Neubauer, Andreas B.; Scott, Stacey B.; Sliwinski, Martin J.; Smyth, Joshua M.

How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
Journal of personality and social psychology 119 (2020) 1, S. 185-203, 10.1037/pspp0000248

How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span

Andreas B. Neubauer[1,2], Stacey B. Scott[3], Martin J. Sliwinski[4], & Joshua M. Smyth[4]

[1]DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

[2]Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany

[3]Stony Brook University

[4]The Pennsylvania State University

Author Note

Andreas B. Neubauer, Education and Human Development, DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany and Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany; Stacey B. Scott, Department of Psychology, Stony Brook University, New York, USA; Martin J. Sliwinski, Department of Human Development and Family Studies and Center for Healthy Aging, The Pennsylvania State University, Pennsylvania, USA; Joshua M. Smyth, Departments of Biobehavioral Health and Medicine, The Pennsylvania State University and Hershey Medical Center, Pennsylvania, USA.

Correspondence concerning this article should be addressed to Andreas B. Neubauer, Education and Human Development, DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany. E-Mail: neubauer.andreas@dipf.de

Acknowledgments

J.M. Smyth and M.J. Sliwinski acknowledge support from the National Institute on Aging grants number R01-AG039409 and R01-AG026728.

Abstract

Daily-diary studies and experience sampling studies examine day-to-day variations in affect using different rating types: The former typically collect retrospective affect reports at the end of the day, whereas the latter collects multiple momentary assessments across the day. The present study examined the convergence of (aggregated) momentary assessments collected repeatedly within a day and retrospective assessments collected at the end of the day. Building on prior research on the memory-experience gap and the peak-and-end rule we predicted that participants would report more intense retrospective affect than aggregated momentary affect, and that retrospective affect would be biased towards the peak and the most recent affect of the day. Based on socio-emotional selectivity theory and the strength and vulnerability integration model, age differences in these convergence indicators were expected. Findings from two age-heterogeneous ecological momentary assessment/daily-diary hybrid studies (*N*=242, 25-65 years; and *N*=175, 20-79 years) revealed (a) a memory-experience gap for negative affect (more intense retrospective ratings than aggregated momentary ratings) that is attenuated with advancing age, (b) only a small memory-experience gap for positive affect for very old adults (66-79 years), but not younger adults, (c) relatively high convergence of aggregated momentary ratings and retrospective ratings despite (d) small biases of retrospective negative affect ratings towards peak and most recent negative affect. Findings suggest that both rating types can discriminate “good days” from “bad days” and provide overlapping but not necessarily exchangeable information.

Keywords: daily diary; ambulatory assessment; well-being; measurement burst; affect

How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span

Affective experiences such as anger, sadness, joy or happiness change dynamically within individuals across various time scales (Augustine & Larsen, 2012). Although quickly changing affect dynamics receive an increasing amount of attention in the psychological literature (e.g., Kuppens, 2015), there is also a large body of research targeting slower, day-to-day variations in affective experiences. In these daily-diary studies, participants are asked once a day (typically at the end of the day) to report their daily experiences as well as their affective experiences for that day (e.g., Almeida, 2005; Brose, Schmiedek, Lovden, & Lindenberger, 2011). There are many advantages of investigating affect using this methodology (Augustine & Larsen, 2012; Schwarz, 2012), one of which is that it should mitigate retrospective biases compared to studies using longer recall periods. Arguably, reporting on affect experienced today places less demand on memory and cognitive heuristics than, for example, recalling, aggregating, and reporting on affective experiences in the last week, month or year. Nevertheless, inquiring about affect today still requires engagement of memory and heuristic processes, opening the possibility that recollection of today's affect will not perfectly align with experienced affect throughout the day.

The aim of the present study is to examine the within-person convergence of aggregated momentary affect and end-of-day retrospective affect. That is, we examined whether end-of-day assessments and aggregated momentary assessments are both capable of discriminating "good days" from "bad days" within a person. Furthermore, we examined potential age differences in these convergences using age-heterogeneous samples from two ecological momentary assessment/daily-diary hybrid studies.

Differences between Momentary and Retrospective Affect Ratings

In their accessibility model of emotional self-report, Robinson and Clore (2002) postulated that momentary reports of an emotional state ("How do you feel right now?") differ qualitatively from retrospective self-reports. Specifically, momentary reports can be made based on experiential knowledge that is available in the current moment only and cannot be "re-lived" after the affective episode has passed. Because emotional experiences are fleeting, retrospective assessments rely on reconstructive processes of the prompted event ("How did you feel when X happened?) or time frame ("How did you feel last week?"). When the prompted emotional report relates to a specific episode or to a short time frame, episodic memory processes are engaged that retrieve information on the event or time frame. The more specific the prompted event is or the shorter the recall period, the more accurate retrospective assessments are expected to be (Schwarz, 2012). Retrospective reports of daily affect ("How did you feel today?") should therefore reasonably accurately reflect the "true" experiences of the day (i.e., be highly related with aggregated momentary experiences), given the rather short retrieval time of one waking day.

Divergences between aggregated momentary affect ratings and retrospective ratings have sometimes been considered as evidence for biases in retrospective assessments. Alternatively, however, it could also be speculated that these differences might be driven by unprecise assessment of momentary affect: For example, asking individuals many times for their current affective state might interrupt their daily routines, causing them to carelessly respond in some cases (Aldwin, 2010). Therefore, differences between aggregated momentary ratings and retrospective ratings should not offhandedly be considered as evidence for psychometric superiority of momentary assessments. Similarly, Conner and Barrett (2012) suggested that differences between these rating types should not be taken as evidence that aggregated momentary ratings have, in general, higher validity than retrospective ratings: although momentary ratings might be more valid representations of current subjective experiences than retrospective ratings (construct validity), their ability to predict future behavior (predictive validity; Cronbach & Meehl, 1955) might not necessarily be higher. Indeed, some research indicates that retrospective affect ratings may actually in some cases predict future behavior better than aggregated momentary affect. For example, retrospective assessments of a vacation were more strongly related to the desire to repeat the experience compared to aggregated momentary assessments taken during the vacation (Wirtz, Kruger, Napa Scollon, & Diener, 2003), and retrospective pain ratings at the end of a colonoscopy (but not momentary pain ratings during the procedure) predicted return rates for a follow-up colonoscopy (Redelmeier, Katz, & Kahneman, 2003). That is, when it came to the intention to repeat the experience (vacation; colonoscopy) memories of the experience were better predictors than experiences collected in the moment.

Conner and Barrett (2012) argue that the reason that aggregated momentary assessments and retrospective assessments are related to different outcomes is that they tap different aspects of the self. Specifically, these authors distinguish the experiencing self from the remembering self. The experiencing self can be assessed using momentary experiences and is hypothesized to be more closely related to physiological states and processes whereas the remembering self, which is measured using retrospective ratings, is often more predictive of deliberative future behavior (Conner & Barrett, 2012). Similar to Robinson and Clore's (2002) accessibility model of emotional self-report, the differentiation between experiencing self and remembering self proposed by Conner and Barrett (2012) predicts less-than-perfect congruence between momentary ratings and retrospective ratings of affective states. The core difference between these two accounts is that the accessibility model attributes differences between the rating types to specific knowledge systems (experiential knowledge vs. episodic knowledge) whereas the account by Conner and Barrett (2012) assumes that these differences are primarily driven by the involvement of different (neuro-)biological structures that map more on the experiencing self (e.g., the autonomic nervous system and the core affect network) or the remembering self (the default/long term memory network), respectively (see Conner & Barrett, 2012, for a detailed discussion of similarities and differences between these two accounts).

In summary, it can be expected that aggregated momentary ratings and retrospective ratings of affective experiences will not perfectly converge. Nevertheless, these two rating types typically show some degree of convergence, as has also been shown in previous research (e.g., Röcke, Hoppmann, & Klumb, 2011). In the present work, we differentiate two types of convergence indicators that we discuss in the following sections: level convergence and correspondence convergence.[1]

Level Convergence

Level convergence relates to the degree to which momentary ratings and retrospective ratings concur with regard to the mean levels. Deviations from perfect level convergence result in what Miron-Shatz, Stone, and Kahneman (2009) referred to as the memory-experience gap, that is, a "discrepancy between the average of experienced emotions and the overall evaluation of the experience, which is usually more intense than the averaged emotions" (p. 885). Prior research has largely reported results consistent with this discrepancy (Ben-Zeev, Young, & Madsen, 2009; Ganzach & Yaor, 2019; Lay, Gerstorf, Scott, Pauly, & Hoppmann, 2016; Miron-Shatz et al., 2009; Parkinson, Briner, Reynolds, & Totterdell, 1995; Thomas & Diener, 1990).

Thus far, the exact mechanism underlying this phenomenon is unclear. Regarding the memory-experience gap for negative affect (NA), one possible explanation could be a negativity bias in retrospective assessments, that is, a tendency to utilize negative information more than positive information in one's judgement (Vaish, Grossmann, & Woodward, 2008).

[1] We borrowed this terminology from Stone, Broderick, and Schwartz (2010) who used the terms *level differences* and *convergence differences* in their work.

This bias might be more pronounced in cases when semantic memory processes are involved than when it is primarily episodic or experiential information that is used for individuals' judgements. Support for this prediction comes from the study by Lay et al. (2016) who reported a positive association between neuroticism and the memory-experience gap for NA. However, a general negativity bias cannot account for findings reporting a memory-experience gap for positive affect (PA). Notably, evidence for a memory-experience gap for PA is more inconsistent than for NA, which could suggest that person- or situation-specific moderators might be important here. Assuming that positive information is actually more relevant for retrospective judgements for some individuals (i.e., a positivity effect for some individuals) could explain part of the inconsistencies in previous studies. Largely in line with this assumption are results in the aforementioned study by Lay et al. (2016), showing that extraversion was positively associated with the memory-experience gap for PA. Taken together, a general tendency towards utilizing either positive or negative information for retrospective assessments could be one explanation for the memory-experience gap.

A second, possibly complementary, explanation for the memory-experience gap is that retrospective ratings might be biased towards the peak affect experienced during the recall period (e.g., Ganzach & Yaor, 2019). When retrospectively judging the affective value of an event, individuals do not exclusively rely on the average affect they experienced during this episode but they also utilize the peak intensity of the emotion. Peak affect is salient and is therefore used in retrospective judgements of affective experiences (Stone, Schwartz, Broderick, & Shiffman, 2005). This is consistent with the finding that the peak of momentary affect ratings was associated with the memory-experience gap for high arousal positive and negative affect (Lay et al., 2016), suggesting that retrospective ratings are "pulled" towards the highest momentary affect rating in the observation period.

According to Fredrickson (2000) the impact of peak affect on retrospective ratings is adaptive because peak affect carries important personal meaning: Peak affect informs individuals about the capacities that were required to experience and endure the episode. When retrospectively assessing how demanding a situation was, it is, of course, informative to know what the average demand was, but the information on the maximum demand of this situation is also important. For example, if an individual decides whether she undergoes a second colonoscopy, her decision will be partly influenced by how painful the first colonoscopy was remembered (Redelmeier et al., 2003). The peak unpleasantness of the first colonoscopy informs the individual about the maximum pain she will endure and provides information that goes beyond the mean pain for the whole procedure. In addition to the peak, a second phase of the momentary affect distribution, the affect at the end of the episode, has also been postulated to predict retrospective ratings beyond the mean of the momentary experience (a phenomenon termed the "peak-and-end rule"; see Fredrickson, 2000). Similarly to peak affect, recent affect is hypothesized to be more salient and therefore utilized in arriving at retrospective evaluations of momentary states.

Correspondence Convergence

The second convergence indicator investigated in the present research, correspondence convergence, relates to the relative convergence between retrospective assessments and aggregated momentary assessments. Hence, in contrast to level convergence (which targets mean level differences between rating types, and hence, the memory-experience gap), correspondence convergence can be operationalized as a correlation between rating types.[2] An exemplary study investigating correspondence convergence has been reported by Röcke et al. (2011) who collected momentary affect ratings (up to 150 momentary ratings collected across one year) and retrospective affect ratings at the end of the

[2] Note that conceptually, level convergence and correspondence convergence can be unrelated. For example, retrospective ratings and aggregated momentary ratings could be perfectly correlated (perfect correspondence convergence), but their means could differ (memory-experience gap; lack of level convergence).

study period (recalled affect for the past twelve months) in a sample of older adults (72 to 91 years). These authors aggregated the momentary ratings into a mean and correlated this variable with the individual's retrospective assessment. Correlations ranged from .55 (NA) to .63 (PA) indicating good (but not perfect) correspondence convergence even for retrospective ratings across one year.

Notably, these results are based on between-person correlations – these analyses target the question if inter-individual differences in experienced affect (aggregated momentary affect) are manifested in inter-individual differences in remembered affect (retrospective assessments) as well. In other words: They allow for assessing *between-person correspondence convergence,* that is, the question whether persons who felt better are also persons who remember to have felt better. Although an interesting question in itself, these analyses do not allow examining whether aggregated momentary ratings and retrospective ratings are equally well suited to discriminate days with more versus less PA (NA) within individuals. It remains an open question if both rating types capture the same within-person, across-day variation in affect (discriminate "good days" from "bad days" rather than "happy persons" from "unhappy persons").

In the present work, we focus on this *within-person correspondence convergence* of PA and NA. This requires that both momentary ratings and retrospective ratings of affect are assessed repeatedly within the same individuals. Parkinson et al. (1995) attempted to investigate within-person correspondence convergence using momentary affect ratings collected seven times per day for 14 consecutive days as well as retrospective affect ratings collected at the end of each day. To investigate within-person correspondence convergence these authors computed within-person correlations of aggregated momentary ratings (each day's mean momentary affect rating) with retrospective ratings. On average, the within-person correlations were reasonably high (.66 for PA and .68 for NA), suggesting good within-person correspondence convergence, but the authors also reported substantial inter-individual differences in the within-person correlations (ranging from -.23 to .94). Although informative, these results need to be considered with two major limitations in mind: First, the sample comprised only 30 participants limiting its potential to explore predictors of inter-individual differences in within-person correspondence convergence. Second, although within-person correlations are conceptually an estimate for within-person correspondence convergence, there are methodological problems associated with this approach because each within-person correlation was based on a maximum of 14 data points only, resulting in a rather imprecise estimate (see Mejía, Hooker, Ram, Pham, & Metoyer, 2014). Modern data analytic tools such as multilevel models (MLM) are well suited to investigate both the average within-person association of aggregated momentary ratings and retrospective ratings and potential inter-individual differences therein. Furthermore, MLM also allow for investigating whether peak and most recent affect predict retrospective affect beyond mean momentary ratings (as would be predicted by the peak-and-end rule). A recent study (Ganzach & Yaor, 2019) has taken this data analytic approach, predicting retrospective, end-of-day affect ratings from the aggregate of eight momentary PA and NA ratings per day, as well as peak and last affect of the day in a MLM. Only peak, but not last affect of the day predicted retrospective affect at the end of the day (within persons) above the day's aggregated momentary affect.

With the present study we sought to examine the degree of level convergence and within-person correspondence convergence of momentary and retrospective affect. To that end, we first examined mean level differences (as a measure of lack of level convergence) between today's aggregated momentary affect (the mean of ratings collected throughout the day; "How do you feel right now?") and retrospective affect with respect to this day ("How did you feel today?"). Next, we determined the degree of within-person correspondence convergence by assessing the within-person association of aggregated momentary assessments and retrospective assessments. Third, we examined whether within-person variations in peak and most recent momentary affect predicted retrospective affect beyond the mean of this day's momentary affect. Finally, we examined inter-individual differences in these effects and investigated one person-level variable that is predicted to be associated with these inter-individual differences: participants' age.

Age Differences in Convergence Indicators

There are both theoretical considerations and empirical data suggesting that age might be associated with inter-individual differences in level convergence. For example, in a study on adults (25-74 years) by Charles et al. (2016), age differences in NA were reported to be stronger with increasing recall period (e.g., larger age differences for the assessment of NA last month compared to NA today). Furthermore, Lay et al. (2016) reported a smaller memory-experience gap for NA (but not PA) with advancing age. They interpreted their results as consistent with the perspective of a reduced negativity bias with advancing age (see Grühn, Scheibe, & Baltes, 2007; Mather & Carstensen, 2005). These considerations are further in line with the notion of Socioemotional Selectivity Theory (SST; Carstensen, Isaacowitz, & Charles, 1999): According to this account, old age is associated with an increased motivation to pursue emotionally meaningful goals. Associated with this shift in motivational priorities is a change in information processing, manifested, for example, in better memory for positive emotional stimuli compared to negative stimuli. It is, however, unclear, whether this altered information processing is associated with a reduction in negativity (less preference for negative information), an increase in positivity (more preference for positive information), or both. For the memory-experience gap, an (age-related) decrease in negativity should lead to a reduction of this gap for NA (reduced negativity in retrospective assessments), whereas an (age-related) increase in positivity should be associated with an increase in this gap for PA with advancing age. Based on considerations of age-related reduced negativity *and* heightened positivity, we expected a negative association of age with the memory-experience gap for NA (i.e., with older age, a smaller memory-experience gap), and a positive association of age with the memory-experience gap for PA (i.e., with older age, a greater memory-experience gap).

We had no specific expectations whether age should predict within-person correspondence convergence (the size of the within-person association between aggregated momentary ratings and retrospective ratings) and we are not aware of any studies or theoretical accounts that have addressed this issue thus far. Regarding a different within-person effect, the impact of peak affect on retrospective affect, we expected moderation by age. This hypothesis was derived from the Strength And Vulnerability Integration model (SAVI; Charles, 2010). SAVI stresses that older adults' life-time of experience helps them in more effectively regulating their emotions by employing effective emotion regulation strategies when they are required. However, this effective emotion regulation is counteracted in situations of high physiological arousal due to heightened vulnerability and reduced flexibility in physiological stress responses (see also Wrzus, Muller, Wagner, Lindenberger, & Riediger, 2013). This implies that for older adults, it becomes particularly important to judge the capacity requirements for the situations they are about to enter. As explained above, peak affect carries this information, and retrospective assessments of affective experiences are influenced by peaks based on the assumption that they provide important salient information about capacity requirements (Fredrickson, 2000). Hence, for older adults, information on peak affect might be particularly important and, hence, be more strongly associated with retrospective affect assessments.

The Present Study

In the present work, we investigated the convergence of retrospective assessments of positive and negative affect at the end of the day with the aggregated momentary assessments provided by the study participants in their daily lives. To that end, we employed two ecological momentary assessment (EMA)/daily-diary hybrid studies which followed participants for 14 (Study 1) and 21 (Study 2) days. In both studies, participants were asked to report their current affect five times per day, and at the end of each day they were asked to provide retrospective assessments of PA and NA of the current day. Based on prior research on the memory-experience gap (e.g., Miron-Shatz et al., 2009) we expected that retrospective assessments of both PA and NA would be higher than the mean of the momentary affect ratings collected on the respective day (Hypothesis 1). Furthermore, age effects on this difference were expected in a way that this gap would be reduced for NA, but exacerbated for PA, with advancing age (Hypothesis 2). We also assessed the degree of within-person correspondence convergence between aggregated momentary assessments and retrospective assessments. Building on early work by Parkinson et al. (1995) we expected high within-person correspondence convergence on average (Hypothesis 3). We further expected, based on the peak-and-end rule (Fredrickson, 2000), that peak momentary affect (Hypothesis 4a) and last momentary affect of the day (Hypothesis 4b) would predict retrospective affect beyond this day's mean momentary affect. Age differences in the impact of peak affect were expected based on SAVI (Charles, 2010), in a way that the impact of peak affect would be higher with advancing age (Hypothesis 5); age differences in the effect of last momentary affect were investigated in an exploratory fashion. Finally, inter-individual differences in both convergence indicators and the impact of peak affect were explored.

Study 1

Method

The protocol of this study was approved by the Institutional Review Board at Albert Einstein College of Medicine (#2010-353; Study ESCAPE).

Participants. Data were drawn from the first wave of a measurement burst study that was still ongoing when this research was conducted. Participants for this study were systematically sampled from registered voter lists in a specific zip code in the Bronx, NY. The neighborhood happens to be diverse (age, socioeconomic status, ethnicity), and this diversity is reflected in the sample that completed the study. Participants were sent letters of introduction in which the study protocol and the inclusion criteria were explained to them. These inclusion criteria were (1) age between 25 and 65 years, (2) ambulatory, (3) fluent in English, (4) no visual impairment that would interfere with operating the smartphones used for data collection, and (5) ability to operate the study smartphone throughout the day. Two-hundred and forty-five participants started the study; three participants did not provide information relevant for the present work yielding a final sample of 242 participants (M_{age} = 46.2, SD_{age} = 11.1; 64.9% female). Most participants (63.0%) identified themselves as Non-Hispanic Black, 8.9% as Caucasian, 17.5% as White Hispanic, 4.9% as Black Hispanic, and 4.1% as Other. Approximately one quarter (24.4%) had received no college education (high school degree or less), 59.1% had received some college education or a college degree, and 16.5% had a graduate or professional degree.

Procedure. Upon recruitment and providing informed consent, participants were invited into the laboratory for cognitive testing and assessment of various questionnaires. During this baseline session participants were acquainted with the study smartphones that were used for the EMA. Participants followed the protocol (described below) on study smartphones for a two-day run-in period. After these two days, participants returned to the laboratory and could ask questions about the study procedure that arose in this training phase. Only those participants who had more than 80% compliance during the two-day run-in period were invited to the full 14-day protocol. During the 14-day EMA, participants were instructed to complete a morning assessment upon awakening (this assessment is not relevant for the present research), five prompted assessments throughout the day (further referred to as beep assessments), and a participant initiated evening assessment prior to going to bed. On some days (*n* = 105), the evening assessment was completed prior to at least one momentary assessment; these days were removed from all further analyses to ascertain that all momentary ratings were taken before the respective retrospective ratings. The five beep assessments occurred in semi-randomized time intervals throughout participants' waking days; beep assessments were scheduled between two and three hours apart. Of the 16,940 possible beep assessments (242 participants, 14 days, five assessments per day), 14,101 (83.2%) were completed; a similar compliance rate was obtained for the evening assessments (83.4% of the 3,388 questionnaires were completed). Only the measures relevant for the present study are reported (see Scott et al., 2015, for a description of the full study protocol). Data from this study have been reported in prior research targeting different research questions (Graham-Engeland et al., 2018; Hyun, Sliwinski, & Smyth, 2019; Majd et al., 2018; Mathur et al., 2018; Scott, Kim, Smyth, Almeida, & Sliwinski, 2019; Scott, Munoz et al., 2018; Scott, Ram, Smyth, Almeida, & Sliwinski, 2017; Scott, Sliwinski et al., 2018; Slavish et al., 2018; Sliwinski et al., 2018; Stawski et al., 2018; Zawadzki et al., 2019).

Measurements. *Momentary affect* was assessed at each randomly prompted signal throughout the day. Participants were asked "How ___ do you feel right now?;" they were queried separately for nine adjectives. Five adjectives assessed momentary NA (tense/anxious; angry/hostile; depressed/blue; frustrated; unhappy) and four assessed momentary PA (happy; pleased; much enjoyment/fun; joyful). Responses were given on a visual analog scale ranging from 0 ("not at all") to 100 ("extremely"). The five negative affect items were averaged into a momentary NA score (within-person α = .83; see Geldhof, Preacher, & Zyphur, 2014), and the four positive affect items were averaged into a momentary PA score (within-person α = .89). *Retrospective affect* was assessed at the end of each day during the evening assessments by asking participants "How ___ did you feel today?". The same nine adjectives as in the momentary assessments were used. Again the five NA items were averaged (within-person α = .85) as were the four PA items (within-person α = .89).

Momentary ratings throughout the day were summarized into three indices: average momentary ratings, peak ratings, and last ratings. For each participant and day, the mean of all momentary PA items (NA items) was computed which resulted in one average momentary rating of PA and one average momentary rating of NA for each participant and each day. For each participant and day, we also extracted the highest rating of all up to five momentary PA (NA) ratings as an indicator for peak PA (peak NA) on this day. Additionally, for each participant on each day we extracted the last rating on the respective day as an indicator for last PA (last NA) to characterize the "end" affect for that day. If a participant missed the fifth beep on a given day, we used the rating provided on the fourth beep; if the fourth beep was missing as well, last affect on the respective day was set to missing for this individual.[3] In summary, for every participant's day we had four PA variables (retrospective PA, average momentary PA, peak PA, and last PA) and four NA variables (retrospective NA, average momentary NA, peak NA, and last NA).

Data Analysis. Our analyses focused on the two indicators of correspondence between aggregated momentary and retrospective assessments of affect: level convergence and within-person correspondence convergence. To investigate level convergence we employed bivariate multilevel models for each affect dimension. To that end, for each day t of each participant i, the aggregated affect rating across all momentary ratings was computed

[3] Across all observations, last affect was operationalized as affect at the fifth beep on 86.6% of all days, and as affect at the fourth beep in 8.0% of the days, respectively. No information on last affect was available on the remaining 5.4% of the days.

and stacked onto this person's respective retrospective affect rating of the same day. By that, each person's daily affect contained two observations: one observation which contained this day's average momentary rating, and one observation which contained this day's retrospective rating. A dummy coded variable (ratingtype_{it}) was created indicating which observation contained the aggregated momentary affect (coded as $\text{ratingtype}_{it} = 0$) and which observation contained the retrospective rating ($\text{ratingtype}_{it} = 1$). Next, we set up the first model in which the affect ratings were predicted by the dummy variable only. For NA (Model 1), the multilevel equations were:

Level 1:

$$NA_{it} = \beta_{0i} + \beta_{1i} \cdot \text{ratingtype}_{it} + \varepsilon_{it} \quad (1)$$

Level 2:

$$\beta_{0i} = \gamma_{00} + \upsilon_{0i} \quad (2)$$

$$\beta_{1i} = \gamma_{10} + \upsilon_{1i} \quad (3)$$

Random intercept (υ_{0i}) and random slope (υ_{1i}) were allowed to covary across individuals (unstructured G-matrix). Separate Level-1 residual variances for aggregated momentary ratings and retrospective ratings were estimated. We then entered the main effect of age into the model as Level-2 predictor and controlled for participants' gender at Level-2 (to account for potential differences in affective well-being between men and women) and the day of the measurement at Level-1 (to account for changes across the observation period; Model 2). In the final model (Model 3), the interaction of age with rating type was added in order to investigate if differences between aggregated momentary ratings and retrospective ratings (the memory-experience gap) were associated with age:

Level 1:

$$NA_{it} = \beta_{0i} + \beta_{1i} \cdot \text{ratingtype}_{it} + \beta_2 \cdot \text{day}_{it} + \varepsilon_{it} \quad (4)$$

Level 2:

$$\beta_{0i} = \gamma_{00} + \gamma_{01} \cdot \text{age}_i + \gamma_{02} \cdot \text{gender}_i + \upsilon_{0i} \quad (5)$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11} \cdot \text{age}_i + \upsilon_{1i} \quad (6)$$

The same models were run for PA as well (Models 4 through 6).

To investigate within-person correspondence convergence, we first set up a series of multilevel models to determine the within-person associations of retrospective assessments and aggregated momentary assessments by predicting retrospective ratings from aggregated momentary ratings. We used an empty model as baseline to determine the within-person variance of retrospective NA / PA. In the next model, only one of the three predictors (mean momentary rating, peak momentary rating, last momentary rating) was entered as Level-1 predictor into the model (predictors were centered on the person mean of all momentary NA/PA ratings). In the following models two predictors were entered simultaneously, before all three predictors were entered at once. Models were compared using likelihood ratio tests. Furthermore, we computed Level-1 pseudo-R^2 (Xu, 2003) as a measure of within-person variance explained by the model. This measure estimates the proportion of within-person variance that is explained by the predictors in the model (relative to the within-person variance in an empty model). Random slopes were omitted in these models to obtain interpretable R^2 estimates (for a discussion of these issues see LaHuis, Hartman, Hakoyama, & Clark, 2014). To examine the unique contributions of the three predictors in more detail, we estimated two more models: First, retrospective assessments were predicted by the three momentary affect predictors (mean momentary rating, peak momentary rating, last momentary rating), age, gender, and study day. Random slopes for the three momentary predictors were estimated in these models to explore inter-individual differences. Finally, the two-way interactions of these three predictors with age were added. For negative affect the resulting equations are:

Level 1:

$$\text{NA.retro}_{it} = \beta_{0i} + \beta_{1i} \cdot \text{NA.mean}_{it} + \beta_{2i} \cdot \text{NA.peak}_{it} + \beta_{3i} \cdot \text{NA.last}_{it} + \beta_4 \cdot \text{day}_{it} + \varepsilon_{it} \quad (7)$$

Level 2:

$$\beta_{0i} = \gamma_{00} + \gamma_{01} \cdot \text{age}_i + \gamma_{02} \cdot \text{gender}_i + \upsilon_{0i} \quad (8)$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11} \cdot \text{age}_i + \upsilon_{1i} \quad (9)$$

$$\beta_{2i} = \gamma_{20} + \gamma_{21} \cdot \text{age}_i + \upsilon_{2i} \quad (10)$$

$$\beta_{3i} = \gamma_{30} + \gamma_{31} \cdot \text{age}_i + \upsilon_{3i} \quad (11)$$

Results

We used Mplus version 8.1 to compute correlations separately for the within-person and between-person level; variance components were estimated via empty multilevel models. As can be seen in Table 1 (upper diagonal), on the between-person level, average momentary ratings, retrospective ratings, peak ratings, and end ratings within each affect dimension were almost perfectly correlated, $r > .90$ for all. That is, participants who on average reported higher momentary NA (PA) also reported higher retrospective NA (PA). Within-person correlations (lower diagonal in Table 1) were also large and statistically significant, but somewhat lower compared to the corresponding between-person correlations. That is, on days when a participant reported higher momentary NA (PA), the participant also reported higher retrospective NA (PA).

Insert [Table 1 here]

Mean Level Comparisons (Level Convergence). To investigate the memory-experience gap, we ran a multilevel model predicting affect from rating type. Results can be found in Table 2 (Model 1a) and Table 3 (Model 4a). For NA, retrospective ratings were higher by three scale points than the aggregated momentary ratings, $b = 3.052$, $p < .001$, whereas there was no significant difference in mean PA for these two rating types, $b = 0.406$, $p = .285$. Results from likelihood ratio tests showed that the random effect for rating type was significant for NA, $\chi^2(2) = 20.56$, $p < .001$, but not for PA, $\chi^2(2) = 0.44$, p = .802. Hence, there were significant inter-individual differences in the memory-experience gap for NA, but not PA.

To investigate age differences in affect ratings, we added participants' age as additional predictor (as well as gender and study day as covariates). We found no main effect for age on NA, $b = -0.353$, $p = .722$, but a positive effect on PA, $b = 2.587$, $p = .035$, indicating that older age was associated with higher PA, but unrelated to NA (see Models 2a and 5a; Table 2 and Table 3). In a final step, two-way interactions of rating type with age were added; this step allowed us to assess if the difference in mean level between aggregated momentary ratings and retrospective ratings was associated with age. Only for NA, $b = -0.810$, $p = .024$, but not PA, $b = -0.307$, $p = .420$, was there a statistically significant age x rating type interaction: This interaction indicates that the discrepancy between aggregated momentary affect ratings and retrospective ratings for NA was lower with advancing age.[4] To illustrate the findings, Figure 1 depicts NA ratings separately for aggregated momentary ratings and retrospective ratings; age was divided into tertiles (25-40 years, 41-52 years, 53-65 years; see Tables S1a-S1c in the online supplemental material, section A, for descriptive statistics by age group). As can be seen from this figure, retrospective NA ratings were higher than aggregated momentary ratings, but this difference was attenuated with advancing age. Follow-up analyses showed that for all three age groups, the difference between retrospective NA and aggregated momentary NA was statistically significant: 25-40 years: $b = 4.297$, $p < .001$; 41-52 years: $b = 3.293$, $p < .001$; 53-65 years: $b = 1.564$, $p = .003$. For PA, there were no differences between rating types in any age group, $|b| < 1.111$, $p > .064$ for all.

[4]We also tested whether there might be nonlinear age effects on the memory-experience gap by adding higher order polynomials of age (quadratic, cubic, quartic) as main effects and interactions with rating type. This did not improve model fit for either PA or NA, $\chi^2(2) < 4.13$, $p > .126$.

Including interactions of rating type with gender and day did not alter the results; none of these interactions were statistically significant (see Table S3 in the online supplemental material, section B).

Insert [Table 2 here]

Insert [Table 3 here]

Insert [Figure 1 here]

Predicting End-of-Day Affect (Correspondence Convergence). In the next set of analyses, we investigated the degree of relative convergence between aggregated momentary affect ratings and retrospective end-of-day ratings. Findings (see Table 4) suggest substantial overlap between the mean of momentary ratings and retrospective ratings on the within-person level: Pseudo-R^2 = .353 (NA) and Pseudo-R^2 = .384 (PA). For NA, both peak and last momentary ratings uniquely predicted retrospective ratings above mean ratings. For PA, only the last rating, but not the peak rating, predicted retrospective ratings beyond the mean of today's momentary ratings.

Insert [Table 4 here]

To further illuminate the unique contributions of aggregated momentary ratings, peak and last affect, as well as potential inter-individual differences therein, we added random effects for these predictors. Tables 5 and 6 show the results of these models (see Model 7a and Model 9a). For both PA and NA, last momentary ratings predicted end-of-day ratings beyond the mean momentary ratings. Peak ratings predicted retrospective ratings for NA, but not PA. Notably, random effects suggest that there were substantial inter-individual differences in the effect of mean, peak, and last momentary ratings on end-of-day retrospective assessments. To explore if these differences could be explained by age, we included interactions of these three predictors with age. However, this did not improve model fit for either NA or PA, $\chi^2(3) < 4.51$, $p > .211$ for both, and none of the interaction effects was statistically significant (see Tables 5 and 6, Model 8a and Model 10a).

We also examined whether convergence increased across time in study by entering interactions of aggregated momentary affect, peak affect, and most recent affect with study day. Including these three interactions did not improve model fit for either PA, $\chi^2(3) = 7.24$, $p = .065$, or NA, $\chi^2(3) = 6.76$, $p = .080$. For both outcomes, the aggregated momentary affect x study day interaction was not statistically significant, $b = 0.011$, $p = .220$ (PA), and $b = 0.015$, $p = .206$ (NA), respectively, providing no evidence for a change in within-person correspondence convergence across the observation period.

Insert [Table 5 here]

Insert [Table 6 here]

Exploratory Analyses: Inter-individual Differences. We estimated for each participant three parameters per affect dimension: an estimate of the memory-experience gap; an estimate of the (zero-order) within-person association between aggregated momentary affect and retrospective affect; and an estimate of the incremental effect of peak affect on retrospective affect (above mean affect). These estimates were obtained as empirical Bayes estimates from multilevel models without covariates (for details see supplemental online material, section C). There was a small, statistically significant correlation between the two memory-experience gaps, $r = -.19$, $p = .003$. Within the affect dimensions, the memory-experience gap was unrelated to the within-person correspondence convergence indicators, $|r| < .03$, $p > .642$, but positively correlated with the incremental effect of peak affect, $r > .24$, $p < .001$. The within-person correspondence convergence indicator for PA was positively correlated with the within-person correspondence convergence indicator for NA, $r = .52$, $p < .001$. Furthermore, the incremental effect of peak PA was not correlated with the incremental effect of peak NA, $r = -.10$, $p = .110$ (for further results see Table S4 and Figure S1 in the supplemental online material, section C).

Discussion

Results from Study 1 revealed the expected memory-experience gap for NA, but not for PA. Furthermore, inter-individual differences in this gap could only be found for NA as well. That is, participants on average reported higher NA in retrospective ratings compared to momentary ratings, but individuals differed in the size of this discrepancy. Part of these inter-individual differences could be explained by participants' age: Consistent with expectations derived from SST, the memory-experience gap for NA was reduced with advancing age. Taken together, our research hypotheses with regard to level convergence (Hypotheses 1 and 2) were confirmed with regard to NA but not PA.

Regarding within-person correspondence convergence, results showed strong within-person associations of aggregated momentary ratings and retrospective ratings: Transformed into a correlation metric, the average within-person associations were .60 for PA and .59 for NA. This demonstrates, in support of Hypothesis 3, that both ratings types shared a substantial amount of within-person variance and were suited to differentiate "good days" from "bad days". Our results provided evidence for incremental effects of the last affect reported on the current day: For both PA and NA, including most recent momentary affect improved the prediction of retrospective affect. In contrast, peak affect only predicted retrospective NA, but not PA, above the mean momentary ratings. That is, predictions made based on the peak-and-end rule were fully supported for NA (Hypotheses 4a and 4b), but only partially for PA (Hypothesis 4b). Finally, we found substantial inter-individual differences in the effects of mean, peak and recent affect on end-of-day assessments; hence, individuals differed in the extent to which their retrospective ratings were affected by mean momentary ratings, peak momentary ratings and last momentary ratings, respectively.

Contrary to Hypothesis 5, age did not moderate the effect of peak ratings on retrospective ratings. Exploratory analyses did not reveal any age differences in the impact of mean ratings or most recent ratings.

Further exploratory analyses of inter-individual differences in the respective indicators showed that level convergence and within-person correspondence convergence were uncorrelated, emphasizing the importance to consider both indicators when investigating convergence of rating types. Furthermore, level convergence was positively correlated with the peak effects, which is in line with the assumption that the memory-experience gap is driven by peak effects. Finally, the within-person correspondence convergence indicators were positively correlated, indicating that participants who showed larger correspondence convergence in PA, also showed larger correspondence convergence in NA. This positive association might be driven by inter-individual differences in variables associated with the ability to correctly recall or report emotional states such as, for example episodic memory. For level convergence, the association was reversed: Participants with a larger NA memory-experience gap tended to have a smaller PA memory-experience gap.[5]

With Study 2 we sought to extend these findings in two ways: First, by including a wider age range, we examined potential age differences in level convergence or within-person correspondence convergence beyond the age of 65 years which was the oldest age in Study 1. Second, we employed a measurement burst design (Nesselroade, 1991; Sliwinski, 2008). In such a design, intensive longitudinal bursts (such as EMA) are repeated several times with the same individuals over a longer time frame which allows examining within-person changes in intra-individual effects. In Study 2 we employed three bursts which enabled us to investigate potential within-person changes the convergence indicators across approximately 18 months.

[5] Results involving inter-individual differences in the memory-experience gap for PA need to be interpreted very carefully, because these differences were only small and not statistically significant.

Study 2

Method

The protocol of the study was approved by the Institutional Review Board at Syracuse University (#08-2193; Study Cognition, Health and Aging Project II).

Participants. Data were used from a project aiming at examining the relationships between daily experiences of stress, affect, and cognitive functioning. Participants for this study were community-residing adults recruited via newspaper ads, flyers, and letters describing the project to potentially eligible participants in Syracuse, NY. Inclusion criteria were (1) aged between 20 and 80 years, (2) fluent in English, (3) having a daily waking schedule after 4 a.m. but before 11 a.m., (4) physical ability to operate a palmtop computer, and (5) lack of major cognitive impairment. Of the 214 participants who expressed initial interest, 22 did not meet all inclusion criteria and 12 were no longer interested after receiving detailed descriptions of the study protocol. Five participants did not provide information on relevant study variables resulting in a final sample of 175 participants. Ninety participants were female; the average age of the sample was 49.7 years (*SD* = 17.2; *min* = 20, *max* = 79). Slightly more than half of the participants (58.3%) identified themselves as White, 30.9% as Black, 3.4% as Hispanic, and 7.4% as Other. Approximately one third (32.6%) reported a high school degree as their highest degree, 16.6% held a Bachelor's degree, 9.7% a Master's or Doctorate, 12.0% had no degree, and 29.1% indicated Other.

Procedure. The study started with participants completing a baseline assessment in the laboratory that included assessment of various questionnaires and cognitive testing. During this session, participants also received training on using the mobile devices (Palm Tungsten E2 palmtop computer) to complete the EMA/daily-diary hybrid protocol, described in detail below. After this session, participants followed the protocol at home for a two-day run-in period. After these two days, participants returned to the laboratory for additional assessments and they were given the opportunity to discuss questions about the study and the devices with trained research assistants. After this session, participants were asked to complete the protocol on the mobile device for seven consecutive days and to complete up to seven assessments each day. In this EMA/daily-diary hybrid protocol, participants completed a brief survey upon waking each day, which is not included in the present work. Throughout the day, five random momentary assessments were prompted at semi-randomized time-points (spaced approximately 2-3 hours apart). At the end of each day, participants completed retrospective ratings of the day overall. They were instructed to self-initiate these retrospective surveys within 30 minutes before bed. Data were removed on days when time stamps indicated that the retrospective evening assessment was completed before at least one momentary assessment of the day (n = 14). This procedure (in-lab assessments and EMA/daily-diary hybrid) was repeated approximately nine and eighteen months later which resulted in a measurement burst design with three measurement bursts. Data from 171, 137, and 111 participants were available at the first, second, and third burst, respectively. Within the three bursts, compliance rates were satisfactory for the random assessments (burst 1: 92.8%; burst 2: 94.1%; burst 3: 93.2%) and the evening assessments (burst 1: 86.0%; burst 2: 83.9%; burst 3: 85.7%), respectively. Only the measures relevant for the present study are reported (for additional information on study materials see Neubauer, Smyth, & Sliwinski, 2018). Data from this study have been utilized in previous research; prior publications using these data have targeted different research questions (Mogle, Muñoz, Hill, Smyth, & Sliwinski, 2019; Neubauer et al., 2018, 2019; Scott, Sliwinski et al., 2018; Stawski et al., 2018; Zawadzki et al., 2019; Zhaoyang, Sliwinski, Martire, & Smyth, 2018).

Measurements. As in Study 1, *momentary affect* was assessed at each beep during the waking day. Participants were asked "How ___ do you feel right now?", with eight adjectives, of which four assessed NA (tense; sad; upset; disappointed) and four assessed PA (happy; enthusiastic; content; excited). Responses were given on a scale ranging from 1 (not at all) to 7 (extremely). Responses were averaged into one momentary NA score (within-person $\alpha = .86$) and one momentary PA score (within-person $\alpha = .83$). *Retrospective affect* was assessed using the same adjectives; the time frame of the stem question was changed to "How ___ did you feel today?". The four PA items were averaged into a retrospective PA score (within-person $\alpha = .82$), and the four NA items were averaged into a retrospective NA score (within-person $\alpha = .86$), respectively. Average momentary ratings, peak ratings, and last ratings were created as in Study 1, again resulting in four PA variables (retrospective PA, average momentary PA, peak PA, and last PA) and four NA variables (retrospective NA, average momentary NA, peak NA, and last NA) for each day.

Data Analysis. Overall, the data analytic strategy was similar as in Study 1. The only difference was that we extended the models to three-level models. In order to account for the dependence of observations within a burst, we added random intercepts at the burst level (Level 2); random slopes for substantive Level-1 predictors were only estimated at the person-level (Level 3).

Results and Discussion

Descriptive statistics, correlations, and variance component estimates are depicted in Table 7. Overall, correlations were very similar compared to the results in Study 1, with high between-person correlations of aggregated momentary ratings, peak ratings, most recent ratings, and retrospective ratings within the affect dimensions of $r > .90$, and somewhat lower correlations on the within-person level.

Insert [Table 7 here]

Mean Level Comparisons (Level Convergence). We again ran a multilevel model predicting affect from rating type (see Tables 2 and 3). As in Study 1, retrospective ratings of NA were higher than aggregated momentary ratings, $b = 0.139$, $p < .001$. Additionally, there was a small but statistically significant difference for PA as well: here, too, retrospective ratings were higher than aggregated momentary ratings, *b* = 0.050, *p* = .004. As in Study 1, the random slope variance associated with the predictor rating type was significant for NA, $\chi^2(2) = 17.01$, p < .001, but not PA, $\chi^2(2) = 0.98$, p = .612. Entering gender, age, study day and burst revealed age differences in PA (higher with advancing age, *b* = 0.183, *p* = .003) and NA (lower with advancing age, *b* = -0.159, *p* = .004). Entering two-way interactions of rating type with age revealed significant age x rating type interactions for both NA, *b* = -0.065, *p* = .003, and PA, *b* = 0.035, *p* = .048. Figure 2 shows NA and PA ratings separately for aggregated momentary ratings and retrospective ratings; we used the same age boundaries as in Study 1 and added the age range older than 65 years as a fourth age group in order to make the results more comparable with the results from Study 1. As in Study 1, retrospective NA ratings were higher than aggregated momentary ratings, and this difference was attenuated with advancing age. For the three youngest age groups, the difference between retrospective NA and aggregated momentary NA was statistically significant: 25-40 years: *b* = 0.214, *p* < .001; 41-52 years: *b* = 0.181, *p* < .001; 53-65 years: *b* = 0.089, *p* = .013. For the oldest age group (66-79 years), the difference failed to reach statistical significance, *b* = 0.060, *p* = .079.

In contrast to Study 1, there was an age x rating type interaction for PA. Follow-up analyses revealed that there was no evidence for a memory-experience gap for PA in the youngest three age groups, *b* < 0.032, *p* > .387, but a statistically significant difference between aggregated momentary ratings and retrospective ratings in the oldest age group, *b* = 0.155, *p* < .001. In follow-up analyses, we found a significant quadratic age x rating type interaction for PA.[6] Inspection of the regression coefficients revealed a significant age squared x rating type interaction, b = 0.070, p < .001, suggesting that the memory-experience gap for PA was accelerated with advancing age.

[6] We again tested for nonlinear age effects by adding higher order polynomials of age into the model. This did not improve model fit for NA, $\chi^2(2) < 4.15$, *p* > .125. For PA, model fit improved when adding quadratic age, $\chi^2(2) = 13.19$, *p* = .001.

There was no burst x rating type interaction for either PA or NA, $p > .770$, hence providing no evidence for within-person change in the memory-experience gap across the 18 month observation period.[7] There were no gender x rating type, $p > .114$ for all, and day x rating type interactions either, $p > .207$ for all, and including these interactions did not alter the conclusions drawn (see Table S3 in the supplemental online material, section B).

Insert [Figure 2 here]

In post-hoc analyses, we repeated the analyses separately for high arousal PA (mean of the items enthusiastic and excited) and low arousal PA (mean of the items happy and content). Results (see supplemental online material, section D, Table S5) showed that there was no main effect of rating type, and no age x rating type interaction for low arousal PA, $p > .226$. For high arousal PA, the pattern of results was the same as for overall PA: There was a main effect of rating type, $b = 0.075$, $p < .001$, that was moderated by age $b = 0.051$, $p = .014$. Following up the interaction showed that only for the oldest age group, $b = 0.199$, $p < .001$, but not for the three other age groups, $b < 0.040$, $p > .322$, was there evidence for a memory-experience gap for high arousal PA.

In summary, in support of research Hypothesis 1 there was a memory-experience gap for both NA and PA in Study 2. Further, age was positively associated with the memory-experience gap for PA and negatively associated with the memory-experience gap for NA, in line with Hypothesis 2.

Predicting End-of-Day Affect (Correspondence Convergence). Aggregated momentary ratings in NA explained 44.3% of the within-person variance of retrospective NA ratings. For PA, the estimate was comparable, pseudo-$R^2 = .386$ (see Table 4). Further including peak affect and the last available affect ratings increased explained variance for NA, $\chi^2(2) = 89.59$, $p < .001$, $\Delta_{R^2} = 2\%$. For PA, the increase in model fit was statistically

[7] We also explored potential age-dependent within-person changes: there was, however, no age x burst x rating type interaction for either PA, $b = 0.006$, $p = .790$, or NA, $b = -0.001$, $p = .970$.

significant, as well, $\chi^2(2) = 6.19$, $p = .045$, but the increase in Pseudo-R^2 was small, $\Delta_{R^2} = .05\%$. Investigating the unique effects of aggregated momentary affect, peak affect, and most recent affect on retrospective ratings showed that, as in Study 1, recent NA, $b = 0.158$, $p < .001$, and peak NA, $b = 0.077$, $p = .012$, predicted end-of day NA beyond this day's mean (see Table 5, Model 7b). Again, none of the interactions of age with momentary affect ratings were significant, $|b| < 0.016$, $p > .580$. For PA (see Table 6, Model 9b), neither the unique effects of peak affect, $b = -0.014$, $p = .716$, nor recent affect, $b = 0.033$, $p = .190$, were significant. Including the interactions with age revealed a significant age x peak PA interaction, $b = 0.113$, $p = .006$. To follow up this interaction, we split the sample by age groups. Results revealed no significant peak effects for the youngest three age groups, $|b| < 0.117$, $p > .056$ for all, but a significant peak effect for the oldest age group, $b = 0.294$, $p < .001$.

In the final models, we examined if within-person correspondence convergence changes either across days or across bursts. Including the interactions of aggregated momentary affect, peak affect, and most recent affect with day did not improve model fit for either PA, $\chi^2(3) = 0.98$, $p = .806$, or NA, $\chi^2(3) = 3.84$, $p = .279$. The aggregated momentary affect x day interaction was not statistically significant for either PA, $b = -0.012$, $p = .496$, or NA, $b = -0.019$, $p = .306$, yielding no evidence for a change in within-person correspondence convergence across days within a burst. Including two-way interactions of burst with aggregated momentary affect, peak affect, and most recent affect, respectively, did not improve model fit for PA, $\chi^2(3) = 5.18$, $p = .159$; the burst x aggregated momentary affect interaction was not significant, $b = -0.024$, $p = .620$. For NA there was a statistically significant improvement in model fit, $\chi^2(3) = 9.21$, $p = .027$. Of the three interaction effects, only the burst x most recent affect interaction was significant, $b = 0.058$, $p = .022$, indicating that the impact of most recent NA ratings on retrospective NA increased somewhat across the

18 month study period. There was, however, no evidence for a change in within-person correspondence convergence across bursts since the burst x aggregated momentary affect interaction was not significant, $b = 0.039$, $p = .445$.

Taken together, results yielded support for a robust within-person association between aggregated momentary affect and retrospective affect, in line with Hypothesis 3. Peak and end effects (Hypotheses 4a and 4b) were found for NA, but not PA. Age differences in the incremental effect of peak affect on retrospective affect were found for PA, but not NA. Hence, Hypothesis 5 was only partially supported by results of Study 2.

Exploratory Analyses: Inter-individual Differences. Person-specific estimates were obtained for level convergence, within-person correspondence convergence, and the effect of peak affect on retrospective affect as in Study 1. Results (see Table S4 and Figure S2 in the supplemental online material, section C) revealed the same pattern among these indicators as in Study 1: The two memory-experience gaps were negatively correlated, $r = -.46$, $p < .001$. Within each affect dimension, the memory-experience gap was unrelated to the within-person correspondence convergence indicators, $|r| < .11$, $p > .140$, and positively correlated with the effect of peak affect, $r > .20$, $p < .008$. The two within-person correspondence convergence indicators were positively correlated, $r = .36$, $p < .001$, and the peak affect effects were uncorrelated, $r = -.12$, $p = .130$. By and large, these results supported the conclusions drawn in Study 1: (a) indicators of level convergence and within-person correspondence convergence were uncorrelated; (b) level convergence was positively correlated with the peak effects (suggesting an important role of peak affect for the memory-experience gap); (c) indicators of within-person correspondence convergence were positively correlated with each other, indicators of level convergence were negatively correlated with each other.[8]

General Discussion

[8] As in Study 1, results involving inter-individual differences in the memory-experience gap for PA need to be interpreted carefully, because these differences were only small and not statistically significant.

Within-person fluctuations in affective experiences have recently gained increasing attention in the empirical literature (e.g., Augustine & Larsen, 2012; Brans, Koval, Verduyn, Lim, & Kuppens, 2013). Two methods frequently employed in this context are ecological momentary assessments (which assess momentary affect, often several times per day) and daily-diary methods (which typically assess retrospective affect once at the end of the day). Based on theoretical accounts of reconstructive processes (Robinson & Clore, 2002), differences between the experiencing self and the remembering self (Conner & Barrett, 2012) and the peak-and-end rule (Fredrickson, 2000), differences between retrospective ratings ("How did you feel today?") and aggregated momentary ratings (the mean of ratings throughout the day; "How do you feel right now?") can be expected, but the degree of these divergences has not been thoroughly investigated thus far. Additionally, the degree of convergence might vary across individuals, potentially as a function of person-level characteristics. In the present study, we examined two indicators of convergence in two EMA/daily-diary hybrid studies: level convergence (defined as the absence of mean level differences between the retrospective and aggregated momentary ratings) and within-person correspondence convergence (defined as the within-person association of retrospective and aggregated momentary ratings). Furthermore, we examined age as one potential predictor of inter-individual differences in these convergence indicators. We will first discuss the findings on level convergence before we will turn to the findings on correspondence convergence. We will then integrate these findings and discuss potential mechanisms underlying the convergence/divergence of momentary and retrospective affect measures, as well as implications for studies investigating within-person fluctuations in affective experiences.

Level Convergence: The Memory-Experience Gap

In both studies, we found evidence for a memory-experience gap for NA, supporting Hypothesis 1. That is, retrospective NA ratings were higher than the average momentary ratings provided throughout participants' daily lives. This finding is in line with previous research (Ben-Zeev et al., 2009; Lay et al., 2016; Miron-Shatz et al., 2009; Thomas & Diener, 1990). Most of these studies (e.g., Ben-Zeev et al., 2009; Thomas & Diener, 1990; Lay et al., 2016) collected momentary ratings in participants' daily lives, but retrospective ratings were not collected regularly at the end of each day but only at the end of the data collection period. Miron-Shatz et al. (2009) assessed retrospective affect at the end of the day, but "momentary" ratings were obtained via the day reconstruction method and not collected in an EMA. In contrast, our study compared momentary ratings collected in an EMA to retrospective assessments collected at the end of each day, thereby going beyond previous research on this issue. Our findings further provided evidence consistent with Hypothesis 2 for NA: Results revealed a linear decrease in the memory-experience gap for NA with increasing age in both studies.

For PA, there was only limited evidence for a memory-experience gap in the present study: Retrospective PA ratings did not differ from aggregated momentary ratings in Study 1, and they were only slightly higher in Study 2. Follow-up analyses showed that this small difference was driven by the oldest age group (66-79 years), suggesting that only the oldest adults investigated in the present research reported more intense retrospective PA ratings than momentary ratings. Hence, Hypotheses 1 and 2 received mixed support for PA in our analyses. Previous findings on the PA memory-experience gap are rather mixed, too. Although both Parkinson et al. (1995) and Miron-Shatz et al. (2009) reported higher retrospective PA ratings than aggregated momentary ratings in their studies, in the latter study this gap was smaller than the respective gap for NA. Furthermore, Lay et al. (2016) reported a PA memory-experience gap only for high arousal PA. Taking a closer look at the items used to assess PA in the present study, differences in arousal captured by these items might help explain differences in the results obtained in the two present studies. Whereas none of the four items assessing PA in Study 1 (happy; pleased; enjoyment/fun; joyful) captured very high arousing PA, two of the items used in Study 2 (enthusiastic; excited) are face-valid high arousal indicators. Lay et al. (2016) suggested that remembering high arousal affect would be more relevant from an evolutionary perspective than remembering low arousal affect, so the memory-experience gap (which the authors suggested to be a consequence of peak effects) should be larger for high arousal than for low arousal affect. Results from exploratory analyses in Study 2 were in line with this post-hoc explanation: Only for high arousal PA, but not for low arousal PA was there evidence for a memory-experience gap that was exacerbated with advancing age.

Correspondence Convergence – Discriminating Good Days from Bad Days

In addition to the memory-experience gap, we also investigated the degree of correspondence convergence (relative convergence) of aggregated momentary affect ratings and retrospective affect ratings. Our findings showed that for both PA and NA, retrospective ratings collected at the end of the day were strongly associated with aggregated momentary ratings on both the between-person level and the within-person level. The former finding replicates and extends results reported in prior research (Ben-Zeev et al., 2009; Röcke et al., 2011) and suggests that both ratings types are very well suited to distinguish persons with higher PA (NA) from persons with lower PA (NA). Extending prior research to the within-person level, our results showed—in line with Hypothesis 3—a substantial within-person association of retrospective and aggregated momentary ratings: These two rating types shared between 35% and 44% of within-person variance, corresponding to a within-person correlation of .59 to .66. This high overlap suggests that both rating types are capable of differentiating days with relatively higher PA (NA) from days with relatively lower PA (NA); that is, both aggregated momentary and retrospective daily ratings can reasonably discriminate “good days” from “bad days”.

This finding conveys important information for researchers working with daily-diary studies: One advantage often claimed for this type of research is that it captures “life as it is lived” (Bolger, Davis, & Rafaeli, 2003). However, when asking study participants to report their affective experiences retrospectively (“today”), the obtained ratings might still be confounded by retrospective biases.[9] Because emotional experiences cannot be re-lived they need to be reconstructed, opening up the possibility for biased assessments (Robinson & Clore, 2002).

Peak and end effects might be one source of such biases: Instead of accurately reporting how they felt on average, individuals might rely on the most intense or most recent affect experienced today (Fredrickson, 2000). Across both studies, retrospective NA was predicted by both peak NA and most recent NA above and beyond mean momentary NA, supporting Hypotheses 4a and 4b for NA. Hence, although there was a strong association of mean momentary ratings and retrospective ratings for NA, retrospective ratings were slightly biased in the direction of peak and most recent affect experienced. Although our results were consistent with predictions made by the peak-and-end rule for NA, evidence was mixed for PA: recent PA predicted retrospective PA ratings only in Study 1 but not in Study 2, lending only partial support to Hypothesis 4a for PA. Hypothesis 4b received no support for PA, because peak PA had no overall effect on retrospective PA in either Study 1 or Study 2. Despite some incremental effects of peak and most recent affect, we hasten to add that these effects were quite small in magnitude (less than 3% increase in explained variance above mean momentary ratings). Some of the remaining discrepancies between aggregated momentary affect and retrospective affect might be explained by the impact of salient daily

---

[9] However, end-of-day assessments are probably less prone to retrospective biases than ratings of a longer recall period (e.g., “How did you feel in the last four weeks?”). Hence, rather than as a dichotomy, the relative contributions of the remembering self / experiencing self (Conner & Barrett, 2012) or episodic / semantic memory processes (Robinson & Clore, 2002) to rendering the response to affect measures might be better represented as a continuum: End-of-day assessments probably engage less semantic processes than retrospective assessments of longer recall periods, but more semantic processes than momentary assessments.

events. Salience is one of the key mechanisms discussed to explain peak and end effects (e.g., Lay et al., 2016), but these two moments are probably not the only salient events an individual encounters in daily life. For example, when reflecting about today's affect, an individual might recall her affect during an unusually pleasant interaction she has had with her supervisor today. This salient memory might bias her retrospective affect assessment towards the affect experienced during this interaction, leading to a less-then-perfect within-person correspondence convergence that cannot be explained by peak and end effects.

In our final hypothesis (Hypothesis 5), we predicted that the effect of peak affect on retrospective affect would be more pronounced with advancing age. This hypothesis was derived from SAVI (Charles, 2010). According to this model, old age is associated with vulnerabilities in the regulation of affect in situations of high levels of emotional arousal. The postulated adaptive value of peak affect – providing incremental information on the maximal capacity requirements of a situation (Fredrickson, 2000) – might therefore be particularly important with advancing age. Our data were only partially consistent with this prediction. Only for PA, and only in Study 2, did we find evidence for age-associated changes in the effect of peak ratings on retrospective ratings. Peak PA predicted retrospective PA only in the oldest age group investigated (66-79 years).

Mechanisms and Implications

Overall, a clear pattern of results was visible across both studies and both PA and NA. First, peak effects and the memory-experience gap were in all instances observed together. Peak NA consistently predicted retrospective NA above and beyond mean NA on this day, and NA consistently exhibited evidence for a memory-experience gap. For PA, which exhibited a memory-experience gap for the oldest adults only, we too found support for a peak effect only for this age group. Second, results of the analyses investigating inter-individual differences in convergence indicators revealed that participants who exhibited stronger effects of peak affect on retrospective affect (i.e., participants who were more strongly “pulled” towards their peak affect) also had larger memory-experience gaps. These results are consistent with the assumption of peak effects driving the memory-experience gap proposed by Lay et al. (2016).

Furthermore, results were more supportive of our research hypotheses for NA than for PA. This is largely in line with previous research that has, for example, often reported a memory-experience gap for NA, but not (or smaller) for PA (Ben-Zeev et al., 2009; Miron-Shatz et al., 2009), and more consistent peak and end effects for NA than for PA (Ganzach & Yaor, 2019; Röcke et al., 2011). On a conceptual level, the differences between NA and PA in this regard have often been explained by the “bad-is-stronger-than-good” hypothesis (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). For example, Ganzach and Yaor (2019) suggested that peak NA is more salient than peak PA, and that therefore, peak effects should be stronger for the former than the latter affect dimension (and, consequently, the memory-experience gap should be larger for NA than for PA). If salience is indeed the mechanism driving peak effects and the memory-experience gap, this would suggest that these effects are modulated when there is a shift towards higher salience of positive information. According to SST (Carstensen et al., 1999), such a shift occurs with advancing age: The perception of reduced time to live prompts individuals to increasingly turn towards valuing emotional well-being over gaining information which leads to an altered information processing preference favoring positive (relative to negative) information. Whether this change in information processing style can be best understood as a decrease in negativity, an increase in positivity, or both, is debated. A reduction in negativity would entail that negative information becomes less salient, leading to a smaller memory-experience gap for NA, and a reduced impact of peak NA on retrospective NA. An increase in positivity should be associated with increased salience of positive information, and hence a larger memory-experience gap for PA and a larger impact of peak PA on retrospective PA.

When interpreting our findings in light of these considerations, they hint towards two complementary processes that differ in their age-associated dynamics: First, the quadratic age effect on the memory-experience gap for PA suggests that positivity might not increase in a linear fashion across the life span, but only be observed among very old adults. Our results further indicate that only for older adults, peak PA is salient enough to "pull" retrospective PA ratings away from the aggregated momentary ratings. Second, the linear decrease in the memory-experience gap for NA across the investigated age range might indicate a reduction of negativity over a large part of the adult life span. However, there was no decrease in the effect of peak NA, suggesting that peak NA remains salient across the life span. Hence, the reduction (and virtual elimination) of the memory-experience gap for NA in very old adults cannot be explained by a reduced peak effect.

If not salience of peak experiences, then other factors have to be considered to explain the age-associated reduction in the NA memory-experience gap. A decrease in episodic memory performance (leading to a reduced ability to remember today's NA) that is compensated by recruiting semantic memory processes could be one alternative explanation: Given restrictions in episodic memory, older adults might use semantic information ("How do I think I felt?") more so than episodic information ("How did I feel today?"). This would result in a weaker association between aggregated momentary assessments and retrospective assessments with advancing age. However, this seems to be an unlikely explanation for the present results, because there was no effect of age on the within-person association between aggregated momentary NA and retrospective NA. Because our sample comprised cognitively healthy individuals (major cognitive impairment was an exclusion criterion for participation in Study 2), cognitive decrements might not have been severe enough to lead to a worse convergence in the older adults investigated here, but this could be an interesting avenue for future research.

Instead, age-associated changes in reappraisal of daily events seem to be a better explanation for the present findings: With advancing age, participants might evaluate a negative experience (e.g., their NA experienced today) as less negative than younger individuals, leading to a smaller memory-experience gap with older age. This is in line with findings on the reported severity of stressful events. For example, older adults often rate the severity of a minor stressor in their daily lives as less severe than younger adults (Almeida & Horn, 2004; Neubauer et al., 2019). Notably, such differences emerged even though the stressors reported by younger and older adults were of the same "objective severity" (as obtained by observer ratings; Almeida & Horn, 2004). According to SAVI, the increasing use of efficient reappraisal strategies is one of the strengths associated with older individuals' life long experience in emotion regulation. This use of reappraisal strategies (or other efficient emotion regulations strategies) might explain the age-associated reduction in the NA memory-experience gap despite an unaltered peak NA effect.

In summary, our results in combination with these elaborations suggest that the PA memory-experience gap might increase with advancing age because peak PA is more salient in older adults (and hence, "pulls" retrospective affect ratings more towards peak PA). This increased salience of peak PA could be a result of an age-associated positivity bias that is only observed in old adults. The NA memory-experience gap decreases with older age, which might be a consequence of age-associated improvements in emotion regulation (e.g., reappraisal processes) that lead to a less exaggerated retrospective assessment of daily NA.

Relevance for Intensive Longitudinal Studies

For future studies investigating affective dynamics within-individuals, it is important to note that the association between mean momentary ratings and retrospective ratings was substantial, but far from perfect. On the one hand, this indicates that both assessment types capture common sources of variation that can be used to distinguish "good days" from "bad days". On the other hand, results also suggest that each assessment type has unique sources of variation. On a conceptual level, momentary ratings are well-suited to capture temporally fine-grained predictors and consequences of fluctuations in affect, whereas end-of-day retrospective assessments obscure temporality throughout the day. A further difference between these two rating types relates to the notion that retrospective assessments are, in some cases, expected to be more predictive of deliberate future behavior than aggregated momentary ratings (Redelmeier et al., 2003; Wirtz et al., 2003), whereas momentary assessments would often be expected to be more closely related to physiological indicators such as cortisol secretion or heart rate (see Conner & Barrett, 2012). Taken together, momentary ratings and retrospective ratings provide overlapping but not necessarily exchangeable information. Researchers interested in intra-individual fluctuations of affect need to weigh the costs and benefits of using either retrospective assessments in daily-diary studies (which might be slightly biased towards peak and recent affect, or other salient events of the day, and influenced by the use of reappraisal processes, but more readily and easily collected) or EMA type of data collection (which increases participant burden and might be more costly, but allows insight into the temporal unfolding of ebbs and flows, as well as predictors and consequences of within-day affect). These findings cannot provide clear cut recommendations regarding which procedure should be preferred, and this decision will certainly need to be made on a case-by-case basis in accordance with research questions and goals. They do, however, suggest that daily retrospective affect assessments are not heavily biased and can therefore provide similar (although not identical) insights into the *day-to-day* fluctuations of affect as do aggregated momentary affect ratings.

Future research might consider expanding the present results in several ways. First, it might be worthwhile to examine the effect of different time frames for the recall period on convergence indicators. Building on results by Charles et al. (2016) and Walentynowicz, Schneider, and Stone (2018), the memory-experience gap may increase when comparing aggregated momentary ratings to retrospective ratings over a longer time frame (e.g., retrospective assessments after one week instead of one day). It remains unclear, however, how a longer recall period might affect within-person correspondence convergence and age differences in the convergence indicators. Second, the results of the present study are based on intensity ratings of affective states, and it remains an open question whether the same pattern of results would emerge when using frequency ratings for the assessment of momentary and retrospective affect. Third, findings on inter-individual differences in convergence indicators suggest that level convergence and within-person correspondence convergence are largely unrelated. Hence, some individuals might, for example, consistently overestimate their NA (low level convergence), but at the same time be quite good at distinguishing better days from worse days in their retrospective assessments (high within-person correspondence convergence) or vice versa. Future studies should examine, what processes account for these different convergence indicators (e.g., episodic and semantic memory; memory biases due to motivational preferences; salience) to better understand the properties of aggregated momentary and retrospective affect ratings.

Strengths and Limitations

Results of the present study are based on two age-heterogeneous samples that were assessed in an intensive longitudinal design for up to 14 (Study 1) and 21 days (Study 2). By assessing study participants' current affect several times per day and retrospective affect at the end of each day, we were in a position to examine both between- and within-person associations of aggregated momentary assessments with retrospective assessments for both

PA and NA. To the best of our knowledge, this is the first study that approached the question of convergence of momentary ratings and retrospective ratings in large, heterogeneous samples using this strong methodological approach. Despite these strengths, a number of limitations must be considered. First, although both samples were heterogeneous with respect to age, the age range was restricted and did not cover childhood and adolescence or very old adulthood (e.g., Riediger, Schmiedek, Wagner, & Lindenberger, 2009). Different patterns might be observed when targeting, for example, a population of very old adults (80 years and older) for whom accumulated losses and physiological restraints overpower age-associated strengths in emotion regulation (Charles, 2010). Second, the age-related effects we found were based on cross-sectional age associations. We note that such cross-sectional effects do not allow firm conclusions about developmental changes because they confound developmental effects with cohort effects (e.g., Lindenberger, von Oertzen, Ghisletta, & Hertzog, 2011). Although we investigated longitudinal change in these effects in Study 2, the follow-up period was rather short (18 months), in particular compared to the cross-sectional age differences in our sample (20-79 years). Measurement burst studies with longer inter-burst intervals are required to shed light onto the developmental dynamics at play. Third, momentary affect was assessed only five times per day and more intense sampling schemes might better capture the “true” mean and peak affect on a day. Our sampling scheme might have missed high intense affective states, in particular on high stress days, which might contribute to the less-than-perfect association between mean momentary ratings and retrospective ratings. Future studies should consider including more momentary ratings each day to better accommodate the highly volatile affective experiences in daily life. Future research might also consider examining the role of within-day affect variability for the convergence between the rating types (see Stone et al., 2005). We note, however, that obtaining reliable estimates of within-day variance in affective experiences would likely require more than five daily assessments (e.g., Estabrook, Grimm, & Bowles, 2012; Wang & Grimm, 2012). Fourth, the reports of momentary affect during the day may have reinforced memory for end of day recall. That is, our results might be biased in favor of convergence due to the fact that each momentary assessment provides a prompt not only to report their current state, but also to encode and remember it for recall later in the day. Future studies might consider examining the effects of repeated momentary samples throughout the day on end-of-day assessments (e.g., by comparing a group of participants investigated in an EMA/daily-diary hybrid design to a group reporting only retrospective assessments at the end of the day). Furthermore, our results showed somewhat larger effect sizes for most recent affect (vs. peak affect). This finding might in part be driven by state dependent memory (given that the last assessment was close in time to the retrospective assessment). We aimed to reduce this overlap by collecting these measures at different occasions on the same day (they were on average about 3 hours apart), but future research might consider to separate them further (e.g., by assessing retrospective affect on the next morning).

Conclusions

Asking individuals to recall, summarize, and report how they felt over a certain period of time can result in divergence from the experiences the participants actually had during the targeted period – even over relatively short periods of time, such as one day. Previous findings on the memory-experience gap suggest that individuals overstate both positive and negative affect when asked to judge them retrospectively. Biasing effects of peak ratings have been postulated, but these peak effects have previously not been thoroughly examined from a within-person perspective. Our findings from two studies employing ecological momentary assessment/daily-diary hybrid studies comparing retrospective end-of-day assessments to momentary assessments collected throughout the day showed (a) a memory-experience gap for negative affect and (only for very old adults) positive affect, (b) very strong between-

person correspondence among retrospective and aggregated momentary ratings, and (c) acceptable within-person correspondence of these two rating types. Within-day peak and most recent negative affect consistently predicted daily retrospective negative affect. Results further suggested inter-individual differences in the convergence of these two rating types that were partially associated with participants' age. Overall, findings provide valuable information that helps inform affect research, aging theory, and methodological approaches to studying intra-individual processes at the day and within-day levels.

References

Aldwin, C. M. (2010). *Stress, coping, and development: An integrative perspective* (2nd ed.). New York: Guilford.

Almeida, D. M. (2005). Resilience and vulnerability to daily stressors assessed via diary methods. *Current Directions in Psychological Science*, *14*, 64–68. https://doi.org/10.1111/j.0963-7214.2005.00336.x

Almeida, D. M., & Horn, M. C. (2004). Is daily life more stressful during middle adulthood? In O. G. Brim, C. D. Ryff, & R. C. Kessler (Eds.), *How healthy are we? A national study of well-being at midlife* (pp. 425–451). Chicago, IL: University of Chicago Press.

Augustine, A. A., & Larsen, R. L. (2012). Emotion research. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 497–510). New York: Guilford Press.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370. https://doi.org/10.1037/1089-2680.5.4.323

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

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Brans, K., Koval, P., Verduyn, P., Lim, Y. L., & Kuppens, P. (2013). The regulation of negative and positive affect in daily life. *Emotion*, *13*, 926–939. https://doi.org/10.1037/a0032400

Brose, A., Schmiedek, F., Lovden, M., & Lindenberger, U. (2011). Normal aging dampens the link between intrusive thoughts and negative affect in reaction to daily stressors. *Psychology and Aging*, *26*, 488–502. https://doi.org/10.1037/a0022287

Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist*, *54*, 165–181. https://doi.org/10.1037/0003-066X.54.3.165

Charles, S. T., Piazza, J. R., Mogle, J. A., Urban, E. J., Sliwinski, M. J., & Almeida, D. M. (2016). Age differences in emotional well-being vary by temporal recall. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *71*, 798–807. https://doi.org/10.1093/geronb/gbv011

Charles, S. T. (2010). Strength and vulnerability integration: a model of emotional well-being across adulthood. *Psychological Bulletin*, *136*, 1068–1091. https://doi.org/10.1037/a0021232

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine*, *74*, 327–337. https://doi.org/10.1097/PSY.0b013e3182546f18

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. https://doi.org/10.1037/h0040957

Estabrook, R., Grimm, K. J., & Bowles, R. P. (2012). A Monte Carlo simulation study of the reliability of intraindividual variability. *Psychology and Aging*, *27*, 560–576. https://doi.org/10.1037/a0026669

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

Ganzach, Y., & Yaor, E. (2019). The retrospective evaluation of positive and negative affect. *Personality & Social Psychology Bulletin, 45, 93–104*. https://doi.org/10.1177/0146167218780695

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*, 72–91. https://doi.org/10.1037/a0032138

Graham-Engeland, J. E., Sin, N. L., Smyth, J. M., Jones, D. R., Knight, E. L., Sliwinski, M. J., . . . Engeland, C. G. (2018). Negative and positive affect as predictors of inflammation: Timing matters. *Brain, Behavior, and Immunity*, *74*, 222–230. https://doi.org/10.1016/j.bbi.2018.09.011

Grühn, D., Scheibe, S., & Baltes, P. B. (2007). Reduced negativity effect in older adults' memory for emotional pictures: the heterogeneity-homogeneity list paradigm. *Psychology and Aging*, *22*, 644–649. https://doi.org/10.1037/0882-7974.22.3.644

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications (2. ed.). Quantitative methodology series.* New York, NY: Routledge.

Hyun, J., Sliwinski, M. J., & Smyth, J. M. (2019). Waking up on the wrong side of the bed: The effects of stress anticipation on working memory in daily life. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *74*, 38–46. https://doi.org/10.1093/geronb/gby042

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, *7*, 297–300. https://doi.org/10.1177/1754073915590947

LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods*, *17*, 433–451. https://doi.org/10.1177/1094428114541701

Lay, J. C., Gerstorf, D., Scott, S. B., Pauly, T., & Hoppmann, C. A. (2016). Neuroticism and extraversion magnify discrepancies between retrospective and concurrent affect reports. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12290

Lindenberger, U., Oertzen, T. von, Ghisletta, P., & Hertzog, C. (2011). Cross-sectional age variance extraction: what's change got to do with it? *Psychology and Aging*, *26*, 34–47. https://doi.org/10.1037/a0020525

Majd, M., Graham-Engeland, J. E., Smyth, J. M., Sliwinski, M. J., Lipton, R. B., Katz, M. J., & Engeland, C. G. (2018). Distinct inflammatory response patterns are evident among men and women with higher depressive symptoms. *Physiology & Behavior*, *184*, 108–115. https://doi.org/10.1016/j.physbeh.2017.11.009

Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: the positivity effect in attention and memory. *Trends in Cognitive Sciences*, *9*, 496–502. https://doi.org/10.1016/j.tics.2005.08.005

Mathur, A., Graham-Engeland, J. E., Slavish, D. C., Smyth, J. M., Lipton, R. B., Katz, M. J., & Sliwinski, M. J. (2018). Recalled early life adversity and pain: The role of mood, sleep, optimism, and control. *Journal of Behavioral Medicine*, *41*, 504–515. https://doi.org/10.1007/s10865-018-9917-8

Mejía, S., Hooker, K., Ram, N., Pham, T., & Metoyer, R. (2014). Capturing intraindividual variation and covariation constructs: Using multiple time-scales to assess construct reliability and construct stability. *Research in Human Development*, *11*, 91–107. https://doi.org/10.1080/15427609.2014.906728

Miron-Shatz, T., Stone, A., & Kahneman, D. (2009). Memories of yesterday's emotions: Does the valence of experience affect the memory-experience gap? *Emotion*, *9*, 885–891. https://doi.org/10.1037/a0017823

Mogle, J., Muñoz, E., Hill, N. L., Smyth, J. M., & Sliwinski, M. J. (2019). Daily memory lapses in adults: Characterization and influence on affect. *The Journals of Gerontology.*

Series B, Psychological Sciences and Social Sciences, *74*, 59–68. https://doi.org/10.1093/geronb/gbx012

Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. S. Palermo (Eds.), *Visions of aesthetics, the environment, and development: The legacy of Joachim F. Wohwill* (pp. 213–240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Neubauer, A. B., Smyth, J. M., & Sliwinski, M. J. (2018). When you see it coming: Stressor anticipation modulates stress effects on negative affect. *Emotion*, *18*, 342–354. https://doi.org/10.1037/emo0000381

Neubauer, A. B., Smyth, J. M., & Sliwinski, M. J. (2019). Age differences in proactive coping with minor hassles in daily life. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *74*, 7–16. https://doi.org/10.1093/geronb/gby061

Parkinson, B., Briner, R. B., Reynolds, S., & Totterdell, P. (1995). Time frames for mood: Relations between momentary and generalized ratings of affect. *Personality and Social Psychology Bulletin*, *21*, 331–339.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.

Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain*, *104*, 187–194. https://doi.org/10.1016/S0304-3959(03)00003-4

Riediger, M., Schmiedek, F., Wagner, G. G., & Lindenberger, U. (2009). Seeking pleasure and seeking pain: Differences in prohedonic and contra-hedonic motivation from adolescence to old age. *Psychological Science*, *20*, 1529–1535. https://doi.org/10.1111/j.1467-9280.2009.02473.x

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*, 934–960.

Röcke, C., Hoppmann, C. A., & Klumb, P. L. (2011). Correspondence between retrospective and momentary ratings of positive and negative affect in old age: Findings from a one-year measurement burst design. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *66*, 411–415. https://doi.org/10.1093/geronb/gbr024

Schwarz, N. (2012). Why researchers should think "real-time". A cognitive rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 22–42). New York: Guilford Press.

Scott, S. B., Graham-Engeland, J. E., Engeland, C. G., Smyth, J. M., Almeida, D. M., Katz, M. J., . . . Sliwinski, M. J. (2015). The Effects of Stress on Cognitive Aging, Physiology and Emotion (ESCAPE) Project. *BMC Psychiatry*, *15*, 146. https://doi.org/10.1186/s12888-015-0497-7

Scott, S. B., Kim, J., Smyth, J. M., Almeida, D. M., & Sliwinski, M. J. (2019). Additive effects of forecasted and reported stressors on negative affect. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *74*, 29–37. https://doi.org/10.1093/geronb/gby068

Scott, S. B., Munoz, E., Mogle, J. A., Gamaldo, A. A., Smyth, J. M., Almeida, D. M., & Sliwinski, M. J. (2018). Perceived neighborhood characteristics predict severity and emotional response to daily stressors. *Social Science & Medicine*, *200*, 262–270. https://doi.org/10.1016/j.socscimed.2017.11.010

Scott, S. B., Ram, N., Smyth, J. M., Almeida, D. M., & Sliwinski, M. J. (2017). Age differences in negative emotional responses to daily stressors depend on time since event. *Developmental Psychology*, *53*, 177–190. https://doi.org/10.1037/dev0000257

Scott, S. B., Sliwinski, M. J., Zawadzki, M., Stawski, R. S., Kim, J., Marcusson-Clavertz, D., . . . Smyth, J. M. (2018). A coordinated analysis of variance in affect in daily life. *Assessment*. Advance online publication. https://doi.org/10.1177/1073191118799460

Slavish, D. C., Sliwinski, M. J., Smyth, J. M., Almeida, D. M., Lipton, R. B., Katz, M. J., & Graham-Engeland, J. E. (2018). Neuroticism, rumination, negative affect, and sleep: Examining between- and within-person associations. *Personality and Individual Differences*, *123*, 217–222. https://doi.org/10.1016/j.paid.2017.11.023

Sliwinski, M. J. (2008). Measurement-burst designs for social health research. *Social and Personality Psychology Compass*, *2*, 245–261. https://doi.org/10.1111/j.1751-9004.2007.00043.x

Sliwinski, M. J., Mogle, J. A., Hyun, J., Munoz, E., Smyth, J. M., & Lipton, R. B. (2018). Reliability and validity of ambulatory cognitive assessments. *Assessment*, *25*, 14–30. https://doi.org/10.1177/1073191116643164

Stawski, R. S., Scott, S. B., Zawadzki, M. J., Sliwinski, M. J., Marcusson-Clavertz, D., Kim, J., . . . Smyth, J. M. (2018). Age differences in everyday stressor-related negative affect: A coordinated analysis. *Psychology and Aging.* Advance online publication. https://doi.org/10.1037/pag0000309

Stone, A. A., Broderick, J. E., & Schwartz, J. E. (2010). Validity of average, minimum, and maximum end-of-day recall assessments of pain and fatigue. *Contemporary Clinical Trials*, *31*, 483–490. https://doi.org/10.1016/j.cct.2010.06.004

Stone, A. A., Schwartz, J. E., Broderick, J. E., & Shiffman, S. S. (2005). Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. *Personality & Social Psychology Bulletin*, *31*, 1340–1346. https://doi.org/10.1177/0146167205275615

Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology*, *59*, 291–297. https://doi.org/10.1037/0022-3514.59.2.291

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*, 383–403. https://doi.org/10.1037/0033-2909.134.3.383

Walentynowicz, M., Schneider, S., & Stone, A. A. (2018). The effects of time frames on self-report. *PloS One*, *13*, e0201655. https://doi.org/10.1371/journal.pone.0201655

Wang, L. P., & Grimm, K. J. (2012). Investigating reliabilities of intraindividual variability indicators. *Multivariate Behavioral Research*, *47*(5), 771–802. https://doi.org/10.1080/00273171.2012.715842

Wirtz, D., Kruger, J., Napa Scollon, C., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science*, *14*, 520–524. https://doi.org/10.1111/1467-9280.03455

Wrzus, C., Muller, V., Wagner, G. G., Lindenberger, U., & Riediger, M. (2013). Affective and cardiovascular responding to unpleasant events from adolescence to old age: complexity of events matters. *Developmental Psychology*, *49*, 384–397. https://doi.org/10.1037/a0028325

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, *22*, 3527–3541. https://doi.org/10.1002/sim.1572

Zawadzki, M. J., Scott, S. B., Almeida, D. M., Lanza, S. T., Conroy, D. E., Sliwinski, M. J., . . . Smyth, J. M. (2019). Understanding stress reports in daily life: A coordinated analysis of factors associated with the frequency of reporting stress. *Journal of Behavioral Medicine.* Advance online publication. https://doi.org/10.1007/s10865-018-00008-x

Zhaoyang, R., Sliwinski, M. J., Martire, L. M., & Smyth, J. M. (2018). Age differences in adults' daily social interactions: An ecological momentary assessment study. *Psychology and Aging*, *33*, 607–618. https://doi.org/10.1037/pag0000242

Table 1

Descriptive Statistics (Study 1).

		Correlations											
		2	3	4	5	6	7	8	9	10	*Mean*	*SD* (across participants)	ICC
1	Age	-.02	-.05	-.09	-.03	-.06	.14*	.12*	.14*	.12*	46.2	11.1	-
2	Gender[a]		.06	.06	.10	.04	-.08	-.05	-.05	-.07	0.65	0.48	-
3	Mean Momentary Negative Affect			.97**	.96**	.99**	-.46**	-.52**	-.41**	-.47**	22.0	15.4	.551
4	Retrospective Negative Affect		.60**		.96**	.97**	-.47**	-.54**	-.41**	-.48**	24.7	17.0	.581
5	Peak Negative Affect		.84**	.52**		.94**	-.51**	-.55**	-.40**	-.51**	32.8	18.1	.554
6	Last Negative Affect		.67**	.49**	.56**		-.44**	-.51**	-.40**	-.46**	21.3	15.6	.551
7	Mean Momentary Positive Affect		-.65**	-.46**	-.55**	-.43**		.97**	.96**	.99**	61.5	18.9	.574
8	Retrospective Positive Affect		-.46**	-.60**	-.37**	-.38**	.63**		.94**	.98**	62.1	18.8	.598
9	Peak Positive Affect		-.47**	-.35**	-.28**	-.34**	.80**	.53**		.96**	72.3	16.9	.636
10	Last Positive Affect		-.46**	-.39**	-.39**	-.61**	.69**	.55**	.59**		62.2	18.9	.563
	SD (across days)	-	7.5	14.5	16.2	14.1	9.5	15.4	12.8	16.7			
	SD (within days)	-	11.9	-	-	-	13.7	-	-	-			

Note. Table depicts descriptive statistics on the between-person level (upper diagonal) and the within-person level (lower diagonal). ICC = intra-class correlation (estimated from a two-level model as the proportion of between-person variance to the total variance). [a]0=male; 1=female. *p < .05; **p < .001. N = 242.

Table 2

Mean Level Comparisons: Level Convergence (Negative Affect).

	Negative affect (Study 1)			Negative affect (Study 2)		
	Model 1a	Model 2a	Model 3a	Model 1b	Model 2b	Model 3b
	Fixed Effects			Fixed Effects		
Intercept	22.040** (1.004)	21.379** (1.695)	21.378** (1.695)	2.005** (0.056)	2.121** (0.082)	2.119** (0.083)
Rating type[a]	3.052** (0.360)	3.048** (0.360)	3.065** (0.357)	0.139** (0.022)	0.137** (0.022)	0.142** (0.021)
Gender[b]	-	1.615 (2.074)	1.617 (2.073)	-	-0.142 (0.107)	-0.142 (0.107)
Age	-	-0.353 (0.992)	-0.717 (1.005)	-	-0.159* (0.054)	-0.178* (0.055)
Day	-	-0.062 (0.038)	-0.061 (0.038)	-	-0.015* (0.005)	-0.015* (0.005)
Burst	-	-	-	-	0.007 (0.033)	0.008 (0.033)
Age x Rating type[a]	-	-	-0.810* (0.358)	-	-	-0.065* (0.022)
	Random Effects (Standard Deviations)			Random Effects (Standard Deviations)		
Intercept (Across Persons)	15.36	15.33	15.32	0.637	0.607	0.607
Rating Type (Across Persons)	2.11	2.11	1.97	0.105	0.103	0.087
Intercept (Across Bursts)	-	-	-	0.488	0.488	0.488
Residual (Momentary Affect)	9.91	9.91	9.91	0.603	0.602	0.602
Residual (Retrospective Affect)	14.48	14.47	14.47	0.824	0.823	0.823

Note. Table depicts point estimates (standard errors for fixed effects in brackets). Age was z-standardized prior to the analyses; day was centered on the first day; burst (Study 2) was centered on the first burst. Number of observations: 5,777 (Study 1), 5,313 (Study 2); number of participants: 242 (Study 1), 175 (Study 2).

[a]0=aggregated momentary affect; 1=retrospective affect; [b]0=male; 1=female. $^*p < .05$; $^{**}p < .001$.

Table 3

Mean Level Comparisons: Level Convergence (Positive Affect).

	Positive affect (Study 1)			Positive affect (Study 2)		
	Model 4a	Model 5a	Model 6a	Model 4b	Model 5b	Model 6b
	Fixed Effects			Fixed Effects		
Intercept	61.238** (1.235)	63.604** (2.070)	63.602** (2.071)	4.282** (0.062)	4.285** (0.091)	4.286** (0.091)
Rating type[a]	0.406 (0.380)	0.398 (0.381)	0.405 (0.380)	0.050* (0.017)	0.051* (0.017)	0.048* (0.017)
Gender[b]	-	-2.562 (2.544)	-2.563 (2.544)	-	0.058 (0.121)	0.058 (0.121)
Age	-	-2.587* (1.217)	2.682* (1.223)	-	0.183* (0.061)	0.170* (0.062)
Day	-	-0.111* (0.044)	-0.111* (0.044)	-	0.003 (0.004)	0.003 (0.004)
Burst	-	-	-	-	-0.059* (0.029)	-0.059* (0.029)
Age x Rating type[a]	-	-	-0.307 (0.381)	-	-	0.035* (0.018)
	Random Effects (Standard Deviations)			Random Effects (Standard Deviations)		
Intercept (Across Persons)	18.91	18.66	18.66	0.749	0.738	0.737
Rating Type (Across Persons)	1.43	1.46	1.40	0.020	0.000	0.011
Intercept (Across Bursts)	-	-	-	0.423	0.418	0.418
Residual (Momentary Affect)	11.86	11.86	11.86	0.533	0.533	0.533
Residual (Retrospective Affect)	15.45	15.43	15.43	0.690	0.690	0.689

Note. Table depicts point estimates (standard errors for fixed effects in brackets). Age was z-standardized prior to the analyses; day was centered on the first day; burst (Study 2) was centered on the first burst. Number of observations: 5,777 (Study 1), 5,313 (Study 2); number of participants: 242 (Study 1), 175 (Study 2).

[a]0=aggregated momentary affect; 1=retrospective affect; [b]0=male; 1=female. $^{*}p < .05$; $^{**}p < .001$.

Table 4

Explained Within-Person Variance in Retrospective Affect: Within-Person Correspondence Convergence.

	Negative Affect		Positive Affect	
	Study 1	Study 2	Study 1	Study 2
Predictor included				
Mean	35.3%	44.3%	38.4%[a]	38.6%[c]
Peak	27.1%	33.6%	26.6%	23.4%
Last	23.2%	26.2%	28.8%	16.0%
Mean + Peak	35.5%	44.6%	38.5%[a]	38.6%[c]
Mean + Last	36.7%	45.9%	40.9%[b]	38.7%[d]
Peak + Last	32.6%	39.6%	35.2%	27.9%
Mean + Peak + Last	36.9%	46.2%	40.9%[b]	38.7%[d]
Unique effects[1]				
Mean	**4.3%****	**6.6%****	**5.7%****	**10.8%****
Peak	**0.2%****	**0.3%****	<0.1%	<0.1%
Last	**1.4%****	**1.6%****	**2.4%****	**0.1%***

Note. Table depicts Level-1 R^2 estimates according to Xu (2003) for retrospective affect reported at the end of the day predicted by the variables in the leftmost column. Within each column, percentages marked with the same superscript are not different to a statistically significant degree, $p >$ .05; all other likelihood ratio tests were significant, $p < .045$.

[1]Computed as the difference of the variance explained by the full model (all three predictors) minus the variance explained by the model including only the other two predictors.* $p < .05$; **$p < .001$.

Table 5

Predicting End-of-Day Negative Affect.

	Negative affect (Study 1)		Negative affect (Study 2)	
	Model 7a	Model 8a	Model 7b	Model 8b
	Fixed Effects		Fixed Effects	
Intercept	23.769** (1.823)	23.714** (1.821)	2.214** (0.090)	2.214** (0.090)
Mean Momentary Rating	0.579** (0.071)	0.576** (0.071)	0.691** (0.055)	0.693** (0.056)
Peak Momentary Rating	0.082* (0.036)	0.087* (0.036)	0.077* (0.031)	0.076* (0.031)
Last Momentary Rating	0.183** (0.028)	0.181** (0.028)	0.158** (0.028)	0.159** (0.028)
Gender[a]	0.907 (2.210)	0.964 (2.207)	-0.175 (0.119)	-0.174 (0.119)
Age	-0.948 (1.056)	-1.318 (1.077)	-0.217** (0.060)	-0.210** (0.062)
Day	-0.094 (0.055)	-0.097 (0.055)	-0.009 (0.006)	-0.010 (0.006)
Burst	-	-	-0.005 (0.017)	-0.004 (0.017)
Age x Mean Momentary Rating	-	-0.083 (0.073)	-	0.005 (0.058)
Age x Peak Momentary Rating	-	-0.008 (0.035)	-	-0.009 (0.032)
Age x Last Momentary Rating	-	0.050 (0.029)	-	0.016 (0.028)
	Random Effects (Standard Deviations)		Random Effects (Standard Deviations)	
Intercept	15.64	15.61	0.755	0.756

(Across Persons)				
Mean Momentary Rating (Across Persons)	0.575	0.571	0.358	0.358
Peak Momentary Rating (Across Persons)	0.288	0.282	0.246	0.246
Last Momentary Rating (Across Persons)	0.196	0.191	0.196	0.197
Intercept (Across Bursts)	-	-	0.076	0.076
Residual	10.30	10.30	0.547	0.547

Note. Table depicts point estimates (standard errors for fixed effects in brackets). Age was z-standardized prior to the analyses; day was centered on the first day; burst (Study 2) was centered on the first burst. Mean momentary ratings, peak momentary ratings, and last momentary ratings were centered on the person mean of all momentary ratings. Number of observations: 2,591 (Study 1), 2,386 (Study 2); number of participants: 242 (Study 1), 174 (Study 2).

[a]0=male; 1=female. *$p < .05$; **$p < .001$.

Table 6

Predicting End-of-Day Positive Affect.

	Positive affect (Study 1)		Positive affect (Study 2)	
	Model 9a	Model 10a	Model 9b	Model 10b
	Fixed Effects		Fixed Effects	
Intercept	62.940** (2.170)	62.980** (2.172)	4.298** (0.094)	4.299** (0.093)
Mean Momentary Rating	0.607** (0.054)	0.608** (0.054)	0.849** (0.049)	0.845** (0.048)
Peak Momentary Rating	0.000 (0.042)	-0.003 (0.042)	-0.014 (0.039)	-0.008 (0.039)
Last Momentary Rating	0.204** (0.027)	0.203** (0.027)	0.033 (0.025)	0.032 (0.025)
Gender[a]	-1.373 (2.576)	-1.375 (2.579)	0.082 (0.124)	0.083 (0.124)
Age	2.500* (1.230)	2.968* (1.343)	0.172* (0.062)	0.121 (0.065)
Day	-0.098 (0.057)	-0.099 (0.057)	0.003 (0.006)	0.003 (0.006)
Burst	-	-	0.014 (0.016)	0.014 (0.016)
Age x Mean Momentary Rating	-	0.019 (0.054)	-	-0.095 (0.050)
Age x Peak Momentary Rating	-	-0.035 (0.041)	-	0.113* (0.041)
Age x Last Momentary Rating	-	0.003 (0.026)	-	-0.011 (0.025)
	Random Effects (Standard Deviations)		Random Effects (Standard Deviations)	
Intercept	19.59	19.59	0.777	0.777

(Across Persons)				
Mean Momentary Rating (Across Persons)	0.479	0.477	0.314	0.303
Peak Momentary Rating (Across Persons)	0.324	0.320	0.285	0.274
Last Momentary Rating (Across Persons)	0.215	0.215	0.180	0.179
Intercept (Across Bursts)	-	-	0.082	0.085
Residual	10.68	10.68	0.519	0.518

Note. Table depicts point estimates (standard errors for fixed effects in brackets). Age was z-standardized prior to the analyses; day was centered on the first day; burst (Study 2) was centered on the first burst. Mean momentary ratings, peak momentary ratings, and last momentary ratings were centered on the person mean of all momentary ratings. Number of observations: 2,591 (Study 1), 2,386 (Study 2); number of participants: 242 (Study 1), 174 (Study 2).

[a]0=male; 1=female. $*p < .05$; $**p < .001$.

Table 7

Descriptive Statistics (Study 2).

		Correlations											
		2	3	4	5	6	7	8	9	10	*Mean*	*SD* (across participants)	ICC
1	Age	-.01	-.24*	-.29**	-.29**	-.27**	.21*	.24*	.12	.20*	49.7	17.1	-
2	Gender[a]		-.10	-.11	-.06	-.13	.02	.06	.06	.02	0.51	0.50	-
3	Mean Momentary Negative Affect			.97**	.95**	.99**	-.40**	-.44**	-.33**	-.37**	1.99	0.64	.377
4	Retrospective Negative Affect		.74**		.95**	.95**	-.43**	-.48**	-.35**	-.41**	2.13	0.72	.423
5	Peak Negative Affect		.83**	.64**		.94**	-.44**	-.47**	-.29**	-.41**	2.73	0.86	.401
6	Last Negative Affect		.70**	.60**	.59**		-.43**	-.47**	-.37**	-.41**	1.92	0.62	.358
7	Mean Momentary Positive Affect		-.51**	-.45**	-.45**	-.34**		.98**	.94**	.99**	4.33	0.77	.487
8	Retrospective Positive Affect		-.42**	-.49**	-.37**	-.34**	.70**		.92**	.98**	4.38	0.77	.537
9	Peak Positive Affect		-.37**	-.31**	-.24**	-.23**	.82**	.57**		.91**	4.99	0.67	.533
10	Last Positive Affect		-.34**	-.32**	-.33**	-.47**	.68**	.50**	.52**		4.23	0.78	.498
	SD (across bursts)	-	0.45	0.49	0.49	0.48	0.40	0.42	0.43	0.43			
	SD (across days)	-	0.49	0.84	1.06	0.85	0.42	0.70	0.62	0.78			
	SD (within days)	-	0.74	-	-	-	0.71	-	-	-			

Note. Table depicts descriptive statistics on the between-person level (upper diagonal) and the within-person level (lower diagonal). ICC = intra-class correlation (estimated from a two-level model as the proportion of between-person variance to the total variance). [a]0=male; 1=female. $^{*}p < .05$; $^{**}p < .001$. $N = 175$.

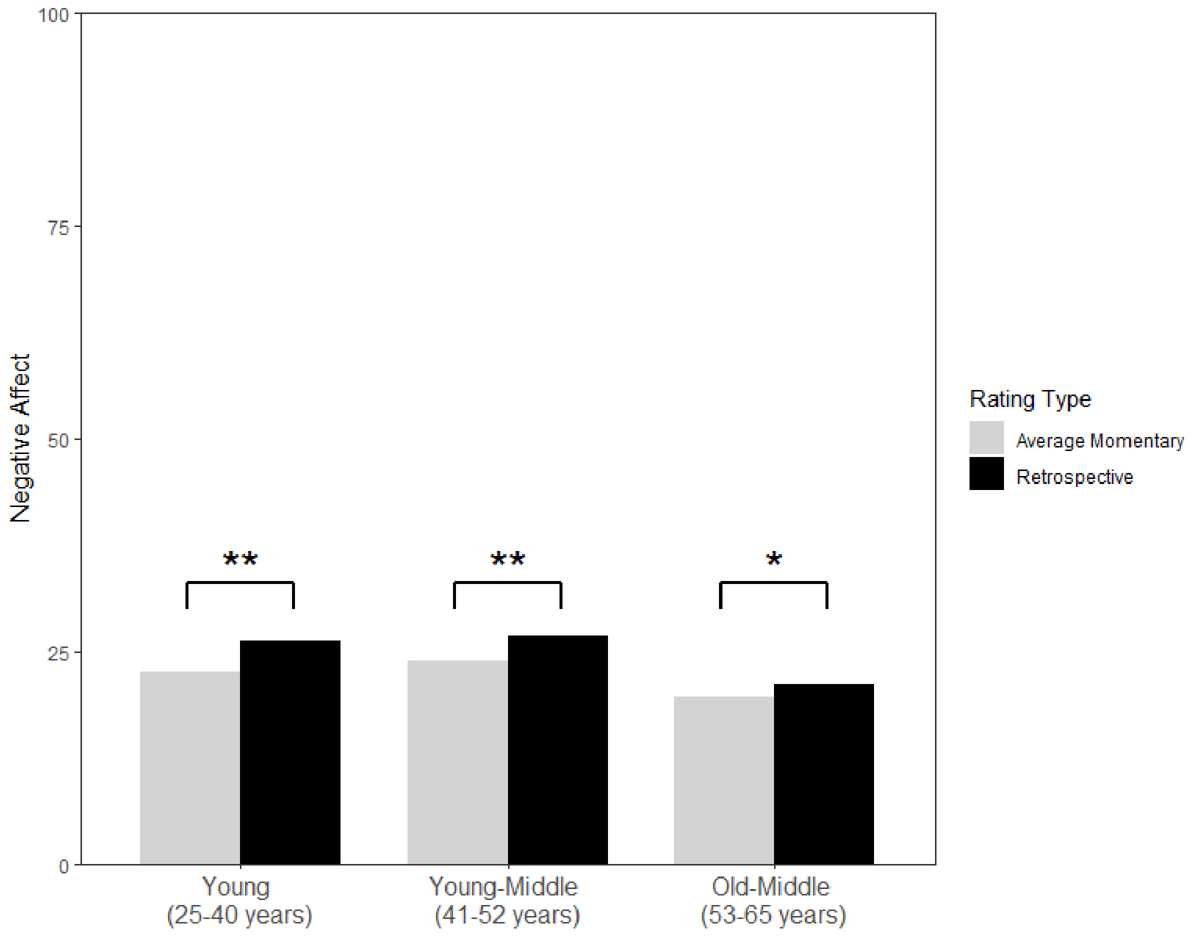

Figure 1. Figure depicts negative affect (Study 1) as a function of rating type (average of momentary ratings across the day vs. retrospective ratings) and participants' age.*$p < .05$; **$p < .001$.

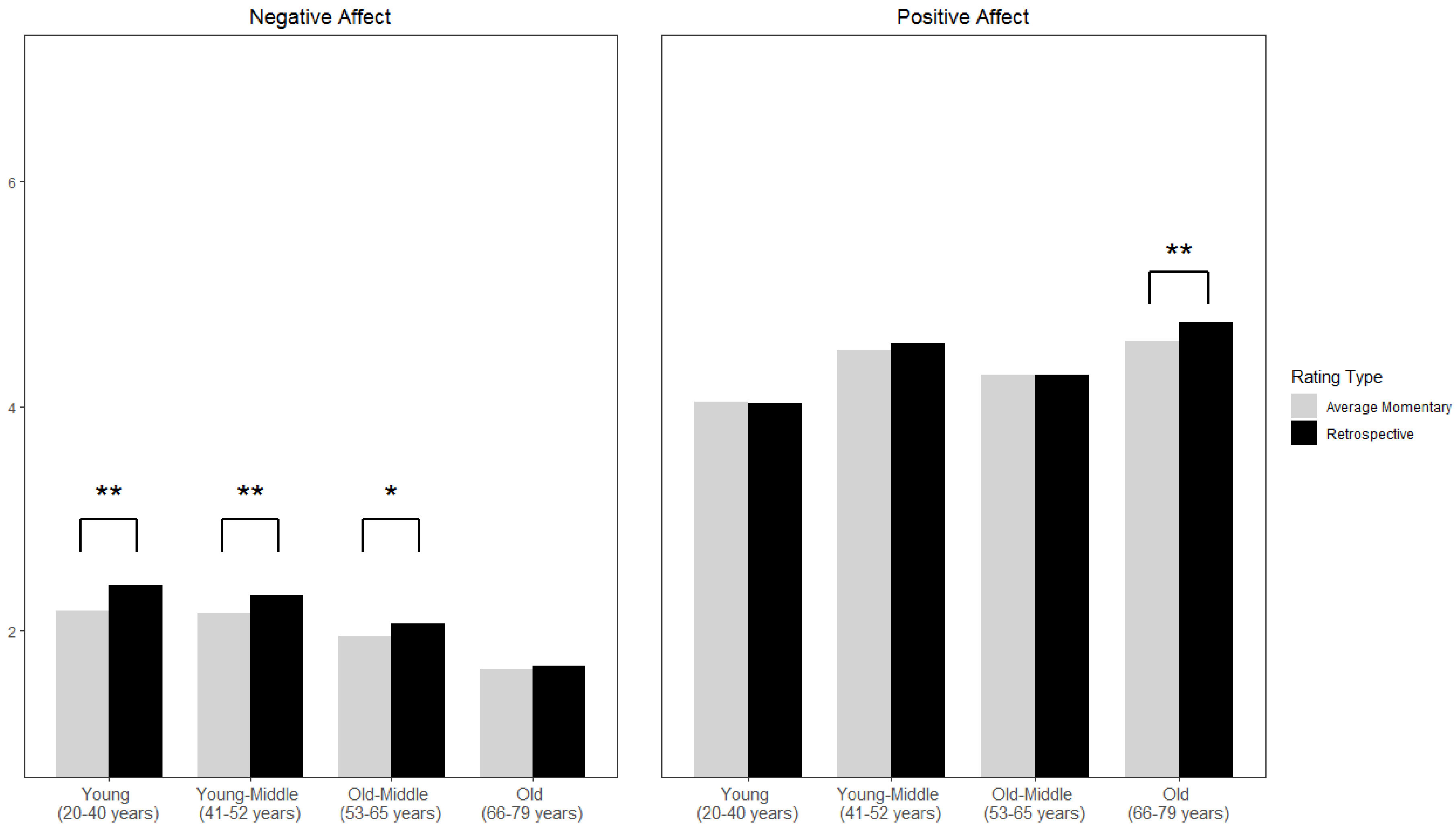

Figure 2. Figure depicts negative affect and positive affect (Study 2) as a function of rating type (average of momentary ratings across the day vs. retrospective ratings) and participants' age. *p < .05; **p < .001.