Supplemental Material

The Emotion Filmmaker: Temporal Memory, Time–Emotion Integration, and Affective Style

Mengsi Li* and Regina C. Lapate*

Department of Psychological and Brain Sciences, University of California, Santa Barbara

*Corresponding Author:
Mengsi Li
mengsili@ucsb.edu
Department of Psychological and Brain Sciences
University of California, Santa Barbara
CA 93106, U. S

Regina C. Lapate
lapate@ucsb.edu
Department of Psychological and Brain Sciences
University of California, Santa Barbara
CA 93106, U. S

Preregistered *vs.* Reported Analysis

Any discrepancies between reported analyses and the preregistered analysis plan (https://osf.io/3h8ns) are detailed below.

Retrospective emotion biases. First, in the preregistration, we planned to establish the best-fitting model predicting retrospective emotion ratings by including average, peak, end, duration-weighted average, duration-weighted peak, and duration-weighted end in *all possible orders*. However, to reduce the severity of multiple comparison corrections and more directly test the source(s) of retrospective emotion biases, we tested duration neglect by comparing the average *vs.* duration-weighted average models. We tested the end effect as well as the peak effect by entering end and (negative and positive) peaks in a *step-wise manner* (as opposed to using all orders) (Scheibehenne & Coppin, 2020). Note that our preregistration focused on novel hypotheses tested by this study. Prior to testing our novel hypothesis, we first examined whether we replicated well-known retrospective emotion biases that serve as the basis for the pre-registration and the remainder of the manuscript (peak, end, and duration neglect).

Second, the analysis to disentangle the contributions of positive *vs.* negative peaks to retrospective emotion biases was not included in the preregistration as it was motivated by recently published research (Scheibehenne & Coppin, 2020) that the authors became aware after they preregistered the present study.

Capacity limits of temporal processing & retrospective emotion. We preregistered analyses centering on sequence length (i.e., the number of movies in a sequence) and retrospective emotion biases. However, in light of recent work underscoring dissociations across different temporal memory metrics, including discrepancies when examining ordinal *vs.* continuous temporal effects, greater stability of temporal memory duration metrics across experimental contexts (Wen & Egner, 2022), and evidence of the impact of emotional challenges on temporal duration memory (for a recent review, see Wang et al., 2022), we modeled sequence-duration effects (reported in the main manuscript)

separately from sequence length effects. For completeness, we also report results examining sequence length in the *Supplemental Material*, including the preregistered analysis predicting trial-wise retrospective emotion bias with sequence length as a predictor in a mixed-effect model framework (see *Testing capacity limits: sequence length* below).

Temporal duration errors. In our preregistration, we also proposed to calculate temporal duration errors as the difference between participants' recollected and the actual duration ratios of subjectively experienced negative emotion for each sequence based on participants' *momentary* ratings in the emotional sequences task. However, following internal piloting, we opted not to use this metric because participants were ultimately instructed to recall the duration ratio of *negative movies* (as opposed to the duration ratio of their own subjectively experienced negative emotion), a question that participants found easier to understand and respond to. Furthermore, several aspects could render subjective emotion duration unrepresentative of participants' perception of the movie duration. Emotional responses induced by a movie are dynamic and can persist beyond the initial provocation window. Yet, participants' continued emotional responses to a movie do not necessarily imply that participants perceive the movie itself as lasting longer. As mentioned, our temporal memory task instruction asked participants to recall the duration of the movies, as opposed to the duration of their subjective feelings. We therefore opted to calculate temporal duration errors using participants' recalled negative *vs.* positive durations relative to the movies' objective durations.

Dispositional NA. To test the hypothesis that dispositional NA is associated with greater retrospective emotion biases and temporal memory indices, we pre-registered using mixed-effect models. However, as each participant provides only one dispositional NA score, we opted to use an across-subjects (i.e., individual-differences) analysis instead (Pearson's r) for parsimony and ease of interpretation.

Data exclusion criteria. Participants' data that were lost (partially or fully) due to technical issues (likely due to suboptimal internet connection) were excluded from analysis. This exclusion criterion was not preregistered as we did not foresee this technical issue.

Supplemental Methods

Task Design: Emotional-Movie Sequences

To examine whether increasing demands on temporal processing capacity produced retrospective emotion biases, we manipulated sequence length (i.e., the number of movies per sequence) using 4 levels (4, 5, 6, *vs*. 7 movies per sequence). Average movie normative valence and arousal ratings were matched across sequence positions (ps > 0.051), and across sequence-length levels (ps > 0.08). The number of positive *vs.* negative movie clips was matched across sequences, as follows: In 4- and 6-movie sequences, equal numbers of positive and negative movies were shown; in 5- and 7-movie sequences, half the sequences included one extra positive (*vs.* negative) movie. Emotional-movie sequences included both positive and negative movie clips. To examine duration memory, we manipulated the duration of negative *vs.* positive movies in each sequence such that negative-movie duration was longer than positive for half of the sequences. We also ensured that the duration ratios of negative movies were not significantly different from 50% relative to the total sequence duration (ps > 0.45).

Supplemental Procedure

Emotional sequences task. Momentary ratings were recorded every 3 frames (~50ms on a 60Hz computer).

Rating data processing. As mentioned in the main manuscript, we binned raw continuous emotion ratings to minimize noise and maximize the reliability of the estimates of momentary emotions and emotional changes, while retaining adequate inter-individual variability. Specifically, the raw, high-resolution continuous scale is

susceptible to capturing noise caused by participants' unintentional movements of their mouse/trackpad in our unmonitored online study. After data collection, prior to analyzing the data, we adopted a data-driven approach to determine the optimal number of rating points to bin continuous ratings. We calculated and plotted emotion rating variability across participants as a function of the rating-point count (ranging from 2 to 30). Specifically, after converting raw continuous ratings to a binned scale, we computed the standard deviation (SD) of ratings across participants at each timepoint within each sequence, and then averaged the SDs across all timepoints of all sequences. To make these averaged SDs on different scales comparable, we standardized them by multiplying each by a scaling factor (1/Range of the scale). We plotted these scaled-averaged SDs against the number of rating points. The resulting plot revealed that, as expected, emotion-rating variability across participants initially decreased with an increasing number of rating points—but, after reaching 9 points, further decreases in variability were rather negligible (see **Fig. S1**). This suggested that a 9-point rating scale conversion provided reliable estimates of momentary emotions across participants. Beyond a 9-point scale, this analysis suggested that reliability plateaued, with the trade-off of increasing sensitivity to noise and reducing inter-individual variability. In light of these insights, considered together with the ubiquity of the 9-point scale in emotion research—which is well-validated and widely used in the field (Bradley & Lang, 1994) —we settled on binning the ratings using a 1-9 point scale.

Temporal memory tasks. In the *temporal order* memory task, representative movie images from each previously-watched sequence (one image per movie) were quasi-randomly placed on the screen (with the constraint that they were shown in a different order from the original presentation).

Missing values for the temporal order memory metrics. For the temporal memory task, missing values (i.e., when/if participants did not provide a temporal order response for a given movie clip in a sequence) were present in 5.02% of all movie sequences. To

circumvent reliability issues with the estimation of temporal order scores for sequences of different lengths with a variable number of missing values, sequences with missing values were excluded from further analysis of temporal order accuracy.

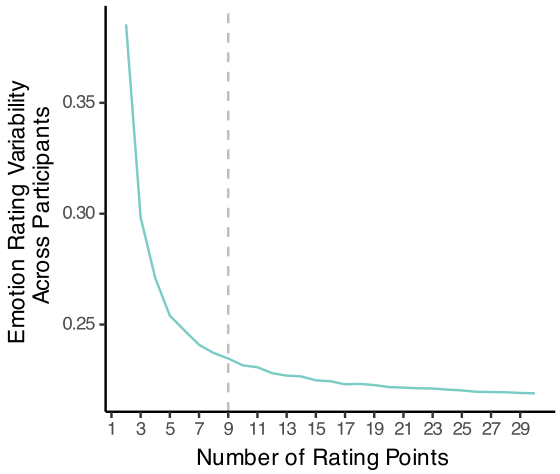


Fig. S1. Across-participants emotion rating variability is plotted as a function of the scale resolution.

Mood Questionnaires

In addition to the mood scales described in the main manuscript (ATQ-NA, BDI, and STAI, see *Procedure: Mood questionnaires*), we also administered one additional preregistered scale--the effortful control (EC) subscale of ATQ (Evans & Rothbart, 2007), which measures attentional, inhibitory, and activation control, and is commonly used in our laboratory in the context of a larger investigation of affective style and cognitive control. Although not the focus of the current manuscript, for completeness, we note that ATQ-EC was not associated with retrospective emotion biases, temporal order memory, or duration memory (ps > 0.1).

Supplemental Analyses & Results

Retrospective Emotion: Ascertaining Specificity of "Peak" and "End" Contributions

Our main analyses revealed significant *peak* and *end* biases from negative peaks in retrospective emotion. To reiterate, we found that including *end* and *negative peak* momentary ratings significantly improved the model fits compared to models that only included the duration-weighted average (see *Retrospective ratings reveal a (negative) peak-end effect* in the main manuscript). To ascertain the specificity of these findings to peak and end momentary ratings—and rule out the possibility that including *any* momentary rating would have similarly improved model fits—we tested (a) whether a random rating significantly predicts retrospective emotion ratings and (b) whether the inclusion of end or negative peak moments improved the model fit to a greater extent than the inclusion of a random rating (as indexed by the Bayes ΔBIC between models) (Scheibehenne & Coppin, 2020; Wagenmakers et al., 2018). The random rating was randomly selected from all timepoints within a sequence (with the constraint that the randomly-selected rating was not itself the peak or end rating).

To test the end effect, we compared the model including duration-weighted average and *end* (m2) against the model including duration-weighted average and a *random rating* (m2.1). The random rating did not significantly predict retrospective emotion ($\beta = 0.02$, SE = 0.01, $t = 1.34$, 95% CI [-0.009, 0.05], $p = 0.18$). Furthermore, the end model significantly outperformed the random-rating model (Bayes factor on ΔBIC of m2 *vs.* m2.1: ΔAIC = -519, ΔBIC = -519, BF > 100). To test the negative peak effect, we compared the model including duration-weighted average, end, *negative peak*, and positive peak (m3) against the model replacing the negative peak with a random rating (m3.1). Similarly, the random rating was not a significant predictor for retrospective emotion ($\beta = 0.02$, SE = 0.01, $t = 1.45$, 95% CI [-0.007, 0.005], $p = 0.15$), and the negative-peak model yielded a better model fit than the random-rating model (Bayes factor on ΔBIC of m3 *vs.* m3.1: ΔAIC = -955, ΔBIC = -956, BF > 100).

Testing Capacity Limits: The Positive Peak

As mentioned, our results indicated that *negative* peaks in momentary ratings predicted retrospective emotion judgments ($\beta = 0.14$, SE = 0.03, $t = 4.26$, 95% CI [0.07, 0.20], $p < 0.001$), whereas *positive* peaks did not ($\beta = 0.03$, SE = 0.03, $t = 0.86$, 95% CI [-0.03, – 0.09], $p = 0.39$) (see Model 3 (m3) in *Retrospective ratings reveal a (negative) peak-end effect* in the main manuscript). For completeness, here we also examined whether the positive peak was sensitive to the increase in emotional-sequence duration. To do so, we tested the interaction term of positive peak by sequence duration following the model testing the contribution of peak ratings. We found a marginally significant interaction effect, suggesting that the positive peak effect tended to become larger as emotional-sequence duration increased ($\beta = 0.009$, SE = 0.005, $t = 1.77$, 95% CI [-0.001, 0.02], $p = 0.08$).

Testing Capacity Limits: Sequence Length

In a recent study, Asutay et al. (2021) revealed that increasing the number of emotional images in a sequence, as well as the total duration of emotional sequences, can both produce a 'peak' effect in retrospective emotion ratings. Therefore, to test putative capacity limits in temporal information processing, we examined whether retrospective emotion biases increased as a function of sequence duration (reported in the main manuscript) and sequence length, as reported next.

We found that sequence length also interacted with 'end' momentary ratings in predicting retrospective emotion judgments ($F = 3.56$, $p = 0.01$; **Fig. S2**). Pairwise comparisons suggested greater end effects at sequence length '7' *vs.* '4' ($t = -2.92$, $p = 0.02$), although not at other contrasts (ps > 0.11). Thus, these results partially align with sequence-duration findings reported in the main manuscript (see section *The 'end' effect increases with sequence duration*). The negative peak did not interact with sequence length ($F = 0.38$, $p = 0.17$). The positive peak marginally interacted with sequence length in predicting retrospective emotion ratings ($F = 2.29$, $p = 0.08$), but pairwise comparisons of sequence length levels were non-significant (ps > 0.1).

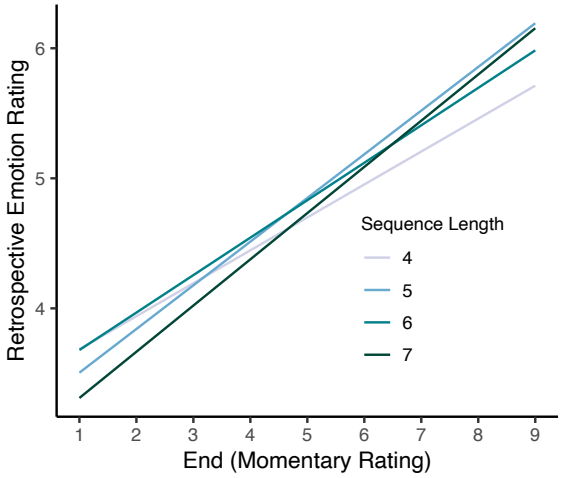


Fig. S2. Sequence length interacted with end (momentary ratings) to predict retrospective emotion ratings ($F = 3.56$, $p = 0.01$).

Following our pre-registration, we examined whether sequence length predicted retrospective emotion biases. To do so, we predicted trial-wise (sequence-wise) indices of retrospective emotion bias from sequence length in a mixed-effects model. Results indicated that retrospective emotion biases did not differ across sequence length levels (ps > 0.1).

Temporal Memory Accuracy

We examined whether temporal duration memory and temporal order memory performance were significantly above chance using permutation tests.

Temporal duration memory. In the main analysis, we computed the temporal duration error (recalled *minus* actual negative duration ratio). To compare temporal duration memory performance against chance, we converted the error metric into temporal duration accuracies by subtracting the absolute temporal duration error from 1. For each participant, we randomly shuffled sequence labels 1,000 times, computing the temporal duration accuracy for each permutation, which resulted in a distribution of empirically derived chance performance for temporal duration memory accuracies,

against which participants' average temporal memory was compared. We found that temporal duration memory was significantly higher than chance ($p < 0.001$; see **Fig. S3**).

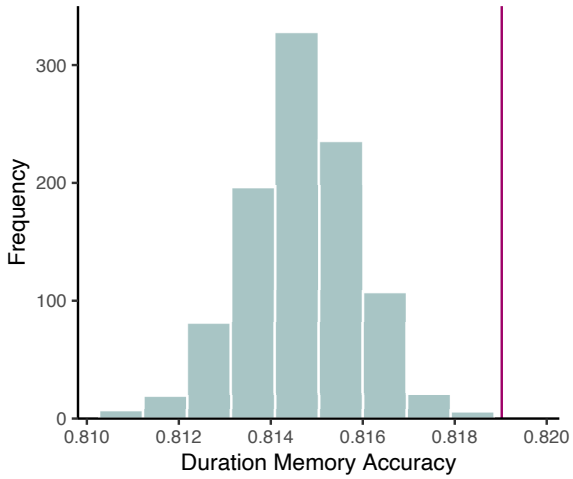


Fig. S3. ***Temporal duration memory.*** A permutation test indicated that participants' memory of the relative temporal duration of prior emotional events (vertical violet line) was significantly better than chance ($p < 0.001$).

Temporal order memory. For every participant, we shuffled their remembered movie order ranks for each sequence 1,000 times. We then computed absolute as well as relative order memory scores (see *Temporal memory metrics*) for each permutation, which formed the empirically-derived distributions of chance-level temporal order memory performance. Participants' absolute order memory was significantly above chance ($p < 0.001$; see **Fig. S4a**). Similarly, participants' relative order memory was also significantly above chance ($p < 0.001$; see **Fig. S4b**).

The Dissociation Between Temporal Order Memory and Duration Memory

The correlations between temporal duration errors and temporal order memory were non-significant (absolute: $r = 0.01$, $t(58) = 0.09$, 95% CI [-0.24, 0.26], $p = 0.93$; relative: $r = 0.11$, $t(58) = 0.84$, 95% CI [-0.15, 0.35], $p = 0.41$).

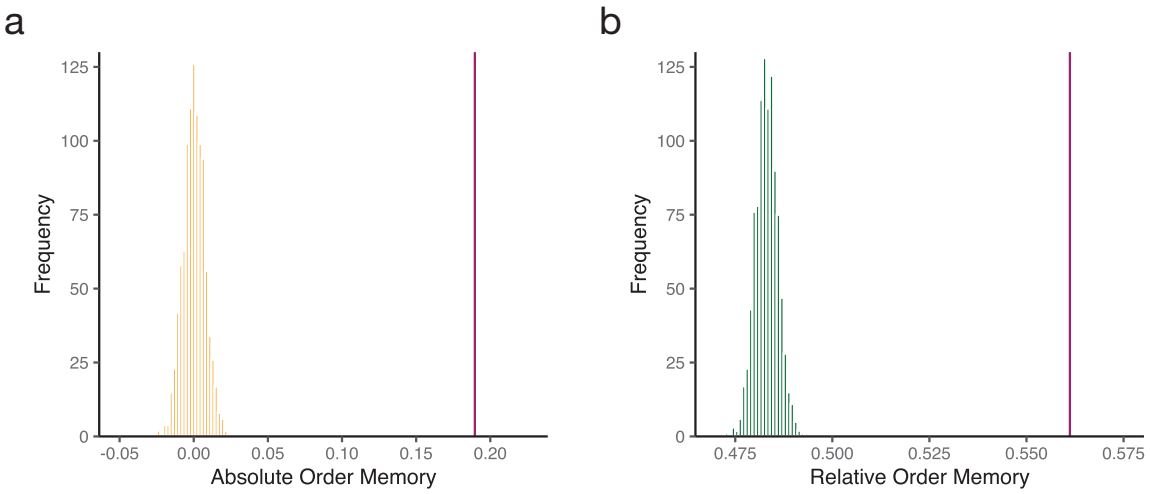


Fig. S4. ***Temporal order memory.*** **(a)** Permutation testing revealed that participants' memory for the absolute temporal order of movies within a sequence (vertical violet line) was significantly better than chance ($p < 0.001$). **(b)** Relative temporal order memory (vertical violet line) was likewise above chance ($p < 0.001$).

Associations Between Temporal Memory and Retrospective Emotion Biases: Individual Model Fit

For the primary analysis examining the association between temporal memory and retrospective emotion biases (see the main manuscript), we computed a trial-wise metric of retrospective emotion bias by taking the absolute difference between participants' retrospective emotion ratings and the duration-weighted average (per sequence). This metric therefore includes retrospective biases from all possible sources, such as peak biases, end biases, and duration neglect. To disentangle which of these distinct sources of retrospective emotion biases potentially underlay the (trial-wise) association between temporal duration memory and retrospective bias we uncovered, we next adopted an individual-model-fit comparison approach. For each participant, we independently fit the following linear regression models to predict retrospective emotion ratings: a **duration-weighted average model** ('duration-weighted average' as the predictor), an **average model** ('average' as the predictor), an **end model** ('end' as the predictor), and a **negative peak model** ('negative peak' as the predictor). (We also fit a **positive peak model**, the results of which are reported below). As mentioned, we next

computed the difference between model fits (i.e., ΔBIC) to quantify the model evidence for these different possible drivers of retrospective emotion biases, as follows.

Duration neglect was indexed by the ΔBIC between the duration-weighted average *vs.* the average model. As *bigger* BICs indicate *worse* model fits, a worse fit of the duration-weighted-average *vs.* average model suggests larger ***duration neglect*** (ΔBIC[Duration-weighted average – Average]). We defined the ***end effect*** as the ΔBIC between the average *vs.* end model (ΔBIC[Average – End]), such that a larger *end effect* corresponded to a worse model fit between average *vs.* end models. We operationalized the ***negative peak effect*** as the ΔBIC between the average *vs.* negative peak models (ΔBIC[Average – Negative peak]), such that a larger negative peak effect corresponded to a worse model fit of the average *vs.* negative peak model. Similarly, the ***positive peak effect*** was quantified as the ΔBIC between the average *vs.* positive peak models (ΔBIC[Average – Positive peak]), such that a larger ***positive peak effect*** reflected a worse model fit of the average *vs.* positive peak model.

To determine the sources of retrospective emotion biases that were associated with temporal memory, we next correlated temporal duration errors with ΔBICs of models indexing the strength of duration neglect, negative peak as well as end effects (as explained above), while partialing out individual differences due to noise as reflected by differences in ΔR^2 (for a similar approach, see Ballard et al., 2019). The results from this analysis are reported in the *Temporal memory and retrospective emotion biases: model comparison* section in the main text**.** (Note that the association between positive peak bias temporal duration (memory) errors, which we report for completeness, was positive, $r = 0.32$, $p = 0.01$).

Associations Between Dispositional NA and Retrospective Emotion Biases: Individual Model Fit

Dispositional NA did not correlate with the ΔBIC-indexed individual biases, including duration neglect, end bias, negative peak bias, and positive peak bias (rs < 0.16, ps > 0.09).

Results On Raw Continuous Emotion Ratings

For reasons detailed in *Rating data processing* in the manuscript and *Supplemental Material*, we converted the raw emotion ratings from the original 0-100 scale to a 1-9 scale. To ensure that this rating scale conversion did not influence our findings, we re-analyzed the data using the original raw continuous ratings. The results of this reanalysis were fully consistent with our original results.

Retrospective ratings reveal a (negative) peak-end effect

The end effect. We observed an 'end' effect as indicated by the significant fit of 'end' momentary ratings in predicting retrospective emotion (β = 0.30, SE = 0.01, t = 24.97, 95% CI [0.28, 0.33], p < 0.001). Moreover, adding 'end' as a predictor significantly improved the model fit (likelihood ratio test on m2 *vs.* m1.2: $\chi^2(1)$ = 553, ΔAIC = -551, ΔBIC = -545, ΔLL = 277, p < 0.001).

The peak effect: Negative and positive peaks. We found that the inclusion of positive and negative peaks significantly improved the model fit compared to the 'end'-only model above, therefore indicating a peak effect (Bayes factor on ΔBIC of m3 *vs.* m2: ΔAIC = -2940, ΔBIC = -2929, ΔLL = 1749, BF > 100). Moreover, negative peaks ('troughs') exerted a greater influence on retrospective emotion compared to positive peaks. Both negative and positive peak momentary ratings predicted retrospective emotion judgments, but the effect of positive peak was close to marginal (negative peak: β = 0.15, SE = 0.03, t = 5.64, 95% CI [0.10, 0.20], p < 0.001; positive peak: β = 0.05, SE = 0.03, t = 1.98, 95% CI [0.0003, 0.11], p = 0.048). Excluding the negative peak from the simultaneous model resulted in a greater decrease in model fit (ΔAIC = 1537, ΔBIC = 1531), compared to excluding the positive peak (ΔAIC = 1321, ΔBIC = 1315). In addition, the effect of the positive peak was

non-significant when it was fitted without the negative peak ($\beta = 0.004$, SE = 0.02, $t = 0.18$, 95% CI [-0.04, 0.05], $p = 0.86$).

Duration neglect. Both average and duration-weighted average made significant contributions to retrospective emotion ratings (average: $\beta = 0.91$, SE = 0.03, $t = 35.23$, 95% CI [0.86, 0.96], $p < 0.001$; duration-weighted average: $\beta = 0.89$, SE = 0.02, $t = 42.52$, 95% CI [0.85 0.93], $p < 0.001$). The inclusion of duration weights to a simple averaging mechanism improved the prediction of retrospective emotion, as revealed by a better fit of the duration-weighted average compared to the average model—thereby underscoring a significant role for the duration of emotional experiences in sculpting retrospective ratings (as opposed to "duration neglect") in our sample (Bayes factor on ΔBIC of m1.2 *vs.* m1.1 favored model m1.2; ΔAIC = -403, ΔBIC = -403, ΔLL = 202, BF > 100).

The 'end' effect increases with sequence duration

We found that longer sequences produced a larger *end* effect, as evidenced by a significant interaction of the 'end' momentary rating and sequence duration ($\beta = 0.008$, SE = 0.002, $t = 3.41$, 95% CI [0.003, 0.01], $p < 0.001$). Neither negative peak nor positive peak effects reliably differed as a function of sequence duration (negative peak: $\beta = 0.003$, SE = 0.004, $t = 0.67$, 95% CI [-0.006, 0.01], $p = 0.50$; positive peak: $\beta = 0.007$, SE = 0.004, $t = 1.66$, 95% CI [-0.001, 0.01], $p = 0.10$). These findings highlight the importance of the final moments of complex emotional experiences in determining retrospective emotion when temporal processing capacity is challenged by long-lasting events.

Temporal duration errors predict retrospective emotion biases

Duration memory and retrospective emotion biases. Temporal duration errors, which were driven by a relative dilation of the remembered duration of negative *vs.* positive emotional movies, correlated with larger retrospective emotion biases ($\beta = 3.87$, SE = 1.05, $t = 3.67$, 95% CI [1.70, 6.05], $p < 0.001$).

Next, to uncover the putative sources of the association between retrospective emotion biases and temporal duration errors, we applied the individual-model-fit approach as detailed in *Associations between temporal memory and retrospective emotion biases: individual-model-fit* above. Consistent with the results originally obtained using binned emotion ratings, we found that greater temporal duration errors were associated with a larger duration neglect across participants (ΔBIC[Duration-weighted average model – Average model]; $r = 0.34$, $t(58) = 2.73$, 95% CI [0.09, 0.54], $p = 0.008$). Temporal duration errors were also associated with a *lower* negative peak effect (ΔBIC[Average model – Negative peak model]; $r = -0.42$, $t(58) = -3.55$, 95% CI [-0.61, -0.19], $p < 0.001$). Temporal duration errors did not correlate with end biases (ΔBIC[Average model – End model]; $r = 0.10$, $t(58) = 0.73$, 95% CI [-0.16, 0.34], $p = 0.47$).

Order memory and retrospective emotion biases. Temporal order memory was not associated with retrospective emotion biases (absolute order memory: $\beta = 0.03$, SE = 0.10, $t = 0.26$, 95% CI [-0.17, 0.22], $p > 0.99$; relative order memory: $\beta = -0.18$, SE = 1.23, $t = -0.23$, 95% CI [-1.73, 1.36], $p > 0.99$).

References (Supplementary Material)

Asutay, E., Genevsky, A., Barrett, L. F., Hamilton, J. P., Slovic, P., & Västfjäll, D. (2021). Affective calculus: The construction of affect through information integration over time. *Emotion*, *21*(1), 159–174. https://doi.org/10.1037/emo0000681

Ballard, I. C., Wagner, A. D., & McClure, S. M. (2019). Hippocampal pattern separation supports reinforcement learning. *Nature Communications*, *10*(1), 1073. https://doi.org/10.1038/s41467-019-08998-1

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Evans, D. E., & Rothbart, M. K. (2007). Developing a model for adult temperament. *Journal of Research in Personality*, *41*(4), 868–888. https://doi.org/10.1016/j.jrp.2006.11.002

Scheibehenne, B., & Coppin, G. (2020). How does the peak-end rule smell? Tracing hedonic experience with odours. *Cognition and Emotion*, *34*(4), 713–727. https://doi.org/10.1080/02699931.2019.1675599

Wang, J., Tambini, A., & Lapate, R. C. (2022). The tie that binds: Temporal coding and adaptive emotion. *Trends in Cognitive Sciences*, S1364661322002224. https://doi.org/10.1016/j.tics.2022.09.005

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wen, T., & Egner, T. (2022). Retrieval context determines whether event boundaries impair or enhance temporal order memory. *Cognition*, *225*, 105145. https://doi.org/10.1016/j.cognition.2022.105145