply of the building, enabled a precise and constant delivery of air at 0.8 l/min. In between the odour sequences, the olfactometer delivered clean air, so that there was no detectable flow variation when odours were sent. This helped us avoid contamination or habituation effects, as well as potentially distracting tactile stimulation inside the nostrils during the odour sequences. To avoid distracting odours and control for other external influences, the experiment took place in a well-ventilated, windowless room, and participants sat in a comfortable chair. The olfactometer itself stood in a separate room to avoid distraction from the clicking noises of its valves. The odours themselves were provided by Firmenich, SA, Geneva, Switzerland and dissolved in propylene glycol in order to be perceived as equi-intense. Solutions (6 ml) of these odorants were enclosed in each glass tube.

Moment-to moment evaluations

The beginning of each sequence was announced on a computer screen in front of participants. During each sequence, participants assessed their subjective experience of the odours on a moment-to-moment basis by continuously positioning a slider bar placed on a table in front of them. The slider ranged from "very unpleasant" on the left side to "very pleasant" on the right. In between each sequence, participants were instructed to reposition the slider back to the "neutral" middle position that was clearly marked. The experimental software sampled the exact slider position with a rate of 4 Hz and internally mapped it onto a scale ranging from −100 (very unpleasant) to +100 (very pleasant). The analysis of the online slider data started with a time lag of 1 s relative to the diffusion of odours by the olfactometer. We chose this lag to account for participants' response latency and a small latency between opening the valve on the olfactometer and the delivery of the odours in participants' noses (Ischer et al., 2014).

Retrospective evaluations

Right after the end of each odour sequence, participants submitted an overall retrospective evaluation of the sequence as a whole, using the computer mouse as input. The rating scale was presented on the computer screen in front of them and ranged from "very pleasant" to "very unpleasant" with a mark labelled "neutral" in the middle. The exact phrasing of the retrospective rating question (translated from French) was "Overall, how do you evaluate this olfactory sequence?" As a plausibility check for the retrospective ratings, we also asked participants how the sequence compared to the one they had encountered previously. Participants made this comparison on a rating scale anchored from "much more unpleasant" to "much more pleasant" with a mark labelled "similar" in the middle.

Manipulation check

Finally, as a manipulation check at the end of the experiment, participants were again presented with each of the 16 odours separately through the olfactometer and were asked to rate each of them with respect to its pleasantness (from "very unpleasant" to "very pleasant"), familiarity (from "very unfamiliar" to "very familiar"), and intensity (from "odorless" to "very intense"). This manipulation check was important because odour-evoked affective reactions are known to be highly variable across individuals (e.g. Ferdenzi et al., 2013).

Statistical analyses

To statistically test our research questions at hand we estimated a multilevel linear regression using the lme4 package (version 1.1-13, Bates, Maechler, Bolker, & Walker, 2014) in R. Multilevel analyses, sometimes also referred to as mixed effects models, are well suited for analysing repeated measurement data (McElreath, 2016). The model we applied assumed random intercepts for individuals to account for the repeated measurement design. We tested predictor variables by adding them consecutively as fixed effects, starting from a baseline model that included the order in which the sequences were presented, the mean online evaluations for each sequence, and the valence condition (positive vs. negative) as fixed effects. The order was included to account for possible influences of fatigue or habituation during the experiment, the mean was included because the peak-end rule should explain additional variance beyond the mean (e.g. Ganzach & Yaor, 2018; Tully & Meyvis, 2016), and the valence condition was included to capture possible differences in the absolute ratings between positive and negative sequences.

We concluded that a predictor variable had a credible (the Bayesian equivalent of "statistically significant") influence if adding it improved predictive accuracy according to the Bayesian information criterion (BIC).[1] This model comparison approach has several advantages: Information criteria such as BIC account for the fact that more complex models (i.e. regressions with more beta coefficients) will always provide a better fit to the data (e.g. higher R^2 or higher likelihood) but not necessarily make better prediction out-of-sample (McElreath, 2016). Besides controlling for overfitting, the difference in BIC values between any two models can be further transformed into a Bayes factor that provides an intuitive interpretation of the results (Kass & Raftery, 1995). For example, a Bayes factor of three indicates that the model with the lower BIC is three times more likely than the model with the higher BIC value, given the observed data.

Results

Manipulation check

Figure 2 plots the mean online slider ratings across all participants for the different odours presented within the sequences. As shown in this Figure, unpleasant odours received on average negative ratings and vice versa for pleasant odours, indicating that the valence manipulation was successful. We obtained similar results based on the separate pleasantness ratings for each odour at the end of the experiment. As further shown in Figure 2, the observed online and retrospective pleasantness ratings did not align with the experimental manipulation of low and high intensity. While participants consistently rated intensely negative odours as less pleasant compared to their low-intensity counterparts, this was not the case for positive odours. This impeded our intended trajectory manipulation that relied on differences in odour intensity.[2] We return to this issue below.

In line with previous research (e.g. Delplanque, Coppin, Bloesch, Cayeux, & Sander, 2015), the separate pleasantness ratings for each odour at the end of the experiment were positively correlated with the subjective familiarity ratings. The mean correlation across individuals was $r = .5$, with a 95% confidence interval ranging from .25 to .68. The separate intensity ratings at the end of the experiment were credibly lower for the low-intensity than for the high-intensity odours, indicating participants discerned the odours' intensities (BF > 10.000; $p < .001$). Overall, there were considerable individual differences in the evaluations, which is expected for odours (e.g. Ferdenzi et al., 2013).

Peak-end rule and duration neglect

Across all participants, the mean online ratings for each sequence were highly correlated with the corresponding retrospective ratings. The mean and median correlations across participants were .87 and .93 respectively,[3] indicating that the average of the moment-to-moment experiences was already a good predictor of the retrospective ratings. To compare this against the peak-end rule, we calculated a peak-end predictor for each sequence by averaging across the end rating (defined as the mean of the last 5 s) and the peak ratings (defined as the most extreme positive and the most extreme negative rating). The mean and median correlations between this peak-end predictor and the retrospective ratings were .89 and .94, respectively. Thus, the predictions by the peak-end rule were very similar to the predictions based on the mean and median moment-to-moment experiences.

To rigorously test the peak-end rule on statistical grounds, we regressed the retrospective ratings for a

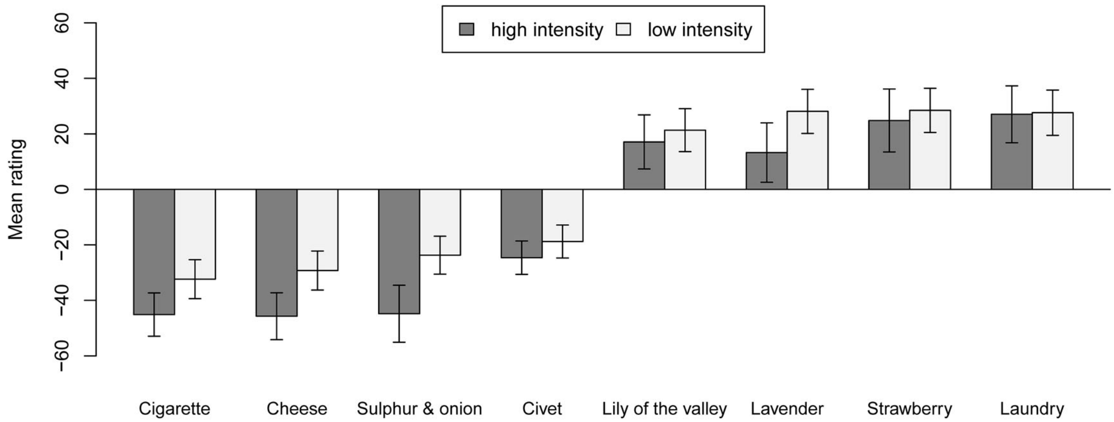

Figure 2. Mean online slider ratings for each odour across all participants. Error bars are 95% confidence intervals around the mean.

given sequence (= dependent variable) onto the moment-to-moment online ratings (= predictor variable) using the multilevel linear regression approach with random individual intercepts outlined above. In particular, we first set up a baseline model (m0) that included fixed effects for the order in which participants in the experiment encountered the sequences, the observed mean online ratings, and the valence condition (positive or negative).

Next, we compared this baseline model with an extended model (m1) that included the duration of the sequence (short vs. long) as an additional fixed-effect predictor variable. Table 1 provides an overview of the estimated models including the fixed-effect beta estimates and the BIC values. Because m1 included an additional predictor variable, it is more complex than m0. However, as shown in Table 1, m1 hardly explains more variance in the observed data compared to m0 (the R^2 values in Table 1 are actually identical for both models due to rounding). The fact that the increased complexity hardly improved the model fit is reflected in the higher (i.e. worse) BIC value of m1 compared to m0. This difference in BIC thus indicates that retrospective ratings did not depend on the duration of the sequence.

In a next step, we included the end (i.e. the last 5 s) of the respective online slider position as an additional fixed effect (m2). As shown in Table 1, adding this variable did improve predictive accuracy as indicated by a smaller BIC value. A direct comparison between m0 and m2 yields a BIC difference of 40. This translates into a Bayes factor of more than 10.000 and hence

Table 1. Overview of the fixed-effect estimates within the hierarchical regression models in Experiment 1.

	Model					
Predictor	m0	m1	m2	m3	m4	m5
(Intercept)	274.73	276.01	283.80	308.35	313.46	339.36
Sequential order	−1.16	−1.16	−1.42	−1.53	−1.77	−1.51
Positive valence	57.77	57.74	50.01	32.50	30.59	−23.35
Mean online rating	2.92	2.92	2.31	1.33		1.30
Long sequence		−2.53				
End rating			0.67	0.55	0.62	0.45
Lowest rating				0.94	1.67	1.37
Highest rating				0.70	1.45	0.49
End × Positive						0.24
Lowest × Positive						−0.86
Highest × Positive						0.39
R^2	.83	.83	.85	.86	.85	.86
Deviance	4316	4316	4269	4245	4273	4235
df	6	7	7	9	8	12
BIC	4352	4358	4311	**4299**	4321	4306

Note: BIC = Bayesian information criterion. Smaller BIC values indicate better model accuracy. The smallest BIC value (model m3) is printed in a bold font.

extreme evidence in favour of m2. The estimated beta coefficient was positive, indicating that participants overweighted the end or most recent experience, as predicted by the peak-end rule.

Adding the peak online ratings within each sequence (m3), defined as the most extreme positive and negative slider positions, led to a further improvement of model accuracy. Here, the comparison between m3 and m2 yield a Bayes factor of 392 in favour of m3. The estimated beta coefficients were again positive, indicating that the most extreme positive and negative momentary evaluations received a higher weight.

As an additional test for the peak-end rule, we assessed if participants' retrospective ratings were better predicted based on mean online ratings (m0) or based on the peak and the end online ratings without the mean ratings (m4). The m4 model explained the observed data better than the m0 model (Bayes factor > 10.000). This result provides further evidence for the importance of the peak and the end moment for retrospective evaluations in the data at hand.

As pointed out by Tully and Meyvis (2016), a possible alternative explanation for the end effect could be that any other position within the sequence would work equally well as a predictor and hence the end has no privileged status. To test this, we repeatedly estimated an alternate version of the m3 model in which we replaced the end predictor (i.e. the mean online rating during the last 5 s of each odour sequence) with other 5-s "windows" of participants' continuous online ratings. In particular, we shifted the window in steps of 1 s, starting from the end of each sequence until its beginning. This yielded 32 alternative models that we could compare to the original m3 model. The result of this comparison showed that all 32 alternative models yielded worse model accuracy (i.e. higher BIC values) than the original m3 model, showing that the end rating indeed stands out. Together, these findings provide strong evidence for the peak-end rule and duration neglect in the context at hand.

Difference between pleasant and unpleasant experiences

To test if the predictive power of the peak-end rule differs between pleasant and unpleasant experiences, we further included the interaction with valence as an additional predictor (m5). Including this interaction did not improve the model relative to m3. The Bayes factor of m3 over m5 was 34. This indicates that the subjective weights for peak and end did not differ between pleasant and unpleasant sequences. Separate analyses for positive and negative sequences confirmed that the end rating was a credible predictor for both valence conditions.

Trajectories

Including the different trajectory conditions as an additional fixed effect did not improve predictive accuracy, indicating that the trajectory manipulation did not affect the retrospective ratings. Further analyses showed that it did not influence participants' mean online ratings either. As mentioned above, a possible reason for these non-significant results was the ineffective manipulation of the trajectories that relied on differences between high- and low-intensity odours.

Discussion Experiment 1

In this first experiment, we found that the retrospective assessments of olfactory sequences systematically depended on their most extreme and last moments, as predicted by the peak-end rule. The results further showed that the retrospective assessments did not depend on the length of the sequences, providing evidence for duration neglect. We did not find evidence of an interaction effect with valence, suggesting that the peak-end rule equally applies to positive and negative experiences. However, such a direct comparison between positive and negative stimuli rest on the assumption that ratings on both sides of the bi-polar valence scale carry the same psychological meaning, which may not necessarily be the case.

While participants seemed to overweight the end of a sequence and hence showed a recency effect, we did not find a difference between the increasing and decreasing sequences. Presumably, this was due to the failed trajectory manipulation. In particular, our trajectory manipulation did not align with participants' actual online ratings. As there was nevertheless variance in the observed online ratings, we could still test the peak-end rule successfully.

To obtain a better manipulation of the trajectories, we ran a second experiment, outlined next, where we aimed at a better alignment between the observed online ratings and the trajectory manipulation. This second experiment also provided a robustness check for the peak-end rule in the olfactory domain at hand. The third goal of this second experiment was to conduct a stronger empirical test for the negativity effect predicting that negative peaks were weighted more heavily than positive peaks.

Experiment 2

Materials and method

Experiment 2 was similar to the first experiment except for two changes. The first change was the mix of pleasant and unpleasant odours within each sequence. To increase the contrasts within each sequence and hence create a stronger trajectory manipulation, we used both pleasant (+) and unpleasant (−) odours in each sequence. This manipulation also provided the basis for rigorously testing the negativity hypothesis because it facilitated both positive and negative peaks within the same sequence. Accordingly, the valence factor was dropped from the experimental design. As a basis for constructing the sequences, we used the low-intensity odours from Experiment 1. As can be seen in Figure 1, these were most comparable in the magnitude of their mean online ratings. This is important because testing the negativity effect is based on a comparison between pleasant and unpleasant odours.

To manipulate the trajectory, increasing sequences consisted of two unpleasant odours followed by four pleasant odours (−−++++). We constructed the decreasing (++++−−), U-shaped (++−−++), and inverse U-shaped (−++++−) sequences in a similar way. Across all sequences, four of the six odours were pleasant so participants were not overwhelmed with unpleasant stimuli. Each participant experienced the four trajectories in a short (36 s) and a long (72 s) version, yielding eight different sequences within participants. Half of participants first evaluated all four short sequences, followed by the four long sequences and vice versa for the other half of participants. Within the short and the long set of sequences, we presented all 24 possible order permutations of the four trajectories twice between participants. Hence, the experimental design again required 48 participants in total.

The other change in comparison to the first experiment was that participants rated each of the eight odours separately at the beginning of the experiment rather than at the end. Briefly experiencing all odours at the beginning allowed participants to familiarise themselves with the range and distribution of the stimuli before they rated the sequences, which presumably increased the reliability of the ratings and reduced order effects.

Participants

Participants were students at the University of Geneva. The mean age was 22 years and 38 of the 48 participants were female.

Results

Manipulation check

As shown in Figure 3, the mean online ratings obtained from the slider box aligned with the experimental valence manipulation. The mean ratings were negative during the diffusion of unpleasant odours and positive during the diffusion of pleasant odours. We obtained similar results for the separate evaluations of the eight odours at the beginning of the experiment.

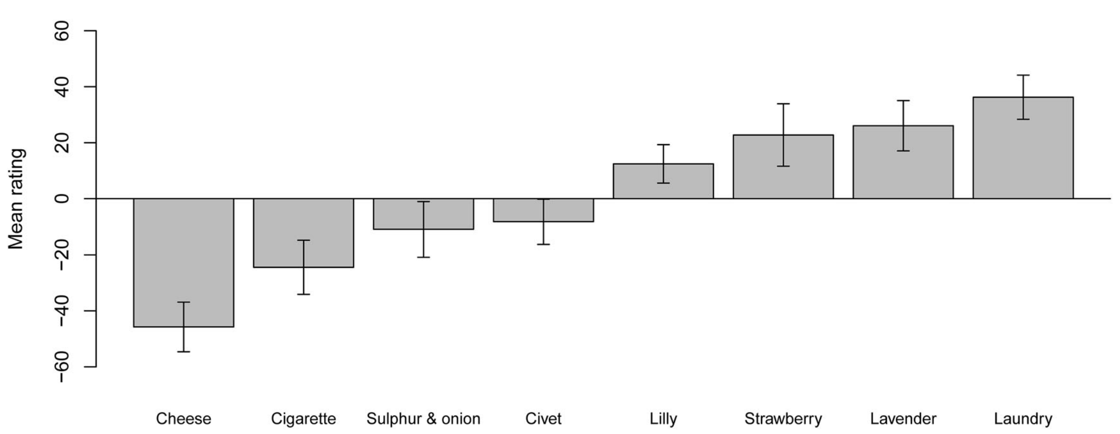

Figure 3. Mean online slider ratings across all participants for each odour. Error bars are 95% confidence intervals around the mean.

As an additional manipulation check, we plotted the median online ratings for the different trajectories separately for long and short odour sequences. Figure 4 shows that on average, the online ratings aligned with the trajectory manipulation but also revealed a fair amount of individual differences. The figure further shows that the ratings always originated from the neutral slider position (i.e. 0), as participants were instructed to reset the slider after each sequence.

Peak-end rule and duration neglect

The mean and median correlations between the average online ratings and the retrospective evaluations across participants were .61 and .65,

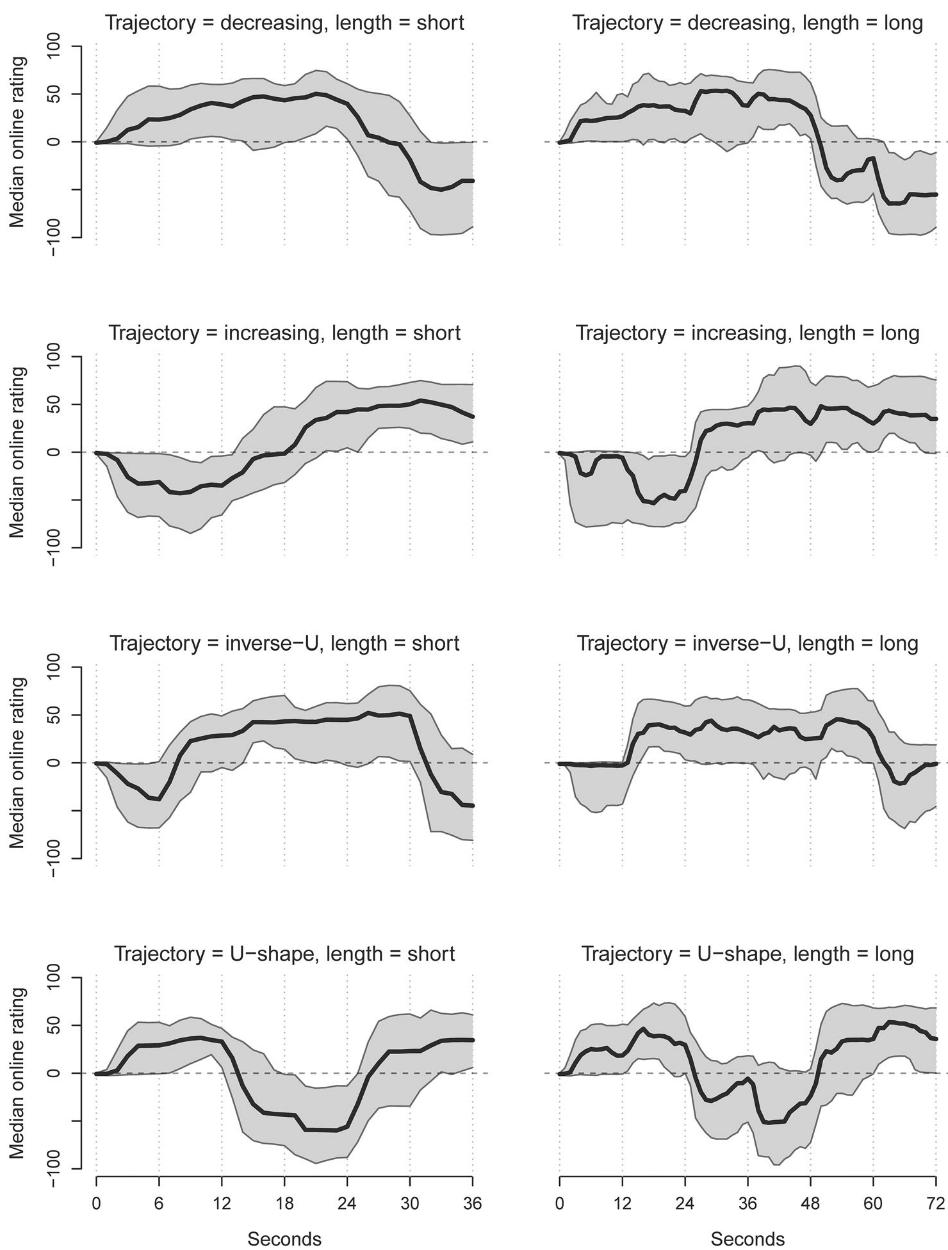

Figure 4. Median online rating across all participants for each trajectory (rows) and sequence length (columns). The grey areas indicate the inter-quartile range (i.e. the middle 50% of all ratings). The dashed horizontal line indicates the neutral (i.e. middle) slider position and the vertical dotted lines indicate the onset of a new odour within the sequence.

respectively. For the peak-end predictions, the mean and median correlations were .52 and .56. The mean correlation between the average online ratings and the peak-end predictions was .38. Compared to the first experiment, these correlations were lower. Presumably, this is because of the higher variance of the online valence ratings within each sequence as compared to Experiment 1.

To test the peak-end rule and duration neglect on statistical grounds, we followed the procedure of the first experiment and set up a baseline multilevel regression model (m2.0) that included random intercepts for individuals and fixed effects for the observed mean online ratings and the order in which participants experienced the odour sequences. Adding the duration of the sequence (short vs. long) as an additional fixed-effect predictor (m2.1) did not improve model accuracy according to BIC, indicating that retrospective evaluations did not depend on the duration of the sequence. The corresponding Bayes factor of the comparison is 19 in favour of m2.0, indicating strong evidence. Table 2 provides an overview of the estimated models including the respective fixed-effect estimates and BICs.

To test the peak-end rule, we included the most extreme positive and negative slider positions as additional predictor variables (m2.2). As shown in Table 2, this model predicts the data better than the baseline model. The corresponding Bayes factor is >10.000, indicating extreme evidence for m2.2 over m2.0. The estimated beta coefficients were positive, showing that the end and the most extreme positive and negative momentary evaluations were overweighed, as predicted by the peak-end rule.

As a stronger test for the peak-end rule, we also compared the baseline m2.0 model that included the mean online ratings as a predictor against a model that included only the peak and end ratings as predictors but not the mean online ratings (m2.3). Results showed that the latter model also explained the observed data better than the baseline model, yielding further evidence for the peak-end rule (BF > 10.000).

To further corroborate the overweighting of the end moment relative to other moments within the odour sequence, we tested a multilevel model that included the end as fixed effect and compared it to alternative models where the 5-s averaging window was systematically shifted through the online rating sequences, using analyses similar to those conducted in Experiment 1. The results of this comparison showed that all but two of the 32 alternative models yielded worse accuracy (i.e. higher BIC values) than the original model. The “windows” of the two alternative models with a better accuracy were the ones next to the end of the odour sequence (i.e. shifted by 1 and 2 s relative to the end rating). This indicates that the final experience within each odour sequence carried a higher weight, as predicted by the peak-end rule.

Negativity effect

To test if negative peaks received higher subjective weights than positive peaks, we estimated a reduced model, labelled m2.2.min, where we removed the highest rating (i.e. the most positive online rating) as a predictor from the original m2.2 model. Next, we compared this model to an alternative model m2.2.max that removed the lowest rating (i.e. the most negative online rating). As shown in Table 2, a comparison of the two models based on BIC revealed that the m2.2.min model predicted the observed data better than m2.2.max. The corresponding Bayes factor

Table 2. Overview of the fixed-effect estimates within the hierarchical regression models in Experiment 2.

	Model					
Predictor	m2.0	m2.1	m2.2	m2.3	m2.2.min	m2.2.max
(Intercept)	322.30	322.81	333.94	337.81	376.46	300.15
Sequential order	−7.29	−7.28	−8.11	−9.75	−7.64	−7.69
Mean online rating	3.16	3.16	1.57		2.33	2.66
Long sequence		−1.09				
End rating			0.33	0.32	0.39	0.44
Lowest rating			1.05	1.76	0.73	
Highest rating			1.00	1.87		0.38
R^2	.59	.59	.65	.60	.65	.63
Deviance	4522	4522	4459	4486	4473	4486
df	5	6	8	7	7	7
BIC	4551	4557	**4507**	4527	4515	4528

Note: BIC = Bayesian information criterion. Smaller BIC values indicate better model accuracy. The smallest BIC value (model m2.2) is printed in a bold font.

was 646, indicating strong evidence that the lowest peak of the online ratings was a better predictor of the retrospective evaluations than the highest peak. These results confirmed the hypothesis of a negativity effect for momentary affective evaluations that consisted of both pleasant and unpleasant episodes.

Influence of trajectories

Despite the successful manipulation of the odour trajectories on average, adding the type of trajectory as a categorical predictor to the regression model did not improve the model relative to the baseline model (m2.0). This indicates that the trajectory manipulation did not systematically influence the retrospective pleasantness ratings. However, contrasting just increasing and decreasing sequences slightly improved model accuracy relative to a null model (m2.0), according to BIC differences. The contrast indicated higher retrospective ratings for increasing relative to decreasing sequences, as predicted by the peak-end rule. The evidence for this contrast effect is not very strong though (BF = 13). Presumably, the relatively high variance between the different trajectories within individual participants diluted the effect of the trajectory manipulation so that the statistical power was not high enough to detect a difference across all trajectories.

Discussion Experiment 2

In this second experiment, we replicated Experiment 1's results showing that the peak-end rule was a good predictor for the retrospective evaluations of the olfactory sequences. Likewise, as in Experiment 1, the retrospective evaluations did not depend on the sequence lengths, indicating duration neglect.

In contrast to the first experiment, participants in the second experiment did overweight negative peaks relative to positive peaks. As this negativity effect occurred only in Experiment 2, it seemed to depend on the co-occurrence of both pleasant and unpleasant odours within the same sequence. This aligns with recent findings by Ganzach and Yaor (2018) who reported a negativity effect for retrospective evaluations of extended experiences that last for days or weeks. As in the first experiment, this comparison between positive and negative rests on the assumption that the negative and the positive online rating ratings were indeed comparable.

General discussion

In the two experiments presented here, participants repeatedly evaluated relatively short odour sequences. Results showed that the peak-end rule and duration neglect applied to positive and negative emotional experiences alike. Using odours was a unique aspect of our design; it allowed the creation of genuinely affective experiences. Moreover, it allowed the comparison of pleasant and unpleasant experiences within the same domain. Our data provides empirical evidence that the peak-end rule and duration neglect do not depend on the valence of the emotional experiences being assessed.

Another advantage of our design is that the onset and change of odours within a sequence were not straightforward to discern. This experimental approach thus fostered a holistic evaluation that presumably facilitated the peak-end rule (Ariely et al., 2000; Ariely & Loewenstein, 2000). Perhaps part of the reason why previous studies using pleasant experiences often did not find evidence for the peak-end rule could be that they relied on discrete or heterogeneous stimuli such as meals (Rode et al., 2007) or activities over the course of a day (Miron-Shatz, 2009) that are difficult to integrate into an overall evaluation.

A disadvantage of our experimental approach was that we determined the number of participants based on pragmatic considerations such as having a balanced design and not on expected effect sizes and statistical power. The latter approach would have been preferred because it helps to ensure sufficient statistical power and hence a lower risk of false positive findings (Brysbaert & Stevens, 2018; Loken & Gelman, 2017).

In both experiments, participants' peak and end evaluations within each odour sequence were highly correlated with the mean of their online evaluations. This correlation was particularly high for the first experiment where the variance within each sequence was lower. To the degree that the momentary evaluations throughout a sequence are similar, the peak and end moments would be as good a predictor as any other segment (e.g. Tully & Meyvis, 2016). Here, we can rule out this alternative explanation because in our data, the peak and end ratings were credible predictors of the retrospective evaluations even after controlling for the mean of the online evaluations. Likewise, alternative linear models that included other momentary evaluations than the end as predictors predicted the data worse.

In our data, the different trajectory conditions (i.e. increasing, decreasing, U-shaped, and inverse U-shaped) had no systematic effect on participants' retrospective evaluations. Odour evaluations are highly variable (e.g. Ferdenzi et al., 2011). Thus, a possible reason why the different trajectories had no effect was the relatively high degree of variance in the observed online ratings between different trajectories within the same individual. In other words, participants' online ratings varied across the course of the sequence, thus providing the basis of testing the peak-end rule, but their ratings did not always follow our trajectory manipulation.

Participants in the second experiment over-weighed negative peaks relative to positive peaks. As this negativity effect did not occur in the first experiment, it seems to depend on the co-occurrence of both pleasant and unpleasant odours within the same sequence (see Ganzach & Yaor, 2018 for similar findings in the context of extended experiences). These findings contribute to a better understanding of the peak-end rule because they render explanations hinging exclusively on negative experiences less likely. For example, the evolutionary hypothesis that the peak is relevant because it signals the maximum damage (e.g. Fredrickson, 2000) cannot fully explain the data at hand. The results accord with theories on human memory predicting that both recent (e.g. Hogarth & Einhorn, 1992) and intense (e.g. Levine, Lench, Karnaze, & Carlson, 2018) emotional moments are better remembered. The finding that negative peaks received a slightly higher weight in mixed sequences suggests an influence of valence on memory that prioritises negative information. A review on the influence of emotions on memory by Kensinger (2009) summarised converging evidence for such a model and also pointed to possible underlying neural processes. However, the finding that negative peaks are over weighted has to be considered with caution, as it occurred in only one of the two experiments we ran. In addition, direct comparisons between valences are limited because they rest on the assumption that ratings on the positive side of the rating scale are comparable. Even when using the same type of stimulus material, this may not necessarily be the case.

Besides these methodological considerations, there is emerging evidence suggesting that relevance, rather than valence, is key in explaining the influence of emotions on memory (Montagrin et al., 2013; see also Talmi et al., 2013). Relevance is independent of valence: Pleasant, unpleasant, but also a priori neutral stimuli conveying goal-, need-, or value-related significance are relevant (Sander et al., 2003). In future experiments investigating the peak-end rule, relevance would consequently be a particularly interesting factor to consider. Here too, odours could be of valuable help: For example, food-related odours are relevant when one is hungry, but much less when satiated. Hence, one could compare the same odour sequences in different physiological states to assess whether relevance plays a role in the peak-end rule.

Notes

1. The raw data and the *R* script used to analyse the data are available online at https://osf.io/h5g4a/.
2. The online supplementary material contains a plot of the median online rating across all participants for each trajectory manipulation showing that the observed trajectory ratings did not correspond well with the ones we intended.
3. For one participant, the correlation was negative, indicating a possible misunderstanding of the task. Removing this participant from the analyses did not lead to a qualitative change in the subsequent results.

Acknowledgements

We would like to thank Sami Agraniou, Dr. Sylvain Delplanque, Dr. Andres Posada and Prof. David Sander for their help with running the experiments, including installing the olfactometer, programming the experimental script, and collecting the data. The authors thank all the members of the Human Perception and Bioresponses Department of the Research and Development Division of Firmenich, SA, for their valuable advice and their theoretical and technical competence and Prof. Claude Messner for providing a guest office to the first author.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The work was supported by a research grant from the Swiss National Science Foundation [grant number 100014_175849] to the first author.

References

Alba, J. W., & Williams, E. F. (2013). Pleasure principles: A review of research on hedonic consumption. *Journal of Consumer Psychology*, *23*(1), 2–18.

Ariely, D., Kahneman, D., & Loewenstein, G. (2000). Joint comment on "when does duration matter in judgment and decision making?" (Ariely & Loewenstein, 2000). *Journal of Experimental Psychology: General*, *129*(4), 524–529.

Ariely, D., & Loewenstein, G. (2000). When does duration matter in judgment and decision making? *Journal of Experimental Psychology: General*, *129*(4), 508–523.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.

Baumgartner, H., Sujan, M., & Padgett, D. (1997). Pattern of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgments. *Journal of Marketing Research*, *34*, 219–232.

Blows, W. T. (2001). *The biological basis of nursing: Clinical observations*. London: Routledge.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*, 1–20.

Cain, W. S. (1974). Perception of odor intensity and the time-course of olfactory perception. *ASHRAE Transactions*, *80*, 53–75.

Chajut, E., Caspi, A., Chen, R., Hod, M., & Ariely, D. (2014). In pain thou shalt bring forth children: The peak-and-end rule in recall of labor pain. *Psychological Science*, *25*(12), 2266–2271.

Chrea, C., Grandjean, D., Delplanque, S., Cayeux, I., Le Calvé, B., Aymard, L., … Scherer, K. R. (2009). Mapping the semantic space for the subjective experience of emotional responses to odors. *Chemical Senses*, *34*(1), 49–62. doi:10.1093/chemse/bjn052

Cojuharenco, I., & Ryvkin, D. (2008). Peak-end rule versus average utility: How utility aggregation affects evaluations of experiences. *Journal of Mathematical Psychology*, *52*(5), 326–335.

Coppin, G., Delplanque, S., Bernard, C., Cekic, S., Porcherot, C., Cayeux, I., & Sander, D. (2014). Choice both affects and reflects preferences. *The Quarterly Journal of Experimental Psychology*, *67*(7), 1415–1427. doi:10.1080/17470218.2013.863953

Coppin, G., Delplanque, S., Porcherot, C., Cayeux, I., & Sander, D. (2012). When flexibility is stable: Implicit long-term shaping of olfactory preferences. *PLoS ONE*, *7*, e37857. doi:10.1371/journal.pone.0037857

Cornil, Y., & Chandon, P. (2016). Pleasure as a substitute for size: How multisensory imagery can make people happier with smaller food portions. *Journal of Marketing Research*, *53*(5), 847–864.

Delplanque, S., Coppin, G., Bloesch, L., Cayeux, I., & Sander, D. (2015). The mere exposure effect depends on an odor's initial pleasantness. *Frontiers in Psychology*, *6*, 920.

Delplanque, S., Grandjean, D., Chrea, C., Aymard, L., Cayeux, I., Le Calvé, B., … Sander, D. (2008). Emotional processing of odors: Evidence for a nonlinear relation between pleasantness and familiarity evaluations. *Chemical Senses*, *33*(5), 469–479. doi:10.1093/chemse/bjn014

Dixon, M. J., Victorino, L., Kwortnik, R. J., & Verma, R. (2017). Surprise, anticipation, and sequence effects in the design of experiential services. *Production and Operations Management*, *26*(5), 945–960.

Do, A. M., Rupert, A. V., & Wolford, G. (2008). Evaluations of pleasurable experiences: The peak-end rule. *Psychonomic Bulletin & Review*, *15*(1), 96–98.

Ferdenzi, C., Roberts, S. C., Schirmer, A., Delplanque, S., Cekic, S., Porcherot, C., … Grandjean, D. (2013). Variability of affective responses to odors: Culture, gender, and olfactory knowledge. *Chemical Senses*, *38*(2), 175–186.

Ferdenzi, C., Schirmer, A., Roberts, S. C., Delplanque, S., Porcherot, C., Cayeux, I., … Grandjean, D. (2011). Affective dimensions of odor perception: A comparison between Swiss, British, and Singaporean populations. *Emotion*, *11*(5), 1168–1181. doi:10.1037/a0022853

Finn, B. (2010). Ending on a high note: Adding a better end to effortful study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *36*(6), 1548–1553.

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion*, *14*(4), 577–606.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*(1), 45–55.

Ganzach, Y., & Yaor, E. (2018). The retrospective evaluation of positive and negative affect. *Personality and Social Psychology Bulletin*. Advance online publication. doi:10.1177/0146167218780695

Geng, X., Chen, Z., Lam, W., & Zheng, Q. (2013). Hedonic evaluation over short and long retention intervals: The mechanism of the peak-end rule. *Journal of Behavioral Decision Making*, *26*, 225–236.

Herz, R. S., Eliassen, J., Beland, S., & Souza, T. (2004). Neuroimaging evidence for the emotional potency of odor-evoked memory. *Neuropsychologia*, *42*(3), 371–378.

Hilgard, J., Weinberg, A., Proudfit, G. H., & Bartholow, B. D. (2015). The negativity bias in affective picture processing depends on top-down and bottom-up motivational significance. *Emotion*, *14*(5), 940–949.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*(1), 1–55.

Hsee, C. K., & Abelson, R. P. (1991). Velocity relation: Satisfaction as a function of the first derivative of outcome over time. *Attitudes and Social Cognition*, *60*(3), 341–347.

Hui, S. K., Meyvis, T., & Assael, H. (2014). Analyzing moment-to-moment data using a Bayesian functional linear model: Application to TV show pilot testing. *Marketing Science*, *33*(2), 222–240.

Ischer, M., Baron, N., Mermoud, C., Cayeux, I., Porcherot, C., Sander, D., & Delplanque, S. (2014). How incorporation of scents could enhance immersive virtual experiences. *Frontiers in Psychology*, *5*, 736.

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*(6), 401–405.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistic Association*, *90*, 773–795.

Kemp, S., Burt, C. D. B., & Furneaux, L. (2008). A test of the peak-end rule with extended autobiographical events. *Memory & Cognition*, *36*(1), 132–138.

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, *1*(2), 99–113.

Langer, T., & Weber, M. (2005). Myopic prospect theory vs. myopic loss aversion: How general is this phenomenon? *Journal of Economic Behavior & Organization*, *56*(1), 25–38.

Levine, L. J., Lench, H. C., Karnaze, M. M., & Carlson, S. J. (2018). Bias in predicted and remembered emotion. *Current Opinion in Behavioral Sciences*, *19*, 73–77.

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, *355*(6325), 584–585.

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2014). Extreme outcomes sway risky decisions from experience. *Journal of Behavioral Decision Making*, *27*(2), 146–156.

McElreath, R. (2016). *Statistical rethinking*. Boca Raton, FL: CRC Press.

Miron-Shatz, T. (2009). Evaluating multiepisode events: Boundary conditions for the peak-end rule. *Emotion*, *9*(2), 206–213.

Mohanty, A., & Gottfried, J. A. (2013). Examining emotion perception and elicitation via olfaction. In J. L. Armony & P. Vuilleumier (Eds.), *The Cambridge handbook of human affective neuroscience* (pp. 241–264). Cambridge: Cambridge University Press.

Montagrin, A., Brosch, T., & Sander, D. (2013). Goal conduciveness as a key determinant of memory facilitation. *Emotion*, *13*(4), 622–628.

Montagrin, A., Sterpenich, V., Brosch, T., Grandjean, D., Armony, J., Ceravolo, L., & Sander, D. (2018). Goal-relevant situations facilitate memory of neutral faces. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(6), 1269–1282.

Oud, B., & Coppin, G. (2012). Extending the study of decision values to cases where options are presented using different sensory modalities. *The Journal of Neuroscience*, *32*(7), 2248–2249.

Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2016). Attentional bias for positive emotional stimuli: A meta-analytic investigation. *Psychological Bulletin*, *142*(1), 79–106.

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*(1), 3–8.

Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain*, *104*(1), 187–194.

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*(1), 18–29.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, *14*(4), 303–316.

Schäfer, T., Zimmermann, D., & Sedlmeier, P. (2014). How we remember the emotional intensity of past musical experiences. *Frontiers in Psychology*, *5*, 911.

Schreiber, C. A., & Kahneman, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology General*, *129*(1), 27–42.

Seta, J. J., Haire, A., & Seta, C. E. (2008). Averaging and summation: Positivity and choice as a function of the number and affective intensity of life events. *Journal of Experimental Social Psychology*, *44*(2), 173–186.

Stussi, Y., Pourtois, G., & Sander, D. (2018). Enhanced Pavlovian aversive conditioning to positive emotional stimuli. *Journal of Experimental Psychology: General*, *147*, 905–923.

Talmi, D., Ziegler, M., Hawksworth, J., Lalani, S., Herman, C. P., & Moscovitch, M. (2013). Emotional stimuli exert parallel effects on attention and memory. *Cognition & Emotion*, *27*(3), 530–538.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 9659–9664.

Tully, S., & Meyvis, T. (2016). Questioning the end effect: Endings are not inherently over-weighted in retrospective evaluations of experiences. *Journal of Experimental Psychology: General*, *145*(5), 630–642.

Vaish, A., Grossman, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*(3), 383–403.

Varey, C., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making*, *5*(3), 169–185.

Von Helversen, B., Coppin, G., & Scheibehenne, B. (2019). *Reality stinks: Testing risk preferences in gambles with real-life consequences* (Working Paper, University of Zürich, Switzerland).

Wirtz, D., Kruger, J., Napa Scollon, C., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science*, *14*(5), 520–524.