To cite this article:

Cloos, L., Vanpaemel, W., Ceulemans, E., & Kuppens, P. (in press). Remembering affect between moments: Assessing peak-end effects in continuous affect measures. *Cognition and Emotion*. https://doi.org/10.1080/02699931.2026.2676719

Remembering Affect Between Moments: Assessing Peak-End Effects In Continuous Affect Measures

Leonie Cloos, Wolf Vanpaemel, Eva Ceulemans, and Peter Kuppens

Faculty of Psychology and Educational Sciences

KU Leuven

Author Note

This article has been accepted for publication in Cognition & Emotion, published by Taylor & Francis. Please be aware that this is the author's version of the manuscript and there may be deviations in the final published version. Readers should consider this when citing or referencing the work. Thank you for your interest and understanding.

Leonie Cloos https://orcid.org/0000-0003-1550-7405

Correspondence concerning this article should be addressed to Leonie Cloos, Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, 3000 Leuven, E-Mail: leonie.cloos@kuleuven.be

Abstract

In experience sampling (ESM) studies, participants are measured repeatedly to track momentary affect changes. The data is collected in specific time intervals, leaving gaps between assessments. To capture affective changes between assessments, we developed a measure where participants retrospectively draw their affect in a continuous drawing. This approach, while more complete, may be influenced by retrospective biases, such as the peak-end effect, where intense or recent moments are remembered disproportionately. In an ESM study, 115 participants rated their momentary positive and negative affect six times daily, drew affect changes between assessments, reported events, and completed one daily measurement burst with up to six momentary ratings. We examined the divergence of the retrospective continuous affect drawings from the momentary affect and event ratings, respectively. For positive affect, intensity and recency were associated with lower divergence; the association was nonlinear. Divergence was greatest at low intensity and reduced at moderate to high levels. For negative affect, intensity was associated with higher divergence and recency with lower divergence. The effect of intensity diminished over time, indicating that divergence is mostly determined by recency. Findings were similar for events, except that for negative events, there was no recency effect. Future research should consider the effect of time and intensity, especially the differential effect of positive and negative intensity, on short retrospective ESM measures.

Keywords: Experience Sampling, Peak-End Rule, Continuous Affect, Retrospective Self-Report, Affect Dynamics

Remembering Affect Between Moments: Assessing Peak-End Effects In Continuous Affect Measures

Affective experiences encompass a range of feelings that vary in intensity and duration over time. These experiences vary over different timeframes, some are transient, whereas others are more enduring and more prolonged. To capture the dynamics of these changes in daily life, researchers typically make use of the Experience Sampling Method (ESM; Csikszentmihalyi & Larson, 1987), where individuals are assessed repeatedly at multiple time points throughout the day in their natural environments. The strength of ESM lies in its ecological validity and momentary nature, providing real-time assessments and minimizing retrospective biases (Shiffman et al., 2008). With this approach, the data that is collected yields snapshots from a movie, individual measurements capturing specific moments (Hollenstein, 2021). While these snapshots provide valuable insights, there is concern that focusing on these snapshot moments may omit affective states and events that occur in between (Cloos et al., 2024; Silvia et al., 2013).

Continuous Affect Drawings

Although a sequence of momentary measurements provides valuable insight into affect dynamics, there is typically little clarity on what information the snapshots contain and what happens before and after. For example, it is unclear whether the snapshot contains the rise, peak, or recovery of an affective episode. The affective ups and downs in between measurements are more like a movie unfolding continuously over time (Hollenstein, 2021). Gaining insight into these continuous feelings provides a wealth of information on how people experience and regulate their emotions. However, achieving this using traditional ESM methods would require drastically increasing the number of momentary measurements, which can become extremely burdensome for participants. To obtain information in between measurements, we have developed a continuous affect measure that prompts participants to report changes in their affect intensity between two momentary assessments using a drawing (Cloos et al., 2024). In contrast to a retrospective

item that asks participants to summarize their affect since the last measurement in a single score, the current approach captures a continuous drawing between two assessments. Specifically, participants are shown their affect intensity of the current and previous moment and *draw* the changes in intensity from the previous to the current moment. The continuous drawings thereby function as an addition to the momentary ratings, capturing information in between snapshots. Although this information is collected retrospectively rather than continuously in real time, we use the term "continuous" to reflect that participants provide an uninterrupted trajectory of affect, reporting changes across the period without breaks.

The drawings are a promising method for researchers who are interested in capturing episodic processes, such as discrete emotional episodes or the recovery after emotional events, on a fine-grained temporal scale. It is especially suited for studies of affect, where emotions may shift rapidly and unpredictably. By enabling retrospective evaluation of recent affective experiences, the method offers a way to capture these chances without increasing sampling frequency. Moreover, it presents a solution for momentary missingness (Rintala et al., 2020): when a momentary rating is skipped, the drawing can still provide information about the affect intensity experienced during that period. Overall, this method is particularly suited for research focused on affect changes that unfold over short timescales and are thus often missed by traditional interval sampling (Revol et al., 2025).

A previous study found that such a drawing provides non-negligible additional information between the discrete moments. Notably, this research revealed that on the basis of this continuous affect drawing, on average, one positive and one negative affective episode were identified per day that would otherwise be missed by traditional ESM (Cloos et al., 2024). This underscores the potential of using a drawing to capture the continuous fluctuations in affective experiences. However, it is crucial to examine the validity of this method and potential biases it may introduce when individuals are asked to

retrospectively draw their affective changes. Since these data are collected retrospectively, it is crucial for the validity of the data to examine memory effects that may lead to deviations from the actual experience in the moment.

Memory Effects

A critique of retrospective evaluations is that they are influenced by cognitive heuristics and memory biases (Schwarz, 2007). Especially, ratings of intensity are argued to be poorly represented in memory. The evaluation of emotional episodes, for example, is based on only a few memorable moments, namely the most intense and the most recent moment of the episode (Fredrickson, 2000). Similarly, affective events are salient and memorable moments that, like the most intense moments in an emotional episode, hold greater significance for retrospective evaluations of affect (Rubin & Umanath, 2015; Talarico et al., 2004). As a result, affective events are likely to carry more weight in retrospective ratings of overall affect. Additionally, asking participants to recall specific events facilitates memory retrieval, where the intensity and timing of the event may play significant roles in shaping retrospective evaluations (Robinson & Clore, 2002b).

This phenomenon, referred to as the peak-end-rule, has been studied in the well-being literature to examine retrospective evaluations of affect and summaries of emotions (Kahneman, 2000). This effect explains, for example, that people undergoing a painful procedure prefer a longer procedure with a milder end over a short procedure with a painful end (Kahneman et al., 1993; Redelmeier & Kahneman, 1996). Studies where people provide online affect ratings while watching short clips, as well as a retrospective global evaluation of the same clip, demonstrate that the duration of a clip does not influence the global evaluation. Instead, the pleasantness or unpleasantness of the most intense moment and the final moment are the most influential in shaping the retrospective evaluation (Fredrickson & Kahneman, 1993). This underlines that retrospective ratings of affective experiences are more complex than the sum of the individual moments alone.

The studies above describe experimental designs in which events were simulated in

an experimental setting. While this seems to have limited generalizability to daily life assessments, similar effects are found in research involving intensive longitudinal data on affect ratings across different timeframes. For example, investigations of end-of-day ratings compared to momentary samples throughout the same day demonstrate small peak and end biases in retrospective negative affect ratings (Neubauer et al., 2020). The influence of salient experiences, such as intense and recent moments, on retrospective ratings is not yet fully understood. A recent study investigated the difference between retrospective summaries and the mean of all momentary ratings over two weeks. The study found significant variation among participants in how close the retrospective rating matched with the average of the experiences. For many participants, the mean provided a good summary score, while for others the extreme moments carried more weight than other moments (Leertouwer et al., 2022).

Divergence of Continuous Drawings from Momentary Experience

The peak-end-rule is specifically applicable to short episodes that are retrievable from episodic memory (Robinson & Clore, 2002a). It is therefore interesting to examine this effect in the continuous drawing, in which the most recent two-hour period of someone's daily life is evaluated retrospectively.

Some studies that examine peak-end effects experimentally manipulate the intensity and duration of an event (e.g., pain) to examine how the peak, end, or the duration of the event determines the overall affect intensity rating in the end. Other studies obtain real-time ratings of affect intensity and ask participants to provide a single retrospective summary rating of the entire experience, to examine how well the peak and end ratings predict the summary evaluation (Ariely & Carmon, 2000; Kemp et al., 2008). The continuous drawing is different from such methods, as there is no summary evaluation of the overall experience. Instead, it obtains a retrospective reconstruction of a period, tracing the changes over approximately two hours.

In the current study, we are interested in how these retrospectively provided

continuous drawings align with the momentary affect experience and whether fluctuations in this alignment follow the peak-end rule. This will give insight into how memory effects influence the evaluation of affect in the continuous drawing, and how peak and recent moments influence the retrospective data.

We investigate this using two parallel sources of information: a) momentary affect intensity scores measured by a high-intensity sampling design (i.e., measurement burst) scheduled in between two fixed assessments once a day (yielding scores alongside a drawing), and b) participant-reported affective events that occurred in between assessments. To examine divergence, these two parallel affective ratings are compared to the intensity of the continuous drawing at that moment or event. This divergence is then related to the intensity and timing of the momentary affect or the event rating. The peak effect is investigated by relating the affect intensity of a momentary or event rating to the divergence of the continuous drawing from the moment or event rating. The end effect is investigated by the time between the momentary rating or event and the continuous drawing to the divergence of the continuous drawing from the moment or event rating.

The Current Study

The goal of this study was to investigate how accurately individuals can recall and depict their experiences retrospectively with a continuous drawing. We aim to determine the validity of the continuous drawing by comparing it to momentary ratings and event ratings that occurred alongside the period of the drawing.

The first research question examines the divergence between the momentary intensity and the intensity of the continuous drawing at the time point of the momentary assessment. To this end, we once daily introduced a high-density sampling period or *burst* in our data collection process. This high-density sampling protocol was scheduled once a day, between two regular momentary assessments. During this period, momentary ratings were taken every twenty minutes until the next regular measurement occasion. After the high-density sampling interval, participants rated their affect over the entire period on the

continuous drawing, thereby also reporting on the same moments recorded in the momentary ratings, but retrospectively. This allows us to examine the divergence between the affect intensity rated in the moment and the corresponding intensity on the continuous drawing at that same point in time. Inspired by the peak-end rule, we relate the divergence between the momentary affect intensity and the corresponding intensity on the continuous drawing to the intensity and timing of the momentary ratings. The divergence between momentary and retrospective affect rating provides a direct comparison of affect on two different timeframes. It can thus be interpreted as measure of retrospective accuracy. We expect that divergence is smaller for more intense levels of momentary affect and for momentary assessments that were timed closer to the retrospective assessment. Additionally, we explore how the interaction between intensity and timing relates to the divergence of the continuous drawing.

The second research question investigates the divergence between affective event intensity and the rating of the continuous drawing at the time point of the event. Affective events are important predictors of affect ratings, since the appraisal of internal and external events determines changes in affect (Kuppens et al., 2012; Scherer, 2005). Moreover, affective events are central to investigate in the context of the retrospective continuous affect drawing, as salient (i.e., intense and recent) events are remembered better and have been shown to disproportionately predict retrospective affect ratings (Fredrickson & Kahneman, 1993). To examine the extent to which retrospective affect ratings are contingent on events, and whether their intensity and recency predict the retrospective drawings, we asked participants to report any positive or negative affective events that occurred between two regular measurement occasions. Participants retrospectively rated the intensity of these events and reported when they had occurred. The timing of the event could therefore be matched with the corresponding time point on the continuous drawing. This enabled us to examine the divergence between the affect intensity rated for the event and the corresponding intensity on the continuous drawing at

that same point in time. To investigate the peak-end rule in the context of affective events, we relate divergence between the event rating and the corresponding intensity on the continuous drawing to the intensity and timing of the events. The divergence reflects how strongly the event is weighed in the retrospective drawings. We expect that more intense and more recent event ratings will be closer to the retrospective affect ratings than less intense events that occurred further in the past. Additionally, we explore how the interaction between intensity and timing of the event relates to the divergence on the continuous drawing.

Overall, these research questions explore the relation between the *intensity* and the *time* (both separately and in interaction) of the momentary ratings or reported events and the divergence of the continuous drawing at the corresponding time point. Based on research reported above, more intense and more recent ratings should increase the salience of the momentary rating or the retrospective event, which, in turn, should increase their congruence with the retrospective affect rating. As a consequence, the continuous drawing should diverge more from the rating of the corresponding momentary assessment or event when the latter were less intense or occurred longer ago. The main goal was to investigate the magnitude of the effect of intensity and time on divergence. Therefore, the direction of potential biases (i.e., overestimation or underestimation) are not taken into account for the main analyses. However, we conduct exploratory analyses to determine whether the divergences are expressed as underestimation or overestimation. Due to the novelty of the method and the lack of similar studies investigating the validity of continuous drawing, all of the analyses remain largely exploratory.

Method

The research questions in this study are investigated with previously collected data that were also reported in (Cloos et al., 2024). The study was approved by the Social and Societal Ethics Committee – KU Leuven (G-2022-4779-R3). The data is publicly available in the Research Data Repository of the KU Leuven at https://doi.org/10.48804/Z2I43Q

(Cloos, Mestdagh, et al., 2025).

Pre-screening

Participants were recruited from a general population sample, without specific requirements, except English language skills and owning a personal smartphone. The study was advertised on social media channels, public flyers, and mouth-to-mouth communication. Interested participants signed up on the online survey-platform Qualtrics for a pre-screening. This consisted of a brief demographic survey (gender, age, language, nationality, current residence), and the Perceived Stress Scale (PSS; Cohen et al., 1983). We recruited and selected participants in two waves. In each sampling wave we computed the quantiles of the recruited sample to group participants into one of four groups and grouped participants into one of four groups (low = 0% to 25%, mild = 26% to 50%, moderate = 51% to 75%, high = 76% to 100%). For each wave, we randomly selected an equal number of participants from each group and contacted them to participate in our study. We sampled participants based on their PSS scores to achieve a representation of individuals with different levels of stress and negative affect in daily life (Mroczek & Almeida, 2004). The two recruitment samples consisted of 869 individuals in total (69.2% female, 1.4% other; $M_{age} = 23.4$, $SD_{age} = 7.5$).

Participants

We conducted a simulation-based power analysis to determine the number of participants needed to estimate the correlation between the high-intensity sampled momentary ratings and the drawing. The results showed that a minimum sample size of 100 participants was needed to estimate an average within-person correlation of 0.1 ($SD = 0.2$) with 96% power ($\alpha = 0.05$).

Our final sample included 115 individuals (58.26% female, one other; $M_{age} = 24.24$, $SD_{age} = 6.68$, *range* = 17–53). All participants were residing in Belgium, but were from 25 different nationalities. The majority of participants were Belgian (53%), the second most common nationalities were Italy and Turkey (each 5.2%), followed by The Netherlands,

India, Poland, and Spain (each 3.5%). The four categories of PSS scores were about equally represented. Twenty-nine participants scored in the low rank ($5 \leq 13.5$), 33 in the mild rank ($13.5 \leq 19$), 25 in the moderate rank ($19 \leq 23$) and 28 in the high rank ($23 \leq 34$). Although the PSS has no official cutoffs, our sample's mean ($M = 18.71$, $SD = 6.68$) is higher than the U.S. means reported in 2006 ($M = 15.31$) and 2009 ($M = 15.84$) (Cohen & Janicki-Deverts, 2012, see). Additionally, a recent Danish adult sample (median age 47) from 2024 had a PSS median score of 15 (interquartile range: 9; 19), which was lower than that of our sample (Pedersen et al., 2024). This comparison suggests that our sample of young adults experienced higher perceived stress levels than recent samples in the literature.

Procedure

Participants who were drafted from the pre-screening were invited to participate in the study. They were asked to participate in a thirty-minute introduction session that took place online or at the university building. During the session, participants were instructed on the use of m-Path, the mobile platform used to collect the data, informed about the procedure, and their consent. The leading experimenter demonstrated the continuous drawing and implementation of the high-intensity sampling scheme and answered participants' questions. The introductory session and the procedure were piloted on six individuals prior to the study to refine the protocol and participant instructions. After the introduction, they received an email with a personal code that anonymized their identity and was used to connect them to the m-Path account of this study. Once connected to m-Path, participants filled in the informed consent and baseline questionnaires. The experience sampling period of 7 days began the following day. At the end of the 7 days, participants were debriefed. Before the start of the study, they were informed that their compensation for participating was calculated based on how many notifications they answered. Participants who answered at least 75% of the notifications received the full reward of 40€. With every 5% decrease in responding, the reward

decreased by 1.50€. For every day in which they missed the first notification, 4.50€ was deducted from the reward. The average compliance to the continuous drawings was 70%, and the average compliance to the burst assessments was 70.31% (SD = 20.52; *range* = 9.52%–100%).

Time-Contingent Sampling

Each day, participants received 6 notifications that were scheduled using a time-contingent sampling scheme, with notifications scheduled every 140 minutes, to assure drawings are roughly the same duration. The first notification of the day only comprised two items for positive and negative *momentary affect*. The following 5 notifications presented the two *momentary affect* items, the *continuous affect drawing* item, and the two *event items*. The questionnaire also included three items to measure rumination, but these are not relevant for the current study. The notifications did not expire until a new notification occurred, thereby the intervals varied in length depending on the reaction times. The average time interval was 146.21 minutes (SD = 37.42). The first notification needed to be completed before noon, as it established the starting point for the continuous affect drawing at the next notification. Failure to respond to the first notification on time resulted in not receiving subsequent notifications for the remainder of the day.

High-Intensity Sampling (Measurement Burst)

During one randomly selected interval each day, participants received a so-called measurement *burst,* in which they rated their *momentary affect* every 20-minutes. This means that in between two fixed notifications, there were additional notifications that started 20 minutes after the fixed notification was completed. Then every 20 minutes, another notification occurred, for a total of six notifications, or until the next fixed notification stopped the sampling burst.

Measures

On the day they signed up and agreed to the informed consent, participants filled in several baseline questionnaires that are not relevant for the current study (see Supplement A).

Momentary Affect

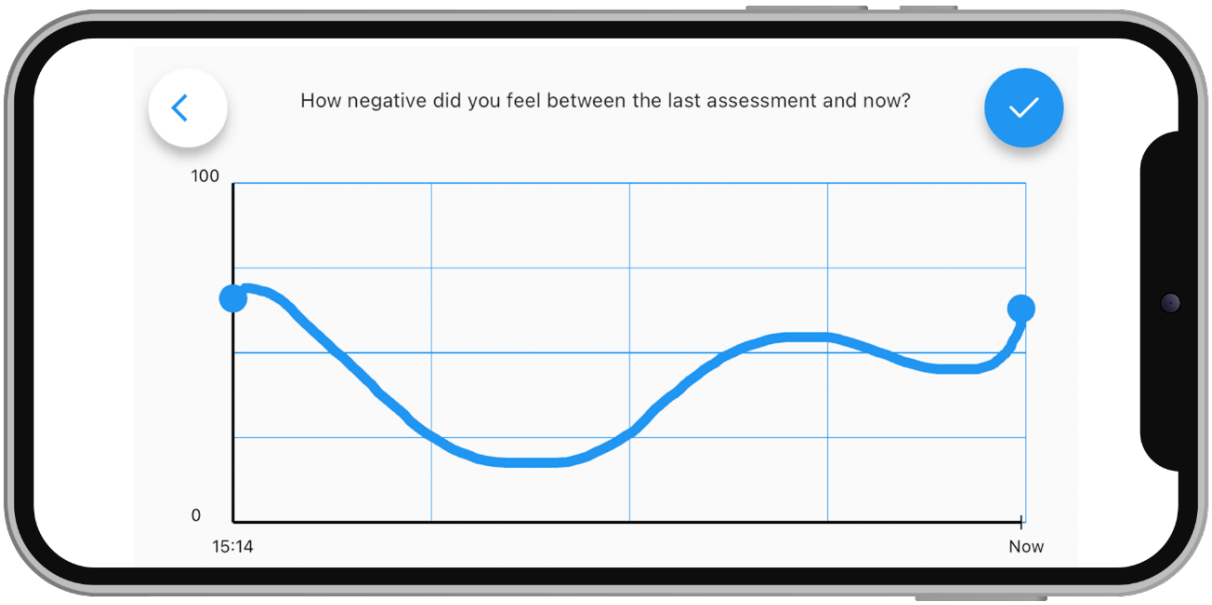
The *momentary affect* items were the first items to be asked at every notification. Participants were asked to rate their *momentary* positive and negative affect with the question: "To what extent do you experience positive [negative] emotions right now?" on a slider scale, ranging from 0 (*I don't experience positive [negative] emotions right now*) to 100 (*I experience very strong positive [negative] emotions*). The item was based on the validation study by Cloos et al. (2023).

Continuous Affect Drawing

From the second notification onward, the participants were asked to provide a positive and a negative continuous affect drawing, after rating their momentary positive and negative affect. The continuous drawings are presented as a two-dimensional figure with time of day on the x-axis and affect intensity on the y-axis, ranging from 0 to 100 (Figure 1). Above the figure, participants read the question "How positive [negative] did you feel between the last assessment and now?". On the x-axis, a timeline was displayed (in hours and minutes of the day) starting at the time of the previous notification and ending at the current time point. The momentary affect score from the last answered notification was used as the starting point of the drawing, and the current momentary affect score as the end point. Participants were asked to draw how their affect changed between these two points over the time span since the last point. Responses were only considered complete when the participants connected both points. More detailed information on the item can be found on the m-path website.

Figure 1

The Continuous Drawing as Currently Implemented in m-Path



Events

Also, from every second notification onward, participants reported the occurrence of positive and negative events. They were asked to answer "Did a positive[negative] event occur since the last assessment?" with *No* (0) or *Yes* (1). If participants responded with *Yes*, a follow-up item asked them to rate the intensity of the event on a scale ranging from 0 (*Neutral*) to 100 (*Very positive [negative]*). Additionally, participants were asked to indicate the event's starting time. They were given response options to determine an approximate time point: *just now* (1), *about 15 minutes ago* (2), *about 30 minutes ago* (3), *about 45 minutes ago* (4), *about 60 minutes ago* (5), *about 90 minutes ago* (6), or *about 120 minutes ago* (7).

Data Analysis

The data was pre-processed so that the drawings yielded minute-to-minute affect intensity scores, and so that these time points were temporally aligned with the data from the measurement bursts and reported events. A detailed description of the data

pre-processing can be found in Supplement B. The preprocessing code can be found on the OSF page.

Data Processing

Divergence Measure

We examined the peak and end effects in continuous affect drawings by assessing their divergence with two variables: a) momentary affect intensity and b) event intensity. Divergence was defined as the absolute difference in intensity between the momentary affect or event rating and the corresponding affect rating on the continuous drawing at the same time point. Because our focus is on the *intensity* of the divergence rather than the direction (i.e., whether the retrospective rating was higher or lower), we computed the absolute difference between the two ratings. This ensured that both types of deviations are treated equally, so that we can understand the overall magnitude of the effect without the need to interpret the direction of the bias. Additionally, we conducted descriptive analyses of the general pattern in which the divergence manifested, to provide context for the results of the main analysis.

The continuous drawing contained minute-to-minute data of affect intensity over the period in between two assessments. In order to compute divergence, we first determined the time point on the continuous drawing that corresponded to the time point of the momentary rating or event occurrence. This time point represented the location on the x-axis of the drawing, with which we then determined the corresponding value on the y-axis of the continuous drawing, representing the intensity rating of the drawing. This allowed us to pinpoint the intensity value of the drawing at the time of the momentary rating or the time of the event. Finally, we subtracted the drawing's intensity at the time of the momentary rating or event from the respective affect intensity or event intensity score and then took the absolute value as a measure of divergence (this is visualized as a gray vertical line in Figure 2 and Figure 3).

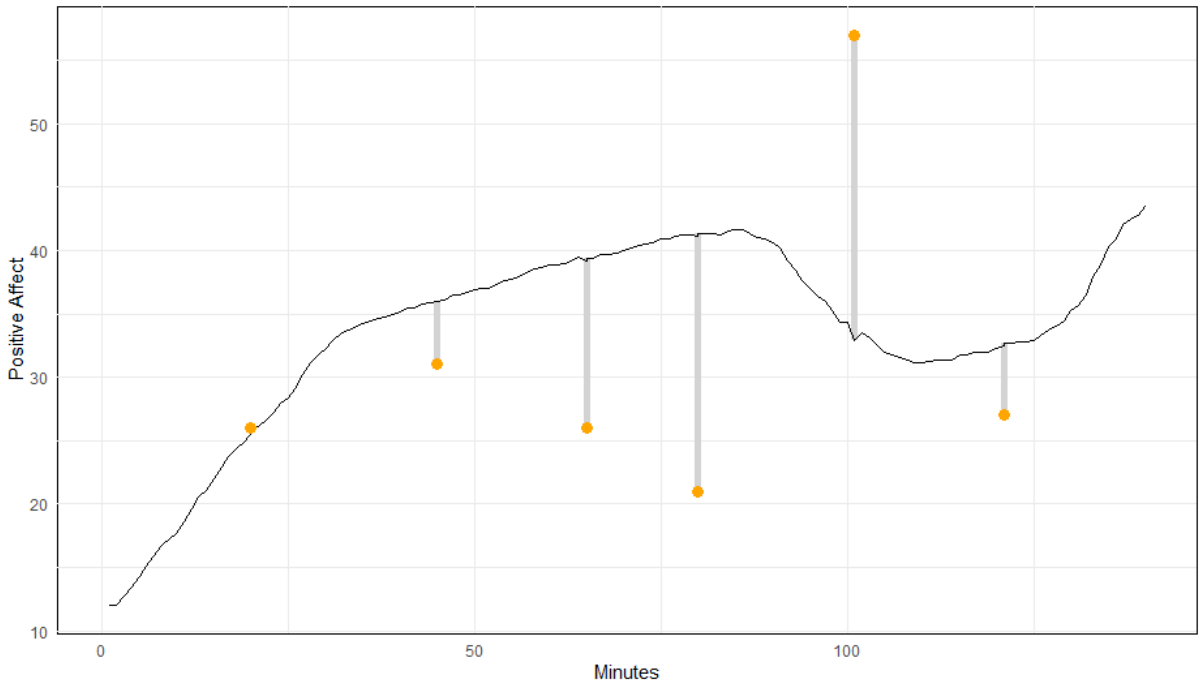
We defined the divergence score for the ith person and the jth time point or j^{th} event

as:

$$Divergence_{ij} = \left| Intensity_{ij} - Drawing\ Intensity_{ij} \right| \quad (1)$$

Figure 2

Example of how Divergence is Quantified for Momentary Intensity and Drawing Values for one Interval Between two Consecutive Assessments



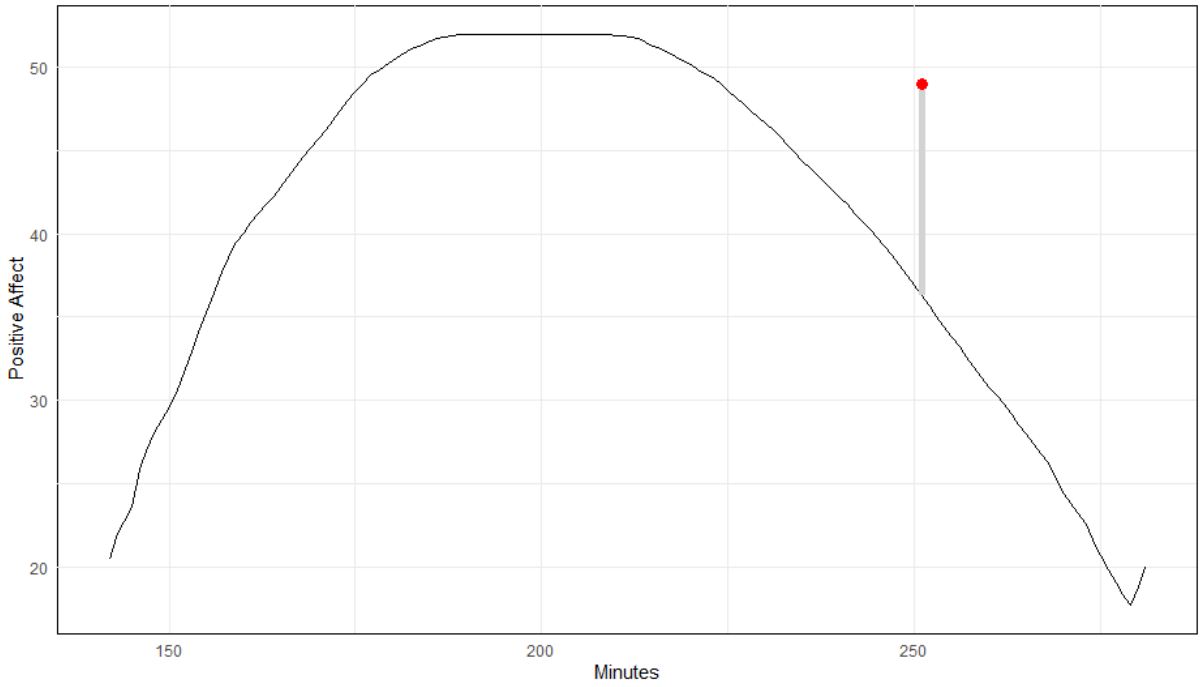
For momentary affect, there was a maximum of six timepoints per drawing, indicated by the orange points on Figure 2. For the events, there was a maximum of one point per drawing, indicated by the red points on Figure 3.

Intensity Ratings

Since the affect and event intensity ratings were already used to compute divergence (see Eq. (1)), it was necessary to transform these variables in order to use affect and event intensity as a predictor in the multilevel regression models. This transformation prevented any overlap between the criterion and predictor variables. We planned two analytic strategies: a person-mean-centered approach and a quartile-based approach. In the person-mean-centered approach, each person's average score on the respective variable

Figure 3

Example of how Divergence is Quantified for Event Intensity and Drawing Values for one Interval in Which an Event was Reported



was subtracted from their time-point-specific observations. In the quartile-based approach, we determined the 25th, 50th, 75th, and 100th percentiles of each variable across observations and categorized the scores into these four categories. The range of scores categorized for the transformed variables is presented in Table 1.

Table 1

Transformation of Original Variable Intensity Ratings into Quartiles

	Momentary Intensity		Event Intensity	
	Positive	Negative	Positive	Negative
1st Quartile	0 - 33	0 - 8	0 - 32	0 - 24
2nd Quartile	33 - 58	8 - 18	32 - 54	24 - 42
3rd Quartile	58 - 74	18 - 34	54 - 71	42 - 63.75
4th Quartile	74 - 100	34 - 100	71 - 100	63.75 - 100

Note: Positive is subsequently transformed (see the following). The 1st and 2nd quartile pertain to the negative side of intensity and the 3rd and 4th quartile pertain to the positive side of intensity.

Assumptions

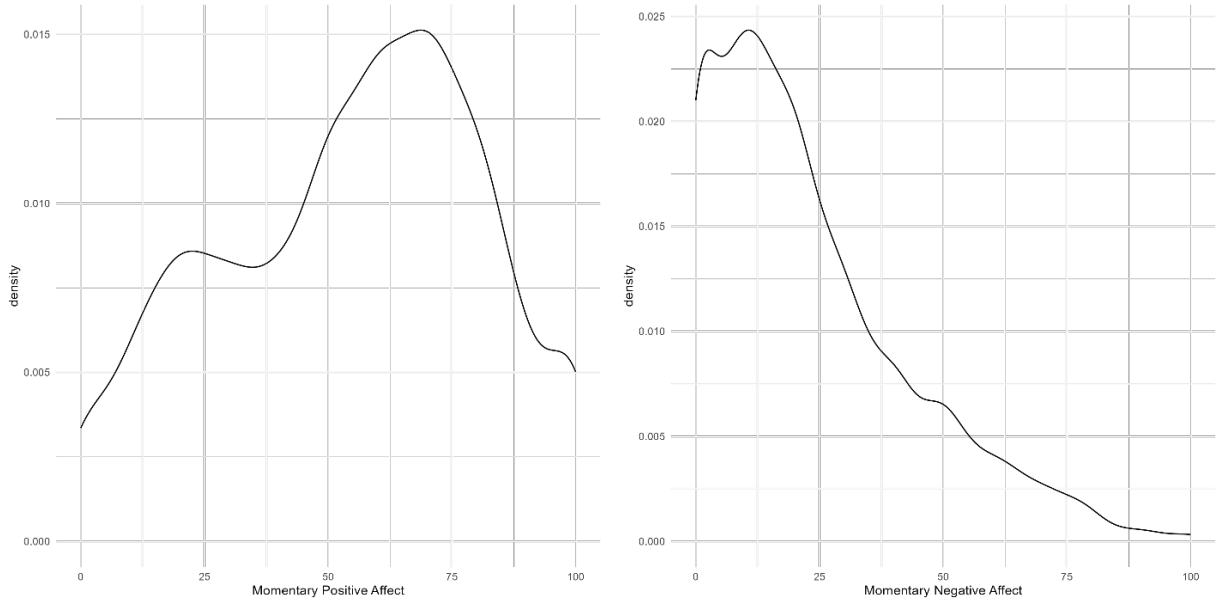
We planned two analyses: a) linear mixed-effects models using the centered predictors and b) non-linear models using the quartile scores. Before model estimation, we evaluated the plausibility of linearity by plotting the relation between the predictor and the outcome as well as the residual distribution of the linear mixed-effects model.

First, we plotted the distributions of the variables across observations. The distribution of momentary positive affect exhibited a bimodal shape (see Figure 4, left), with a clear split at the midpoint, one dense mode above and another below the midpoint.[1] The distribution of momentary negative affect was right-skewed (see Figure 4, right), indicating a low frequency of intense negative affect. For both positive and negative event intensity, the distributions showed slight bimodality as well (Supplementary Figure D17 and Figure D22).

[1]This pattern is often observed in ESM slider scales, as the midpoint may be interpreted as neutral (Cloos, Siepe, et al., 2025; Haslbeck et al., 2023).

Figure 4

Density Distributions of Momentary Affect Items



Note: The positive affect distribution (left) shows two modes. The negative affect distribution (right) is skewed to the right.

To test the linearity assumption, we plotted the relation between the untransformed predictors and the divergence scores across observations and within individuals. A scatter plot of momentary positive affect observations revealed a clear non-linear relationship with divergence (Supplementary Figure D1). Furthermore, most individual participants' scatter plots showed no clear linear pattern (Supplementary Figure D2 - Figure D3). For momentary negative affect, the overall scatterplot showed a non-linear relationship with divergence (Supplementary Figure D5), while the individual-level scatterplots showed more consistent linear trends (Supplementary Figure D6 - Figure D7). The scatterplot of the positive event intensity observations indicated a linear relationship with divergence (Supplementary Figure D18). However, there were several non-linear trends in the individual scatterplots (Supplementary Figure D19 - Figure D20). For negative event intensity, both the overall and individual-level scatterplots indicated linear relationships with divergence (Supplementary Figure D23 - Figure D25).

Finally, we inspected the residuals from the linear-mixed effects model using the centered scores as predictors. The residual plot of momentary positive affect showed an uneven spread, with greater variability at the extreme ends of the predictor (Supplementary Figure D4). This pattern suggests that a single linear effect may not adequately capture the relationship across the full scale. For negative affect, the residual plot showed more clustering around zero, with increased variability at higher values, indicating potential heteroscedasticity (Supplementary Figure D8). The residuals of the linear models using the predictors positive and negative event intensity were normally distributed (Supplementary Figure D21 and Figure D26).

Statistical Analysis

The assumption checks indicated that the relationships between predictors and divergence were not strictly linear. While person-centering and the inclusion of random effects accounted for individual differences, centering did not fully address the observed non-linear patterns. Given the distributional characteristics of the predictors and indications of non-linearity, we prioritized ordered factor models in the main analyses. This approach is robust to non-normal distributions and enables the detection of non-linear patterns, including linear, quadratic, and cubic trends. Ordered factor models offer a flexible and interpretable framework for all predictor variables. Linear models with centered continuous predictors are reported in the Supplement E for completeness and should be interpreted with caution.

Ordered Factor Models

We conducted a series of ordered factor models to examine divergence between retrospective continuous affect drawings and two sources of affective information: momentary affect ratings and reported affective events. To account for the hierarchical structure of time points nested within individuals, we employed multilevel models,

including fixed and random intercepts and slopes, where supported by model convergence.[2] To account for the serial dependence of the data over time, we included an autoregressive (*AR1*) within-person error process.

Affect and event intensity predictors were operationalized as categorical variables using quartile scores. The quartile scores were treated as an ordered factor with four levels ($1^{st} < 2^{nd} < 3^{rd} < 4^{th}$ quartile). Levels were treated as equally spaced, though it was not assumed that the steps between adjacent quartiles were equal (i.e., that moving from quartile 1 to 2 is the same as moving from quartile 2 to 3, and so forth). We applied polynomial contrasts to examine the linear, quadratic, and cubic trends across these levels and assess the changes between successive quartiles. The polynomial contrasts provided a simple interpretation of the relationship between intensity and divergence with a parsimonious set of coefficients for the linear, quadratic, and cubic trends. The linear contrast (β_1) investigated whether divergence changed monotonically with increasing levels of affect or event intensity, the quadratic contrast (β_2) tested for an (inverted) U-shaped relationship between affect or event intensity and divergence, and the cubic contrast (β_3) captured complex non-monotonic relationships.

The predictor time since the affect rating or event was modelled as a continuous predictor. We examined the main effect of time (β_4) and the fixed effects of the interaction between the linear, quadratic, or cubic trend and time (β_5, β_6, β_7). The interaction coefficients determined whether the relationship between and divergence changed as a function of time.

$$\begin{aligned} Divergence_{ij} = \beta_0 &+ \beta_1 \cdot Linear_{ij} + \beta_2 \cdot Quadratic_{ij} \\ &+ \beta_3 \cdot Cubic_{ij} + \beta_4 \cdot Time_{ij} + \beta_5 \ (Linear_{ij} \times Time_{ij}) \\ &+ \beta_6 \left(Quadratic_{ij} \times Time_{ij}\right) + \beta_7 \left(Cubic_{ij} \times Time_{ij}\right) \\ &+ u_i + \epsilon_{ij} \end{aligned} \tag{2}$$

[2]In models including interaction effects, convergence issues prevented the estimation of random interaction slopes (see Supplementary Table F1 for an overview of the random effects in each model).

where u_i represents person-specific random effects and ϵ_{ij} the person and occasion-specific error term.

In all models, we were primarily interested in the fixed effects to understand how intensity of and time since the momentary rating or affective event, separately and in interaction, relate to divergence (on average). Models were fitted iteratively, starting with main effects and adding two-way interactions. For brevity, we present only the final (most complex) model in the main text, equation, and results. The stepwise results are reported in overview tables. The equation displays only the person-specific random-effect (u_i) for simplicity. Random slopes were estimated where supported by model convergence. Details on the estimation of random slopes and (co)variances of the random effects are provided in Supplement F.

Momentary Affect Ratings

In the first set of models, we examined the divergence between *momentary affect ratings* and the continuous drawing's rating at the time of the momentary rating. The outcome variable in these models was the divergence, defined as the absolute difference between the momentary affect rating and the corresponding affect rating on the continuous drawing at the same time point (Figure 2). Separate models were estimated for *positive momentary affect* (Model 1) and *negative momentary affect* (Model 2). Divergence in these models captured the disagreement between a momentary and retrospective affect rating. The continuous predictor *time* was operationalized as the time elapsed between the momentary affect rating and the rating of the continuous drawing.

Reported Affective Event Ratings

In the second set of models, we investigated the divergence between the event intensity ratings and the continuous drawing at the time of the event. The outcome variable was the divergence, or absolute difference between event intensity and the corresponding value on the continuous affect drawing at the time of the event (Figure 3). Separate models were estimated for *positive event intensity* (Model 3) and *negative event*

intensity (Model 4). The outcome variable divergence captured to what extent the retrospective affect drawings are informed by salient events. The predictor time was operationalized as participants' self-reported estimate of how long before the continuous drawing the event had occurred.

Supplementary Analyses

We initially estimated models using centered predictors. However, since the assumption checks indicated non-normal distributions and non-linear relationships between intensity and divergence, this hampered the interpretation of linear effects. Given these diagnostics, we prioritized the ordered factor models in the main analyses, as they provide a more robust and interpretable framework under these conditions. The centered-predictor models are reported in the Supplementary Materials for completeness and transparency (see Supplement E).

Results

Divergence with Momentary Intensity Ratings

Descriptive Statistics

The average momentary positive affect intensity score was 54.24 (SD = 26.24). The average absolute divergence of positive affect intensity between momentary ratings and the retrospective drawings was 12.37 (SD = 12.52). The average negative affect intensity score across all observations was 22.98 (SD = 20.34). The absolute divergence of negative affect intensity of the momentary rating and the continuous drawing was, on average, 11.88 (SD = 13.19).

Model 1. Positive Affect Intensity (Quartile)

Divergence was regressed on momentary positive affect intensity quartile scores and the time since the momentary rating. We estimated a main effects model with polynomial contrasts and time as predictors, and a model including interaction effects between predictors. Table 2 reports all estimates, while the following discusses results from the

interaction model, the random effects are reported in Supplement F.

Table 2

Nested Models of Positive Affect Intensity Quartiles and Time

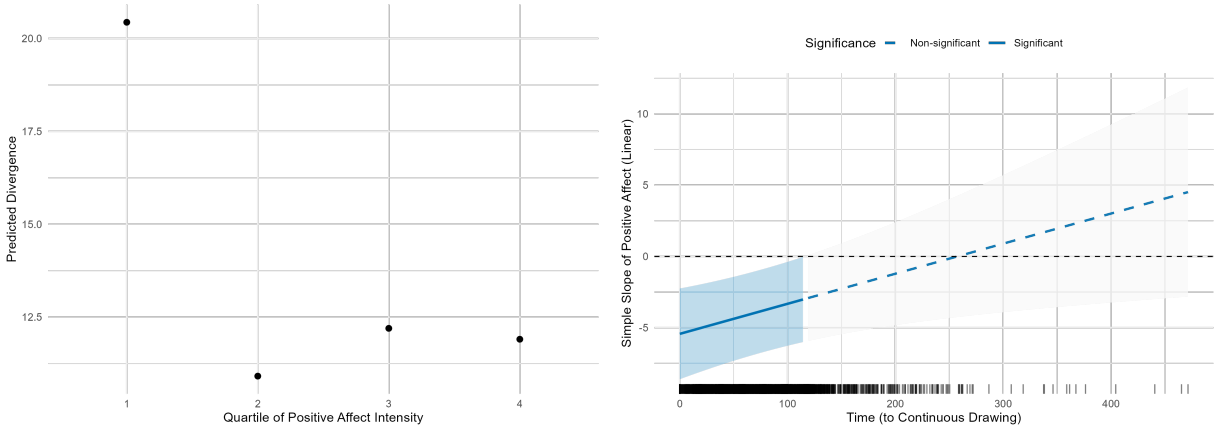
	Main Effects		Two-Way Interaction	
Effect	Estimate	SE	Estimate	SE
Intercept	13.886***	0.619	13.862***	0.625
Linear	-3.789*	1.477	-5.436**	1.624
Quadratic	3.886***	1.007	4.615***	1.191
Cubic	-2.229 ***	0.547	-2.770**	0.803
Time	0.015**	0.005	0.016**	0.005
Linear:Time			0.021*	0.009
Quadratic:Time			-0.009	0.008
Cubic:Time			0.007	0.008
R^2	43.59%		44.28%	

Note: p < .05 *; p < .01 **; p < .001 ***; SE = Standard Error

All polynomial contrasts for positive affect intensity were statistically significant, indicating a non-linear relationship with divergence. Specifically, we observed a significant negative linear effect of positive affect intensity (β_1 = -5.44, p = .001), indicating a decreasing trend in divergence for higher positive affect intensity scores. As shown in Figure 5 (left), when time is extrapolated to zero, the largest divergence occurred in the 1st quartile, representing the lowest positive affect intensity quartile. The quadratic trend was also significant (β_2 = 4.62, p < .001), revealing a U-shaped relationship between divergence and intensity, where divergence initially decreased from the 1st to the 2nd quartile, then increased from the 2nd to the 3rd quartile. The cubic trend was significant and negative (β_3 = -2.77, p = .001), indicating that divergence levelled off and slightly decreased again at the highest positive affect intensity quartiles, with lower divergence in the 4th compared to the 3rd quartile.

Figure 5

Predicted Divergence and Simple Slopes of Linear Effect Positive Affect Intensity (Quartiles) by Time



Note: Left: Main effect when time is extrapolated to 0; Right: Johnson-Neyman plot illustrating how the linear effect of *positive affect intensity quartiles* on *divergence* varies as a function of *time*, with 95% confidence bands. The shaded region represents the time interval during which the effect is statistically significant ($p < .05$). The observed range of time was [0, 471.42] minutes. The distribution of the observations in the variable time is given on the x-axis.

Furthermore, the time that passed between the momentary rating and the drawing was a significant predictor of divergence ($\beta_4 = 0.02$; $p = .001$). This indicates that at the midpoint[3] of the positive affect intensity scale (i.e., between quartiles 2 and 3), divergence increased as more time passed between the momentary and drawing ratings. We also found a significant interaction between time and the linear trend of positive affect intensity ($\beta_5 = 0.02$, $p = .015$), suggesting that the effect of affect intensity on divergence is moderated by time.

To explore the interaction between positive affect intensity and time, we computed the simple slope of the linear contrast of positive affect across the observed range of time, and estimated the 95% Confidence Interval (CI) to identify at which values of the

[3]The main effect of *time* in this interaction model reflects its effect on divergence when positive affect intensity is at its midpoint (i.e., when the linear, quadratic, and cubic contrast terms equal zero).

moderator time the linear effect was significant ($p < .05$). We visualized this in a Johnson-Neyman plot (Figure 5, right), which revealed that the strongest negative linear effect of positive affect intensity on divergence occurred immediately after the momentary rating. At this point, low intensity was strongly associated with high divergence, while high intensity was associated with low divergence. As time increased, the linear effect of positive affect intensity weakened, becoming non-significant after 114 minutes passed between the momentary rating and the drawing. This interaction indicates that divergence becomes less related to positive affect intensity as time passes. In Supplementary Figure G1 we plotted the slope of time by positive affect, to visualize the changes in the effect of time across quartiles. For higher positive affect intensity, the effect of time was stronger, increasing divergence at higher intensity levels as more time passed.

Model 2. Negative Affect Intensity (Quartile)

Divergence was modelled as a function of negative affect intensity and time. The first model included only main effects of polynomial contrasts and time, while the second model additionally included interaction terms between the polynomial contrasts and time. We focus on the interaction model in the following and report complete results in Table 3, the random effects are reported in Supplement F.

Our analysis revealed a significant, positive linear trend ($\beta_1 = 6.76$; $p < .001$), indicating a consistent increase in the divergence between momentary and retrospective reports for increasing levels of negative affect intensity quartiles. The quadratic and cubic trends were significant in the main effect model, but decreased in magnitude when adding the interaction with time. The quadratic trend was positive and significant in the interaction model ($\beta_2 = 2.42$; $p = .033$).

Table 3

Nested Models of Negative Affect Intensity Quartiles and Time

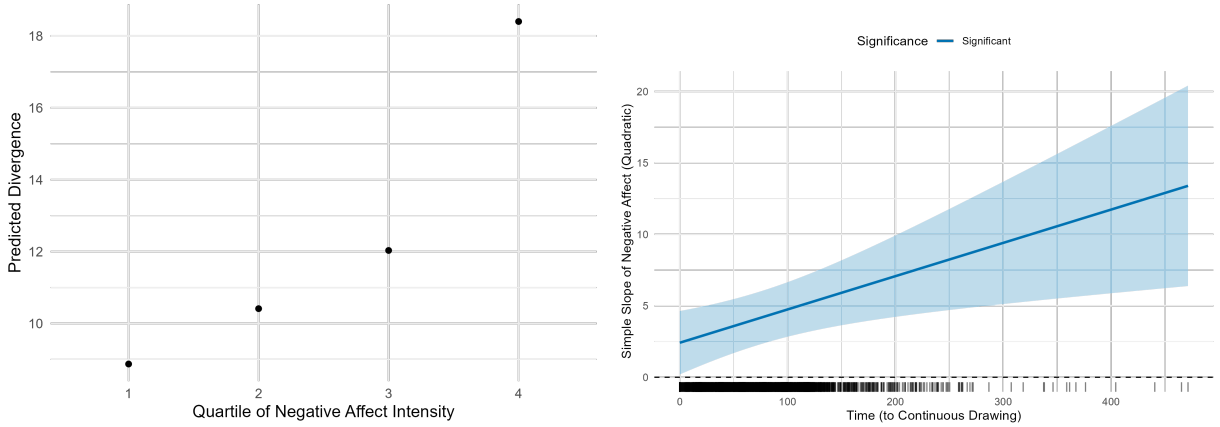
	Main Effects		Two-Way Interaction	
Effect	Estimate	SE	Estimate	SE
Intercept	12.459***	0.624	12.429***	0.619
Linear	5.875***	1.298	6.755***	1.471
Quadratic	4.006***	0.934	2.415*	1.132
Cubic	1.815***	0.482	1.045	0.743
Time	0.017**	0.005	0.019**	0.005
Linear:Time			-0.013	0.010
Quadratic:Time			0.023**	0.009
Cubic:Time			0.010	0.007
R^2	40.33%		40.80%	

Note: p < .05 *; p < .01 **; p < .001 ***; SE = Standard Error

Moreover, we found that the time between the momentary rating and the retrospective rating was a significant positive predictor of divergence ($\beta_4 = 0.02$; $p = .001$). This suggests that, at the midpoint of the contrast-coded variable, the more time there was between the momentary and retrospective rating, the greater the divergence between these ratings. The interaction between the linear trend and time was not significant; however, the quadratic effect significantly interacted with time ($\beta_6 = 0.02$; $p < 0.01$).

Figure 6

Predicted Divergence and Simple Slopes of Quadratic Effect of Negative Affect Intensity (Quartiles)



Note: Left: main effect when Time is extrapolated to 0: Right: Johnson-Neyman plot illustrating how the quadratic effect of negative affect intensity quartiles on divergence varies as a function of time. The blue shaded region indicates the time window during which the effect is statistically significant ($p < .05$). The observed range of *time* was [0, 471.42] minutes.

Figure 6 (left) displays the prediction plot for the main effect of negative affect intensity, illustrating a clear linear trend. As levels of negative affect intensity increase across quartiles, there is a corresponding rise in predicted divergence, with time extrapolated to zero. To further explore the interaction between time and the quadratic trend of negative affect, we conducted a simple slopes analysis across a continuous range of time values. The Johnson-Neyman plot (Figure 6; right) illustrates how the simple slope of the quadratic effect changes over time. The plot indicates that the quadratic trend becomes more pronounced as time increases. The U-shaped relationship between negative affect intensity and divergence strengthens with longer time since the momentary rating. This means that divergence becomes more pronounced at the lowest and highest quartiles of negative affect intensity as time increases, whereas moderate negative affect is less prone to divergence in ratings that occurred longer ago. We plotted the effect of time changes across quartiles (Figure G3), showing an increase in the effect of time for higher intensity

levels.

Direction of Divergence

To further investigate the nature of the divergence, we explored *in what way* the momentary ratings differed from the retrospective drawing. We determined the direction of the divergence (whether drawings over- or underestimated the momentary rating) by removing the absolute term from Eq. (1). The magnitude and direction of divergence were computed by subtracting the drawing value from the momentary rating.

$$Momentary\ Intensity_{ij} - Drawing\ Intensity_{ij}$$

We found that the average difference between momentary positive affect intensity and the corresponding positive affect drawing across all observations was 1.64 (SD = 17.53), indicating that momentary ratings were generally higher than the drawings. For most momentary ratings, positive affect is underestimated in the drawings. To better understand how affect intensity influences the direction and magnitude of divergence, we examined the average difference within each quartile of momentary positive affect (Table 4). In the 1st quartile, divergence was largest and negative, meaning that drawings were, on average, rated higher than the corresponding momentary assessments. In the 2nd quartile, divergence was smallest and already positive, indicating that drawings were rated slightly lower than momentary ratings. In the 3rd and 4th quartiles, divergence increased further, suggesting that higher levels of momentary positive affect were increasingly underestimated in the drawings.

For negative affect intensity, the average difference between momentary ratings and drawings across all observations was -2.95 (SD = 17.51), indicating a general tendency to overestimate negative affect retrospectively. Looking at the average differences across quartiles of momentary negative affect intensity (Table 4), we observed negative differences in the 1st to 3rd quartile, which means that the drawings were scored higher in negative affect intensity than what the moments showed. This indicates an overestimation of

negative affect intensity in retrospective drawings for most momentary ratings. In contrast, the 4th quartile showed a positive difference, indicating that for high levels of negative affect intensity, negative affect was underestimated in drawings.

Table 4

Average Differences Between Momentary and Retrospective Ratings by Quartile of the Overall Distributions of the Momentary Affect Intensity Items

	Positive Affect		Negative Affect	
	Mean	*SD*	*Mean*	*SD*
1st Quartile	-9.29	17.35	-7.46	12.10
2nd Quartile	1.22	15.21	-7.46	14.35
3rd Quartile	6.03	16.10	-4.83	15.97
4th Quartile	8.44	15.88	7.97	21.26

Note: The quartiles and the distributional properties are determined based on the untransformed scores.

Divergence with Reported Event Intensity

Descriptive Statistics

The average positive event intensity score, of 825 reported events, was 51.41 (SD = 24.50). The average absolute divergence between the positive event and the retrospective positive affect drawing at the time of the event was 23.43 (SD = 19.52). For the 418 reported negative events, the average intensity score was 44.11 (SD = 24.78) and the average absolute divergence between the negative event and the retrospective negative affect drawing at the time of the event was 22.07 (SD = 18.11). The absolute divergence between the event ratings and the retrospective drawing was therefore larger than the divergence between the momentary affect ratings and the retrospective drawing.

Model 3. Positive Event Intensity (Quartile)

We examined how the quartiles of positive event intensity and the time since the event related to the divergence between event ratings and retrospective drawings at the

time of the event. We estimated two ordered factor models: one including the polynomial contrasts and the main effect of time, and a second including the interaction parameters Table 5. The results reported below are based on the interaction model, the random effects are reported in Supplement F.

Table 5

Nested Models of Positive Event Quartiles and Time

	Main		Two-Way	
Effect	Estimate	SE	Estimate	SE
Intercept	24.656***	1.177	24.789***	1.174
Linear	-12.314***	2.127	-17.794***	2.578
Quadratic	6.827***	1.257	7.849***	1.879
Cubic	-0.472	1.136	0.471	1.847
Time	-0.013	0.022	-0.022	0.022
Linear:Time			0.162***	0.043
Quadratic:Time			-0.042	0.042
Cubic:Time			-0.027	0.043
R^2	46.11%		47.18%	

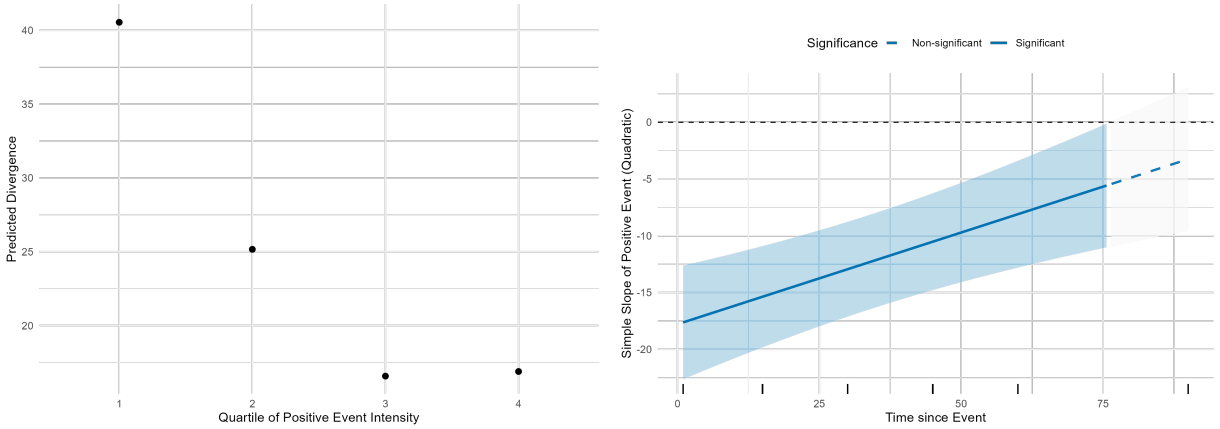
Note: p < .05 *; p < .01 **; p < .001 ***; SE = Standard Error

We found a significant negative linear trend (β_1 = -17.79; $p < .001$), suggesting that divergence decreased for higher levels of positive event intensity at time point zero. Figure 7 (left) visualizes the effect of positive event intensity across the quartiles when time is extrapolated to zero. There is a clear linear relationship between intensity and divergence, with divergence decreasing as positive event intensity increases. The quadratic trend was also significant and positive ($\beta_2 = 7.85$; $p < .001$), indicating that the decrease in divergence was not strictly linear but followed a U-shaped pattern across the quartiles of positive event intensity. This is reflected in the flattening of the effect between the 3rd and 4th quartiles (Figure 7, left). The cubic trend and the effect of time since the event

were not significant.

Figure 7

Predicted Divergence and Simple Slopes of Linear Effect of Positive Event Intensity (Quartiles)



Note: Left: Main effect of intensity when *time* is extrapolated to 0. Right: Johnson-Neyman plot illustrating how the linear effect of positive affect intensity quartiles on divergence varies as a function of time. The blue shaded region indicates the time period during which the effect is statistically significant ($p < .05$). Participants reported time since the event in [0, 15, 30, 45, 60, and 90] minutes. Slopes were interpolated by minutes in between observations.

We also found a significant interaction between time and the linear trend of positive event intensity ($\beta_5 = 0.16$, $p < .001$), suggesting that the effect of event intensity on divergence is moderated by time. To interpret this interaction, we conducted a simple slopes analysis, estimating the slope and its 95% CI to determine the time intervals in which the slope was significant. As illustrated in Figure 7 (right), the slope of the linear effect of positive event intensity on divergence changes as a function of time. Since participants reported time in discrete categories (0, 15, 30, 45, 60, and 90 minutes), the slopes were interpolated to allow for a continuous visualization of how the effect changed over time. The negative effect of positive event intensity is the strongest when the event occurred more recently. The linear trend weakens as time passes, and divergence becomes non-significant after approximately 75 minutes. We plotted how the effect of time changed

over levels of the positive event intensity (Figure G4), indicating a stronger effect of time for higher event intensity.

Model 4. Negative Event Intensity (Quartile)

In the negative event model, we used polynomial contrasts to investigate the effect of negative event intensity across the four quartiles and the effect of time since the event on divergence. We investigated the main effects and the interaction effects between the contrasts and time in iterative models. Table 6 provides all estimates, the random effects are reported in Supplement F.

Table 6

Nested Models of Negative Affect Intensity Quartiles and Time

	Main Effects		Two-Way Interaction	
Effect	Estimate	SE	Estimate	SE
Intercept	23.191***	1.33	23.530***	1.345
Linear	13.211***	2.073	16.578***	2.911
Quadratic	4.574*	1.844	5.189	2.745
Cubic	1.340	1.593	-1.035	2.564
Time	-0.034	0.03	-0.042	0.03
Linear:Time			-0.097	0.061
Quadratic:Time			-0.012	0.06
Cubic:Time			0.069	0.058
R^2	37.04%		37.55%	

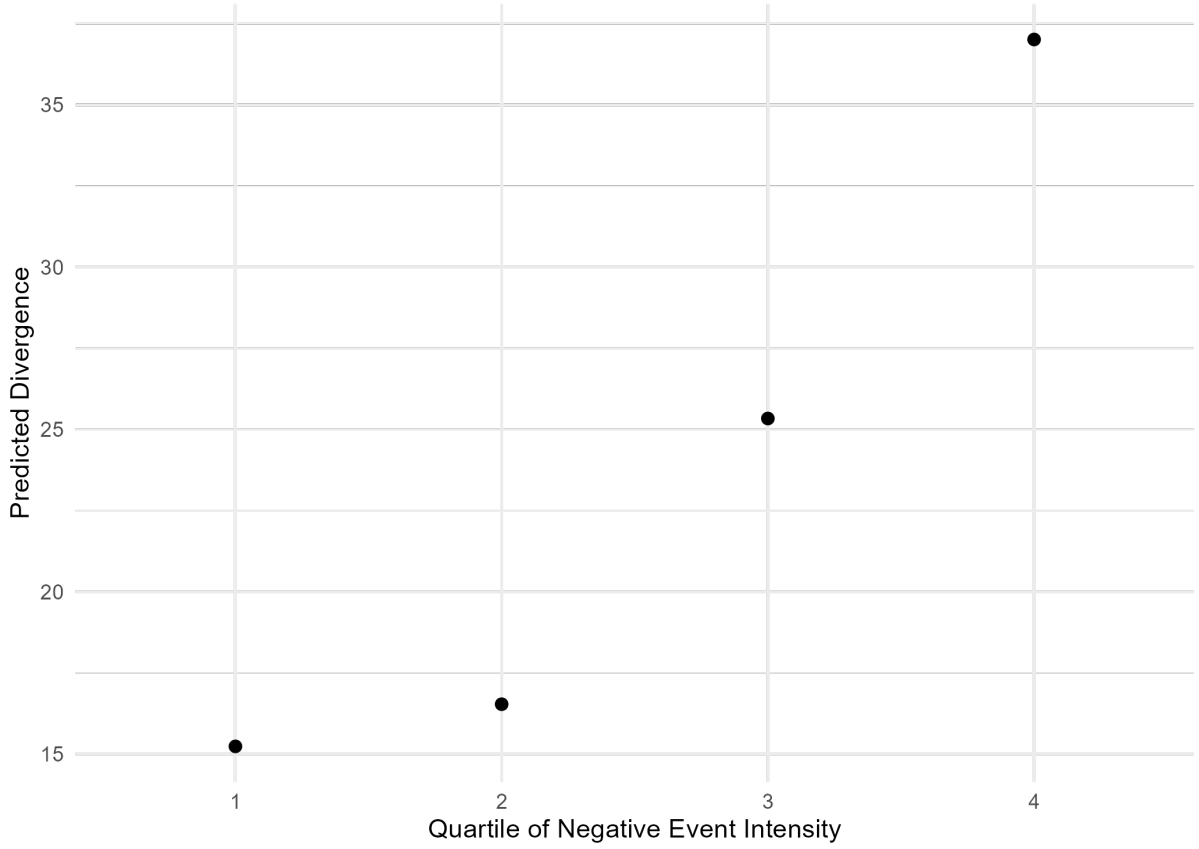
Note: p < .05 *; p < .01 **; p < .001 ***; SE = Standard Error

Firstly, we found that the linear trend of negative event intensity across the four quartiles was statistically significant ($\beta_1 = 16.58$; $p = 0.001$), indicating that at time zero, higher levels of negative event intensity were associated with greater divergence. Secondly, the quadratic trend was significant in the main effects model but not in the interaction model. Thirdly, there was no significant main effect of time and no significant interaction

between the polynomial trends and time since the negative event.

Figure 8

Predicted Divergence by Negative Event Intensity (Quartiles)



Note: Main Effect when Time is Extrapolated to 0

Figure 8 illustrates divergence across increasing negative event intensity quartiles, based on predictions made using the model with interaction effects (albeit non-significant) and time extrapolated to zero. The plot demonstrates that as negative event intensity increases, the divergence between event ratings and retrospective ratings also increases. Between the 1st and 2nd quartile, the increase in divergence is relatively slow, while from the 2nd to 4th, there is a more linear, steeper increase.

Direction of Divergence

As we did for the momentary affect divergence, we further investigated the nature of the divergence, exploring *in what way* event intensity ratings and retrospective drawing ratings differ from each other. We determined the magnitude and direction of the divergence by removing the absolute term from Eq. (1).

$$Event\ Intensity_{ij} - Drawing\ Intensity_{ij}$$

We found that the difference between positive event intensity ratings and the drawing intensity at the time of the event was, on average, -10.91 ($SD = 28.49$). This indicates that for positive events, the continuous drawing rating was generally higher than the event rating. We analyzed the average difference across each quartile of positive event intensity to better understand how these differences manifest across levels of event intensity (Table 7). In the 1st, 2nd, and 3rd quartile, the differences were negative. means that the continuous drawing at the time of the event indicated higher positive affect than the event intensity rating drawing. The event intensity was rated as less positive than the drawing rating at the time of the event, suggesting an overestimation of the event's positivity in the continuous drawing. For the 4th quartile, the continuous drawing rating was on average less positive than the event intensity rating, indicating that high-intensity positive events were underestimated.

For negative events, the difference between the event intensity and the continuous drawing rating at the time of the event was, on average, 7.19 ($SD = 27.65$). Looking at the average differences per quartile of the negative event intensity distribution (Table 7), we found negative differences for the 1st quartile, indicating that lower-intensity negative events were overestimated with more intense negative ratings on the continuous drawing. The higher-intensity quartiles show increasing positive differences. Very high-intensity negative events were severely underestimated in the retrospective drawings.

Table 7

Average and Direction of Divergence Scores per Event Intensity Quantile

	Positive Event		Negative Event	
	Mean	*SD*	*Mean*	*SD*
1st Quartile	-34.44	27.90	-8.77	17.84
2nd Quartile	-13.19	22.73	1.51	21.39
3rd Quartile	-5.93	21.95	6.45	15.13
4th Quartile	8.08	23.13	27.48	30.02

Note: The quartiles and the distributional properties are determined based on the untransformed scores.

Discussion

Continuous affect drawings were developed as a method to capture how continuous affect unfolds between momentary assessments taken in ESM studies. Although the drawings provide information on what happens in between assessments, they still rely on short-term retrospective recall (and as such are a recalled trajectory rather than an online continuous assessment). Therefore, the short-retrospective period captured in the drawing may be influenced by memory effects, particularly the peak-end effect, which might cause the most intense and the most recent moments to be remembered more clearly than other time points. Using data from a validation study that applied this method within an ESM design, we investigated this research question in two ways. Firstly, we investigated whether the intensity and timing of momentary ratings taken in between assessments were associated with their divergence from the continuous drawings. Secondly, we looked at reported events to determine whether the intensity and timing of the affective event ratings were related to their divergence from the affect ratings in the continuous drawing.

We found that divergence patterns differed by valence and as a function of recency. The relation between intensity and the divergence from the affect drawings was moderated by the valence of affect or the event, showing a different effect for positive and negative affect or event intensity.

Positive Affect or Event Intensity

The ordered factor analyses found that higher positive affect intensity was generally associated with lower divergence, supporting the peak-effect hypothesis, that intense positive affect enhances memory accuracy. However, this effect was time-dependent: after approximately two hours, the linear relation between divergence and affect intensity diminished. Thus, while intensity may improve recall, this effect is driven by recency, and fades over longer recall periods between the drawing and the momentary rating.

The same effect was also observed in models focusing on positive event intensity. Higher intensity events were associated with lower divergence, consistent with the peak-effect hypothesis. This indicates that the continuous rating aligned with events, but that the association between event and affect ratings depends on the intensity of the event. Furthermore, this was moderated by time, and we found a decrease in the association of high-intensity events and the continuous drawing as more time passed between the event and the drawing. Additionally, there was a positive main effect of recency, indicating that more recent events are remembered better. While event intensity can initially increase the association between events and drawings, this effect declines over time.

Positive emotions or events are generally remembered better than neutral ones, because they are in line with the generally positive (default) view people have of themselves (especially in Western cultures, Heine, 2004; Taylor & Brown, 1988). Positive experiences are easily encoded and retrieved, since they are consistent with this view. In contrast, low positive or neutral experiences are less salient and fade more quickly from memory than very positive ones (Walker et al., 2003). Furthermore, neutral or low positive affective experiences have a similar cognitive representation as negative memories (D'Argembeau et al., 2003). This may explain why their divergence is similar to the divergence of negative experiences.

Negative Affect or Event Intensity

In the negative affect intensity or negative event models, higher negative affect intensity was associated with increased divergence between momentary negative affect or negative event ratings and the corresponding retrospective continuous drawing ratings. As indicated above, negative experiences are not only perceived as less in line with people's self-view; they are also typically less frequent and more intense than positive experiences (Diener et al., 1985). Looking at the direction of the divergence, we found that for lower momentary negative affect or events, the retrospective ratings were higher. This latter finding is in line with research that suggests that retrospective evaluations of negative affect episodes result in retrospective reports being generally more intense than real-time reports (Robinson & Clore, 2002a, p. 952). Intense negative experiences may be overwhelming due to their uncommonness and are subsequently remembered as more intense than they were at the time. Interestingly, for high negative experiences, the reports in the drawings underestimated the original intensity. Downregulating or reappraising negative emotions is essential to maintain emotional well-being, but may reduce the salience of the emotional experience in memory (Gross, 2002). The observed underestimation of intense negative affect in retrospective reports may reflect memory bias but could also point to an adaptive emotion regulation process (Gross & John, 2002). The observation that high negative events have less overlap with the retrospective affect may be explained by rationalization processes, causing people to minimize the impact of negative events on their affect in retrospect (Taylor, 1991; Wilson et al., 2003).

In the negative affect models, we found a relationship between the recency of momentary assessments and divergence. As expected, more time between the momentary and drawing ratings was associated with greater divergence between these ratings, aligning with our hypothesis based on the end effect. In the negative event models, we did not find this effect. Fredrickson and Kahneman (1993) suggest that the recency of an experience may not necessarily be a reliable predictor of retrospective evaluations for negative

experiences. Other research has shown that retrospective recall of negative affect in non-depressed individuals was better predicted by the average of momentary affect ratings, rather than by the most recent experience (Ben-Zeev et al., 2009).

Momentary Affect Versus Event Measures

We investigated the peak and end effects from the momentary affect ratings as well as the event ratings. Our findings show that the patterns of associations with divergence were the same for both predictors: Positive affect or event intensity predicted a decrease in divergence, negative affect or event intensity predicted an increase in divergence.

However, the divergence between event and continuous drawings was notably larger than that observed for momentary affect versus continuous drawings, even though both event and continuous items were asked retrospectively. This observation underlines that there may not be a direct one-to-one overlap between event ratings and affective experiences. Prior work suggests that these processes rely on different memory systems and may be influenced by different biases (Robinson & Clore, 2002a). As such, affective events and affective experiences may not be reducible to a single evaluative dimension. This distinction underscores that rating an affective event involves more than recalling how one felt in the moment; it also requires interpreting the event's meaning, context, and implications, leading to potential discrepancies between the two (Scherer & Moors, 2019).

The divergence measure should be interpreted differently for the two predictors. While the discrepancy between momentary and retrospective affect can be interpreted as a measure of recall accuracy, the divergence between events and retrospective drawings should be seen as an indication of how much the continuous drawings are informed by salient events. The latter reveals how events and their intensity are integrated into reconstructions of affect over time and how recency shapes this.

Limitations

There are some noteworthy considerations regarding the peak end rule and the methodology we used to investigate it. Firstly, the peak-end rule applies to the comparison

of individual ratings to an overall summary of these ratings. The effect shows that the summary rating does not reflect the average of the individual ratings (Alaybek et al., 2022). In this study, we do not have a summary rating, but take individual ratings and compare them to other individual ratings taken from the drawing. Our comparison measure is thus not a summary measure of the time that elapsed, but in itself an approximation of what happened at the moment in which the individual ratings were taken.

In this regard, a limitation of our study is that, while we have accurate time points for momentary assessments, the corresponding time points derived from the drawings are not as precise. This lack of time sensitivity in the drawings introduces some noise into the divergence scores, making the timing less accurate. Consequently, there is a discrepancy in timing that affects the overall accuracy and reliability of our divergence measurements.

Secondly, it is necessary to note that most of the work on the peak-end rule has been done in the context of duration neglect. For example, that the retrospective summary of a painful procedure is not influenced by how long it was, but by how intense the peak and most recent moments were (Kahneman et al., 1993; Redelmeier & Kahneman, 1996). In this study, we have not discussed this aspect of the literature.

A third limitation of our study is the potential overlap between the predictor used to calculate the divergence measure and the divergence measure itself. This overlap can introduce bias, as the error variance in the predictor may affect the outcome. To address this issue, we transformed the data in two different ways to check the sensitivity of the results. Both transformations pointed in the same direction, suggesting consistency in our findings. However, the overlap of error variance in the predictor and outcome still poses a challenge and could impact the reliability of the divergence measure. Further, addressing the overlap of error variance in predictor and outcome remains critical. Future studies could explore alternative methods for calculating divergence measures that minimize this overlap.

Lastly, we cannot fully disentangle whether observed divergences reflect memory bias or are partially influenced by participants' ability to accurately translate affective experiences into a drawing format. While piloting and debriefing did not suggest major difficulties with the task, we did not account for cognitive, visuospatial, or motor skills. Future research should explicitly examine these alternative explanations, for example, by assessing cognitive or visuospatial abilities at baseline, or by comparing the continuous drawing across populations with differing motor or cognitive profiles. Clarifying these sources of variance is essential for the broader application of continuous drawings in affect research.

Future Research

The findings indicating that momentary ratings and retrospective evaluations differ and that this divergence is associated with affective intensity are crucial for future research. Especially in ESM research, which aims to capture ecologically valid, moment-to-moment data, there is a challenge that momentary reports may carry saliency and accessibility biases (Ram et al., 2017). This means that momentary data may be highly context-specific and does not translate well to later time points. Instead, retrospective data may contain more information, since participants have reflected and embedded the experiences in a more permanent and generalizable rating. There is a risk that, although they are ecologically valid, such context-specific momentary measures may be less relevant for predicting behavior and long-term outcomes, such as well-being and mental health (Dejonckheere et al., 2019). Just as short-term and long-term memory differ (Norris, 2017), the momentary experience may be functionally distinct from what remains in our memory, behavior, and retrospective evaluation of these experiences. Understanding these differences and identifying which types of information predict the outcomes of interest (such as predicting depression, see Wichers & Groot, 2016) is essential for improving affect measurement methodologies in ESM research.

The continuous drawings are useful for researchers who want to follow up on

longer-term processes. They provide insight into emotion regulation, recovery after events, and change processes. In contexts where the underlying temporal process is unknown, continuous drawings may provide empirical insight into temporal dynamics, supporting theory development and informing methodological decisions about assessment timing and model specification (Hopwood et al., 2022). Moreover, the continuous item may be particularly useful in clinical contexts. It may enable long-term follow-up monitoring, allowing clinicians to decrease the sampling rate and the burden on their patients, while still being able to track gradual mood changes. It may also be particularly useful to assess high-intensity episodes (e.g., anxiety-inducing situations) during which immediate reporting is impaired (Murray et al., 2023), but their unfolding could be mapped shortly after using retrospective drawings. The drawing can be applied to collect data on the temporal process of for example episodes of binge-eating (in the context of eating disorder), craving (in the context of addiction), or panic (in the context of anxiety disorders), capturing gradual changes and acute shifts surrounding such episodes with a single report rather than requiring multiple follow-up assessments that may be difficult to attend to (e.g., Smyth et al., 2007). Finally, by reducing reliance on frequent, abstract numeric ratings, continuous affect drawings may be especially useful in contexts where participant burden, literacy, or repeated interruptions pose challenges, although further validation in these contexts is needed (Messick, 1995).

For the future application of the continuous item, it is essential that the retrospective drawings should be evaluated as patterns and not with time-point specific analyses. The identifiable peaks, the steepness of the curve, and the duration from peak to valley in these drawings reveal insights that extend beyond simple momentary assessments (Heylen et al., 2016; Verduyn et al., 2012). These parameters may offer a way forward in utilizing continuous drawings effectively, and may even be informative parameters to study emotion regulation (Koval et al., 2015). Since minute-to-minute ratings in the drawings individually may not provide much interesting information, validating the overall patterns

of the drawings becomes necessary.

Lastly, we have not explored how individual differences might contribute to the observed over- and underestimations. The variations in the direction of divergences are complex, and individual factors could provide additional insights. For instance, individuals with high levels of depression might struggle to acknowledge positive emotions or may discount them, leading to under-reporting of positive affect in retrospective evaluations (Heininga & Kuppens, 2021). Another factor to explore in this context would be neuroticism; individuals high in neuroticism may experience stronger and more prolonged negative reactions and might over-report or exaggerate negative affect in retrospective assessments (Bylsma et al., 2011).

Conclusion

The primary contribution of the continuous drawing lies in expanding the range of research questions that ESM measures can address. In the present study, it was used to examine retrospective affect by providing insight into how affective intensity and the passage of time predict retrospective recall. Our findings indicate that even when the retrospective window is short, affect or affective event intensity is meaningfully associated with recalled affect. Interestingly, this effect is different for positive and negative valence. Particularly, the accuracy of rating high negative affect intensity retrospectively, compared to momentary reports or event ratings, is questionable. The overlap between momentary or event and retrospective ratings tends to be better for positive affect measures. Nonetheless, our findings suggest that negative affect is often reported less accurately in retrospective assessments, which needs to be considered in future studies using this assessment. For future research employing continuous retrospective assessments, it is essential to analyze the entire drawing, as it contains valuable parameters beyond the moment-to-moment time points that are not necessarily time-sensitive. Future work should focus on identifying patterns and indicators within this data that can enhance the utility of this assessment method, particularly in using the information between moments

for substantive research. This approach will help refine the interpretation of continuous data and improve its applicability to investigate substantive research questions.

Disclosures

Conflicts of Interest

The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

The research reported in this paper was supported by a KU Leuven Research Council grant C14/23/062. The first author was funded by grant PDMT2/24/034 from the Research Council of KU Leuven and grant FWO 12ADH26N by the Fonds Wetenschappelijk Onderzoek - Vlaanderen.

Materials and Data

Materials and analysis code for this study are available at the OSF and can be accessed at https://osf.io/wg2zv/. The data is available at doi: 10.48804/Z2I43Q (Cloos, Mestdagh, et al., 2025).

Artificial Intelligence Use

The grammar of selected sections of this manuscript were checked with Microsoft Copilot, under the license of KU Leuven. GPT-4.1 was used to support analyses reported in the supplementary material.

References

Alaybek, B., Dalal, R. S., Fyffe, S., Aitken, J. A., Zhou, Y., Qu, X., Roman, A., & Baines, J. I. (2022). All's well that ends (and peaks) well? a meta-analysis of the peak-end rule and duration neglect. *Organizational Behavior and Human Decision Processes*, *170*, 104149. https://doi.org/10.1016/j.obhdp.2022.104149

Ariely, D., & Carmon, Z. (2000). Gestalt characteristics of experiences: The defining features of summarized events. *Journal of Behavioral Decision Making*, *13*(2), 191–201. https://doi.org/10.1002/(SICI)1099-0771(200004/06)13:2<191::AID-BDM330>3.0.CO;2-A

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Ben-Zeev, D., Young, M. A., & Madsen, J. W. (2009). Retrospective recall of affect in clinically depressed individuals and controls. *Cognition and Emotion*, *23*(5), 1021–1040. https://doi.org/10.1080/02699930802607937

Bylsma, L. M., Taylor-Clift, A., & Rottenberg, J. (2011). Emotional reactivity to daily events in major and minor depression. *Journal of Abnormal Psychology*, *120*(1), 155–167. https://doi.org/10.1037/a0021662

Cloos, L., Ceulemans, E., & Kuppens, P. (2023). Development, validation, and comparison of self-report measures for positive and negative affect in intensive longitudinal research. *Psychological Assessment*, *35*(3), 189. https://doi.org/10.1037/pas0001200

Cloos, L., Mestdagh, M., Vanpaemel, W., Ceulemans, E., & Kuppens, P. (2024). Measuring continuous affect in daily life with intensity profile drawings. *Assessment*, *32*(5), 689–704. https://doi.org/10.1177/10731911241266286

Cloos, L., Mestdagh, M., Vanpaemel, W., Ceulemans, E., & Kuppens, P. (2025). *Continuous Affect Measures for Daily Life*. https://doi.org/10.48804/Z2I43Q

Cloos, L., Siepe, B., Piccirillo, M., Fried, E., Wang, S., Helmich, M., Johnson, S., Hoffart, A., & Ebrahimi, O. (2025, February 6). Accuracy and consistency of visual analog scales in ecological momentary assessment and digital studies. https://doi.org/https://doi.org/10.1525/collabra.142735

Cohen, S., & Janicki-Deverts, D. (2012). Who's stressed? distributions of psychological stress in the united states in probability samples from 1983, 2006, and 2009. *Journal of Applied Social Psychology*, *42*(6), 1320–1334. https://doi.org/10.1111/j.1559-1816.2012.00900.x

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396. https://doi.org/https://doi.org/10.2307/2136404

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, *175*(9), 526–536. https://doi.org/10.1097/00005053-198709000-00004

D'Argembeau, A., Comblain, C., & Van der Linden, M. (2003). Phenomenal characteristics of autobiographical memories for positive, negative, and neutral events. *Applied Cognitive Psychology*, *17*(3), 281–294. https://doi.org/10.1002/acp.856

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, *3*(5), 478–491. https://doi.org/10.1038/s41562-019-0555-0

Diener, E., Larsen, R. J., Levine, S., & Emmons, R. A. (1985). Intensity and frequency: Dimensions underlying positive and negative affect. *Journal of Personality and Social Psychology*, *48*(5), 1253–1265.

Ehring, T., Zetsche, U., Weidacker, K., Wahl, K., Schönfeld, S., & Ehlers, A. (2011). The perseverative thinking questionnaire (PTQ): Validation of a content-independent

measure of repetitive negative thinking. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(2), 225–232. https://doi.org/10.1016/j.jbtep.2010.12.003

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions [Publisher: Psychology Press Ltd]. *Cognition and Emotion*, *14*(4), 577–606. https://doi.org/10.1080/026999300402808

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*(1), 45–55. https://doi.org/10.1037//0022-3514.65.1.45

Garnefski, N., & Kraaij, V. (2007). The cognitive emotion regulation questionnaire. *European Journal of Psychological Assessment*, *23*(3), 141–149. https://doi.org/10.1027/1015-5759.23.3.141

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, *39*(3), 281–291. https://doi.org/10.1017/S0048577201393198

Gross, J. J., & John, O. P. (2002). Wise emotion regulation. In *The wisdom in feeling: Psychological processes in emotional intelligence* (pp. 297–319). The Guilford Press.

Haslbeck, J., Ryan, O., & Dablander, F. (2023). Multimodality and skewness in emotion time series. *Emotion*, *23*(8), 2117–2141. https://doi.org/10.1037/emo0001218

Heine, S. J. (2004). Positive self-views: Understanding universals and variability across cultures. *Journal of Cultural and Evolutionary Psychology*, *2*(1-2), 109–122. https://doi.org/10.1556/jcep.2.2004.1-2.7

Heininga, V. E., & Kuppens, P. (2021). Psychopathology and positive emotions in daily life. *Current Opinion in Behavioral Sciences*, *39*, 10–18. https://doi.org/10.1016/j.cobeha.2020.11.005

Heylen, J., Van Mechelen, I., Verduyn, P., & Ceulemans, E. (2016). KSC-n: Clustering of hierarchical time profile data. *Psychometrika*, *81*(2), 411–433. https://doi.org/10.1007/s11336-014-9433-x

Hollenstein, T. (2021). Affect dynamics and time scales: Pictures of movies. In C. E. Waugh & P. Kuppens (Eds.), *Affect dynamics* (pp. 117–129). Springer International Publishing. https://doi.org/10.1007/978-3-030-82965-0_6

Hopwood, C. J., Bleidorn, W., & Wright, A. G. C. (2022). Connecting theory to methods in longitudinal research. *Perspectives on Psychological Science*, *17*(3), 884–894. https://doi.org/10.1177/17456916211008407

Kahneman, D. (2000, September 25). Evaluation by moments: Past and future. In D. Kahneman & A. Tversky (Eds.), *Choices, values, and frames* (pp. 693–708). Cambridge University Press. https://doi.org/10.1017/CBO9780511803475

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. Retrieved September 8, 2023, from https://www.jstor.org/stable/40062570

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. https://doi.org/10.3758/MC.36.1.132

Koval, P., Brose, A., Pe, M. L., Houben, M., Erbas, Y., Champagne, D., & Kuppens, P. (2015). Emotional inertia and external events: The roles of exposure, reactivity, and recovery. *Emotion*, *15*(5), 625–636. https://doi.org/10.1037/emo0000059

Kuppens, P., Champagne, D., & Tuerlinckx, F. (2012). The dynamic interplay between appraisal and core affect in daily life. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00380

Leertouwer, I., Schuurman, N. K., & Vermunt, J. K. (2022). Are retrospective assessments means of people's experiences? *Journal for Person-Oriented Research*, *8*(2), 52–70. https://doi.org/10.17505/jpor.2022.24855

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Mroczek, D. K., & Almeida, D. M. (2004). The effect of daily stress, personality, and age on daily negative affect. *Journal of Personality*, *72*(2), 355–378. https://doi.org/https://doi.org/10.1111/j.0022-3506.2004.00265.x

Murray, A. L., Brown, R., Zhu, X., Speyer, L. G., Yang, Y., Xiao, Z., Ribeaud, D., & Eisner, M. (2023). Prompt-level predictors of compliance in an ecological momentary assessment study of young adults' mental health. *Journal of Affective Disorders*, *322*, 125–131. https://doi.org/10.1016/j.jad.2022.11.014

Neubauer, A. B., Scott, S. B., Sliwinski, M. J., & Smyth, J. M. (2020). How was your day? convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. *Journal of Personality and Social Psychology*, *119*(1), 185–203. https://doi.org/10.1037/pspp0000248

Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin*, *143*(9), 992–1009. https://doi.org/10.1037/bul0000108

Pedersen, H. S., Christensen, K. S., Prior, A., & Christensen, K. B. (2024). The dimensionality of the perceived stress scale: The presence of opposing items is a source of measurement error. *Journal of Affective Disorders*, *344*, 485–494. https://doi.org/10.1016/j.jad.2023.10.109

Radloff, L. S. (1977). The CES-d scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Ram, N., Brinberg, M., Pincus, A. L., & Conroy, D. E. (2017). The questionable ecological validity of ecological momentary assessment: Considerations for design and analysis. *Research in Human Development*, *14*(3), 253–270. https://doi.org/10.1080/15427609.2017.1340052

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. https://doi.org/10.1016/0304-3959(96)02994-6

Revol, J., Ariens, S., Lafit, G., Adolf, J., & Ceulemans, E. (2025). Episode-contingent experience-sampling designs for accurate estimates of autoregressive dynamics. *Psychological Methods*. https://doi.org/10.1037/met0000758

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2020). Momentary predictors of compliance in studies using the experience sampling method. *Psychiatry Research*, *286*, 112896. https://doi.org/10.1016/j.psychres.2020.112896

Robinson, M. D., & Clore, G. L. (2002a). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934–960. https://doi.org/10.1037/0033-2909.128.6.934

Robinson, M. D., & Clore, G. L. (2002b). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology*, *83*(1), 198–215. https://doi.org/10.1037/0022-3514.83.1.198

Rubin, D. C., & Umanath, S. (2015). Event memory: A rheory of memory for laboratory, autobiographical, and fictional events. *Psychological review*, *122*(1), 1–23. https://doi.org/10.1037/a0037907

Scherer, K. R. (2005). Appraisal Theory. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 637–663). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470013494.ch30

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*, 719–745. https://doi.org/10.1146/annurev-psych-122216-011854

Schwarz, N. (2007, April 19). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. A. Stone (Ed.), *The science of real-time data capture: Self-reports in health research* (pp. 11–26). Oxford University Press, USA. https://doi.org/10.1093/oso/9780195178715.003.0002

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Silvia, P. J., Kwapil, T. R., Eddington, K. M., & Brown, L. H. (2013). Missed beeps and missing data: Dispositional and situational predictors of nonresponse in experience sampling research. *Social Science Computer Review*, *31*(4), 471–481. https://doi.org/10.1177/0894439313479902

Smyth, J. M., Wonderlich, S. A., Heron, K. E., Sliwinski, M. J., Crosby, R. D., Mitchell, J. E., & Engel, S. G. (2007). Daily and momentary mood and stress are associated with binge eating and vomiting in bulimia nervosa patients in the natural environment. *Journal of Consulting and Clinical Psychology*, *75*(4), 629–638. https://doi.org/10.1037/0022-006X.75.4.629

Talarico, J. M., LaBar, K. S., & Rubin, D. C. (2004). Emotional intensity predicts autobiographical memory experience. *Memory & Cognition*, *32*(7), 1118–1132. https://doi.org/10.3758/BF03196886

Taylor, S. E. (1991). Asymmetrical Effects of Positive and Negative Events: The Mobilization-Minimization Hypothesis. *Psychological Bulletin*, *110*(1), 67–95.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193–210. https://doi.org/10.1037/0033-2909.103.2.193

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research*, *27*(3), 247–259. https://doi.org/10.1023/A:1023910315561

Verduyn, P., Van Mechelen, I., & Frederix, E. (2012). Determinants of the shape of emotion intensity profiles. *Cognition and Emotion*, *26*(8). https://doi.org/10.1080/02699931.2012.662152

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant—and memory helps to keep it that way! *Review of General Psychology*, *7*(2), 203–210. https://doi.org/10.1037/1089-2680.7.2.203

Wichers, M., & Groot, P. C. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, *85*(2), 114–116. https://doi.org/10.1159/000441458

Wilson, T. D., Meyers, J., & Gilbert, D. T. (2003). How happy was i, anyway? a retrospective impact bias. *Social Cognition*, *21*(6), 421–446. https://doi.org/10.1521/soco.21.6.421.28688

Remembering Affect Between Moments: Assessing Peak-End Effects In Continuous Affect Measures

Supplementary Material

Supplement A

Additional Measures

Description of measures and items not reported in the main text.

Baseline Trait Measures

Center for Epidemiologic Studies Depression Scale (CES-D).

We administered the Center for Epidemiologic Studies Depression (CES-D) Scale (Radloff, 1977) to measure the degree to which participants have experienced symptoms of depression in the preceding week. The 20-item measure consists of statements that are rated on a scale from 1 (*rarely or none of the time/less than a day*), to 4 (*most or almost all the time/5-7 days*).

Cognitive Emotion Regulation Questionnaire (CERQ)

The CERQ (Garnefski & Kraaij, 2007) is a 36-item self-report questionnaire, designed to assess individual differences in cognitive emotional regulation in response to negative events and stress. The scale comprises nine subscales that assess specific types of emotion regulation strategies with four items each: Self-blame, Acceptance, Rumination, Positive refocusing, Refocus on planning, Positive reappraisal, Putting into perspective, Catastrophizing, Other-blame. Participants rate the items on a 5-point Likert scale, ranging from 1 (*almost never*) to 5 (*almost always*). Subscale scores are obtained by summing the respective items (with scores ranging from 4 to 20), where higher scores indicate a greater utilization of the corresponding emotion regulation strategy.

Perseverative Thinking Questionnaire (PTQ)

This 15-item questionnaire assesses repetitive negative thinking patterns in general population samples, (Ehring et al., 2011). Participants are asked to rate how they typically think about negative experiences on a scale of 0 (*never*) to 4 (*almost always*) for instance, “Thoughts intrude into my mind”. The questionnaire captures three underlying factors: repetitive negative thinking, unproductiveness, and capturing mental resources.

These factors are reflected by nine, three, and three items, respectively.

Ruminative Response Scale (RSS)

Tenynor and colleagues (2003) designed this questionnaire to measure response styles related to depression. The 22 items evaluate individual differences in thinking about their lives and surroundings when they experience negative feelings. Participants rate different statements on a from 1 (*almost never*) to 4 (*almost always*). The sum of the items differentiates the level of ruminative thinking patterns between individuals.

Additional Momentary Measures

Rumination

Additionally the study included three items to measure rumination. Participants were asked "Since the last assessment did you ruminate about your feelings?" with a slider scale ranging from 0 (*Not at all*) to 100 *(A lot*), and they were asked to rate "Since the last assessment did you get stuck in your emotions?" on a scale from 0 (*Not at all*) to 100 (*A lot*). If participants indicated that they ruminated they were asked "How long have you been thinking about your emotions since the last assessment?" rated with the response options *just now* (1), *about 15 minutes ago* (2), *about 30 minutes ago* (3), *about 45 minutes ago* (4), *about 60 minutes ago* (5), *about 90 minutes ago* (6), or *about 120 minutes ago* (7).

Debriefing Questionnaire

System Usability Scale

The system usability scale (Bangor et al., 2008) measures the perceived usability of a product. It includes 10 items, with five positively and five negatively worded statements. We adapted the scale for the purpose of this study and reworded the statements. Eight items measured the ease ("I found the item with the drawing easy to answer") and learnability of the product ("It took me a long time to understand how the item with drawing worked"). Participants rate their agreement with these statements on a scale from

1 (*Strongly Disagree*) to 5 (*Strongly agree*).

Measurement Comparison

We created a set of items that aimed to compare the two methods, drawing and measurement bursts, for measuring affect in daily life. We asked participants to compare the usability, perceived accuracy, burden, memory bias, reactivity and general feasibility of the two methods. Open-ended questions provided participants with the opportunity to express additional insights or discoveries made during the study. The goal of the debriefing was to gather feedback on the strengths and weaknesses of the new method.

Supplement B

Preprocessing

The *continuous drawing* item recorded a series of data points, containing the x- and y-values of the plot that represented *time* and *affect intensity* respectively. To efficiently manage this data, each drawing was saved as a single concatenated string of comma-separated x- and y-values, where the first half of the data points represented the x-values, and the second half the y-values. We split and organized the values into two distinct columns for the x-values (*time*) and the y-values (*affect intensity*), respectively. Since the software implemented the x- and y-axes of the grid on a scale from 0 to 100, the x-values had to be converted back to actual *time* stamps. To achieve this, we multiplied the x-value by the difference between the *time* of the previous assessment and the *time* of the current assessment and divided this value by 100. After splitting the x- and y-values into separate rows they were ordered within each *continuous drawing* by the *time*. However, the number of data points captured in each drawing depended on the response duration, leading to variation in the time intervals and number of data points. To address the variability across assessments we standardized the *time* points to 1-minute intervals. The *time* values were standardized to be equidistant by binning them to 1 minute (usually resulting in no more than one value per *time* point). To fill in any missing *time* points and achieve a minute-to-minute scale for the x-axis we employed linear interpolation. For each interpolated *time* point a row was added to the data and the corresponding y-value was interpolated using a linear approximation between two adjacent *time* points.

Supplement C

Power Analysis

Prior to the study, we conducted a power simulation to determine the number of participants needed to estimate a small average within-person correlation ($\overline{r_i} = 0.1$) between the *drawing* and a measurement *burst*. As explained earlier, measurement bursts are used to collect momentary affect ratings more frequently with a high intensity sampling rate, by reducing the intervals between assessments (e.g., to 20 minutes).

We generated 1,000 datasets with 100 participants each. The *measurement bursts* were simulated with 10 notifications per day for 6 days, but with 50% missing data. For each participant, we simulated measurement bursts with 10 notifications over 6 days and introduced 50% missingness to the data, yielding an average of 30 observations per person. For each participant, we sampled the burst data from a standard normal distribution and generated the *continuous drawing* data based on these bursts. Specifically, burst observations were used to create the *continuous drawing* data, with a small average within-person correlation between the continuous drawing data and the burst data, set at an average correlation of 0.1 ($SD = 0.2$) and varying across individuals.

Results indicated that a sample of 100 participants would ensure 96% power ($\alpha = 0.05$). Consequently, we adjusted our study design to 7 days and 6 notifications per burst, yielding an average of 30 burst observations per person, assuming a compliance rate of 70%.

Supplement D

Assumptions

Affect Models

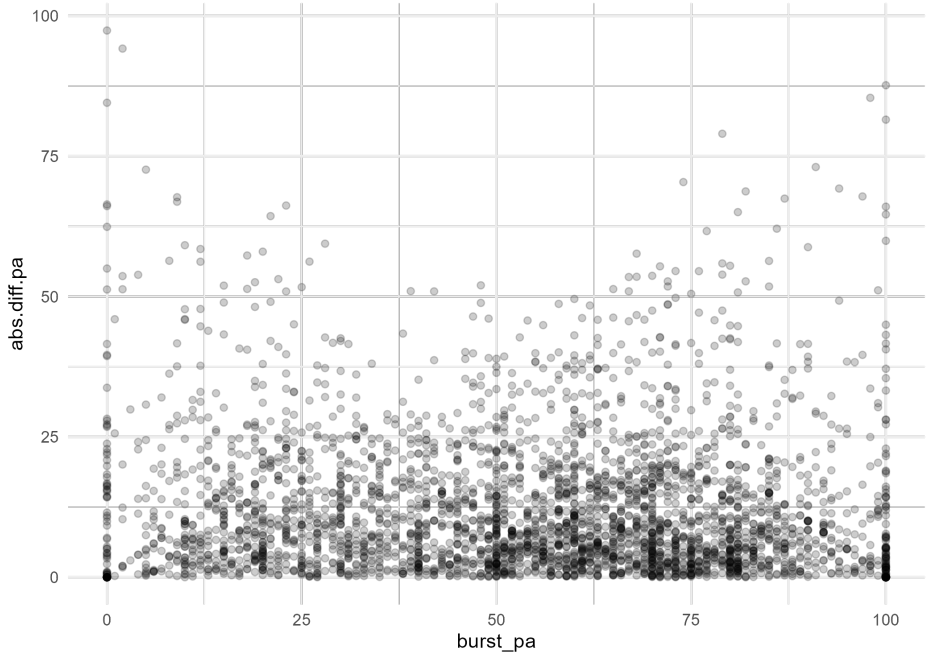
Positive Affect Intensity

To evaluate the suitability of linear mixed-effects modeling for the centered momentary positive affect predictor, we conducted a series of assumption checks, including visual inspection of scatter plots, residual diagnostics, and model fit statistics.

Visual Inspection. We evaluated the linear relation between the predictor positive affect intensity and divergence. We plotted the non-centered momentary positive affect against absolute divergence. The plot revealed no strong linear relationship between positive affect intensity and divergence (Figure D1). At the extreme ends, the data exhibited a curved pattern, with greater divergence at both low and high ends of the predictor range, suggesting potential non-linearity.

Figure D1

Scatter Plot of Positive Affect Intensity Against Divergence



To further explore non-linear patterns, we plotted scatter plots for 115 participants in four panels Figure D2 and Figure D3. These plots showed substantial heterogeneity in the relationship between positive affect intensity and divergence. Most participants did not exhibit a clear linear trend, and several showed U-shaped or S-shaped patterns, reinforcing the presence of non-linearity at the individual level. These patterns suggest that the relationship between positive affect and divergence is not consistently linear across individuals, and may vary in both direction and form. This variability supports the use of random slopes in mixed-effects models and motivates the exploration of non-linear modeling approaches or transformations to better capture the underlying structure of the data.

r

Figure D2

Individual Scatter Plots of Positive Affect and Divergence for all Participants

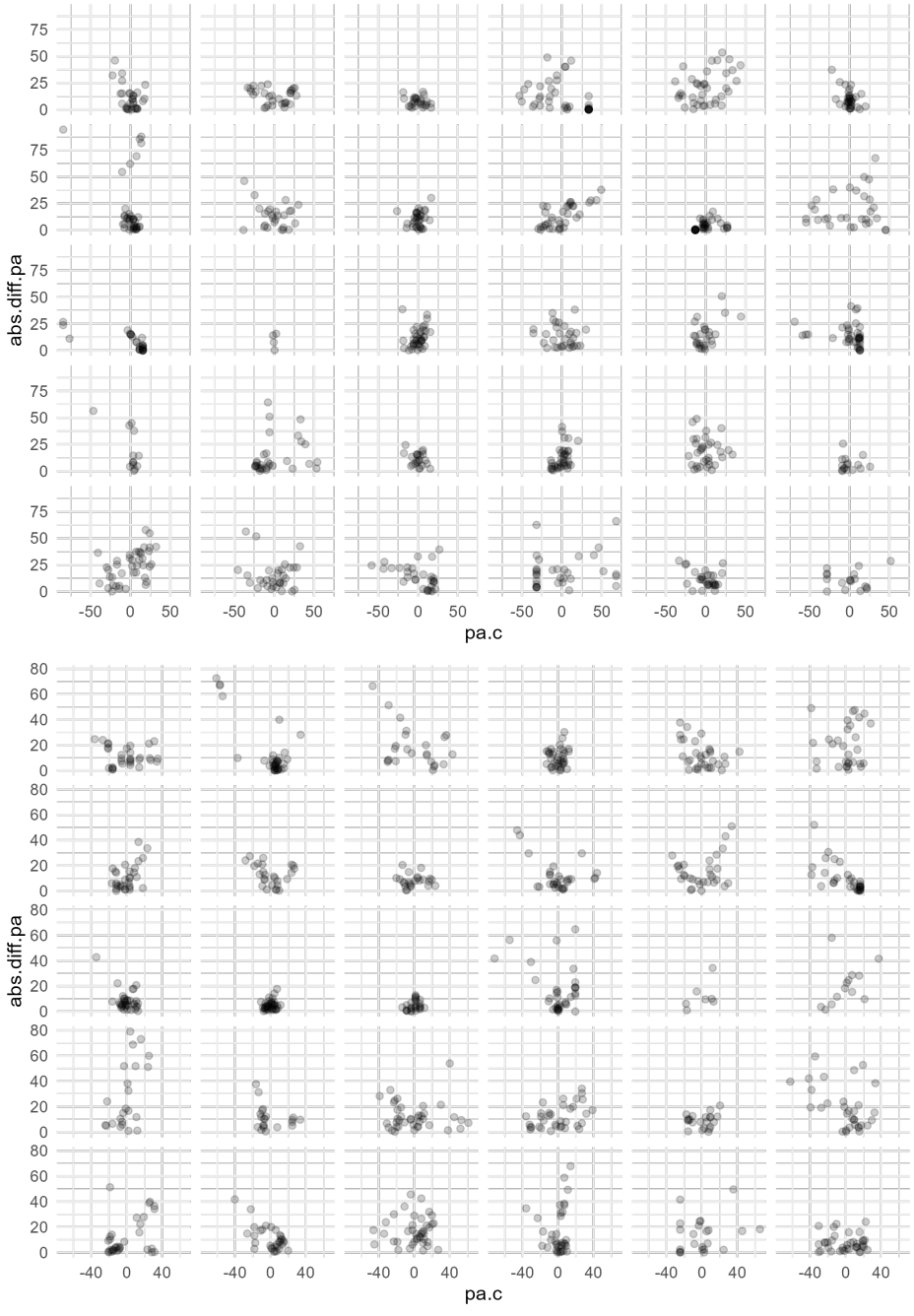
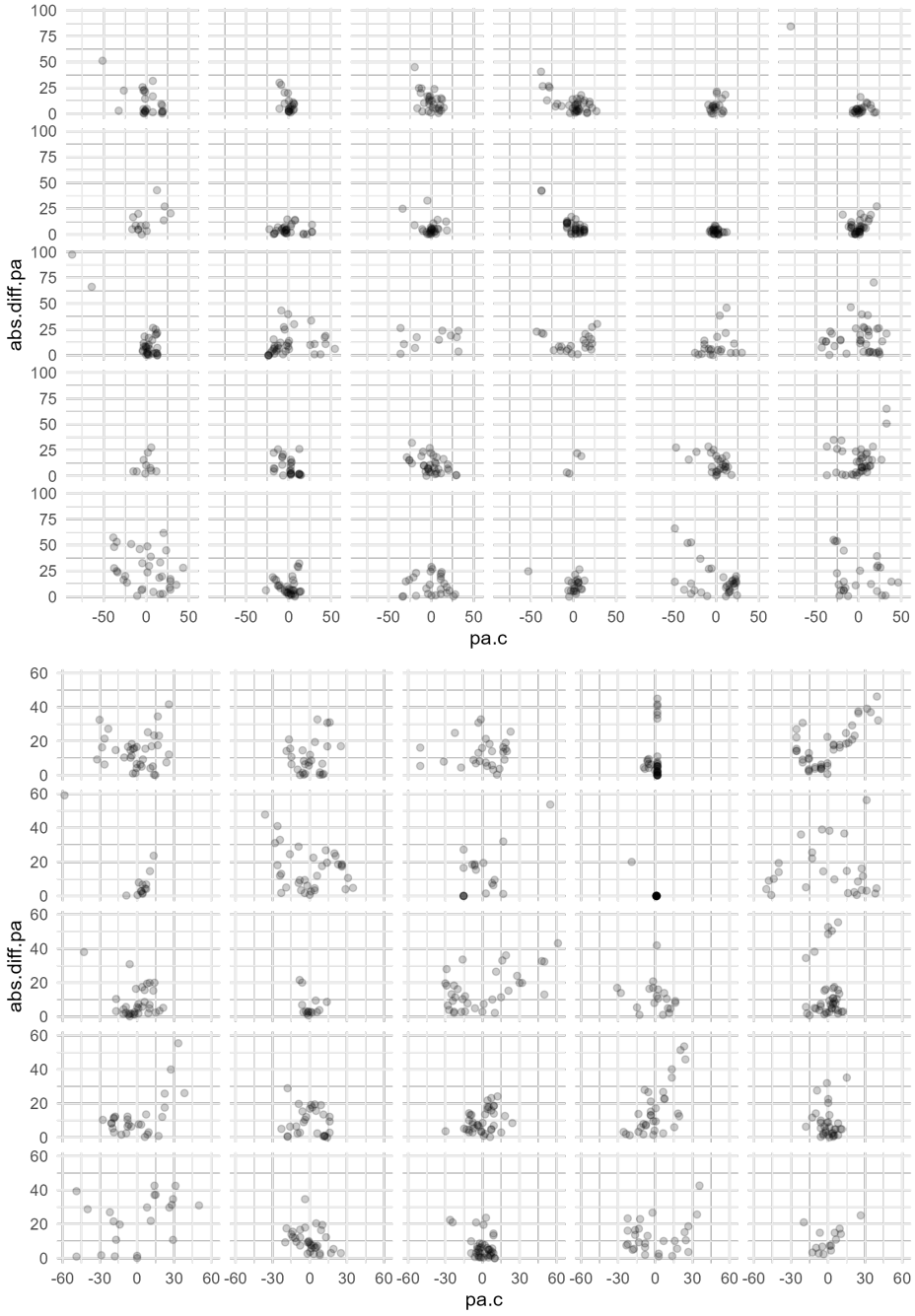


Figure D3

Individual Scatter Plots of Positive Affect and Divergence for all Participants (continued)

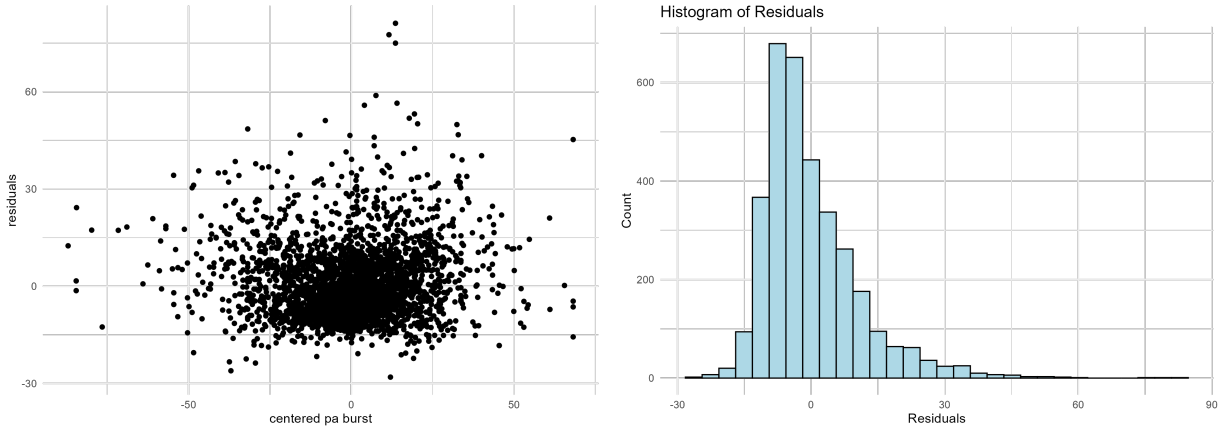


Residual Diagnostics. We next fit a linear mixed-effects model with a random intercept and slope for centered momentary positive affect, using an AR(1) correlation structure to account for autocorrelated errors within participants. Residuals from this

model were extracted and visually inspected (Figure D4). The residual plot showed uneven spread, with increased variance at the extremes of the predictor, indicating heteroscedasticity and further supporting the presence of non-linearity. A histogram of the residuals showed a roughly normal distribution with a longer tail to the right, though with slight deviations from normality.

Figure D4

Residual Plot of Centered Positive Affect on Divergence



Together, these assumption checks suggested that the relationship between momentary affect and divergence is likely non-linear and heteroscedastic, with greater variability and distinct patterns emerging above and below the midpoint of the affect scale. To account for this, we proceeded to split the affect variable at the midpoint and test non-linear models that included separate directional effects for each side of the scale in the main analyses. However, for completeness and transparency, we report the results of the corresponding linear model here.

Negative Affect Intensity

Visual Inspection. A scatter plot of all observations revealed a right-skewed distribution of negative affect intensity, with a high density of observations at lower values Figure D5. This pattern is typical in experience sampling data, where intense negative affect is relatively infrequent. The relationship between negative affect intensity and

divergence did not appear to be strongly linear. While lower levels of negative affect intensity were associated with relatively low and consistent divergence values, higher levels of negative affect intensity showed increased variability in divergence. This suggests that while some participants may experience greater divergence during intense negative affect, others do not, resulting in a fan-shaped pattern indicative of heteroscedasticity. Although the overall trend was weak, the increased spread of divergence at higher levels of negative affect suggests that a simple linear model may not fully capture the complexity of the relationship. However, this pattern does not preclude the use of linear mixed-effects models, particularly when random slopes are included to account for individual differences.

We plotted individual-level scatter plots for all 115 participants, divided into four panels Figure D6 and Figure D7. These plots revealed greater consistency in linear trends compared to the group-level scatter plot. While variability remained, many participants showed a positive linear association between negative affect intensity and divergence, suggesting that increases in negative affect were often accompanied by greater divergence. This pattern supports the use of a linear model for this predictor when accounting for individual differences through random slopes.

Figure D5

Scatter Plot of Negative Affect Intensity Against Divergence

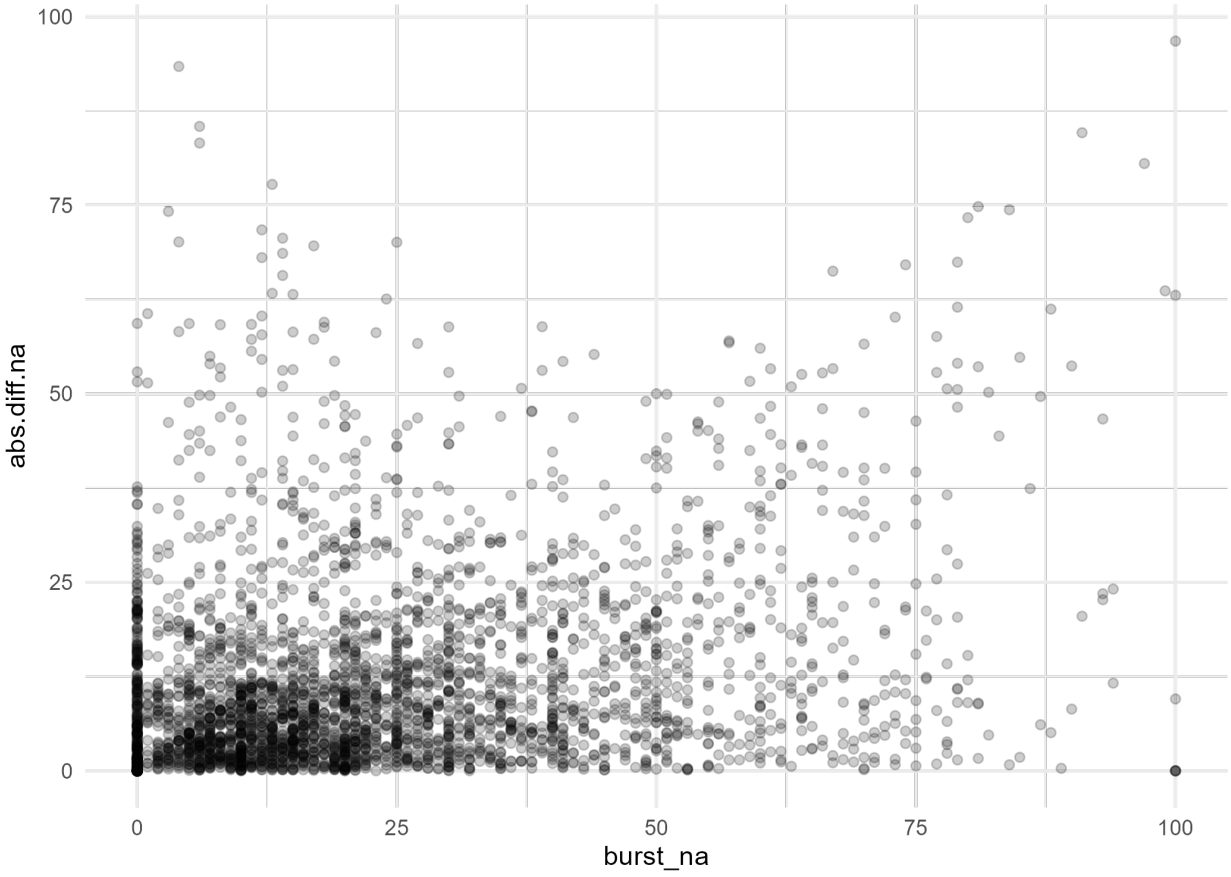


Figure D6

Individual Scatter Plots of Negative Affect and Divergence for all Participants

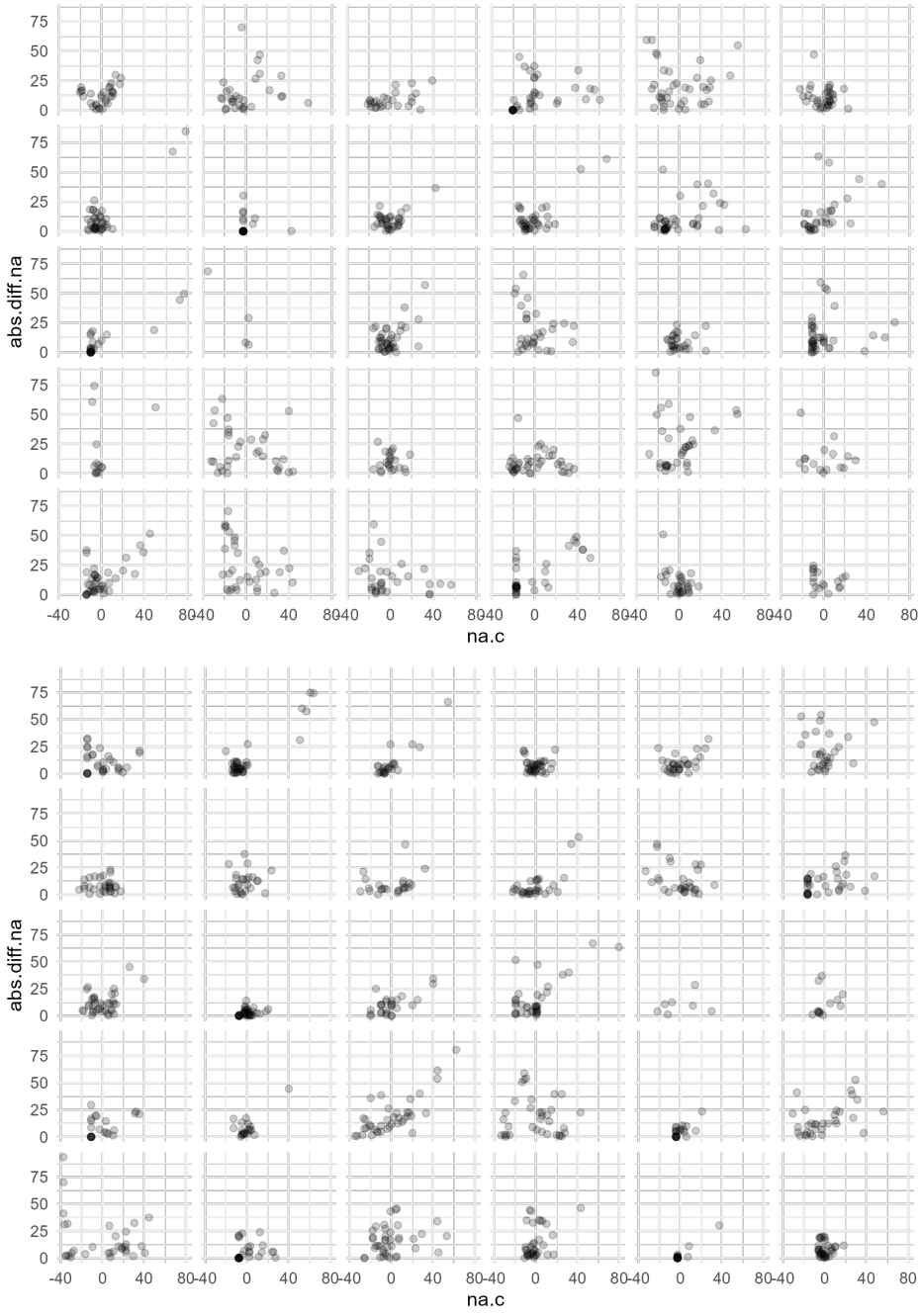
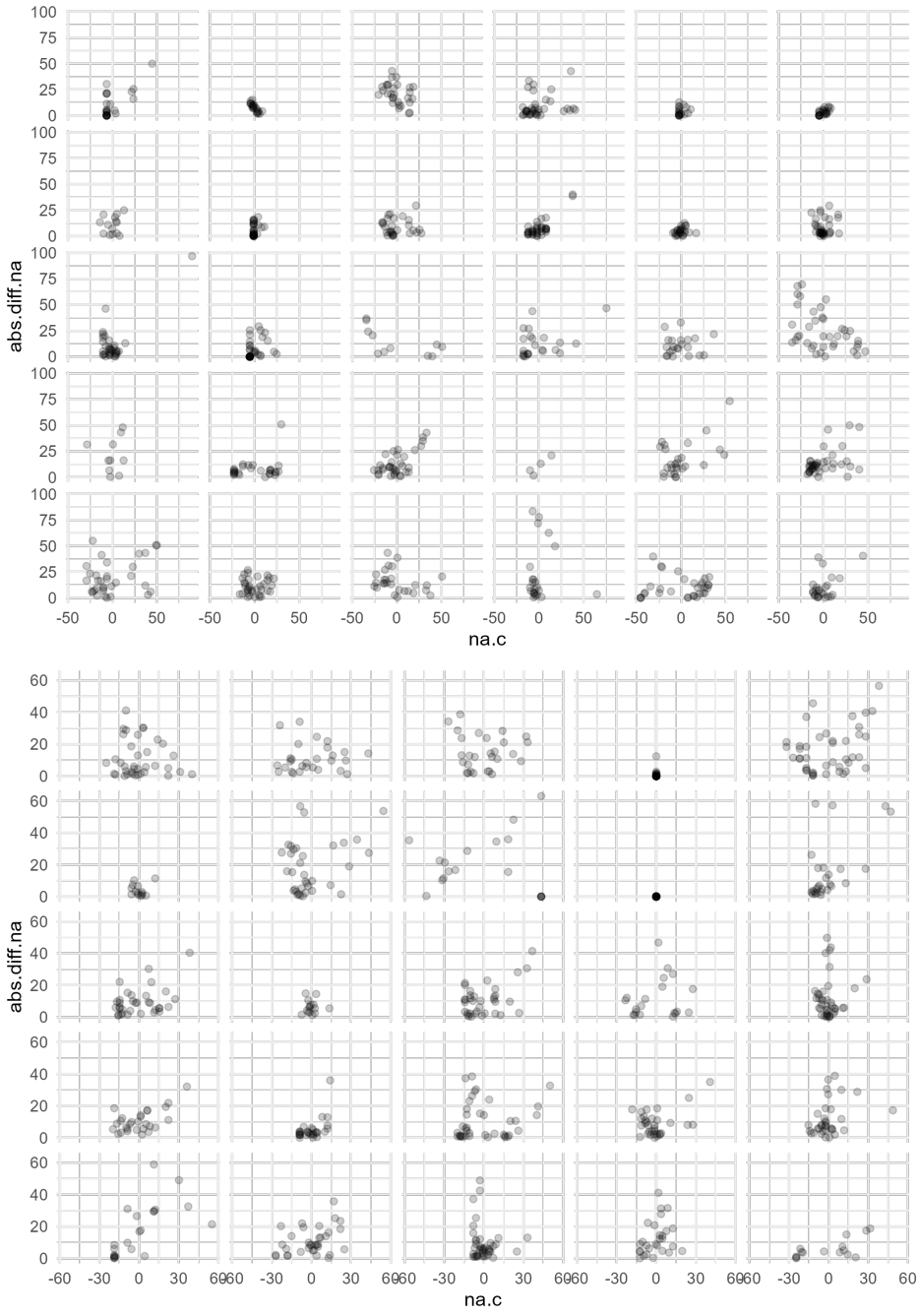


Figure D7

Individual Scatter Plots of Negative Affect and Divergence for all Participants (continued)

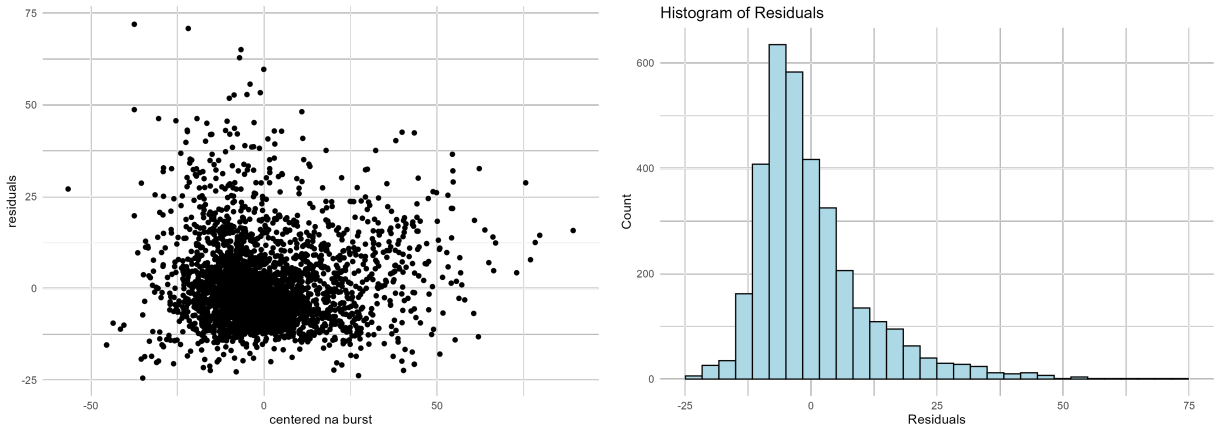


Residual Diagnostics. The residual plot revealed a concentration of residuals around the origin, with increasing dispersion as values of negative affect moved away from zero. This pattern suggests mild heteroscedasticity, particularly at the higher and lower

ends of the predictor range. However, the spread was not extreme, and the residuals appeared to be approximately symmetrically distributed, supporting the assumption of normality. While the residuals did not show a strong systematic pattern, the increased variability at the extremes may indicate that the model does not fully capture the complexity of the relationship for all participants. For this, it is important to include random slopes to account for individual differences.

Figure D8

Residual Plot of Centered Negative Affect on Divergence



Time (Between Momentary & Drawing)

Visual Inspection. A density plot of the variable Time Figure D9 revealed a right-skewed distribution. The distribution reflects the sampling design, where intervals between drawings were 140 minutes. In between this range, the most observations occurred. The bursts were given every 20 minutes, and there are small peaks in the density distribution showing a clustering of time to the drawing around these scheduled points.

To examine the relationship between time and divergence in positive affect and negative affect, we plotted scatter plots of all observations (Figure D10). The plots showed a higher concentration of data points at lower values of both time and divergence. As time increased, the spread of divergence values decreased due to the low number of data points.

Figure D9

Distribution of Time (Between Momentary & Drawing)

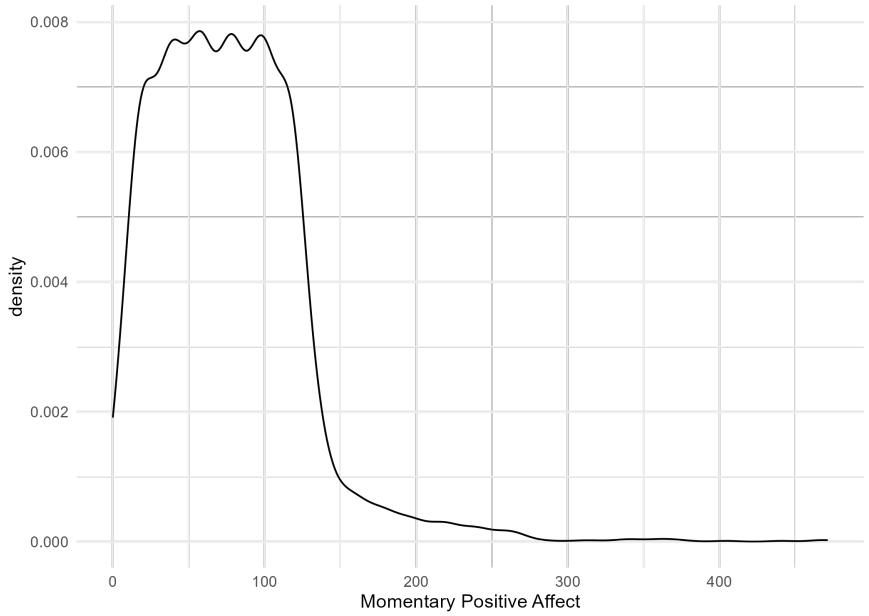
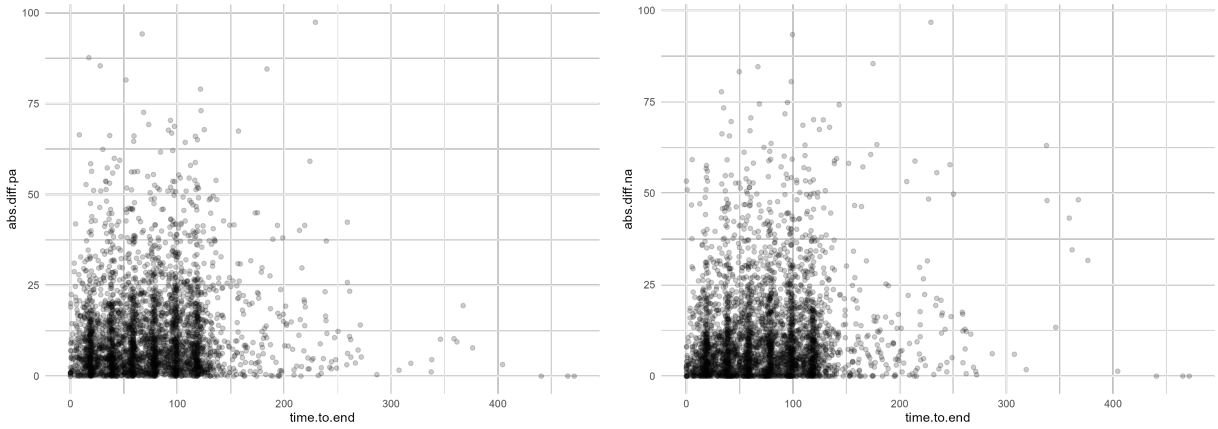


Figure D10

Scatterplots of Positive and Negative Affect Divergence and Time



To further examine the relationship between time and divergence, we plotted individual-level scatter plots of all participants (Figures Figure D11 to Figure D14). The scatter plots revealed substantial variability across individuals. While some participants showed a negative trend, with divergence decreasing over time, others exhibited flat or inconsistent patterns. Furthermore, the range of time between momentary and

retrospective rating varied across individuals but was very consistent within individuals. The individual-level plots suggest that the relationship between time and divergence is not uniform across participants, supporting the inclusion of random slopes for time.

Figure D11

Individual Scatter Plots of Time and Divergence (Positive) for all Participants

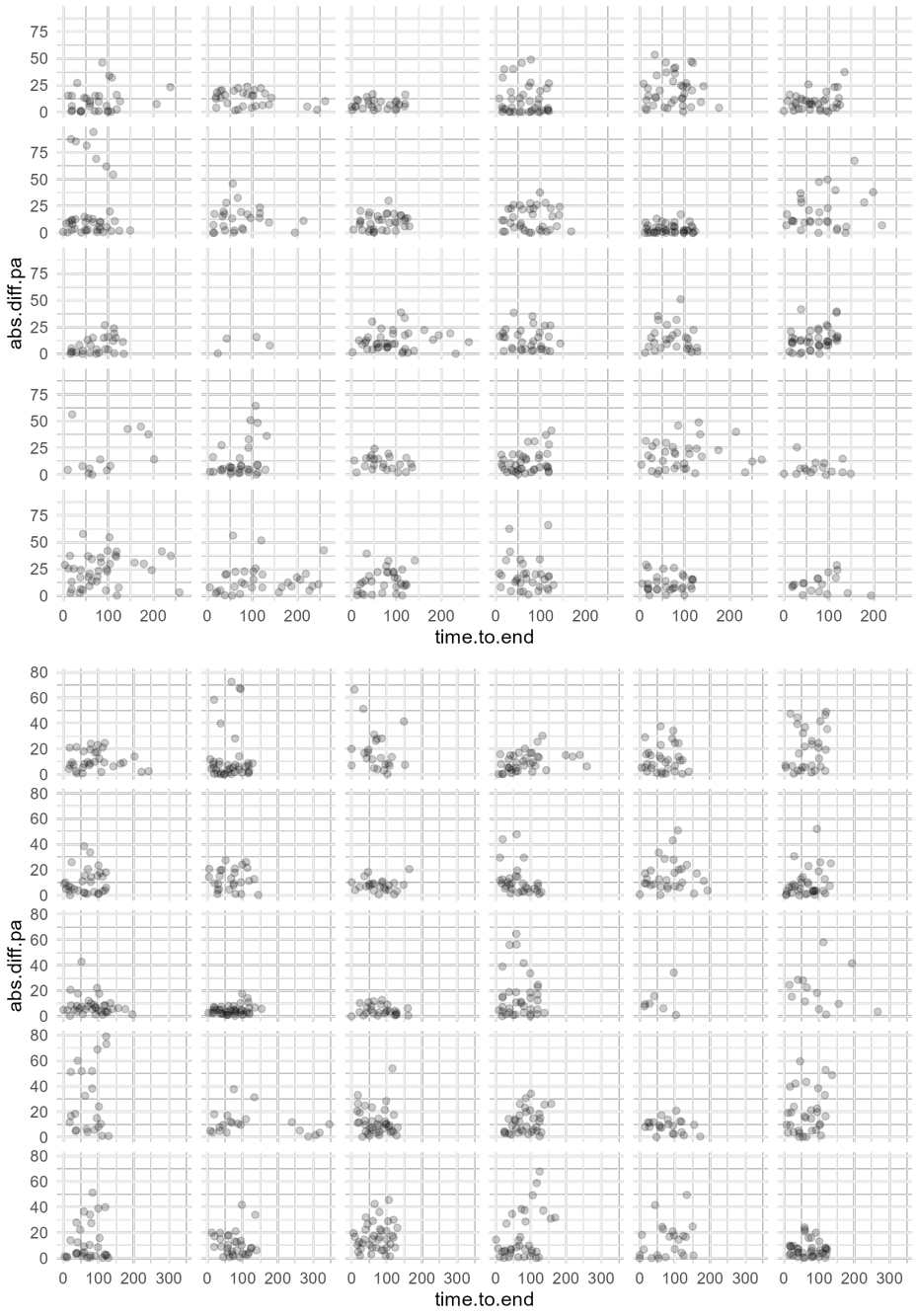


Figure D12

Individual Scatter Plots of Time and Divergence (Positive) for all Participants (continued)

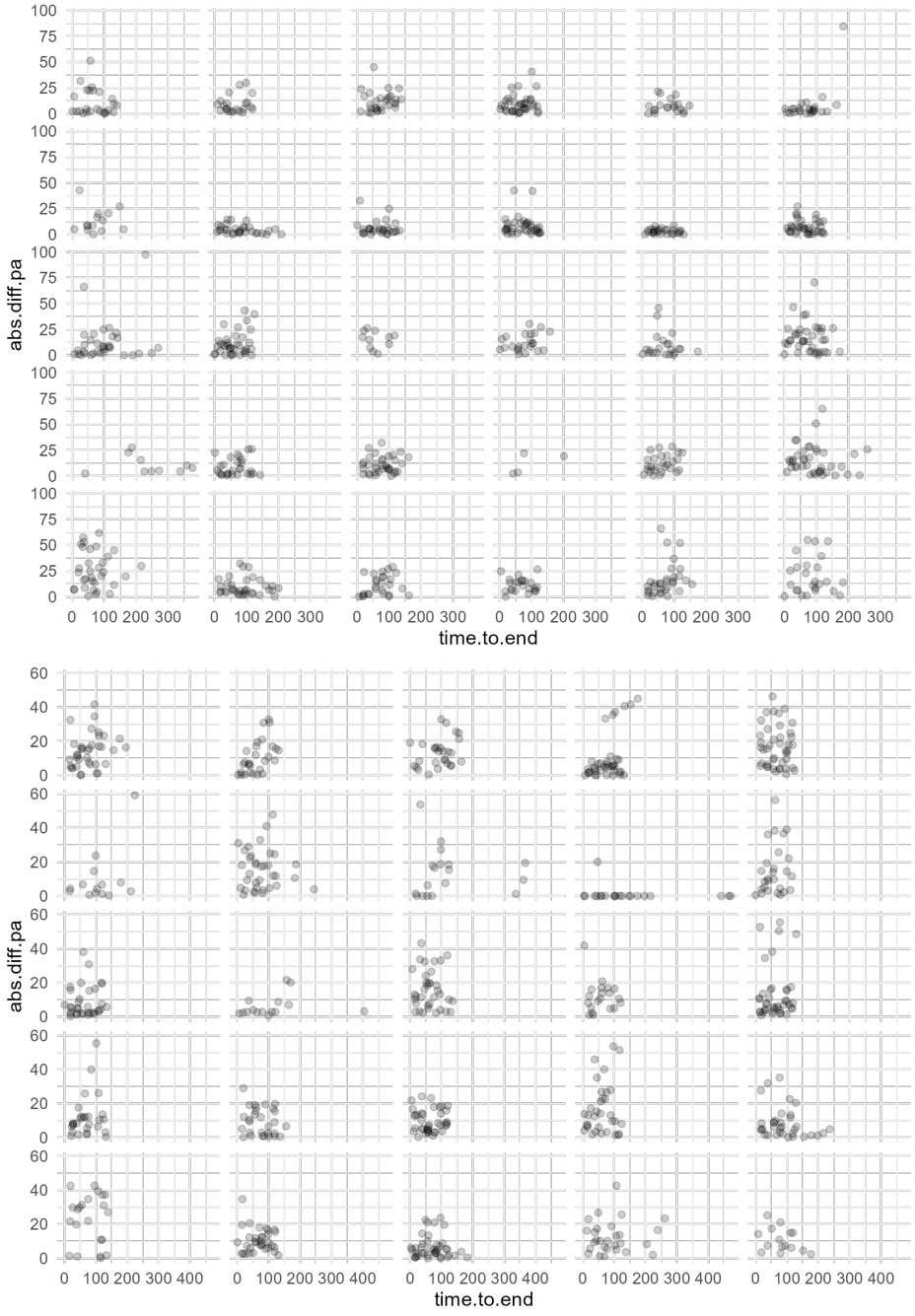


Figure D13

Individual Scatter Plots of Time and Divergence (Negative) for all Participants

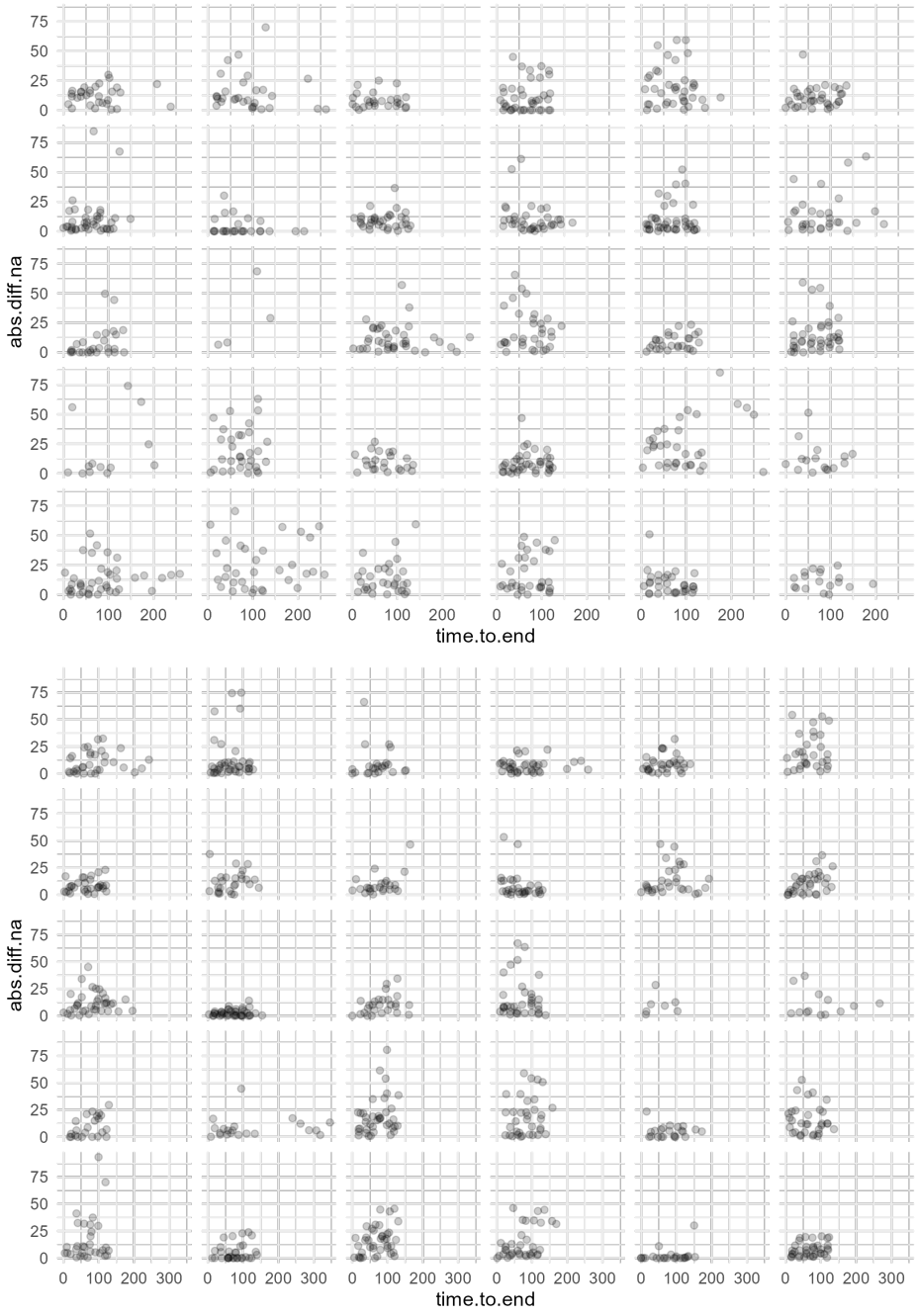


Figure D14

Individual Scatter Plots of Time and Divergence (Negative) for all Participants (continued)

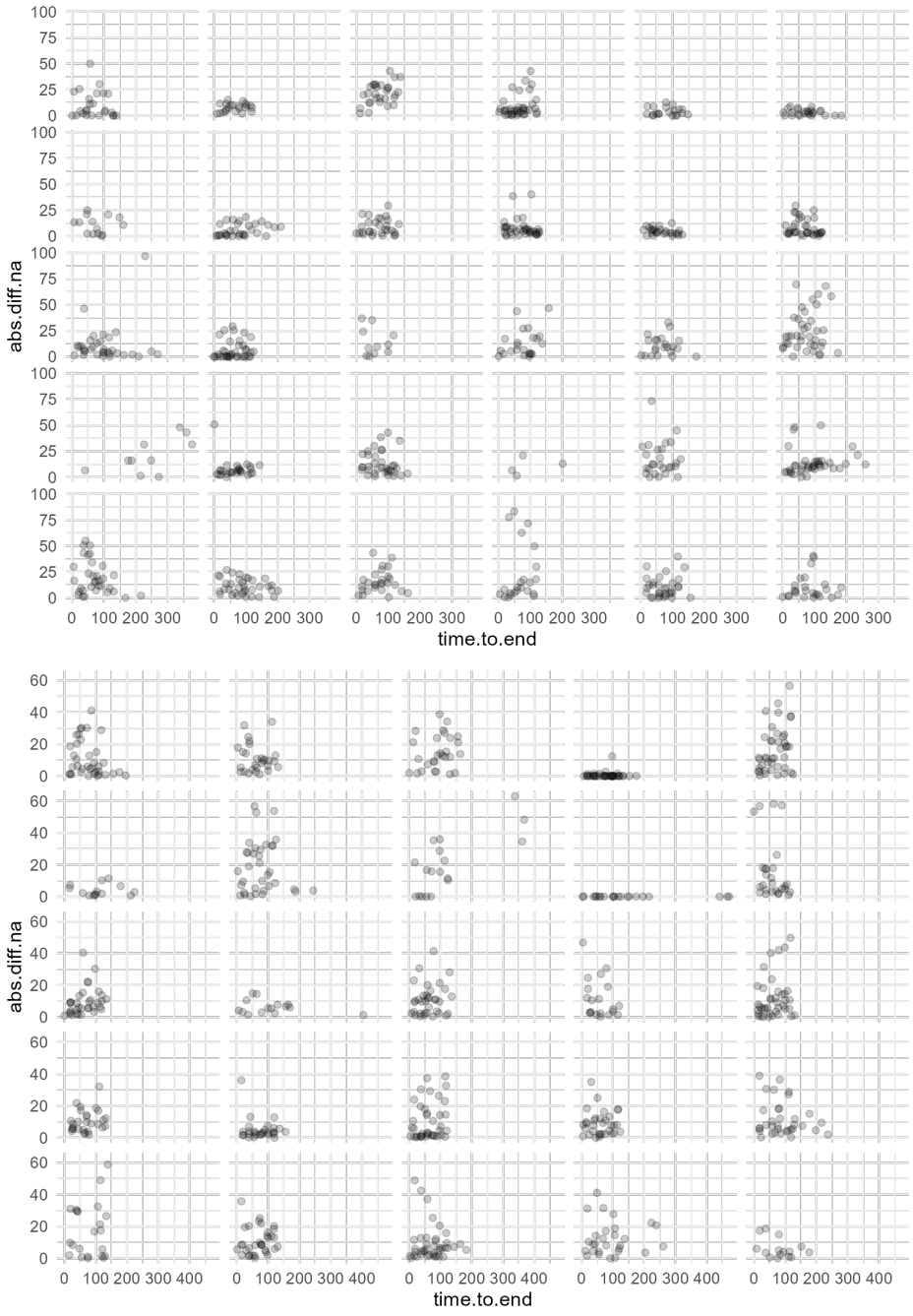


Figure D15

Residual Plot of Time on Divergence (Centered Positive)

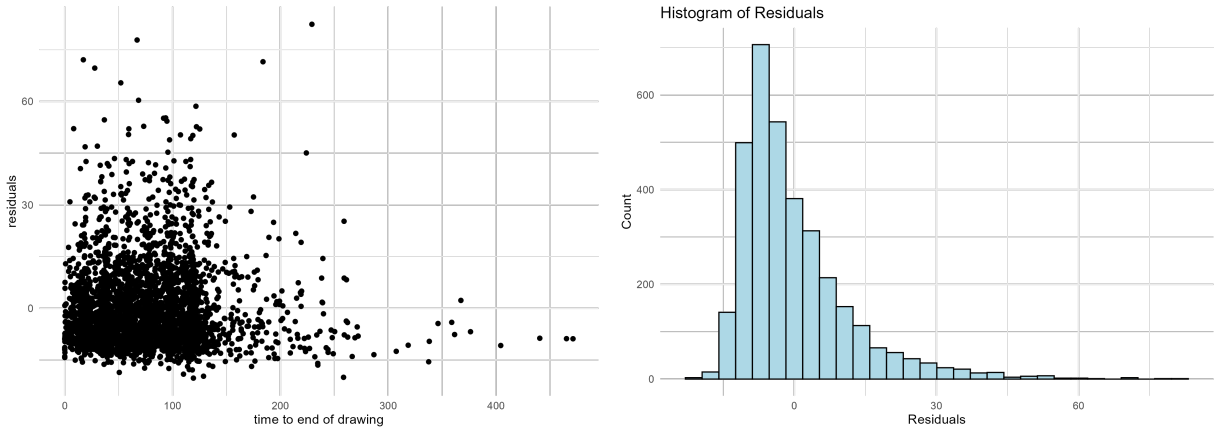
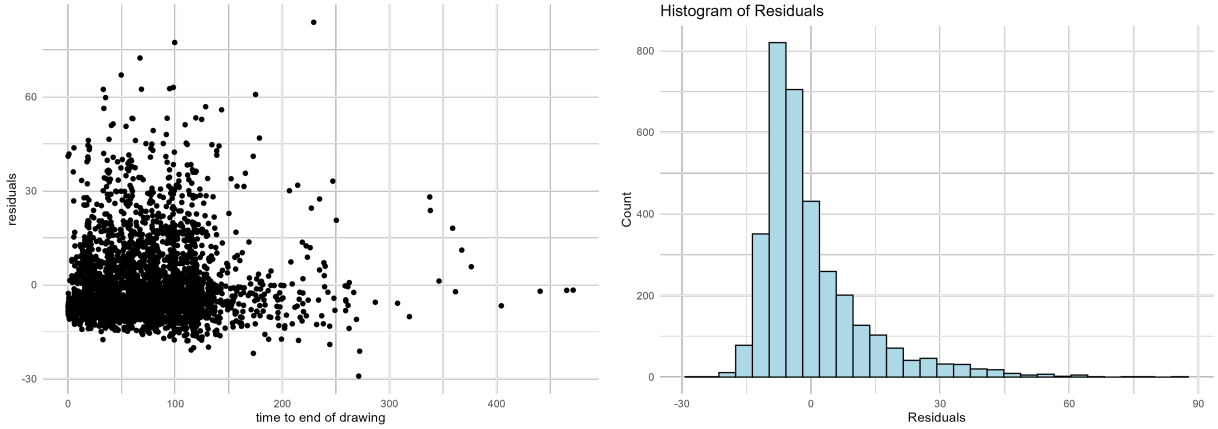


Figure D16

Residual Plot of Time on Divergence (Centered Negative)



Residual Diagnostics. To evaluate the assumptions of the linear mixed-effects models predicting divergence from time, we examined residual plots for both positive and negative affect models (Figure D15 and Figure D8). In both models, the residuals were plotted against time, which is measured on an absolute scale from 0 to the maximum observed interval of 471 minutes. The residual plots revealed a dense clustering of residuals at 0, with increased dispersion as time increases. This pattern suggests mild heteroscedasticity, where the variance of the residuals increases over time. Despite this, the residuals appeared to be centered around zero. The observed heteroscedasticity may reflect

natural variability in affective responses as time increases. Since this heteroscedasticity appeared primarily at the group level, the inclusion of random intercepts and slopes helped account for this variability.

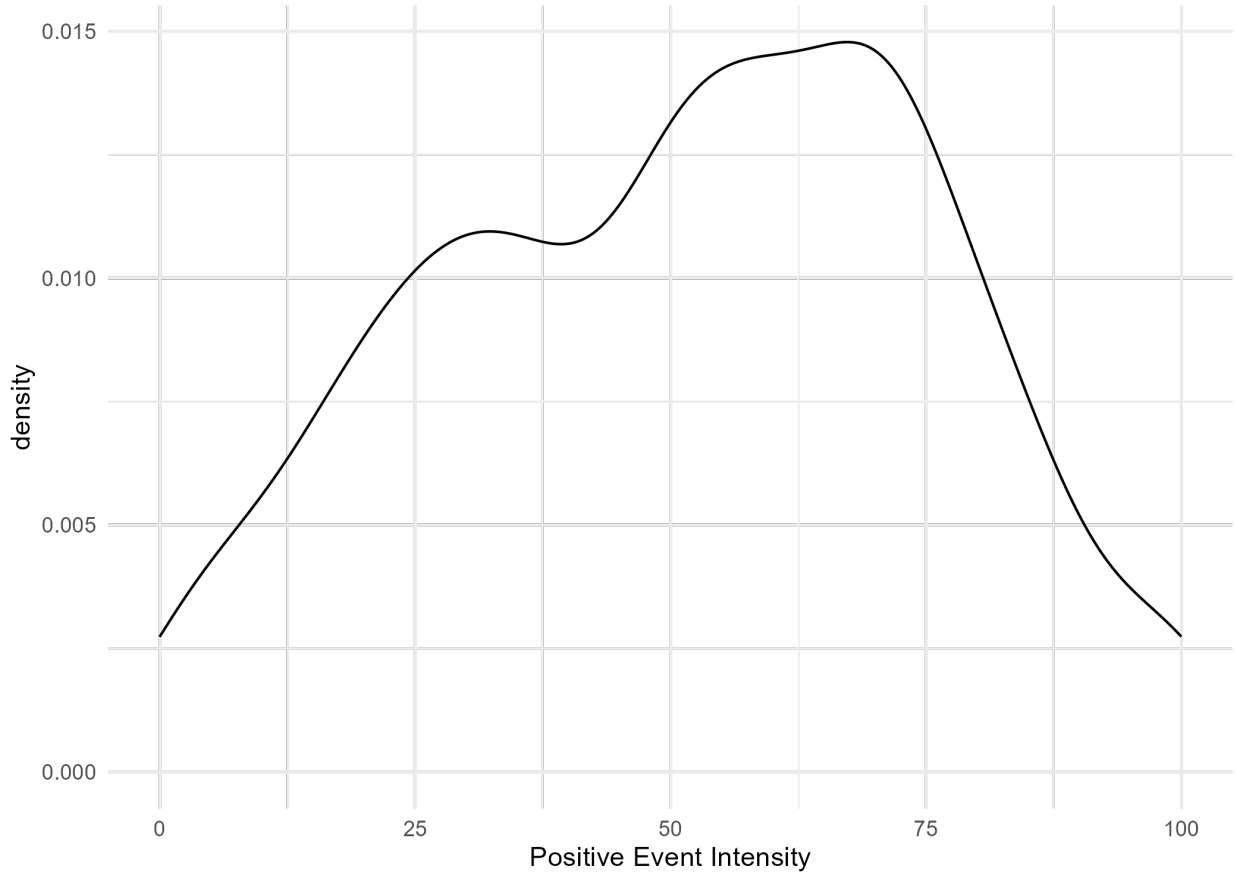
Event Models

Positive Event Intensity

Visual Inspection. As mentioned, the positive event variable was also slightly bi-modally distributed Figure D17, so we inspected the linearity assumption.

Figure D17

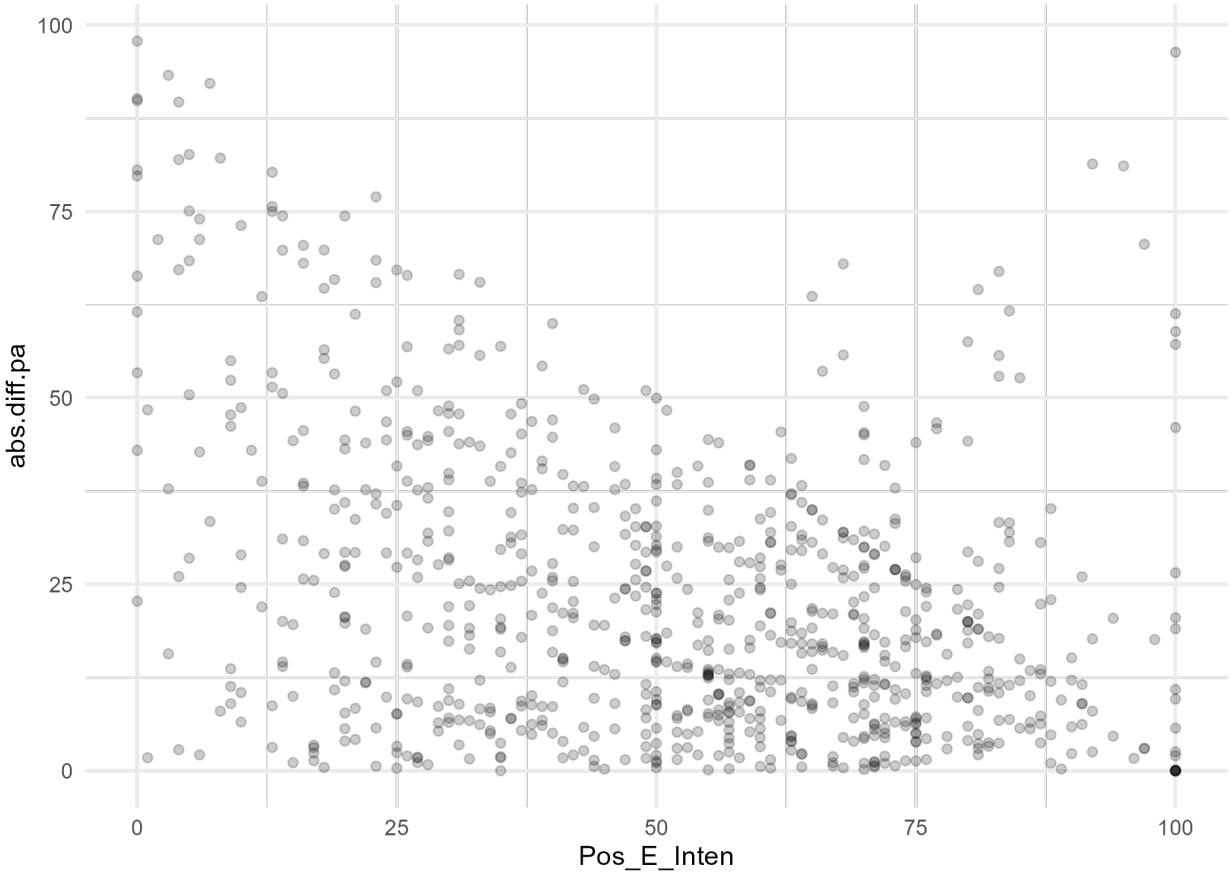
Density of Positive Event Intensity



We explored the relationship between positive event intensity and divergence across all observations Figure D18. There was an overall linear pattern.

Figure D18

Scatter Plot of Positive Event Intensity Against Divergence



To look at potential individual-level heterogeneity, we examined individual-level plots for this model. We modeled the relation between centered positive event intensity and divergence per person. The scatter plots revealed moderate consistency in linear trends across individuals (Figure D19 and Figure D20). Many plots exhibited a negative linear relationship, indicating that higher-than-usual positive event intensity was associated with lower divergence in positive affect. The overall consistency in direction and form supports the use of a linear mixed-effects model for this predictor.

Figure D19

Individual Scatter Plots of Positive Event and Divergence for all Participants

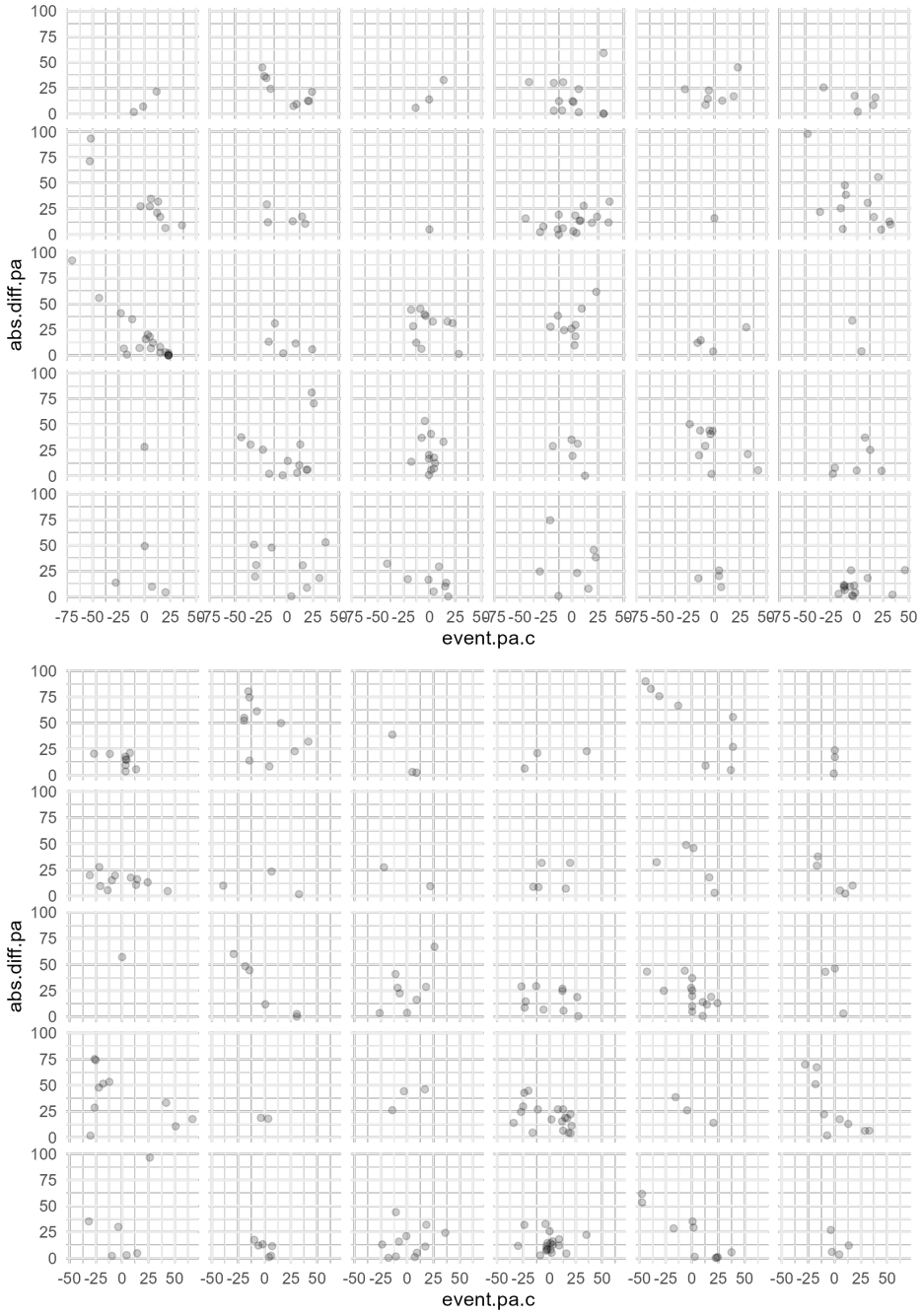
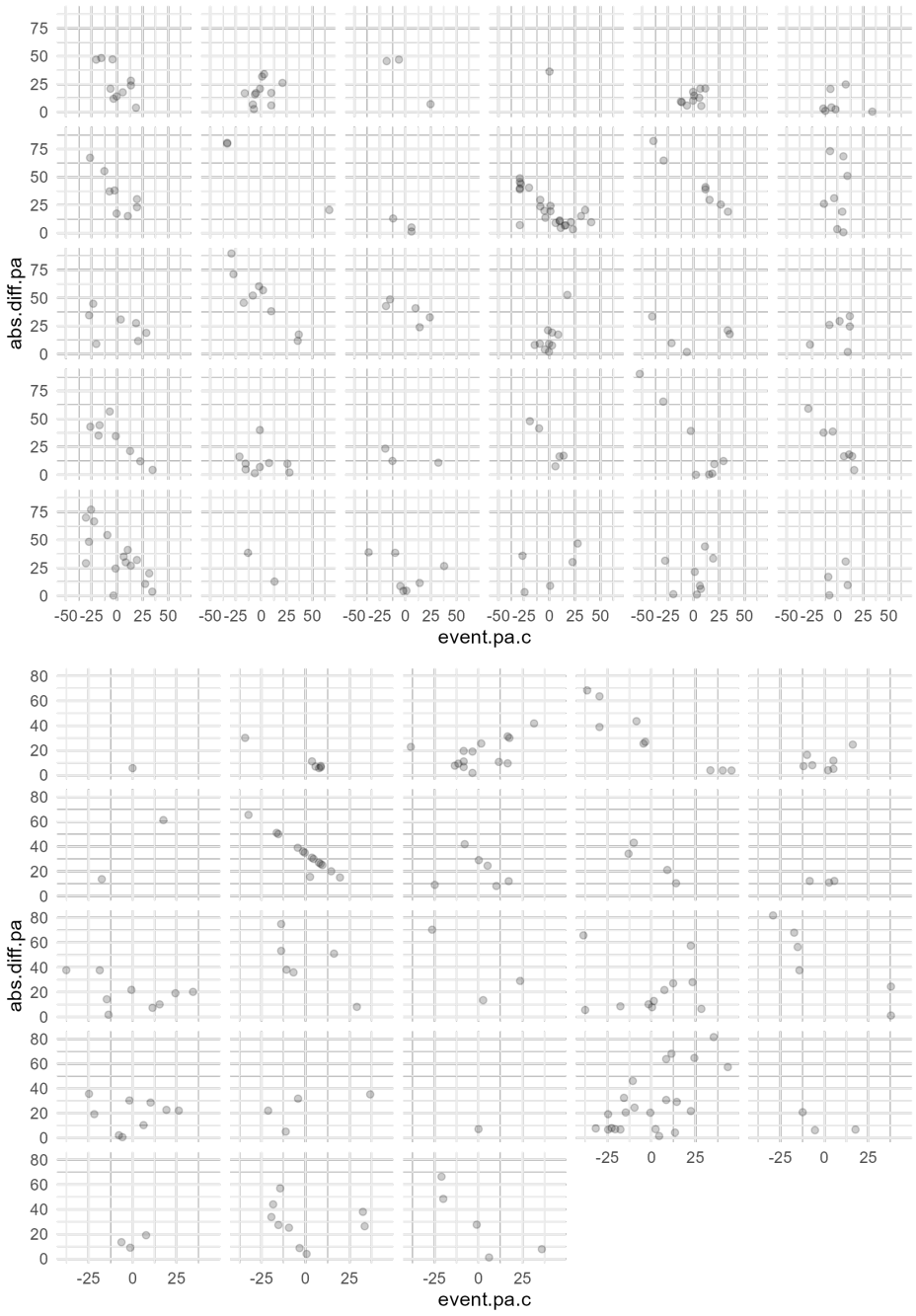


Figure D20

Individual Scatter Plots of Positive Event and Divergence for all Participants

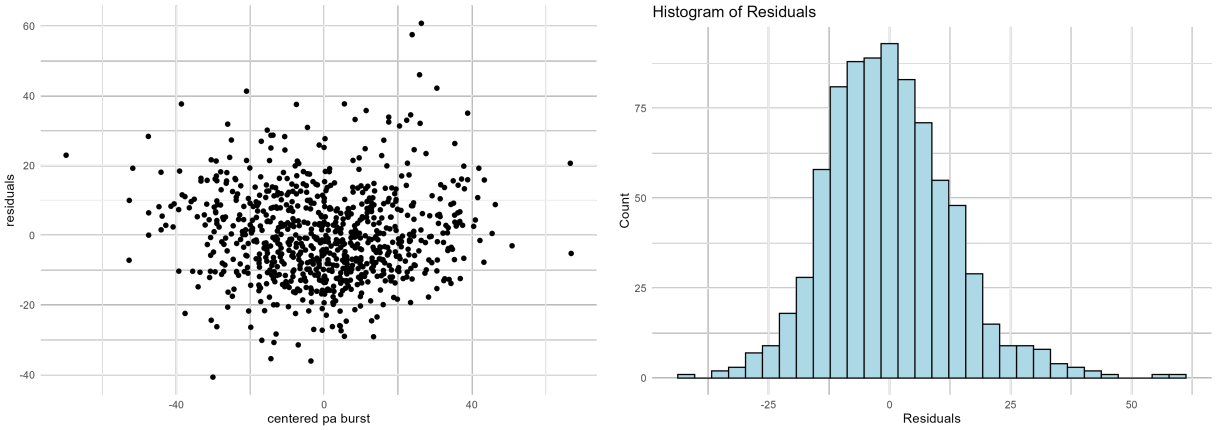


Residual Diagnostics. We fitted a model examining the relationship between centered positive events and divergence. The residual plot Figure D21 showed a wide spread of values centered around zero. There was no clear linear trend, which motivates

the consideration of referring to the non-linear models (i.e., quartile-based) for more robust evidence.

Figure D21

Residual Plot of Centered Positive Event Intensity on Divergence



Negative Event Intensity

Visual Inspection. The overall distribution of Negative Event intensity is also slightly bimodal Figure D22. A scatter plot of all observations explored the relationship between negative event intensity and divergence across all observations Figure D23. The distribution of points suggests a linear positive association, where higher-intensity negative events may be linked to greater divergence in affective responses.

Figure D22

Density of Negative Event Intensity

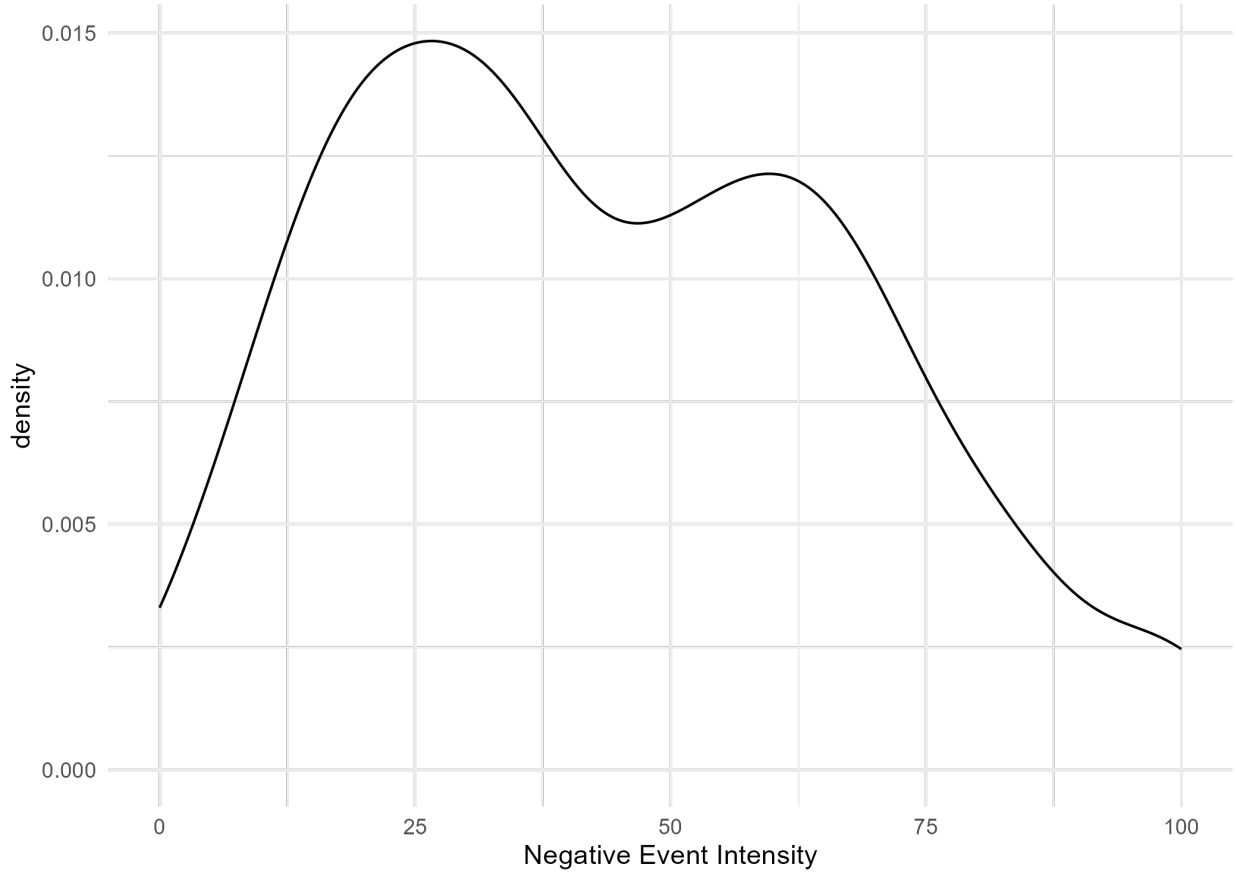
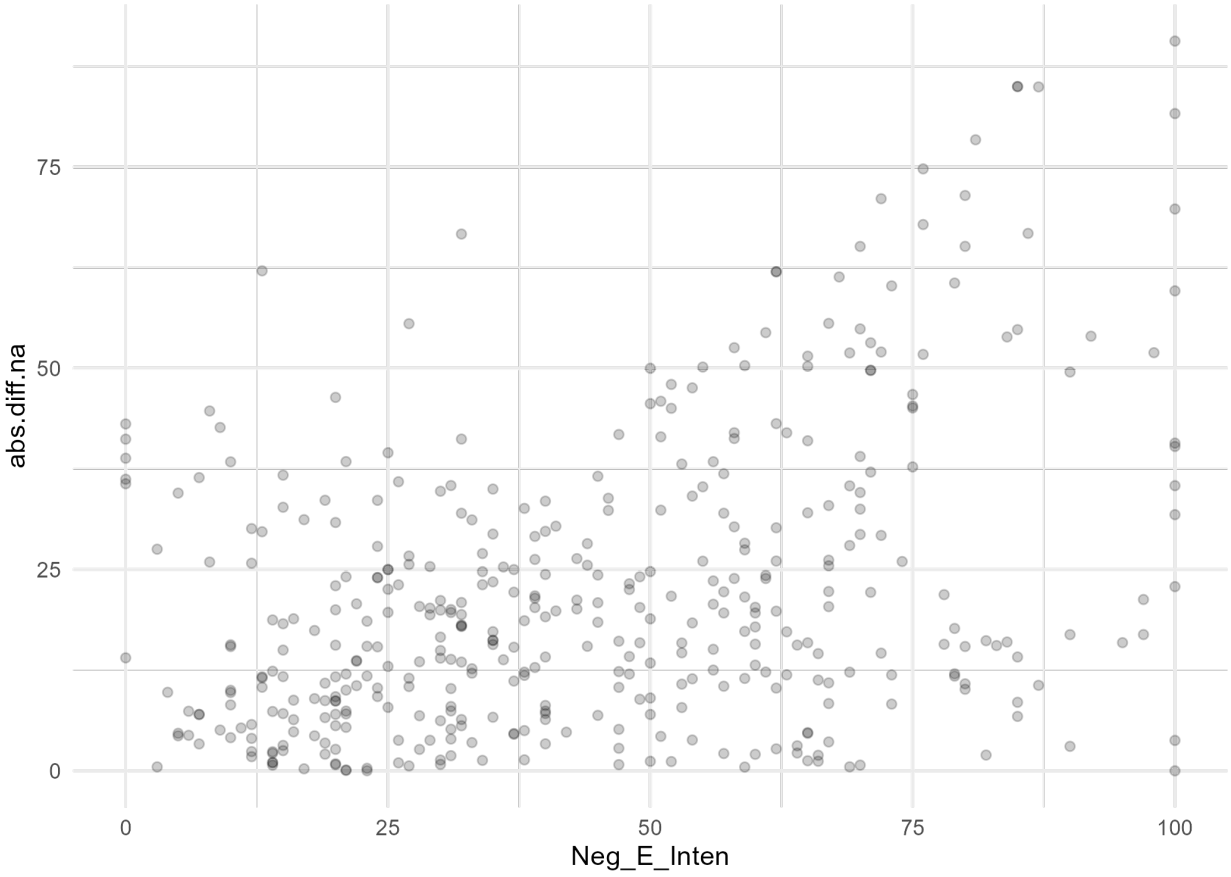


Figure D23

Scatter Plot of Negative Event Intensity Against Divergence



To further explore the relationship between centered negative event intensity and divergence at the individual level, we plotted the scatter plots for this relation for all participants. The scatter plots revealed moderate consistency in linear trends across individuals. Many participants showed a positive linear association (Figure D24 and Figure D25). Overall, the consistency of the positive trend across participants supports the use of a linear mixed-effects model for this predictor, with random slopes to account for individual differences.

Figure D24

Individual Scatter Plots of Negative Event and Divergence for all Participants

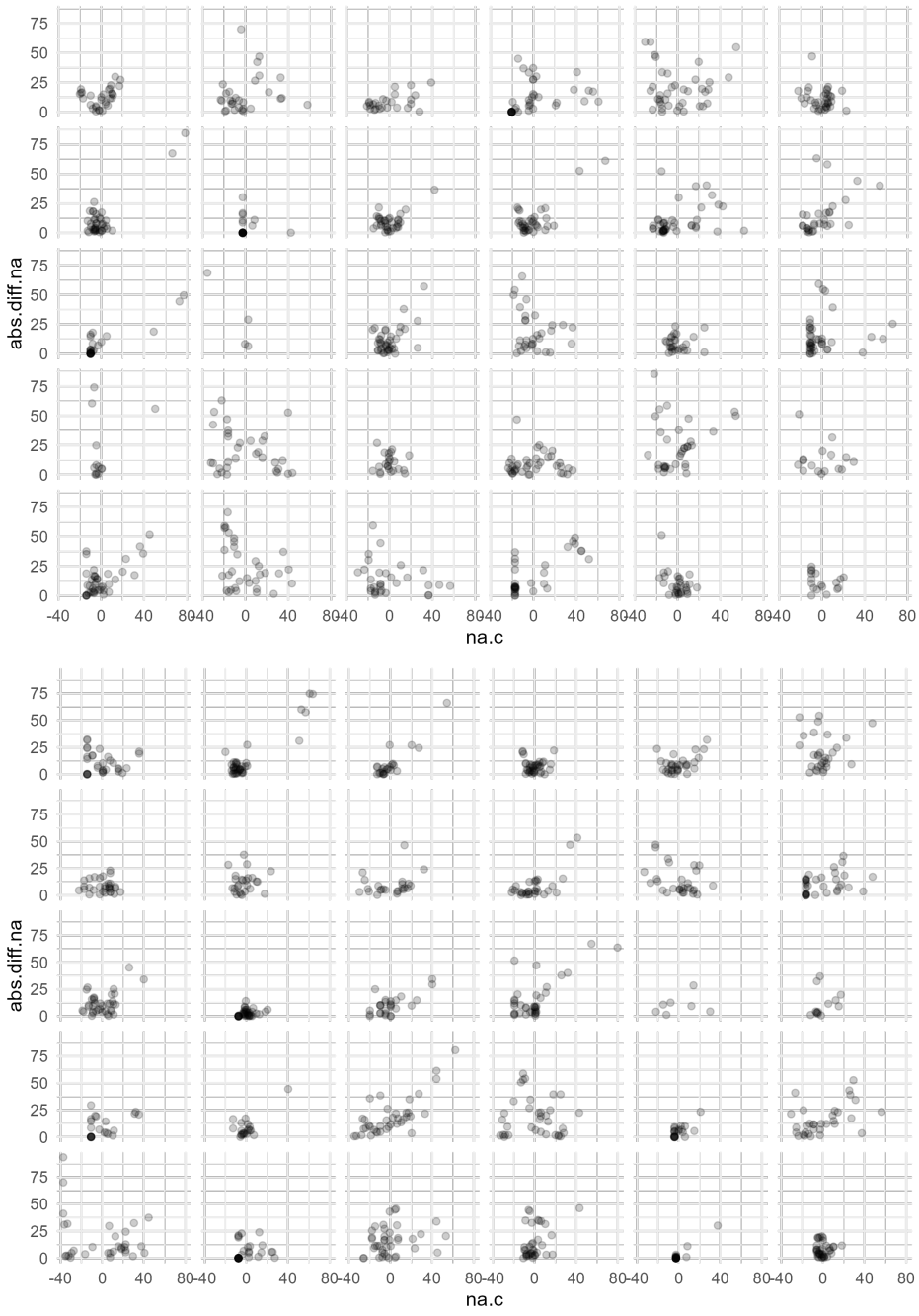
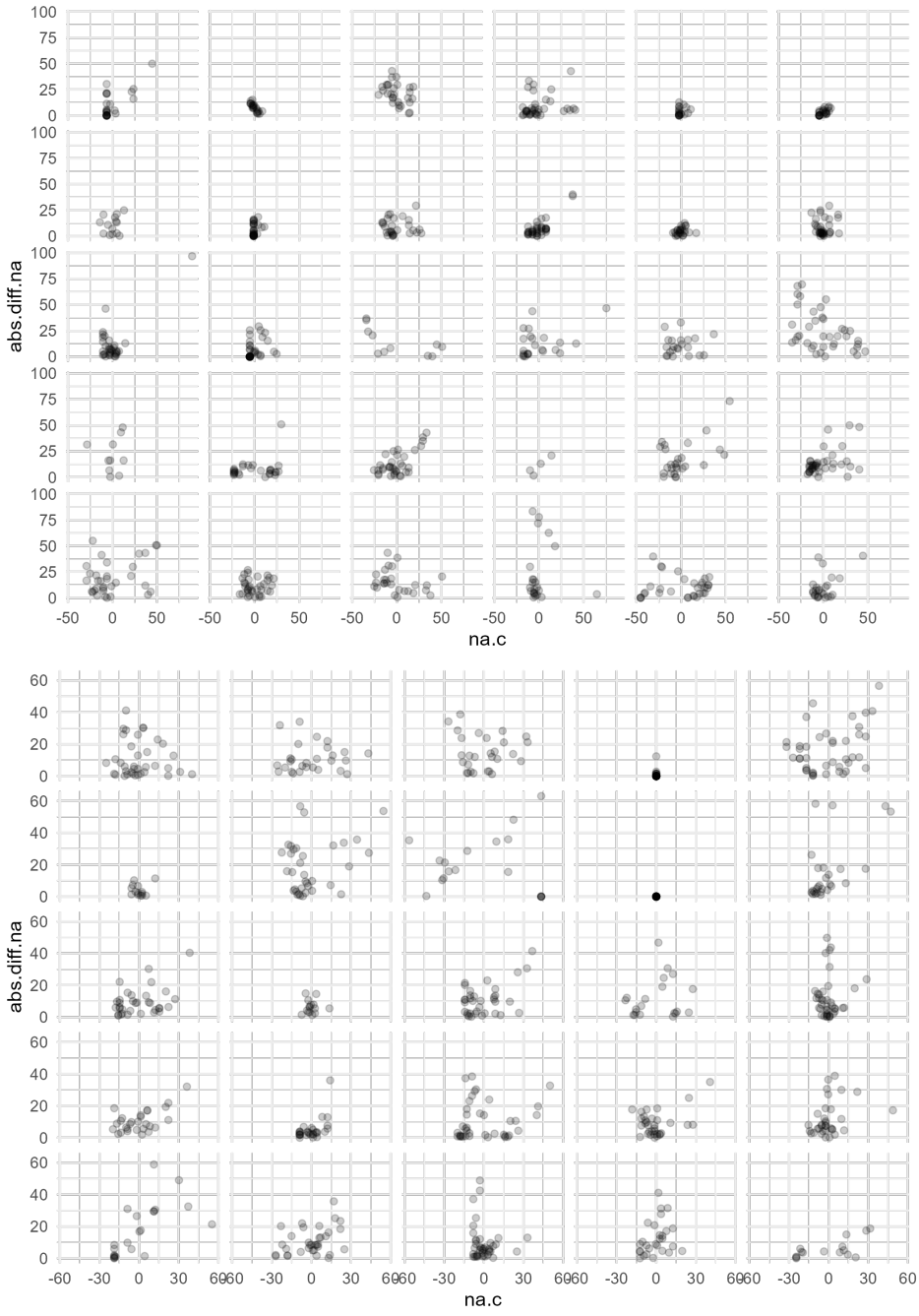


Figure D25

Individual Scatter Plots of Negative Event and Divergence for all Participants

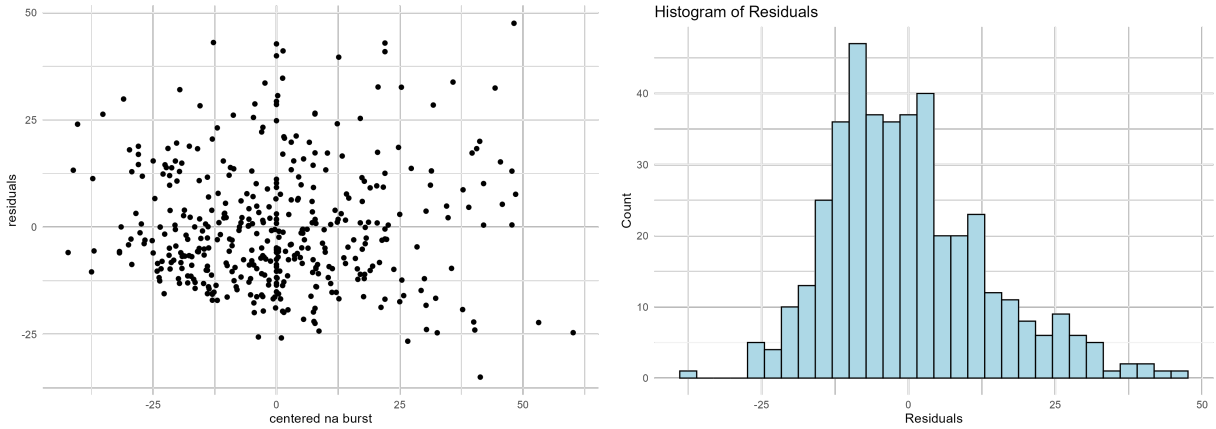


Residual Diagnostics. We examined the residuals from the fitted model, examining the relationship between centered negative event intensity and divergence. The residual plot Figure D26 showed a scattered distribution of residuals across the range of

the predictor, with no clear pattern or systematic trend. The residuals were generally centered around zero, and their spread appeared relatively uniform across the x-axis, suggesting that the assumptions of linearity and homoscedasticity were reasonably met. There was very slight funneling towards the positive end of the variable but the linear centered predictor results meet the assumptions of linearity and normality.

Figure D26

Residual Plot of Centered Negative Event Intensity on Divergence

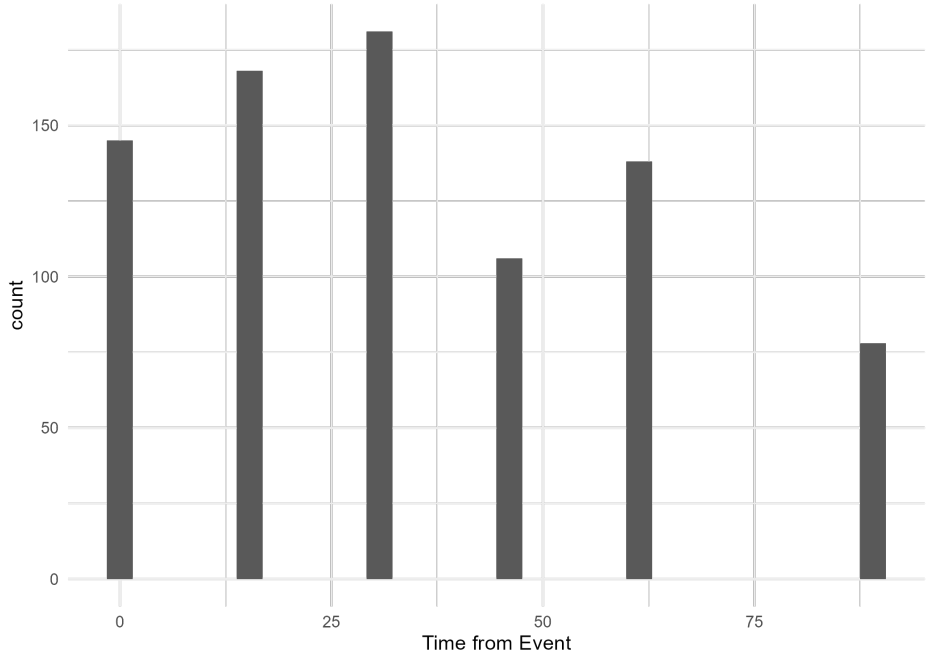


Time (Between Event & Drawing)

Visual Inspection. The variable time was treated as a continuous predictor, representing the time since the event. However, it is important to note that in practice, this variable contained only six distinct values that were given to participants to report how long ago the event occurred Figure D27.

Figure D27

Histogram of Positive Event Intensity



A scatter plot of all observations was used to examine the linear relationship between time and divergence in positive and negative affect Figure D28. The overall distribution of points did not suggest a strong linear trend. Divergence values were widely distributed across all time points, and no consistent upward or downward slope was observed.

Although the variable is treated as continuous, it contains only six distinct values, which limits its granularity. Nevertheless, the individual plots show a substantial linearity in the relationship. Within individuals, the time values are distributed so that linear patterns emerge. However, the scatter plots showed substantial variability across individuals. While some participants showed a slight negative trend, suggesting decreasing divergence over time, others exhibited flat or inconsistent patterns (Figure D29 to Figure D32).

Figure D28

Scatter Plots of Positive and Negative Event Divergence and Time

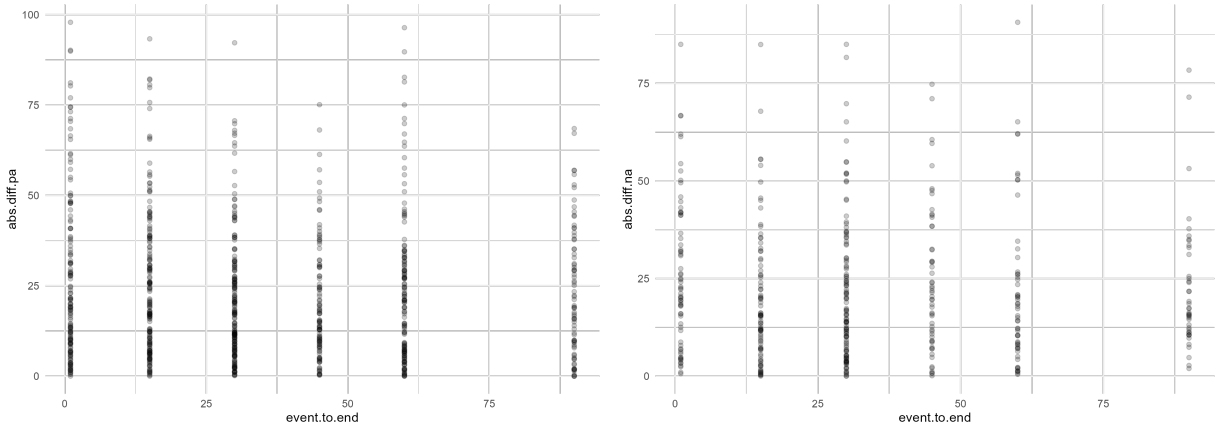


Figure D29

Individual Scatter Plots of Time and Divergence (Positive) for all Participants

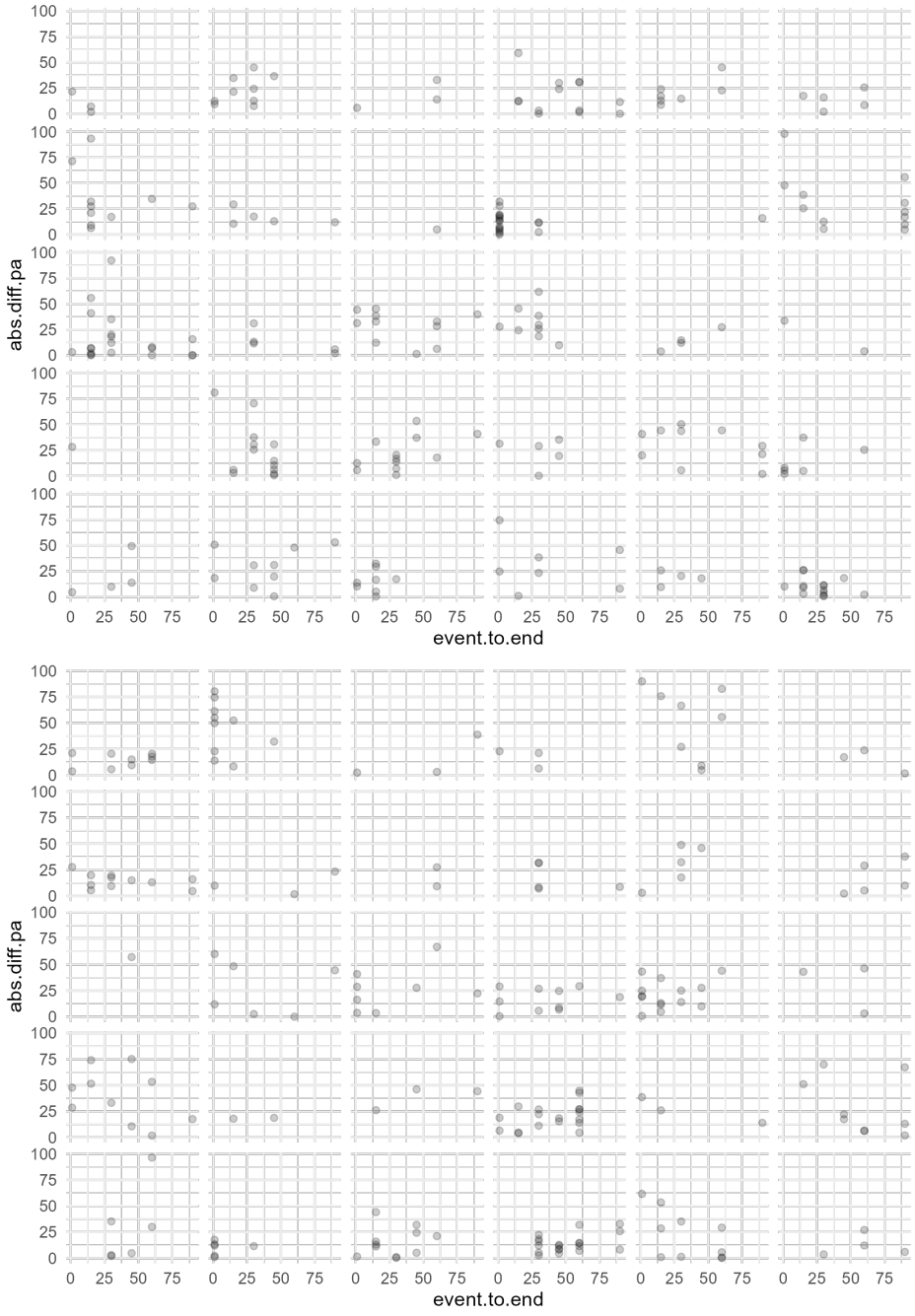


Figure D30

Individual Scatter Plots of Time and Divergence (Positive) for all Participants (continued)

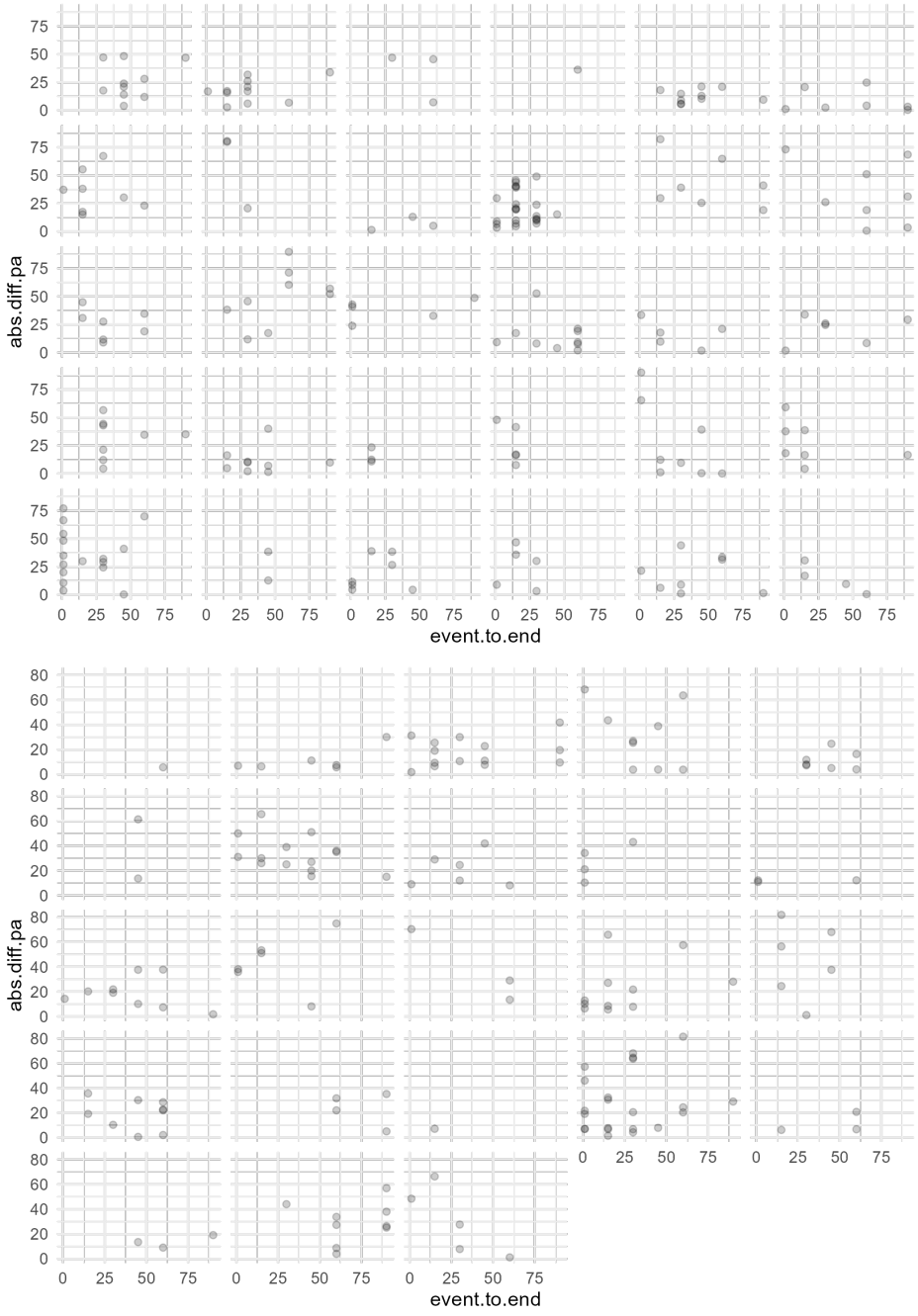


Figure D31

Individual Scatter Plots of Time and Divergence (Negative) for all Participants

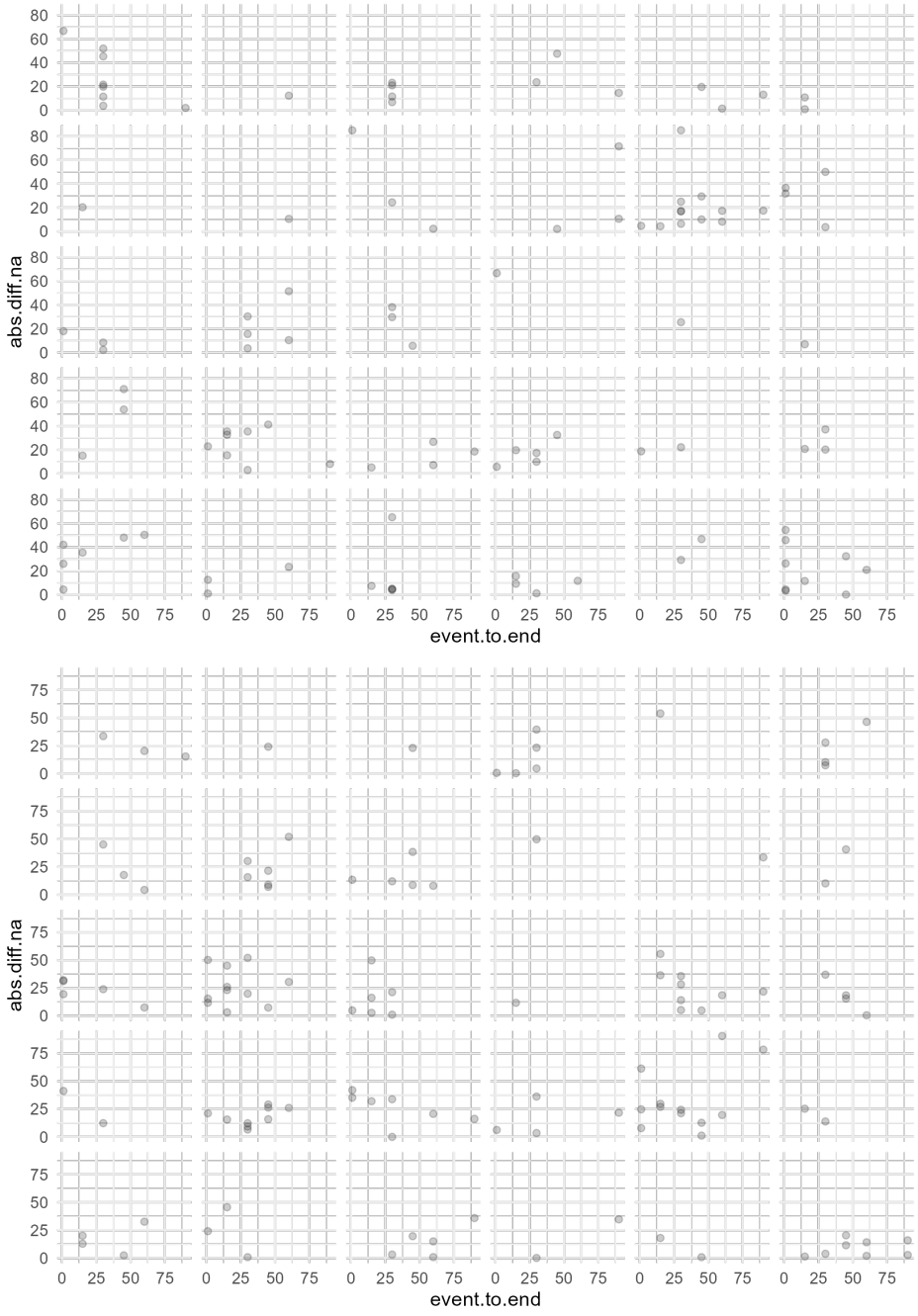
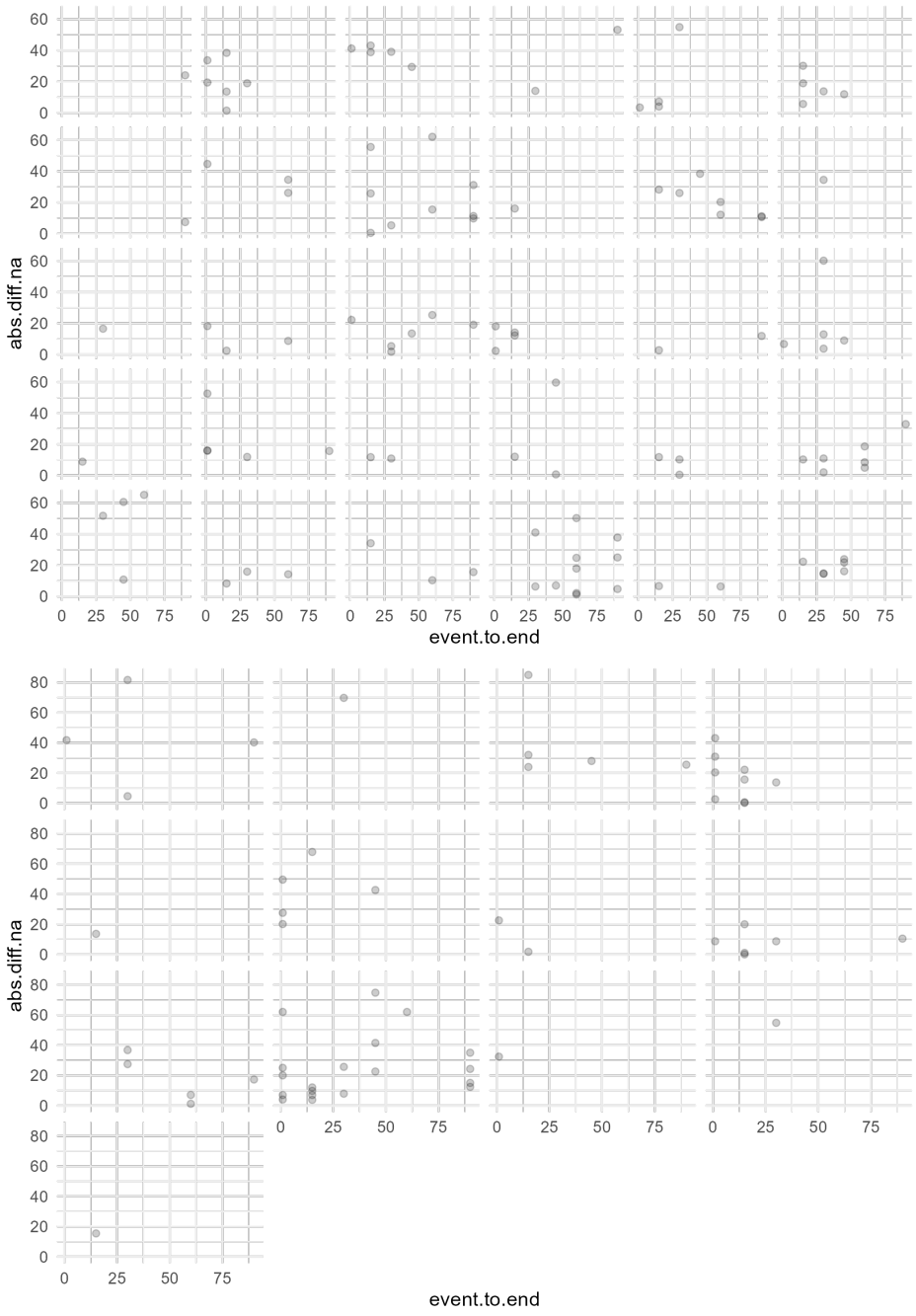


Figure D32

Individual Scatter Plots of Time and Divergence (Negative) for all Participants (continued)



Residual Diagnostics. To assess the assumptions of the linear mixed-effects model predicting divergence in positive events and negative events from time, we examined the residuals from the linear mixed-effects models predicting divergence in positive and

negative affect from time (Figure D33 and Figure D34). Overall, the residual structure supports the use of a linear model for time, particularly when random slopes are included to account for individual variability.

Figure D33

Residual Plot of Time on Divergence (Positive Event Centered)

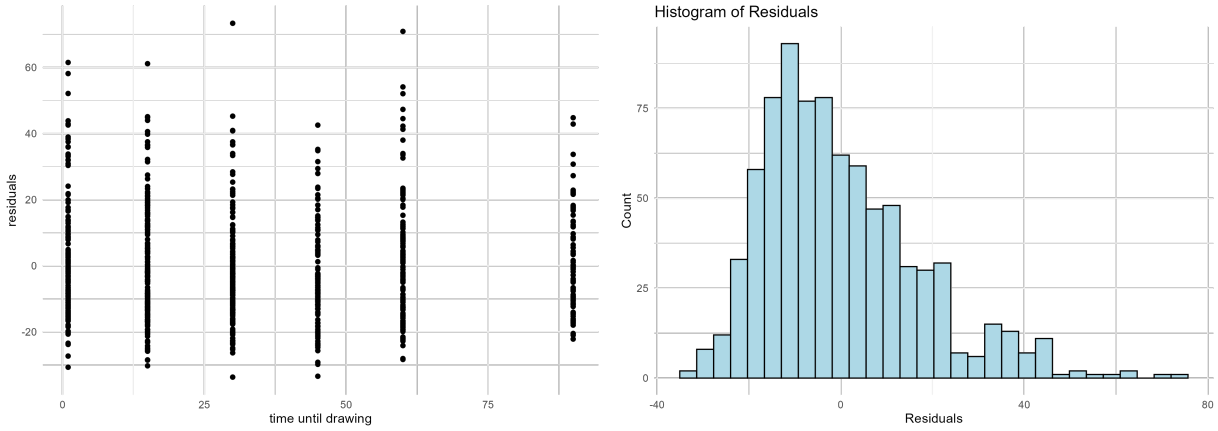
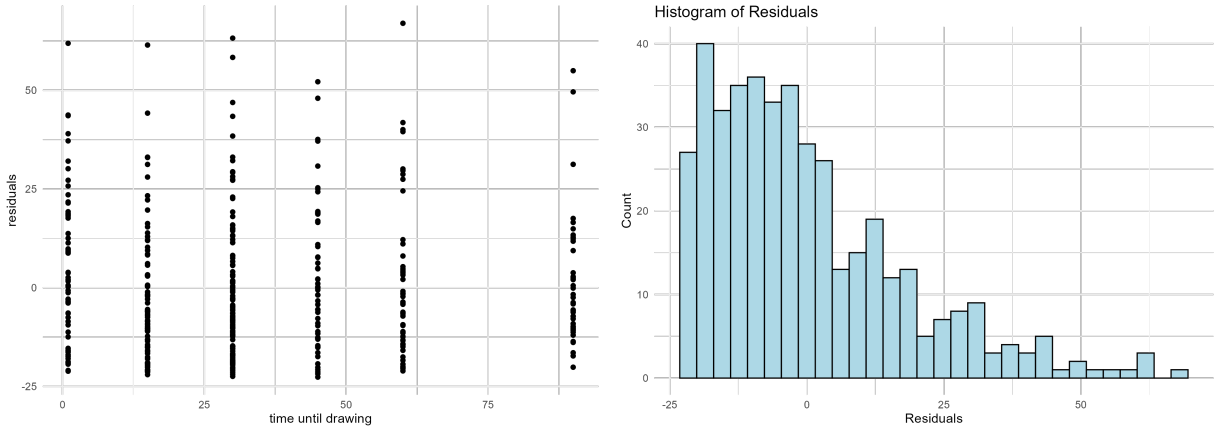


Figure D34

Residual Plot of Time on Divergence (Negative Event Centered)



Supplement E

Models Using Person-Mean Centered Predictors

The originally planned analytic strategy included the use of continuous person-mean-centered variables as predictors. Due to the non-linear associations, we report and interpret the quartile-based ordered factor models in the main manuscript. For completeness and transparency, we report the linear models here.

General Model Specification

For models using person-mean centered predictors, divergence was regressed on momentary affect or event intensity (person-mean centered) and the time elapsed since the momentary rating or event. All models included fixed effects of *intensity* (β_1), *time* (β_2), and their interaction (β_3). Random intercepts and random slopes for intensity, time, and their interaction were included at the person level to account for individual differences, if model convergence allowed for it. Random-effects estimates are reported in Supplement F.

$$\begin{aligned} Divergence_{ij} = \beta_0 + \beta_1\, Intensity_{ij}^{(\mathrm{pc})} + \beta_2\, Time_{ij} \\ + \beta_3 \left(Intensity_{ij}^{(\mathrm{pc})} \times Time_{ij} \right) + u_{0i} + u_{1i}\, Intensity_{ij}^{(\mathrm{pc})} \\ + u_{2i}\, Time_{ij} + u_{3i} \left(Intensity_{ij}^{(\mathrm{pc})} \times Time_{ij} \right) + \epsilon_{ij}. \end{aligned} \tag{E1}$$

Model 1*. Positive Affect Intensity (Person-Mean Centered)

We estimated mixed-effects models including person-mean centered positive affect intensity and time since the momentary rating to predict divergence. We then extended this model by including the two-way interactions among these predictors (Table E1).

Momentary positive affect intensity was significantly associated with lower divergence, $\beta_1 = -0.10$; $p = .031$. Time since the momentary rating was significantly positively associated with divergence, $\beta_2 = 0.02$; $p = .001$, suggesting that retrospective reports tended to diverge more from momentary reports as the day progressed. The interaction between affect intensity and time to end was not significant, $p = .348$.

Table E1

Nested Models of Person-Centred Positive Affect Intensity and Time

	Main Effects		Two-Way Interaction	
Effect	Estimate	SE	Estimate	SE
Intercept	11.209***	0.520	11.080***	0.513
Positive Affect	-0.075*	0.034	-0.095*	0.044
Time	0.016**	0.005	0.017***	0.005
Positive Affect:Time			0	0
R^2	28.20%		28.76%	

Note: p < .05 *; p < .01 **; p < .001 ***; SE = Standard Error

In an additional analysis, we explored whether high and low levels of positive intensity had a different effect on divergence. Due to the bimodal split in the data, we hypothesized that participants interpret the positive scale as bimodal, with a neutral interpretation of the midpoint. This motivated additional analyses that explicitly accounted for the apparent bimodal structure of the affect scale.

Model 1**. Affect Intensity with Direction (Bimodal)

To examine whether the association between affect intensity and divergence differed for values below versus above the scale midpoint, we estimated models that explicitly accounted for the bimodal structure of the positive affect scale. Accordingly, positive affect intensity scores were centered around the midpoint (50): first, we classified them into a categorical predictor indicating *direction* (below 50 => negative direction (0); above 50 => positive direction (1)). Then we computed the absolute distance from the midpoint and subsequently person-mean centered these scores, labelling this variable *affect intensity*. Divergence was regressed on person-mean centered bimodal affect intensity, direction, and time since the momentary rating. Three nested models were estimated: a main-effects model, a model including all two-way interactions, and a model including the three-way interaction (Table E2). Across these models, the estimated effects remained stable in

magnitude and significance. Incorporating the two-way interactions markedly improved the model's explained variance, whereas the addition of the three-way interaction contributed only a modest further increase.

Table E2

Nested Models of Affect Intensity, Direction, and Time

	Main		Two-Way		Three-Way	
Effect	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	15.628***	1.266	14.895***	1.179	15.180***	1.177
Affect Intensity	0.166***	0.036	0.290***	0.073	0.413***	0.080
Direction	-3.554*	1.509	-3.198*	1.470	-3.763**	1.456
Time	0.016**	0.005	0.024**	0.008	0.023**	0.008
Affect:Direction			-0.202**	0.072	-0.468***	0.097
Affect:Time			0.001	0.001	-0.001	0.001
Direction:Time			-0.002	0.011	0.003	0.011
Affect:Direction:Time					0.004***	0.001
R^2	45.12%		49.05%		50.12%	

Note: p < .05 *; p < .01 **; p < .001 ***; SE = Standard Error

The intercept represents the estimated mean divergence for the reference category (negative direction) when bimodal affect intensity and time were at their person-mean centered values ($\beta_0 = 15.18$; $p < .001$). The main effect of direction was negative and significant ($\beta_2 = -3.76$; $p = .01$), indicating lower average divergence for the positive direction compared to the negative direction.

For the reference category negative direction, affect intensity was positively related to divergence ($\beta_1 = 0.41$; $p < .001$). The two-way interaction between affect intensity and *direction* represents the difference between the slopes of the negative direction and the positive direction. The slopes were significantly different from each other ($\beta_4 = -0.47$; $p < .001$), indicating that, for time point zero, the relation between momentary affect intensity and divergence was significantly smaller in magnitude for the positive direction than for

the negative direction. The predicted slope for the positive direction for time point zero, given by $\beta_1 + \beta_4 = -0.06$, indicates a smaller negative relationship between momentary affect intensity and divergence in the positive direction than in the negative direction.

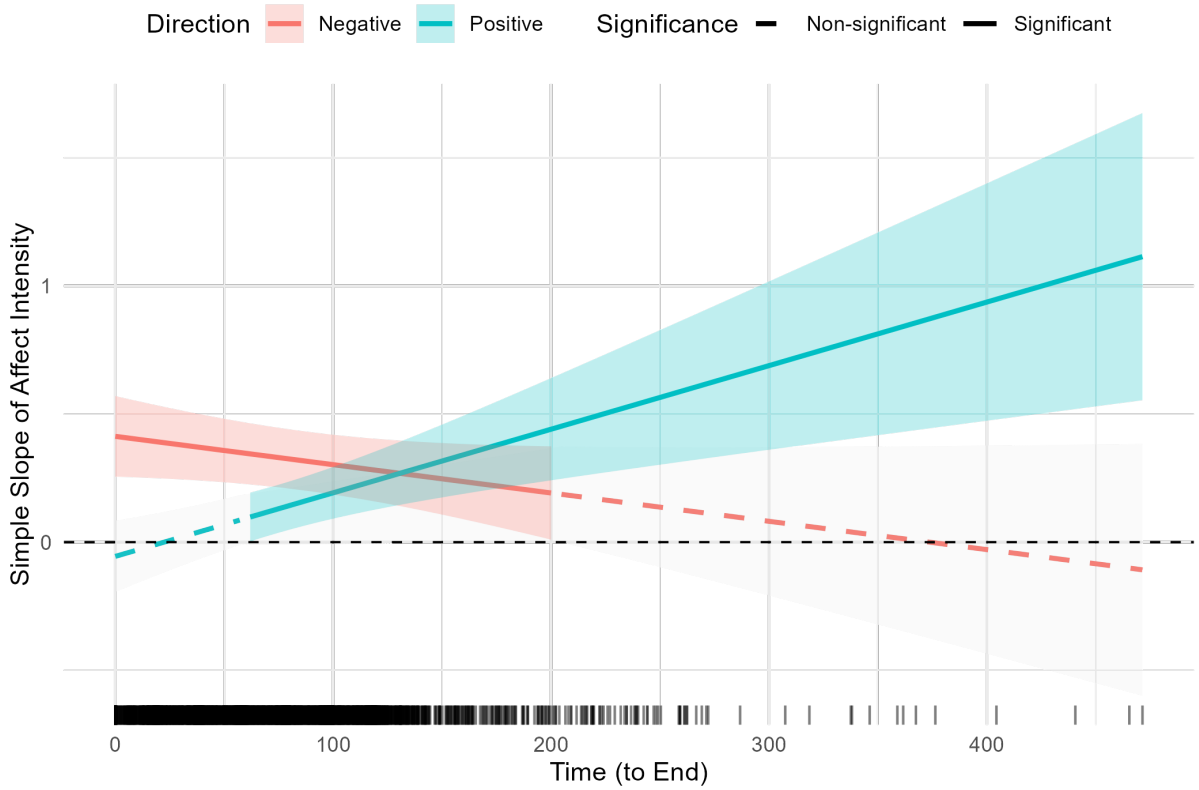
Time elapsed since the momentary rating was a significant predictor of divergence ($\beta_3 = 0.03$; $p = .006$) for the reference category (negative direction). This indicates that, when affect intensity is at the within-person average in the negative direction (as the variable is centered), greater *time* elapsed since the momentary rating is associated with increased divergence.

There was no interaction between affect intensity and time, indicating that the slope of the affective intensity did not change significantly for different scores of time. Moreover, the interaction between direction and time was not significant, indicating that the relation between time and divergence did not differ for the two directions. However, there was a significant positive three-way interaction between affect intensity, direction, and time ($\beta_7 = 0.004$; $p < .001$). This indicates that the interaction between affect intensity and time differed significantly by direction.

To examine how the relationship between affect intensity and divergence varied as a function of time and direction, we computed the simple slopes of affect intensity on divergence across values of time, separately for each direction. We estimated the 95% CIs, identifying regions where the slopes were significantly different from zero ($p < .05$). We visualized this in a Johnson-Neyman plot, depicting how the relationship between affect intensity and divergence varied as a function of time for each direction.

Figure E1

Simple Slopes of Affect Intensity by Direction and Time



Note: Johnson-Neyman plot illustrating how the linear effect of bimodal affect intensity on divergence varies as a function of time, and direction, with 95% confidence bands. The red line indicates the slope of affect intensity for the negative direction, the blue line the slope for the positive direction. The shaded region represents the time interval during which the effect is statistically significant ($p < .05$). The observed range of time was [0, 471.42] minutes. The distribution of the variable time is given on the x-axis.

For the negative direction, the slope of affect intensity on divergence was significantly negative when time between the momentary rating and continuous drawing was less than 200 minutes (see Figure E1, red line). This suggests that the closer in time the momentary rating was to the drawing, the stronger the negative relationship between affect intensity and divergence. In other words, when participants rated the continuous drawing shortly after they reported affect intensity in the negative direction, the intensity

of the rating increased divergence. However, as more time passed, the influence of affect intensity on divergence diminished and became non-significant.

In contrast, for the positive direction, the slope of affect intensity on divergence was significantly positive when the momentary rating was longer than 62 minutes ago (see Figure E1, blue line). Notably, the slope was negative at time 0 and became positive and significant thereafter, as the magnitude of the effect increased over time. This suggests that over time the peak effect gets less salient and the drawings get increasingly discrepant from the momentary affect ratings, especially when the momentary rating was more intense.

There is a direction-dependent effect: in the negative direction, the relation between affect intensity and divergence is strongest when the drawing is made soon after the momentary rating and weakens over time. For the positive direction, the relationship between affect intensity and divergence is initially weak or even negative, but becomes stronger and more positive as time increases. This model shows that assuming a linear or undifferentiated effect of intensity masks important time- and direction-specific non-linear patterns. These findings motivated the use of quartile-based models in the main analyses, which more flexibly capture non-linear, time-dependent patterns without imposing linear or directionally homogeneous assumptions.

Model 2*. Negative Affect Intensity (Person-Mean Centered)

To examine the effect of centered negative affect intensity and time between the momentary and retrospective ratings on divergence, we estimated a main effects model and a model including the two-way interaction (see Table E3).

Table E3

Nested Models of Person-Centered Negative Affect Intensity and Time

	Main Effects		Two-Way Interaction	
Effect	Estimate	SE	Estimate	SE
Intercept	10.629***	0.531	10.599**	0.529
Negative Affect	0.175***	0.034	0.164**	0.052
Time	0.016**	0.005	0.015**	0.005
Negative Affect:Time			0	0
R^2	28.20%		28.76%	

Note: p < .05 *; p < .01 **; p < .001 ***; SE = Standard Error

Firstly, negative affect intensity was a significant positive predictor of divergence (β_1= 0.18; $p < .001$), indicating that increasing levels of negative affect intensity were associated with a higher divergence. Secondly, time was a significant positive predictor of divergence ($\beta_2 = 0.02$; $p = .002$), indicating that as the time since the momentary rating increased, divergence increased as more time passed. Finally, there was no significant interaction between negative affect intensity and time.

Model 3*. Positive Event Intensity (Person-Mean Centered)

The person-mean centered scores of positive event intensity and time since the event occurred, were regressed on divergence between the retrospective event ratings and the retrospective drawing score for the time of the event. We first modelled the main effects and then added the two-way interactions in a series of nested models (Table E4).

Positive event intensity was significantly and negatively associated with divergence (β_1 = -0.39 $p < .001$) at time point zero. This indicates that when participants

experienced more intense positive events, their event ratings are incorporated into the drawings.

Table E4

Nested Models of Person-Centred Positive Event Intensity and Time

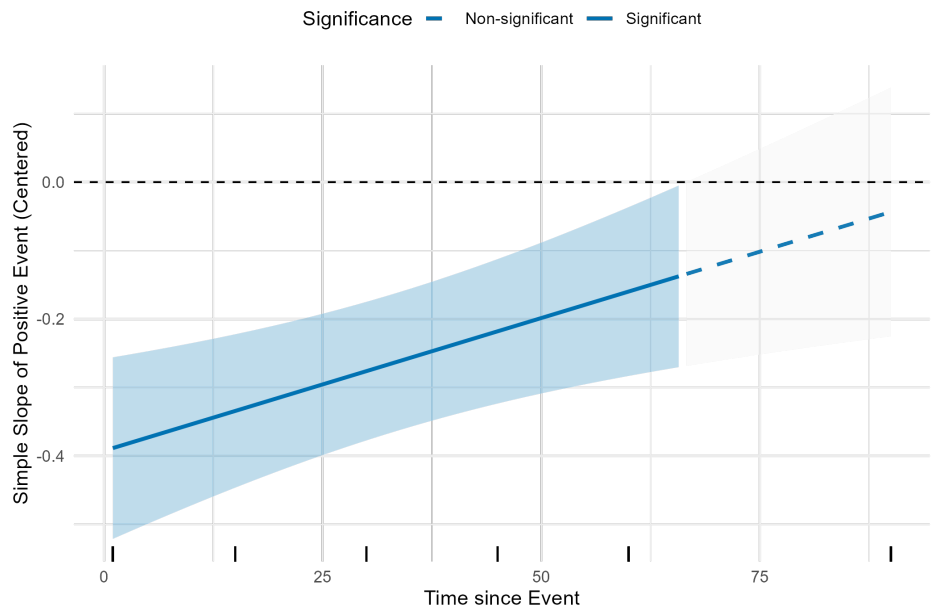
	Main		Two-Way	
	Estimate	SE	Estimate	SE
Intercept	23.217***	1.274	23.199***	1.241
Positive Event	-0.263***	0.051	-0.392***	0.069
Time	0.008	0.022	-0.001	0.022
Positive Event:Time			0.004**	0.001
R^2	42.96%		45.26%	

Note: p < .05 *; p < .01 **; p < .001 ***; SE = Standard Error

There was no significant main effect of time since the affective event on divergence. However, there was a significant interaction between positive event intensity and time (β_3 = 0.004; p = .004), indicating that the effect of positive event intensity on divergence varied depending on how much time had passed since the event. To interpret this interaction, we conducted a simple slopes analysis, estimating the slope of positive event intensity on divergence across a continuous range of time values. Since participants reported time in discrete categories (0, 15, 30, 45, 60, and 90 minutes), we interpolated slopes between these values to visualize the pattern in a Johnson-Neyman plot (see Figure E2). We estimated a 95% CI, which identifies the time range in which the slope is significantly different from zero. The plot includes a CI to test the significant region of the slope on the moderator time. The negative effect of positive event intensity on divergence was strongest immediately after the event, with higher-intensity events being associated with lower divergence when recalled soon after they occurred. As more time passed since the event, this negative relationship weakened, suggesting that the influence of positive event intensity on divergence diminishes over time.

Figure E2

Simple Slopes of Positive Event Intensity by Time



Note: Johnson-Neyman plot illustrating how the effect of positive event intensity on divergence varies as a function of time, with 95% confidence bands. The blue shaded region indicates the time window during which the effect is statistically significant ($p < .05$). Participants reported time since the event in [0, 15, 30, 45, 60, and 90] minutes. Slopes were interpolated between observations.

Model 4*. Negative Event Intensity (Person-Mean Centered)

For negative events, we investigated the effects of centered negative event intensity and time on the divergence between the negative event rating and the negative continuous drawing rating at the time of the event. The nested models included the main effects only and the interaction between negative event and time (Table E5).

Table E5

Nested Models of Person-Centered Negative Event Intensity and Time

	Main		Two-Way	
	Estimate	SE	Estimate	SE
Intercept	23.415***	1.637	23.423***	1.625
Negative Event	0.310***	0.06	0.300***	0.085
Time	-0.04	0.035	-0.041	0.034
Negative Event:Time			0	0.002
R^2	35.00%		34.72%	

Note: p < .05 *; p < .01 **; p < .001 ***; SE = Standard Error

We found a significant positive relationship between negative event intensity and divergence ($\beta_1 = 0.30$; $p < 0.001$), indicating that as the negative event intensity increases, the divergence also increases. We found no significant main effect of *time*, nor was there a significant interaction effect between negative event intensity and *time* on divergence.

Supplement F

Random Effects

All models included random intercepts and random slopes for each predictor; however, the inclusion of random interaction effects varied by model due to convergence issues (see Table F1 for an overview of the random effects structure for the momentary affect models Table F7 provides an overview of the random effects structure across the event intensity models.)

Table F1

Random Effects Structure for All Affect Models

Random Effects Structure	Positive Affect (Quartiles) Model 1		Positive Affect (Centered) Model 1*		Affect Intensity (Bimodal) Model 1**			Negative Affect (Quartiles) Model 2		Negative Affect (Centered) Model 2*	
	Main Effects	2-Way Interactions	Main Effects	2-Way Interactions	Main Effects	2-Way Interactions	3-Way Interactions	Main Effects	2-Way Interactions	Main Effects	2-Way Interactions
Random Intercepts	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Slopes	All	All	All	All	All	All	All	All	All	All	All
Random Interaction Effects	–	Only main	–	All	–	All	Only 2-way	–	Only main	–	All

Note: Overview of the random effects structure included in each model. All models included random intercepts and random slopes for each predictor. Random interaction effects were added when convergence allowed. A dash (–) indicates that no interaction terms were modeled for that condition. "Only main" and "Only 2-way" indicate convergence issues for higher-order random interaction effects.

Model 1. Positive Affect Intensity (Quartiles)

The positive affect ordered factor model incorporated random slopes for linear, quadratic, and cubic trends, along with the intercept and time, to capture individual-level variability in physical activity trajectories. Table F2 shows the variance and correlation structure among these random effects. The model shows decreasing variances across the linear, quadratic, and cubic contrasts, suggesting diminishing between-person variability in higher-order time effects. Strong correlations were observed between the intercept and quadratic term (r = 0.60), and between linear and quadratic terms (r = -0.76), indicating potential collinearity in the random effects structure.

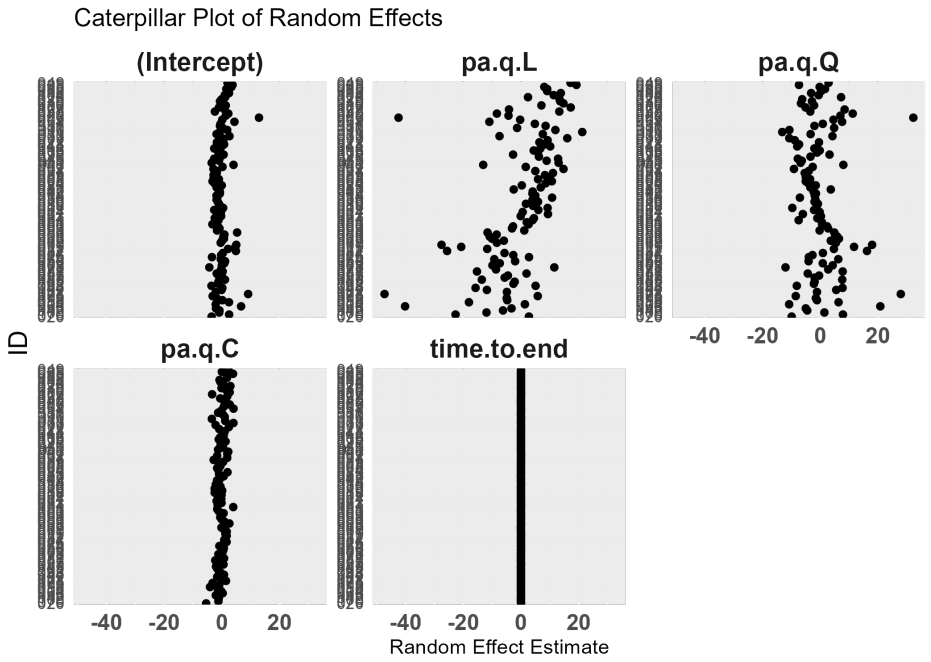
Table F2

Variance and Correlations of Random Effects in Positive Affect Quartile Model

	Variance	Correlations			
Term		**Intercept**	**Linear**	**Quadratic**	**Cubic**
Intercept	14.386				
Linear	189.209	-0.472			
Quadratic	76.625	0.604	-0.759		
Cubic	9.466	0.251	-0.085	0.206	
Time	0.001	-0.168	0.131	-0.127	0.043
Residual	119.885				

Figure F1

Caterpillar Plots of Random Effects (Positive Affect Quartiles: Polynomial Contrasts and Time)



The caterpillar plots (Figure F1) illustrate the distribution of random slopes across individuals. Substantial variability is evident in the linear and quadratic slopes, suggesting

meaningful individual differences in these components of the positive affect trajectory. In contrast, the intercept, cubic term, and time effect show minimal variation, indicating more consistent patterns across individuals for these parameters. These findings support the inclusion of random slopes for the linear and quadratic terms in the final model.

Model 1*. Positive Affect Intensity (Person-Mean Centered)

The linear model with the centered positive affect scores included random effects for the intercept, the slope of positive affect intensity, time and the interaction. Table F3 shows the random effects structure. There is between-person variance in the intercept and the within-person association between positive affect intensity and divergence, even though the predictor was centered. The correlations between the random effects were small, and given the low variance in the interaction slopes, the correlations are unsuited for interpretation.

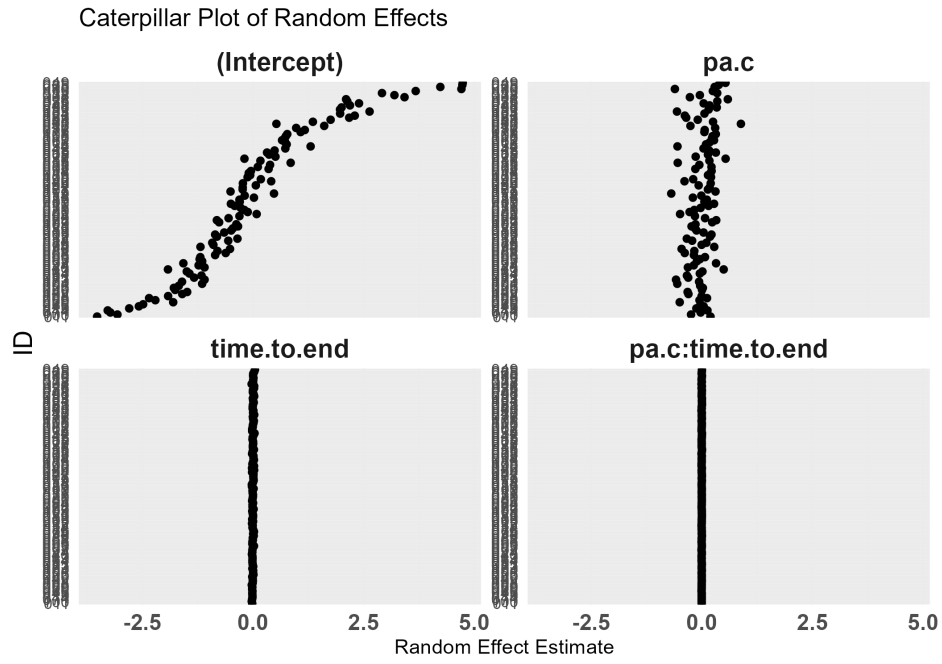
Table F3

Variance and Correlations of Random Effects in Person-Centered Positive Affect Model

	Variance	**Correlations**		
Term		**Intercept**	**Positive Affect**	**Time**
Intercept	7.926			
Positive Affect	0.131	0.056		
Time	0.001	-0.081	0.067	
Positive Affect:Time	0	0.135	-0.529	0.202
Residual	121.389			

Figure F2

Caterpillar Plots of Random Effects (Positive Affect Centered, Time and Interaction)



The caterpillar plots Figure F2 present the participant's random effect estimates plotted from largest to smallest. There was large heterogeneity in the intercept, whereas the other slopes show little variance.

*Model 1**. Affect Intensity with Direction (Bimodal)*

The multilevel model included random slopes for the main effects and two-way interactions to account for individual-level variability. However, the model including the random effects of the three-way interaction failed to converge, suggesting overparameterization or insufficient data to estimate such complexity. Table F4 presents the variance of and correlations among the random effects. There is substantial between-person variance in the slope of direction, and variance in the residuals, suggesting that considerable within-person variability remains unexplained by the model. The correlations indicate substantial interdependence between the intercept and predictors such as direction (r = -0.92) and intensity (r = 0.72), which may contribute to convergence issues.

Table F4

Variance and Correlation of Random Effects in Affect Intensity (Bimodal) Model

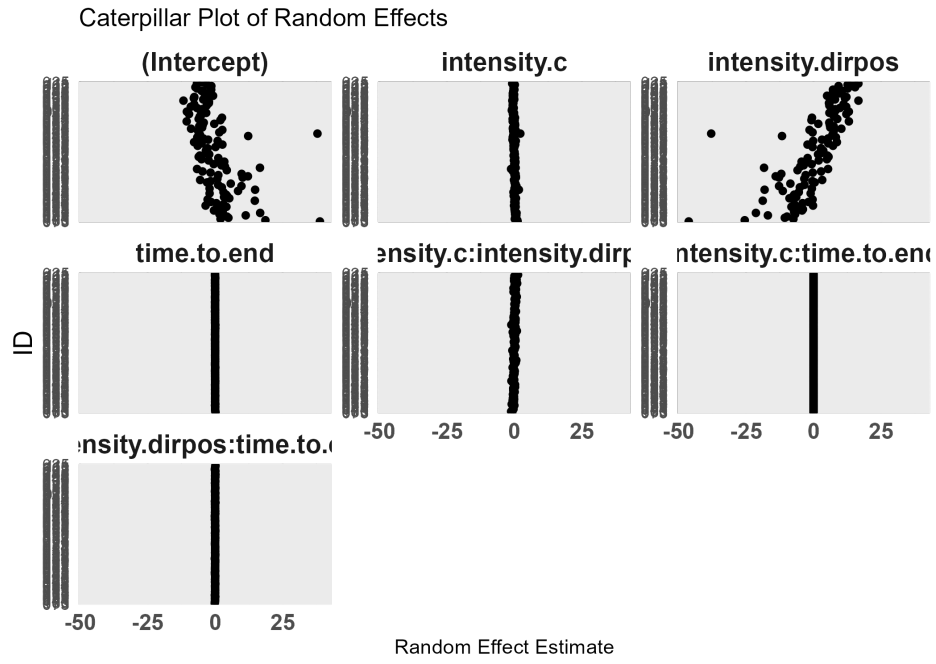
	Variance	**Correlations**					
Term		**Intercept**	**Intensity**	**Direction**	**Time**	**Intensity: Direction**	**Intensity: Time**
Intercept	89.391						
Intensity	0.369	0.715					
Direction	147.954	-0.922	-0.684				
Time	0.002	0.319	0.254	-0.295			
Intensity:Direction	0.333	-0.471	-0.724	0.532	-0.427		
Intensity:Time	0	-0.519	-0.623	0.349	-0.115	0.248	
Direction:Time	0.005	-0.217	-0.298	0.162	-0.822	0.543	0.254
Residual	112.462						

The caterpillar plots (Figure F3) illustrate the variation in slopes across individuals. Notably, only the slope for the main effect of intensity shows meaningful variability, while the other random slopes exhibit minimal spread, suggesting limited individual differences in those effects. This supports the decision to retain only the

random slope for intensity in the final model.

Figure F3

Caterpillar Plots of Random Effects (Intensity, Direction, Time and 2-Way Interactions)



Model 2. Negative Affect Intensity (Quartiles)

The negative affect ordered factor model included random slopes for linear, quadratic, and cubic trends, along with the intercept and time. Attempts to include random interaction terms between negative affect quartiles and time led to non-convergence. Table F5 shows a large variance in the linear effect and strong correlations among the random effects, particularly between the intercept and cubic term (r = 0.94), and between linear and quadratic terms (r = 0.79).

The caterpillar plots (Figure F4) reveal substantial individual variation in the linear and quadratic slopes, while the cubic term and time effect show minimal variability. This suggests that individuals differ meaningfully in their linear and quadratic trajectories of negative affect, but not in higher-order trends or time-related changes.

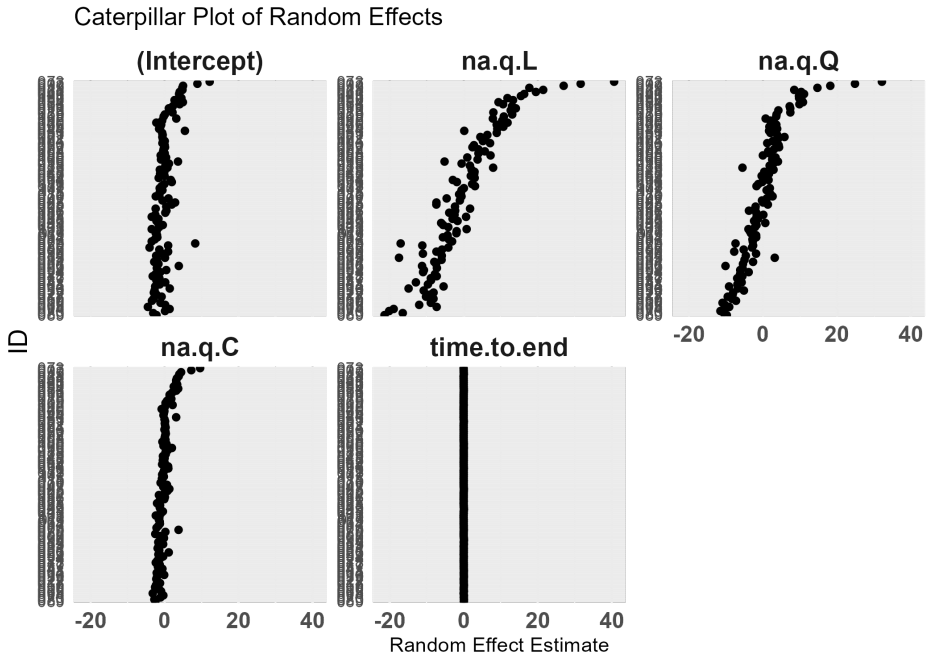
Table F5

Variance and Correlations of Random Effects in Negative Affect Quartile Model

	Variance	Correlations			
Term		**Intercept**	**Linear**	**Quadratic**	**Cubic**
Intercept	15.184				
Linear	143.618	0.421			
Quadratic	67.949	0.533	0.785		
Cubic	7.058	0.94	0.607	0.704	
Time	0.001	-0.208	-0.366	-0.183	-0.238
Residual	129.122				

Figure F4

Caterpillar Plots of Random Effects (Negative Affect Quartiles: Polynomial Contrasts and Time)



Model 2*. Negative Affect Intensity (Person-Mean Centered)

The negative affect model included random slopes for the main effect of negative affect, time, and their interaction, in addition to the intercept. Table F6 presents the variances of and correlations among random effects. The variance of the intercept showed between person differences and a strong negative correlation was observed between negative affect and its interaction with time (r = -0.76), which is not interpretable due to the low variance in the slopes of the interaction.

The caterpillar plots (Figure F5) illustrate the distribution of random slopes across individuals. The intercept shows considerable variability, and the slope for negative affect also demonstrates some individual differences, indicating that the effect of negative affect on the outcome varies across participants.

Table F6

Variance and Correlations of Random Effects in Person-Centered Negative Affect Model

	Variance	**Correlations**		
Term		**Intercept**	**Negative Affect**	**Time**
Intercept	8.712			
Negative Affect	0.195	-0.178		
Time	0.001	0.157	0.028	
Neagtive Affect:Time	0	0.07	-0.762	-0.251
Residual	126.617			

Figure F5

Caterpillar Plots of Random Effect Estimates (Negative Affect Centered, Time, and Interaction)

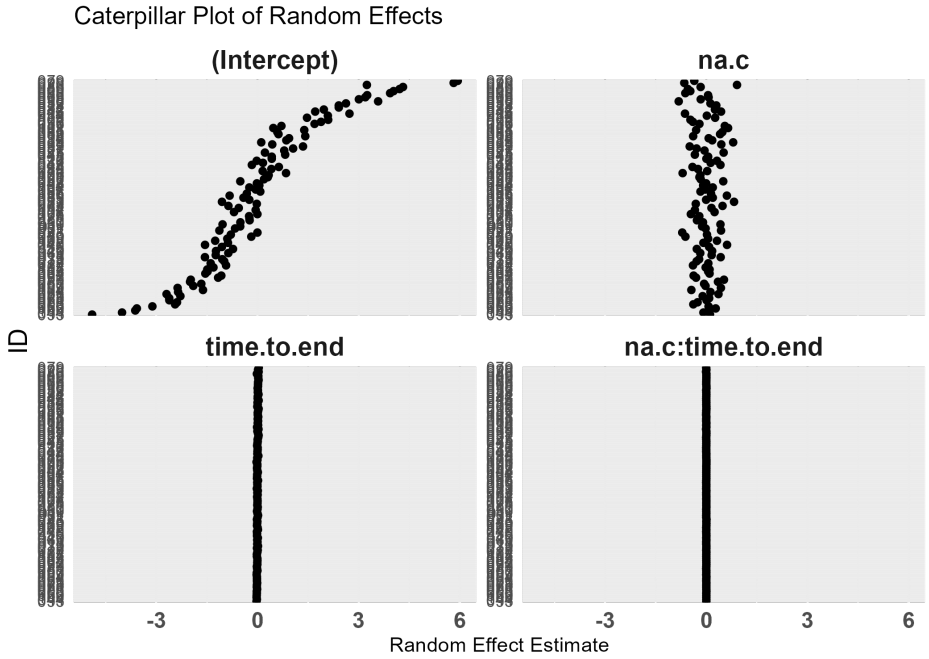


Table F7

Random Effects Structure for All Event Models

Random Effects Structure	Positive Event (Quartiles) Model 3		Positive Event (Centered) Model 3**		Negative Event (Quartiles) Model 4		Negative Event (Centered) Model 4**	
	Main Effects	2-Way Interactions	Main Effects	2-Way Interactions	Main Effects	2-Way Interactions	Main Effects	2-Way Interactions
Random Intercepts	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Slopes	All	All	All	All	All	All	All	All
Random Interaction Effects	–	Only main	–	All	–	Only main	–	Only main

Note: Overview of the random effects structure included in each model. All models included random intercepts and random slopes for each predictor. Random interaction effects were added when convergence allowed. A dash (–) indicates that no interaction terms were modeled for that condition. "Only main" indicate convergence issues for higher-order random interaction effects.

Model 3. Positive Event Intensity (Quartiles)

The model examining positive event quartiles included random slopes for linear, quadratic, and cubic trends, along with the intercept and time. Attempts to include interaction terms led to non-convergence, likely due to the high intercorrelations and complexity of the random effects structure. These random interaction terms were therefore excluded from the final model. Table F8 reveals strong correlations among the random effects, particularly between the intercept and quadratic term ($r = 0.98$), and between linear and cubic terms ($r = 0.96$), indicating substantial multicollinearity that may affect model stability.

The caterpillar plots (Figure F6) show considerable individual variation in the linear and quadratic slopes, while the cubic and time effects exhibit more constrained distributions. This suggests that individuals differ meaningfully in their linear and quadratic responses to positive events.

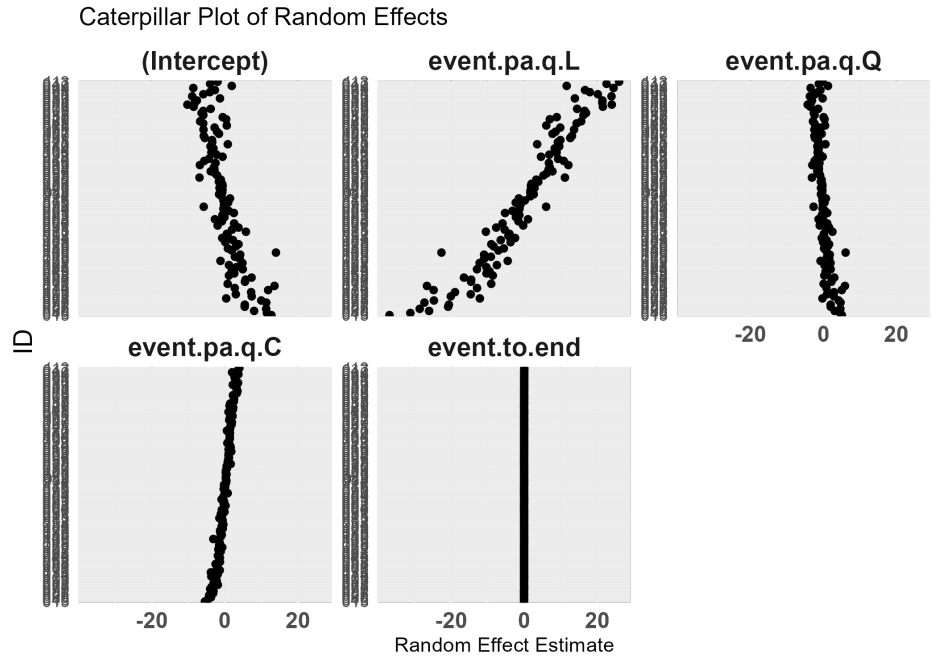
Table F8

Variance and Correlation of Random Effects in Positive Events Quartile Model

	Variance	Correlations			
Term		**Intercept**	**Linear**	**Quadratic**	**Cubic**
Intercept	46.471				
Linear	300.293	-0.822			
Quadratic	8.628	0.977	-0.754		
Cubic	7.312	-0.766	0.964	-0.69	
Time	0.002	-0.186	0.178	-0.12	0.138
Residual	204.138				

Figure F6

Caterpillar Plots of Random Effects (Positive Event Quartiles: Polynomial Contrasts and Time)



Model 3*. Positive Event Intensity (Person-Mean Centered)

The model examining the effects of positive event intensity included random slopes for the main effect, time, and their interaction. Table F9 presents the variances of and correlations among these random effects. There was a large between-person variance in the intercepts, and the high residual variance indicates that a substantial portion of within-person variability remains unexplained. A strong negative correlation was observed between the intercept and positive event intensity (r = -0.66).

Table F9

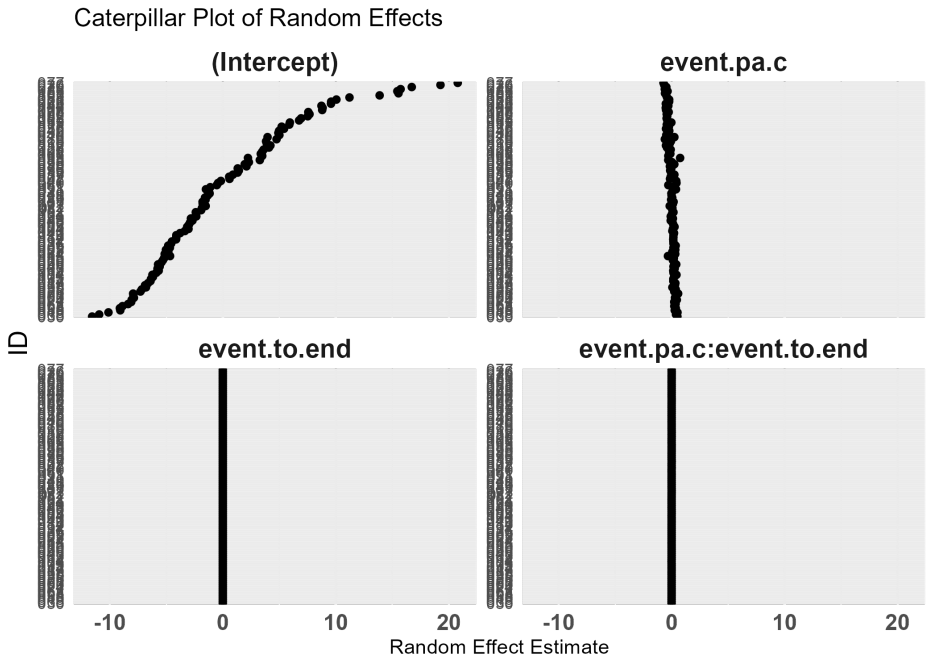
Variance and Correlation of Random Effects in Person-Centered Positive Events Model

	Variance	**Correlations**		
Term		**Intercept**	**Positive Event**	**Time**
Intercept	69.623			
Positive Event	0.182	-0.663		
Time	0	-0.021	0.024	
Positive Event:Time	0	0.26	-0.366	-0.062
Residual	201.056			

The caterpillar plots (Figure F7) show that while the intercept exhibits a wide range of variability across individuals, the slopes for positive event intensity, time, and their interaction are tightly clustered around zero. This indicates that the effect of positive event intensity on the outcome is relatively consistent across individuals.

Figure F7

Caterpillar Plots of Random Effects (Positive Event Centered, Time, and Interaction)



Model 4. Negative Event Intensity (Quartiles)

The negative event ordered factor model included random slopes for linear, quadratic, and cubic trends, along with the intercept and time. Table F10 shows strong correlations among the random effects, particularly between the intercept and cubic term (r = 0.94), and between linear and quadratic terms (r = 0.79), indicating potential multicollinearity. The caterpillar plots (Figure F8) reveal substantial individual variation in the linear effect.

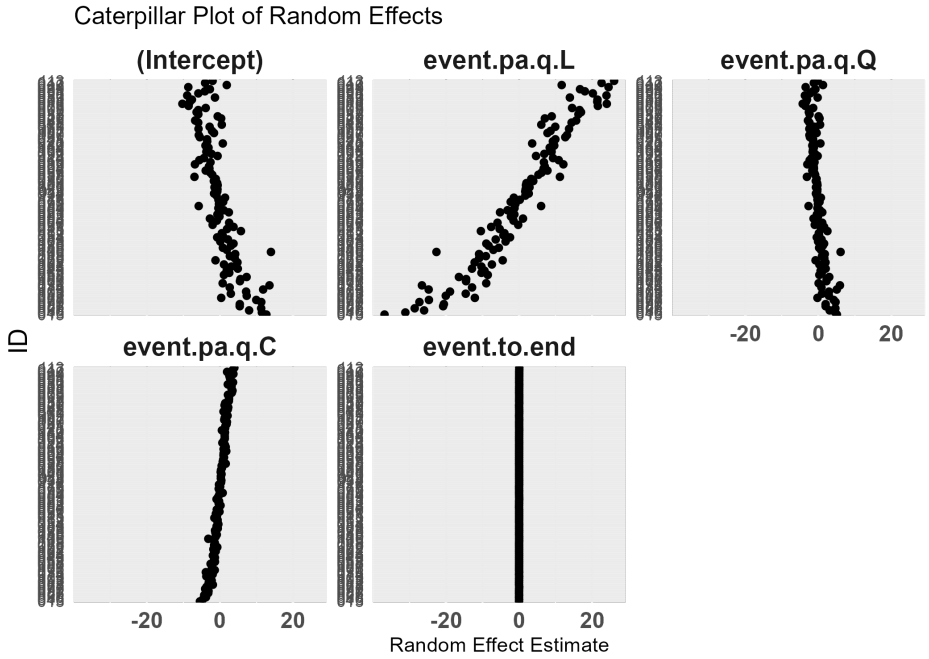
Table F10

Variance and Correlation of Random Effects in Negative Events Quartile Model

	Variance	**Correlations**			
Term		**Intercept**	**Linear**	**Quadratic**	**Cubic**
Intercept	14.222				
Linear	120.116	0.904			
Quadratic	73.414	0.762	0.421		
Cubic	21.634	0.781	0.45	0.996	
Time	0	0	0	0	0
Residual	209.176				

Figure F8

Caterpillar Plots of Random Effects (Negative Event Quartile: Polynomial Contrasts and Time)



Model 4*. Negative Event Intensity (Person-Mean Centered)

We included random slopes for the main effect of negative event intensity, time, and the intercept. Table F11 shows modest correlations among the random effects, with a notable negative correlation between the intercept and time (r = -0.78), although the variance in the slopes is too small for a substantive interpretation of this effect.

Table F11

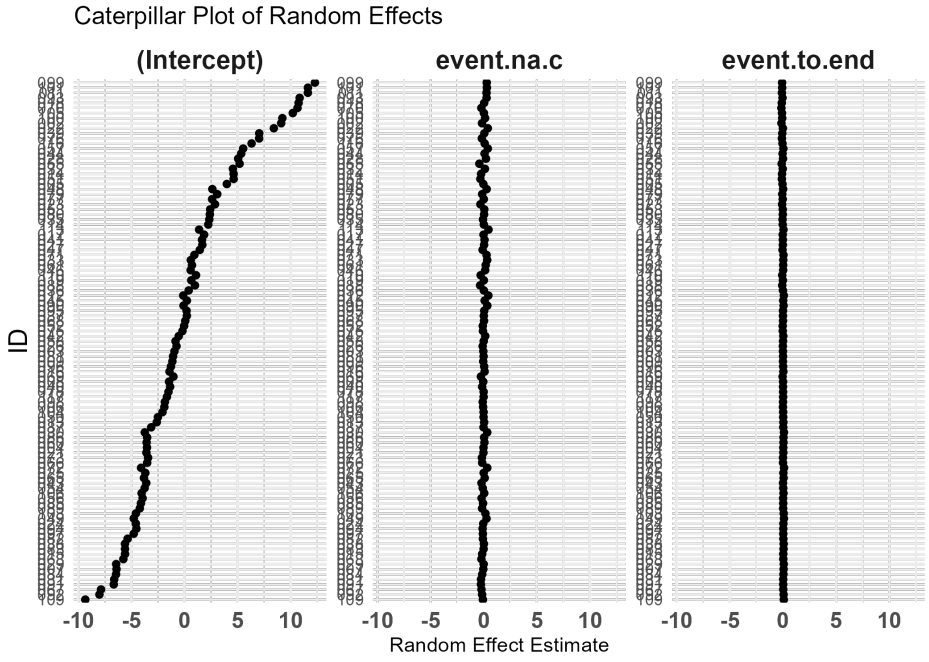
Variance and Correlations of Random Effects in Person-Centered Negative Events Models

	Variance	**Correlations**	
Term		**Intercept**	**Negative Event**
Intercept	72.345		
Negative Event	0.107	0.09	
Time	0.013	-0.781	0.383
Residual	217.836		

The caterpillar plots (Figure F9) reveal substantial individual variability in the intercept. The effects of negative events and time are relatively consistent across individuals.

Figure F9

Caterpillar Plots of Random Effects (Negative Events Centered and Time)



Supplement G

Johnson Neyman Interaction Plots of Slope of Time

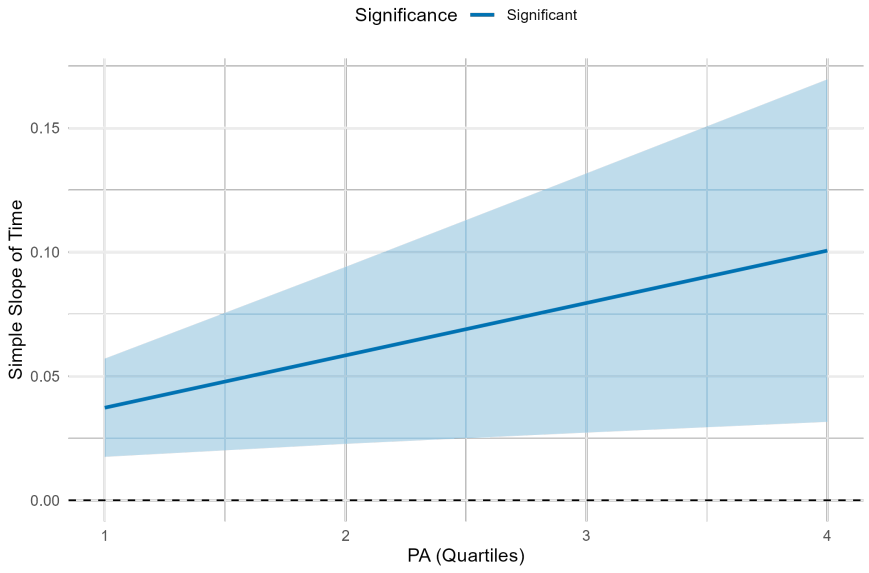
The Johnson-Neyman plots are only given for models in which the interaction terms were significant.

Slope of Time by Affect Intensity

Model 1. Positive Affect Quartiles

Figure G1

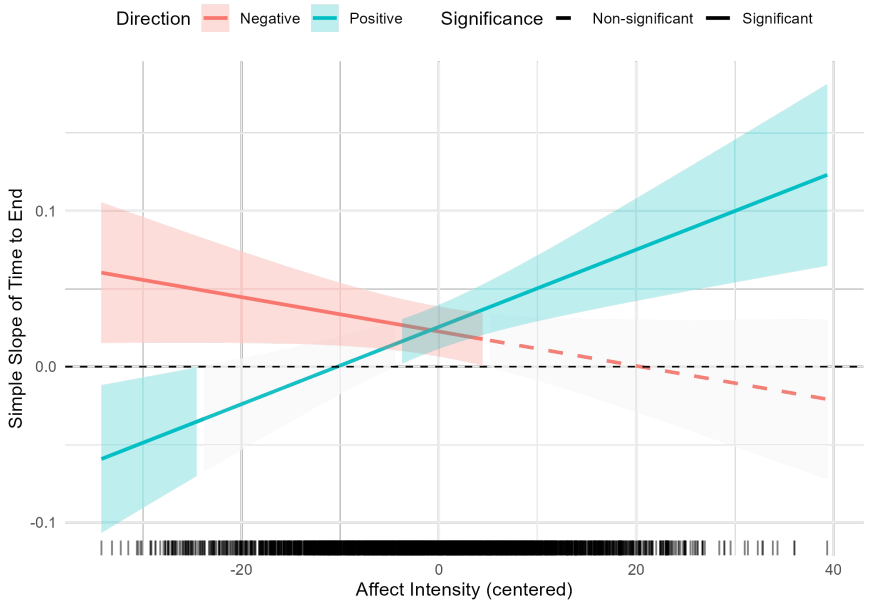
JN Plot Slope of Time by Positive Affect Quartiles (Linear Effect)



*Model 1**. Affect Intensity with Direction (Bimodal)*

Figure G2

JN Plot Slope of Time by Affect Intensity

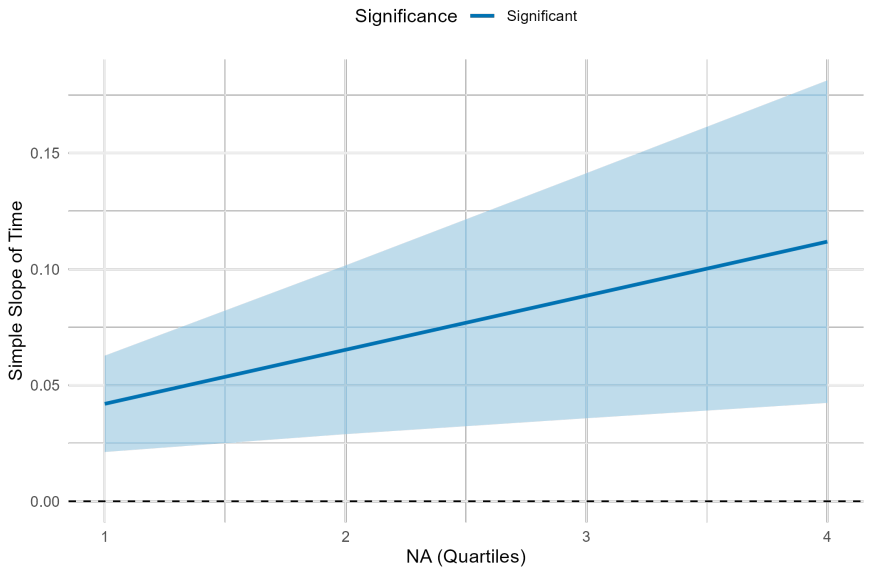


The Johnson-Neyman interaction plots modelling the slope of time over positive event intensity levels (seeSupplement G) revealed a negative recency effect for very low positive affect (divergence was smaller for less recent events). This suggests that low-intensity positive experiences may be remembered more accurately if the event occurred more recently.

Model 2. Negative Affect Quartiles

Figure G3

JN Plot Slope of Time by Positive Affect Quartiles (Quadratic Effect)

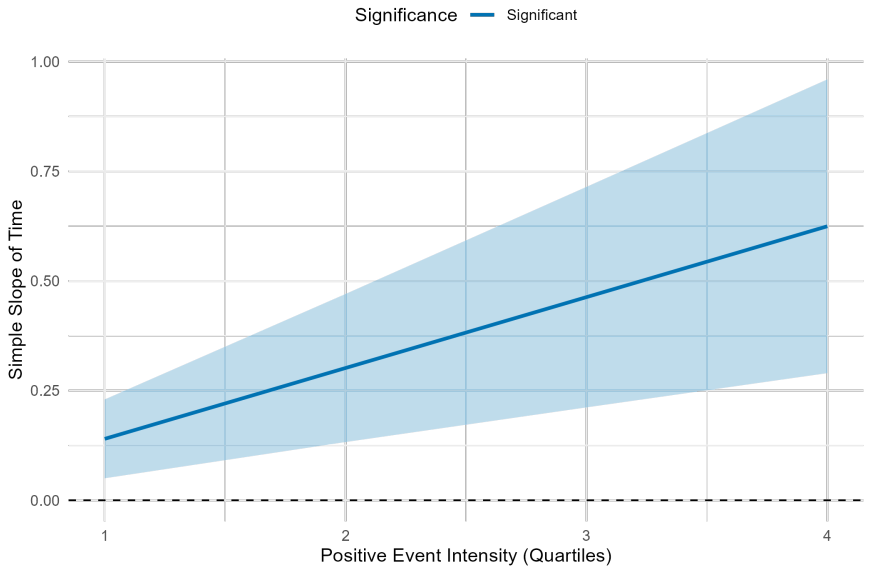


Slope of Time by Affective Event

Model 3. Positive Event Intensity Quartiles

Figure G4

JN Plot Slope of Time by Positive Event Quartiles (Quadratic Effect)



Model 3. Positive Event Intensity*

Figure G5

JN Plot Slope of Time by Positive Event Intensity

