Affect dynamics or response bias? Extreme response style in daily-life assessments

Mirka Henninger https://orcid.org/0000-0003-4676-2361[1]

Sandro Menegola https://orcid.org/0009-0000-3747-4822[1]

Niels Vanhasbroeck https://orcid.org/0000-0002-0056-3183[2]

Francis Tuerlinckx https://orcid.org/0000-0002-1775-7654[3]

[1]Faculty of Psychology; University of Basel

[2]Faculty of Social and Behavioral Sciences; Psychological Methods; University of Amsterdam

[3]Faculty of Psychology and Educational Sciences; Research Group of Quantitative Psychology and Individual Differences; Katholieke Universiteit Leuven

Author Note

Correspondence to this article should be addressed to Mirka Henninger, Faculty of Psychology, University of Basel, Switzerland, mirka.henninger@unibas.ch.

Funding details: NV is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 Excellent Science program (Grant number: 101053880); FT is supported in part by the Research Fund of KU Leuven (Grant number: C14/23/062).

Disclosure statement: The authors report there are no competing interests to declare.

Data availability statement: We reanalyzed existing datasets; most of them are publicly available. Details on data sources and links to the original datasets are provided in the Method section in Table 1.

Authors' contributions: Conceptualization (MH, SM, NV, FT), Methodology (MH, SM, NV, FT), Formal analysis (MH, SM), Data Curation (MH), Writing Original Draft (MH), Writing Review and Edition (MH, SM, NV, FT), Visualization (MH).

Abstract

Intensive longitudinal data (ILD) are increasingly used to study affect dynamics such as intraindividual variability and inertia. These parameters are often interpreted as substantive individual differences, yet they may be biased by stable response style tendencies, in particiular extreme response style (ERS). In the present study, we investigated whether ERS is systematically associated with affect dynamics estimated from multiple experience sampling method (ESM) datasets (total $N = 1{,}254$). Using a joint Bayesian IRTree–location–scale modeling approach, we estimated person-specific ERS alongside affective variability and inertia across multiple ESM datasets and synthesized study-level associations using a meta-analytic approach. We find that higher ERS is consistently associated with greater intraindividual variability. In contrast, we find no evidence that ERS was related to affective or emotional inertia. These results suggest that commonly used measures of intraindividual variability may partly reflect systematic response behavior rather than solely genuine affective fluctuations. Ignoring ERS may therefore bias variability estimates obtained from location–scale and related multilevel dynamic models. Overall, this study highlights the importance of accounting for response styles when modeling affect dynamics and interpreting person-specific variability parameters in ILD research.

Keywords: experience sampling, affect dynamics, variability, inertia, extreme response style, IRTree, location-scale model, research synthesis

Affect dynamics or response bias? Extreme response style in daily-life assessments

Introduction

Intensive longitudinal data (ILD) consist of repeated measurements collected within individuals over time, often several times per day across multiple weeks. In psychological research, ILD are most commonly obtained using self-report questionnaires administered using the experience sampling methodology (ESM; sometimes also referred to as ambulatory assessment, AA, or ecological momentary assessment, EMA). Such data have become increasingly prominent, as they allow researchers to examine not only changes in average levels of psychological states—such as behavior or affect—but also their moment-to-moment fluctuations and temporal dynamics in daily life (e.g., McKone & Silk, 2022; Russell et al., 2007). One domain in which ILD have been particularly influential is the field of emotional and affective dynamics, which revolved around moment-to-moment fluctuation in affect over time (e.g., Bos et al., 2019; Ernst et al., 2021; Koval et al., 2013; Kuppens & Verduyn, 2017; Trull et al., 2008).

Affect dynamics are generally conceptualized as indicators of how individuals respond to and adapt to their environments and everyday events. Importantly, these dynamics show substantial between-person differences and are often regarded as relatively stable individual characteristics (Bos et al., 2019; Carver, 2015; Smit et al., 2023). As such, affect dynamics are considered central markers of psychological health, including constructs such as well-being, emotional flexibility, and self-regulation (Czyz et al., 2021; Franck & De Raedt, 2007; Hamaker et al., 2016; Houben et al., 2015; Koval et al., 2016; Kuppens et al., 2007; Maher et al., 2019; Röcke & Brose, 2013). Moreover, affect dynamics have been linked to a range of clinically relevant outcomes, including depressive symptoms, depression recurrence, and substance use (Jenkins et al., 2024; Mohr et al., 2015; Palmier-Claus et al., 2012; Piasecki et al., 2016; Russell et al., 2007; Trull et al., 2008, 2015). Thus, ILD have substantially advanced the study of emotional functioning by moving beyond traditional analyses that focus solely on mean levels of affect or behavior (Ernst et al., 2021; Scott et al., 2020).

Two of the most widely studied indicators of affect dynamics are intraindividual variability and inertia, which capture conceptually distinct aspects of emotional functioning and are typically only weakly correlated (Jahng et al., 2008; Wang et al., 2012; Wendt et al., 2020). Intraindividual variability, also referred to as dispersion, describes the extent to which an individual's momentary affect deviates from their typical affective level over time. Inertia, by contrast, reflects the degree to which an affective state predicts subsequent affective states and is commonly interpreted as an index of emotional carryover or persistence. High inertia has been associated with reduced emotional adaptability and slower recovery from emotionally challenging situations. Methodologically, intraindividual variability is commonly operationalized as the within-person standard deviation across repeated affect ratings, whereas inertia is typically estimated using first-order autoregressive effects (Bos et al., 2019; Eid & Diener, 1999; Ferrando, 2014; Hedeker & Mermelstein, 2020; Jongerling et al., 2015; Koval et al., 2013, 2016; Kuppens et al., 2007; Larsen & Diener, 1987; Wang et al., 2012; Wichers et al., 2015).

However, given that ILD are collected via self-report questionnaires, observed affect dynamics may reflect not only individuals' true affective experiences but also systematic differences in how respondents react to and use measurement instruments and their response scales. In cross-sectional self-report research, such systematic response tendencies are commonly referred to as response styles: stable individual differences in scale usage that are largely independent of item content (Adams et al., 2019; Baumgartner & Steenkamp, 2001; Bolt & Johnson, 2009; Van Vaerenbergh & Thomas, 2013). Prominent response styles include extreme response style (ERS), the tendency to overselect the extreme categories, midpoint response style (MRS), the tendency to prefer the scale's midpoint, and acquiescent response style (ARS), the tendency to agree with items regardless of their content (Couch & Keniston, 1960; Hamilton, 1968; Henninger & Meiser, 2020a; Paulhus, 1991).

Among these, ERS is particularly prevalent in self-report data and has been shown to affect the estimation of trait means, variances, and intercorrelations (e.g., Adams et al., 2019; Böckenholt & Meiser, 2017; Bolt et al., 2014; De Jong et al., 2008; Falk & Cai, 2016; Henninger, 2021; Henninger & Meiser, 2020b, 2022; Plieninger, 2017; Plieninger & Heck, 2018; Soland & Kuhfeld, 2020; Ulitzsch et al., 2024). Because response styles are considered stable person characteristics, which are consistent across measurement occasions and content domains (Weijters et al., 2010; Wetzel et al., 2013), there is growing consensus that accounting for response style effects is essential for ensuring the quality and validity of self-report data.

In the context of ILD, ERS may pose particular challenges for the measurement of affect dynamics. A strong tendency to endorse extreme response categories can artificially inflate intraindividual variability. As a consequence, affect variability or inertia may not solely reflect genuine affective dynamics over time, but may also partly reflect systematic response behavior, potentially biasing other model estimates, such as regression coefficients and thereby threatening the validity of subsequent modeling and inference.

Empirical research directly examining the association between response styles and affect dynamics remains limited, but initial evidence suggests that response tendencies may be associated with variability estimates. For instance, Baird et al. (2017) showed that response variability in ratings of cartoon characters was comparable to variability in ratings of participants' own personality, despite clear conceptual differences between these constructs. The authors concluded that a substantial proportion of observed variability was likely driven by response-related measurement artifacts rather than genuine psychological variability. Similarly, Deng et al. (2018) demonstrated, using psychometric modeling in longitudinal data, that ERS substantially inflated estimates of variability. However, these studies also exhibit important limitations, including reliance on the same items to assess both variability and response tendencies (Baird et al., 2017; Deng et al., 2018), cross-sectional designs (Baird et al., 2017), or longitudinal designs with relatively few measurement occasions (Deng et al., 2018).

Recently, Henninger et al. (2025) extended this line of research by examining the relationship between ERS and affect dynamics−specifically variability and inertia−using data from a controlled experiment. In this experiment, affect was induced using a probabilistic reward task, allowing for tight control over mean affect levels. The authors found that ERS, assessed using independent background questionnaires, was positively associated with intraindividual variability in repeated affect ratings, whereas no association between ERS and inertia was observed.

While this study had several strenghts, such as using independent items to assess ERS and a large sample size that enabled the examination of between-person effects, the controlled experimental nature of the design limits the generalizability of the findings to naturalistic contexts in which ILD are typically collected. For example, the very short intervals between measurement occasions in the experimental setup may have induced fatigue effects (Jangraw et al., 2023; Vanhasbroeck et al., 2024), and affective experiences in everyday life are influenced by situational factors that are largely absent in laboratory or experimental paradigms. Consequently, the association between ERS and affect dynamics observed under controlled conditions may differ from that observed in daily-life settings.

The present study

The goal of the present study is to extend the work of Henninger et al. (2025) to naturalistic, real-world ILD collected in typical experience sampling studies. If ERS and affect dynamics are associated with each other in data collected in daily life, it would suggest that variability estimates derived from ILD may be systematically inflated by ERS, with important implications for the interpretation of affect dynamics in psychological research and the necessity of novel modeling practices to account for response styles in ILD.

Unlike the experimental data used by Henninger et al. (2025) which included a large sample of more than 1,300 participants and therefore had high statistical power to detect between-person associations, most individual ESM datasets in naturalistic settings are relatively small, typically around 100 participants. Therefore, we expect that a single ESM study would lack sufficient power to reliably estimate associations between stable traits, such as ERS, and affect dynamics. To address this challenge, we will synthesize results from multiple independent ESM datasets, enabling more precise estimation of between-person associations and increasing the generalizability of our findings. Across these datasets, we will examine the relationship between ERS and the previously mentioned two primary indicators of affect dynamics: intraindividual variability and inertia. This approach allows us to investigate whether the association between ERS and intraindividual variability and potentially inertia observed in controlled experimental settings also occurs in daily-life assessments.

Based on prior findings and simulation results reported by Henninger et al. (2025), we first hypothesize that individuals with higher levels of ERS will exhibit greater intraindividual variability in affective measures derived from ILD, even after accounting for prior affective states. Accordingly, our primary hypothesis predicts a positive association between ERS and affective and emotional variability. Second, although Henninger et al. (2025) found no empirical association between ERS and inertia, their simulation results suggested a small negative relationship. Therefore, as a secondary hypothesis, we explore whether ERS is negatively associated with affective inertia in naturalistic ILD.

Methods

Preregistration

Our hypotheses and analysis plan were preregistered on the Open Science Framework (OSF) prior to conducting the main analyses.[1] Before submitting the preregistration, we collected and cleaned the datasets to be analyzed and selected the variables (emotion items and background questionnaires) for the main analysis. We did not run any analysis model on the preprocessed datasets. One exception is the third wave in the data collected by Erbas et al. (2018) on which we piloted the model. In the analyses reported in this article, we only used data from the first wave of this study. All deviations from the preregistered analysis plan are documented in Appendix A.

Datasets

We collected a sample of ten openly available datasets. Datasets were included if they met the following criteria: (a) emotion items measuring were assessed repeatedly multiple times per day over several days, and (b) background questionnaires contained constructs conceptually unrelated to affect, allowing us to estimate trait-level ERS. Deriving ERS from measures independent of affect ensured that ERS estimates were not confounded with affective states or affect dynamics assessed during the ESM phase.

A major source for suitable datasets was the "Everyday Measures of Temporal Emotions" (*EMOTE*) database (Kalokerinos et al., 2023, https://emotedatabase.com/), an open-access repository of experience sampling data on daily emotions. Additional datasets were obtained from open data repositories or via personal communication with the original authors. Participants were predominantly healthy adults, often university students, although several datasets used stratified or targeted recruitment to increase variability in affective experiences (Erbas et al., 2018; Haines et al., 2016; Naragon-Gainey et al., 2023; Niemeijer, 2024). Table 1 summarizes the included datasets and their key characteristics.

Data preprocessing

The datasets differed with respect to the specific emotions that they assessed, as well as in the response scales employed (0-100 visual analogue scale vs. 5-point Likert-type scale; see Table 1). To ensure comparability across datasets, we included only emotions measured in at least three datasets and rescaled all responses to a common 0-1 scale. ILD responses are treated as continuous variables, regardless of their original scale levels. We excluded participants with compliance rates below 50% of measurement occasions in the ESM phase (see e.g., Dejonckheere et al., 2019b).

Aggregated affect scores and individual emotion items

A common practice in ILD research is to aggregate individual emotion items into composite scores of positive affect (PA; e.g., *happy*, *relaxed*, *calm*) and negative affect (NA; e.g., average of *sad*, *angry*, *stressed*) for each participant at each measurement occasion (e.g., Bos et al., 2019; Hamaker et al., 2016; Hedeker et al.,

[1] https://osf.io/paxut/overview?view_only=f03e8e5c50c24e2da342f6fe2bbda150

Table 1

Overview of studies included in the analysis.

Study	Source	Sample Size	ILD	Background Questionnaires	Affect Measures
Cloos et al. (2023)	Author contact	$N = 153$	14 days, 10x/day	Extraversion (8 items); Conscientiousness (9 items); 5-point scale	*happy, relaxed, calm, angry, stressed, irritated, anxious, depressed*; slider scale from 0 to 100
Dejonckheere et al. (2019a)	https://emotedatabase.com/datasets/27	$N = 100$	14 days, 7x/day	Openness (10 items); Conscientiousness (9 items); 5-point scale	*happy, relaxed, sad, angry, stressed*; slider scale from 0 to 100
Erbas et al. (2018)	https://emotedatabase.com/datasets/22	$N = 200$	7 days, 10x/day	Emotion Regulation (10 items; 7-point scale); subset from the Penn State Worry Questionnaire (7 items; 5-point scale)	*happy, relaxed, confident, sad, angry, stressed, depressed, lonely*; slider scale from 0 to 100)
Fried et al. (2022)	https://github.com/jmbh/EmotionTimeSeries	$N = 76$	14 days, 4x/day	Disorganization (subset of 6 items; 5-point scale); Motivation (subset of 7 items; 7-point scale)	*angry, stressed, irritated, anxious, depressed, lonely*; 5-point scale
Grommisch et al. (2020)	https://emotedatabase.com/datasets/21	$N = 175$	7 days, 10x/day	Extraversion (8 items); Conscientiousness (4 items); 5-point scale	*happy, relaxed, confident, sad, angry, stressed*; slider scale from 0 to 100
Haines et al. (2016)	https://emotedatabase.com/datasets/16	$N = 72$	21 days, 9x/day	Agreeableness (9 items) Conscientiousness (9 items); 5-point scale	*happy, calm, confident, sad, angry, anxious*; slider scale from 0 to 100
Knouse et al. (2024)	https://osf.io/a5gbf/ and https://osf.io/r8fkc/	$N = 101$	6 days, 3x/day	Agreeableness (9 items); Conscientiousness (9 items); 5-point scale	*happy, sad, lonely, angry*; 5-point scale
Naragon-Gainey et al. (2023)	https://osf.io/gct7x/ and author contact	$N = 303$	7 days, 6x/day	Agreeableness (9 items); Conscientiousness (9 items); 5-point scale	*excited, sad, irritated, anxious*; 5-point scale
Niemeijer et al. (2023)	Author contact	$N = 104$	21 days, 10x/day	Extraversion (12 items); Openness (12 items); 5-point scale	*happy, relaxed, sad, stressed, anxious, lonely, (tired)*; slider scale from 0 (“not at all”) to 100 (“very much”)
Pavani et al. (2017)	https://osf.io/s3chz/ and author contact	$N = 70$	14 days, 5x/day	Extraversion (subset of 10 items); Conscientiousness (subset of 10 items); 5-point scale	*happy, relaxed, calm, sad, irritated*; 5-point scale

2009b; Jenkins et al., 2024; Koval et al., 2013, 2016; Kuppens & Verduyn, 2017; Maher et al., 2019; Mohr et al., 2015; Palmier-Claus et al., 2012; Russell et al., 2007). To align our analyses with these practices, we also computed such aggregated PA and NA scores. This approach allows us to examine the extent to which ERS is associated with the estimation of affect variability and inertia in the context of commonly used ILD measurement practices.

At the same time, analyzing individual emotion items may provide a more detailed view of how participants use the response scales in ILD and the extent to which ERS plays a role in these measurements. Aggregating items into PA and NA scores can reduce variability in affect measures (Cloos et al., 2023) and potentially mask ERS effects. For example, a participant may report high anger but low sadness at one occasion, and the reverse at another; combining these items into a single PA or NA score could obscure these fluctuations. Accordingly, we replicate our analyses using individual items in addition to aggregated PA/NA scores.

Background questionnaires to measure ERS

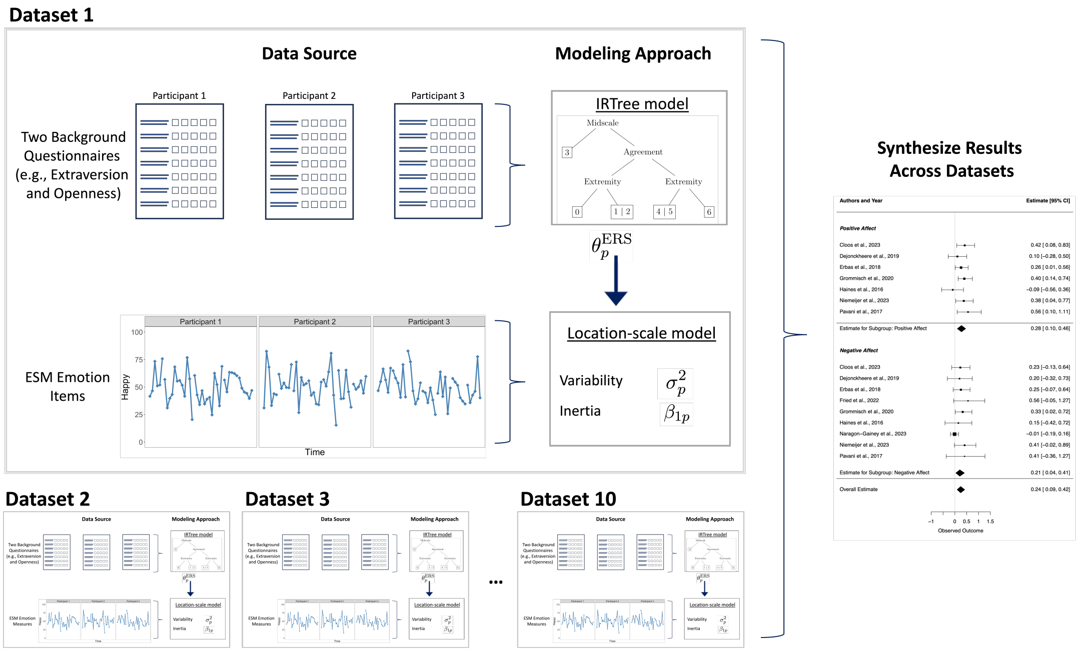
To measure participants' ERS, we selected two scales from the background questionnaires available in each dataset. Questionnaires were included based on the following criteria: (a) use either a 5-point or 7-point Likert-type response scales, (b) demonstrate adequate internal consistency, and (c) be conceptual independent from affect measures. For the latter criterion, we ensured that the selected constructs were distinct from the ESM emotion items. Specifically, we prioritized personality traits, such as the Big Five, which exhitibited low to moderate correlations with aggregated PA and NA scores. This approach allowed us to derive trait-level estimates of ERS that were not confounded with the emotion and affect measures collected during the ESM phase. Using two scales per dataset is recommended for response style modeling, as estimating response styles across multiple content domains facilitates the separation of substantive trait variance from response style variances (e.g., Henninger & Meiser, 2020b, 2022; Wetzel & Carstensen, 2017). A detailed overview of the datasets, included affect measures, selected background questionnaires (along with reasons for inclusion or exclusion), and all preprocessing scripts is available on OSF as part of the preregistration materials.

Analysis Model

Analogous to Henninger et al. (2025), we employed a joint modeling approach that combines a psychometric model for ERS with a multilevel location-scale model to examine its relationship with affect dynamics. Based on the background questionnaire data, ERS was modeled using a two-parameter IRTree model (Böckenholt, 2012), which yields latent trait estimates reflecting individuals' tendencies to provide extreme responses. These latent ERS trait estimates were incorporated as a between-person predictor variable in a multilevel location-scale regression model (see Hedeker & Mermelstein, 2020; Hedeker et al., 2009a; Jongerling et al., 2015; McNeish, 2021) to assess their association with affect variability and inertia (see Figure 1 for a visualization of the modeling procedure).

Figure 1

Modeling procedure combining a psychometric IRTree model for ERS with a multilevel location-scale model for affect and emotion dynamics (variability and inertia). The results from the individual datasets are synthesized using a meta-analytical approach.



Note: IRTree model: Item Response Tree model; ERS: Extreme Response Style; ESM: Experience Sampling Methodology; p indicates the person with $p = 1, ..., N$.

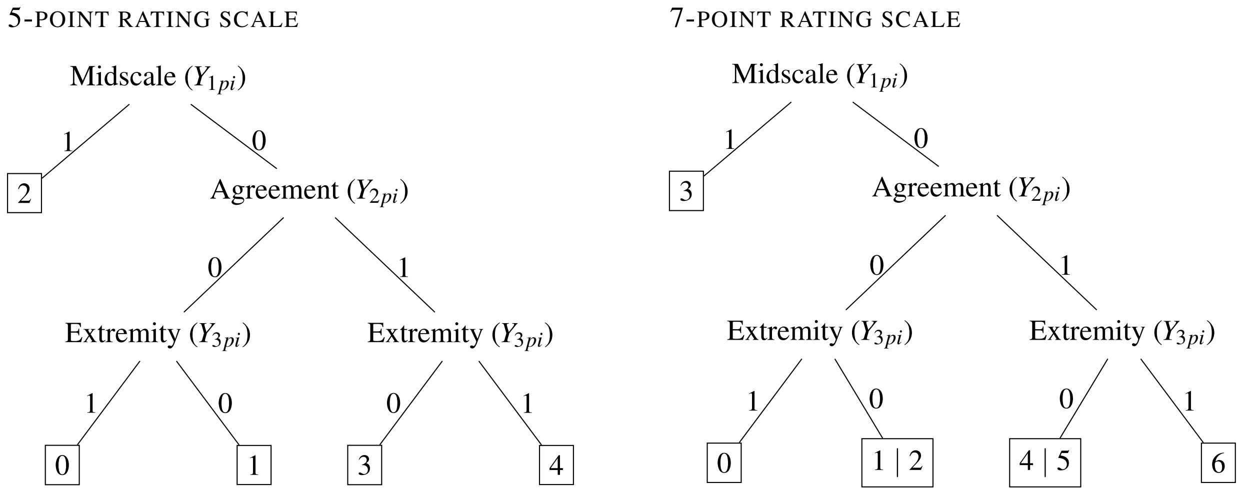
Psychometric model for ERS

Psychometric models based on Item Response Theory (IRT) can be extended to capture systematic response tendencies (see Böckenholt & Meiser, 2017; Henninger & Meiser, 2020a, 2022). A widely used approach is the Item Response Tree (IRTree) model, which decomposes the ordinal responses to Likert-type items into distinct response processes reflecting both the substantive content dimension (e.g., Extraversion) and response styles such as ERS and MRS (Böckenholt, 2012; De Boeck & Partchev, 2012; Khorramdel & von Davier, 2014; Meiser et al., 2019; Plieninger & Meiser, 2014).

Figure 2 illustrates the decomposition from 5-point and 7-point rating scales into three response processes using an IRTree model for response styles. For the 5-point scale ranging from 0 to 4, the responses are decomposed into (a) a midscale response process (category 2 vs. categories 0, 1, 3, and 4), (b) an agreement response process (categories 3 and 4 vs. categories 0 and 1), and (c) an extreme response process (categories 0 and 4 vs. 1 and 2). Each response process is parameterized using a two-parameter logistic (2PL) model, for which the recoded items (the so-called pseudo-items) serve as indicators. The probability of observing a specific response category is obtained multiplicatively by traversing the branches of the tree (see Böckenholt, 2012; Plieninger, 2020, for details on IRTree specifications).

Figure 2

Top: IRTree structure for 5-point (left) and 7-point (right) rating scales with midscale, agreement and extremity response processes. Bottom: Coding scheme for the midscale, agreement, and extremity response processes into pseudo-items and parameterization using the two-parameter logistic (2PL) model.



Response process	Coding scheme for pseudo-items												2PL IRT model
	5-point rating scale					7-point rating scale							
	0	1	2	3	4	0	1	2	3	4	5	6	
Midscale (Y_{1pi})	0	0	1	0	0	0	0	0	1	0	0	0	$\frac{\exp(\alpha_{1i}\theta_{1p}-\delta_{1i})}{1+\exp(\alpha_{1i}\theta_{1p}-\delta_{1i})}$
Agreement (Y_{2pi})	0	0	–	1	1	0	0	0	–	1	1	1	$\frac{\exp(\alpha_{2i}\theta_{2p}-\delta_{2i})}{1+\exp(\alpha_{2i}\theta_{2p}-\delta_{2i})}$
Extremity (Y_{3pi})	1	0	–	0	1	1	0	0	–	0	0	1	$\frac{\exp(\alpha_{3i}\theta_{3p}-\delta_{3i})}{1+\exp(\alpha_{3i}\theta_{3p}-\delta_{3i})}$

We used responses from the background questionnaires to construct pseudo-items for the agreement, extremity, and midscale response processes of the IRTree model. Because we selected two scales (e.g., Extraversion and Openness) for each dataset, we modeled separate agreement dimensions for of the two traits. The response style dimensions, the midscale and extremity response processes, however, were modeled jointly across items from both scales. This model specification follows recommendations in the psychometric literature, as response styles are typically conceptualized as stable person characteristics that are largely independent of item content and thus generalize across content domains. Modeling response styles across multiple domains therefore improves their reliability and validity (e.g., Henninger & Meiser, 2020b, 2022; Wetzel & Carstensen, 2017). The person parameter associated with the extremity response process θ_{3p} was used as a predictor representing the latent ERS trait, denoted θ_p^{ERS}, in the location-scale model estimated within a joint modeling framework.

Location-scale model using ERS as a predictor variable

To study our hypotheses, we employed a multilevel location-scale model (Hedeker & Mermelstein, 2020; McNeish, 2021; Wang et al., 2012), which simultaneously captures interindividual differences in mean affect levels, intraindividual variability in affect, and affect inertia. In the model, Level-1 represents repeated measurement occasions nested within individuals (Level-2).

At Level-1, affect at time point t for person p is modeled as a function of a person-specific intercept and a first-order autoregressive effect (inertia) capturing moment-to-moment dependencies in affect:

$$\text{Level-1:} \qquad y_{tp} = \beta_{0p} + \beta_{1p} \cdot y_{(t-1)p} + \varepsilon_{tp} \tag{1}$$

At Level-2, we specify random effects for both the intercept and the autoregressive slope. These between-person differences in affect levels and inertia are modeled as function of the latent ERS trait:

Level-2 model for location: (2)

$$\text{Between-person intercept:} \qquad \beta_{0p} = \gamma_{00} + \gamma_{01} \cdot \theta_p^{\text{ERS}} + u_{0p}$$

$$\text{Between-person slope of inertia:} \qquad \beta_{1p} = \gamma_{10} + \gamma_{11} \cdot \theta_p^{\text{ERS}} + u_{1p}$$

To obtain a model-based estimate of intraindividual variability, we allow the Level-1 residual variance to vary across individuals

$$\varepsilon_{tp} \sim N(0, \sigma_p^2) \tag{3}$$

These person-specific residual variances are modeled at Level-2 as a function of a fixed intercept, the latent ERS trait, and a random effect capturing residual between-person differences in variability. The exponential link ensures that variances are strictly positive and lognormally distributed across individuals (Hedeker et al., 2008, 2009a):

$$\text{Level-2 model for variability:} \qquad \sigma_p^2 = \exp(\omega_0 + \omega_1 \cdot \theta_p^{\text{ERS}} + u_{2p}) \tag{4}$$

The random effects for affect level (u_{0p}), inertia (u_{1p}), and intraindividual variability (u_{2p}) are assumed to follow a multivariate normal distribution with zero means and variance-covariance matrix Σ_τ:

$$\begin{bmatrix} u_{0p} \\ u_{1p} \\ u_{2p} \end{bmatrix} \sim MVN \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{01} & \tau_{11} & \\ \tau_{02} & \tau_{12} & \tau_{22} \end{bmatrix} \right). \tag{5}$$

Both the IRTree model and the location–scale model were estimated jointly within a single modeling framework, such that uncertainty in the latent ERS trait is propagated to the estimation of affect dynamics. We estimated separate models for each dataset and for aggregated PA and NA scores, as well as for individual emotion items (such as *happy*, *relaxed*, *calm*, *sad*, *angry*, *stressed*).

Model estimation

As in Henninger et al. (2025), all models were estimated in Mplus 8.6 (L. K. Muthén & Muthén, 2017) using Bayesian estimation with the default uninformative priors (see Asparouhov & Muthen, 2023; Hamaker et al., 2018; B. Muthén & Asparouhov, 2012). For the IRTree model, these default priors are normal distributions with mean 0 and standard deviation 5 for discrimination and difficulty parameters α_i and δ_i, and inverse-Wishart priors for variances and covariances ($IW(1,5)$ and $IW(0,5)$, respectively). For the location-scale model, regression parameters γ and ω have univariate normal priors with mean 0 and standard deviation 10^{10}, and the variance-covariance matrix of random intercepts and slopes has an improper inverse-Wishart prior $IW(0,-5)$ (Asparouhov & Muthen, 2023; Leckie et al., 2014; Rast et al., 2012). Prior sensitivity analyses conducted using the pilot data (the third wave of Erbas et al., 2018) confirmed that parameter estimates and inferences were robust to alternative prior specifications.[2]

We used four MCMC chains with a minimum of 30,000 iterations. The first half of the iterations from each chain were discarded as burn-in. Convergence was evaluated using visual inspection of Bayesian traceplots, the potential scale reduction (PSR) statistic, and the effective sample size (ESS; Brooks and Gelman, 1998; Gelman and Rubin, 1992). When convergence criteria were not met, a total of 40,000 iterations (half of them as burn-in) were run.

Effect synthesis across datasets

The Level-2 sample size in these ESM datasets range from $N = 70$ to $N = 303$, and consequently between-person regression coefficients are expected to show a relatively low precision. To obtain more reliable estimates of the association between ERS and affect dynamics, we synthesized study-specific regression coefficients using Bayesian hierarchical meta-analytic models using the `brms` package in R (Bürkner, 2017). Herein, each study was treated as a cluster providing multiple regression estimates (e.g., for PA and NA or for different emotions). Study-specific regression coefficients were treated as observations of underlying true effects, with known sampling variances using the posterior standard deviations obtained from the Bayesian location–scale models. Between-study heterogeneity was modeled through random intercepts of the datasets.

We used uninformative normal priors for regression coefficients ($\mathcal{N}(0,10)$) and a half-Student-*t* prior

---

[2] We conducted sensitivity analyses for the models for *positive affect*, *negative affect*, *happy*, and *sad* by varying the prior specification of the variance–covariance matrices of the latent traits. Specifically, we considered improper inverse-Wishart priors for the variances and covariances in the IRTree model ($IW(1,-5)$ and $IW(0,-5)$, respectively), as well as proper inverse-Wishart priors with different degrees of freedom for the variances and covariances of the random effects in the location–scale model ($IW(0,4)$ and $IW(0,5)$; see Henninger et al., 2025). Across these specifications, posterior estimates showed only minor variation, and posterior evidence for the regression effects, remained essentially unchanged. This indicates that inference was primarily driven by the likelihood rather than the prior and that the results are robust to reasonable prior choices.

for between-study standard deviation (Student-$t^{+}(0, 10)$ with df $= 3$).[3] We additionally conducted subgroup meta-analytic models for PA and NA, as well as positive and negative emotions to examine whether associations between ERS and affect dynamics differed by affective or emotional valence.

Models were estimated using Hamiltonian Monte Carlo sampling via Stan (Carpenter et al., 2017). Four Markov chains were run for 10,000 iterations each, including 4,000 burn-in iterations and an adaptation target acceptance rate of 0.99 to reduce divergent transitions. Convergence was assessed using visual inspection of trace plots, PSR, and ESS.

Software

We used Mplus 8.6 (L. K. Muthén & Muthén, 2017) for the analysis model, and the R packages `dplyr`, `tidyr`, `forcats`, `stringr`, and `ggplot2` (Wickham, 2016, 2025a, 2025b; Wickham et al., 2023, 2024) for data cleaning and plotting, `kableExtra` (Zhu, 2024) for creating tables, `MplusAutomation` (Hallquist & Wiley, 2018) and `misty` (Yanagida, 2025) for reading in the models in R and extracting convergence diagnostics and model coefficients, and the `brms` (Bürkner, 2017) and `metafor` packages (Viechtbauer, 2010) for the meta-analytic results synthesis in R 4.5 (RCoreTeam, 2025). All analysis scripts and Mplus output files are available on OSF.

Results

Descriptive statistics

We descriptively assessed person-specific mean levels of positive and negative affect, as well as of positive emotions, and negative emotions, across datasets (see Figure B1 in Appendix B). Unsurprisingly, participants reported higher levels of positive affect and positive emotions compared to negative affect and negative emotions. Considerable intraindividual variability is evident (grey bars indicate the 20% and 80% person-specific quantiles), alongside plausible differences in mean levels between items−for example, responses to *depressed* tended to be lower than responses to *tired*.

We furthermore evaluated the distribution of person-specific standard deviations for the same affect measures across datasets (see Figure B2 in Appendix B). Across datasets and emotions, most participants exhibit standard deviations between 0.1 and 0.2 (raw values range from 0 to 1). Notably, there are substantial inter-individual differences: some participants show very low variability (around 0.05), whereas others display high variability (above 0.3). In the datasets by Fried et al. (2022), Knouse et al. (2024), and Pavani et al. (2017), some participants show particularly zero variability in specific negative emotions, indicating that these emotions did not fluctuate over the course of the study (e.g., *sad*, *angry*, and *lonely* in the dataset by Knouse et al. (2024)).

[3] We conducted prior sensitivity analyses using more regularizing ($\mathcal{N}(0, 1)$) for regression coefficients and exponential priors ($\exp(1)$) for the between-study standard deviation. Meta-analytic results remained essentially unchanged.

Bayesian model evaluation: Location-scale models

We evaluated the Bayesian models by examining traceplots, as well as PSR and ESS values. Bayesian traceplots indicated good mixing and convergence (see supplementary material on OSF). All PSR values were ≤ 1.05, except the models for *angry*, *lonely*, and *tired* for the data by Fried et al. (2022); further inspection revelead that these models did not converge, even after increasing the number of iterations to 40,000, and were therefore excluded from subsequent analyses.

We further evaluated convergence for the parameters of primary interest, namely the regression coefficients capturing the associations between ERS and affect variability (ω_1), and between ERS and inertia (γ_{11}). For all retained models, tail ESS for these parameters were $\geq 5{,}000$ and bulk ESS were $\geq 2{,}000$, except for a small subset of parameters[4] for which we further increased the number of iterations to 40,000. These values represent the lowest observed ESS; for the majority of models, both bulk and tail ESS were substantially higher ($M_{\text{bulk ESS}} \geq 7{,}143$; $M_{\text{tail ESS}} \geq 15{,}428$). PSR values for these parameters were ≤ 1.02, indicating excellent convergence. Full ESS and PSR values for regression parameters are provided in Appendix C.

Two parameters—specifically, the effect of ERS on inertia for *sad* and *lonely* in the data by Knouse et al. (2024)—did not reach adequate ESS even after increasing the number of iterations. Importantly, this dataset exhibited additional estimation issues (see next paragraph) and was therefore excluded from the primary meta-analytic models.

Parameters of the location-scale models: Intercepts, variability, inertia, and their variances and covariances

Figure 3 shows estimates of intercepts, variability, inertia, as well as of the variance of variability and inertia in the different models, using PA and NA as an example. We can observe the following patterns: (a) the intercepts are higher for PA than for NA, (b) for the datasets by Knouse et al. (2024) and Pavani et al. (2017), the variance of intraindividual variability for NA is substantially higher compared to the other datasets, and (c) the remaining model parameters are highly consistent across the affect variables and datasets. While there is considerable variance in intraindividual variability, the variance in inertia is very small, suggesting that study participants did not vary much in their moment-to-moment affective inertia.

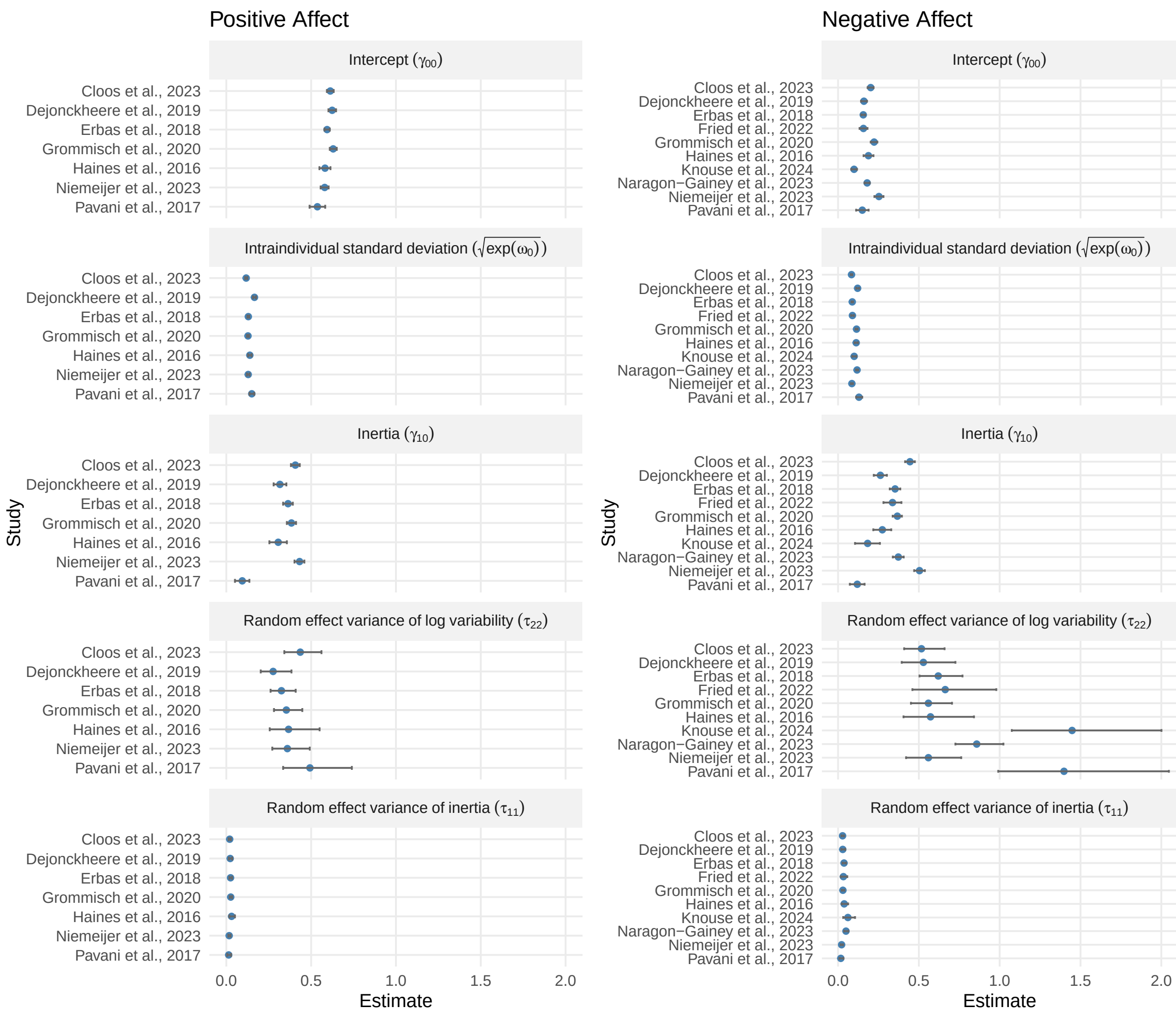
The model parameters for the positive and negative emotions are shown in Appendix D. Similar to PA and NA, there is substantial variance in intraindividual variability, but only small variance in inertia, indicating that individuals differ markedly in their emotional variability but little in their emotional inertia.

Figure 4 shows the estimated correlations between random effects for intercept, variability, and inertia for PA and NA. The correlations for the positive and negative emotions, which revealed similar patterns, are

[4] Specifically, bulk ESS for the effects of ERS on inertia for *sad*, *lonely*, and *angry* in the data by Knouse et al. (2024) and *anxious* in the data by Fried et al. (2022), and the effect of ERS on variability for *relaxed* in the data by Grommisch et al. (2020)

Figure 3

Estimates of intercepts, intraindividual standard deviation, inertia, random effect variances of log variability and inertia for positive and negative affect across datasets. Error bars represent 95% *credible intervals in the Bayesian location-scale models.*

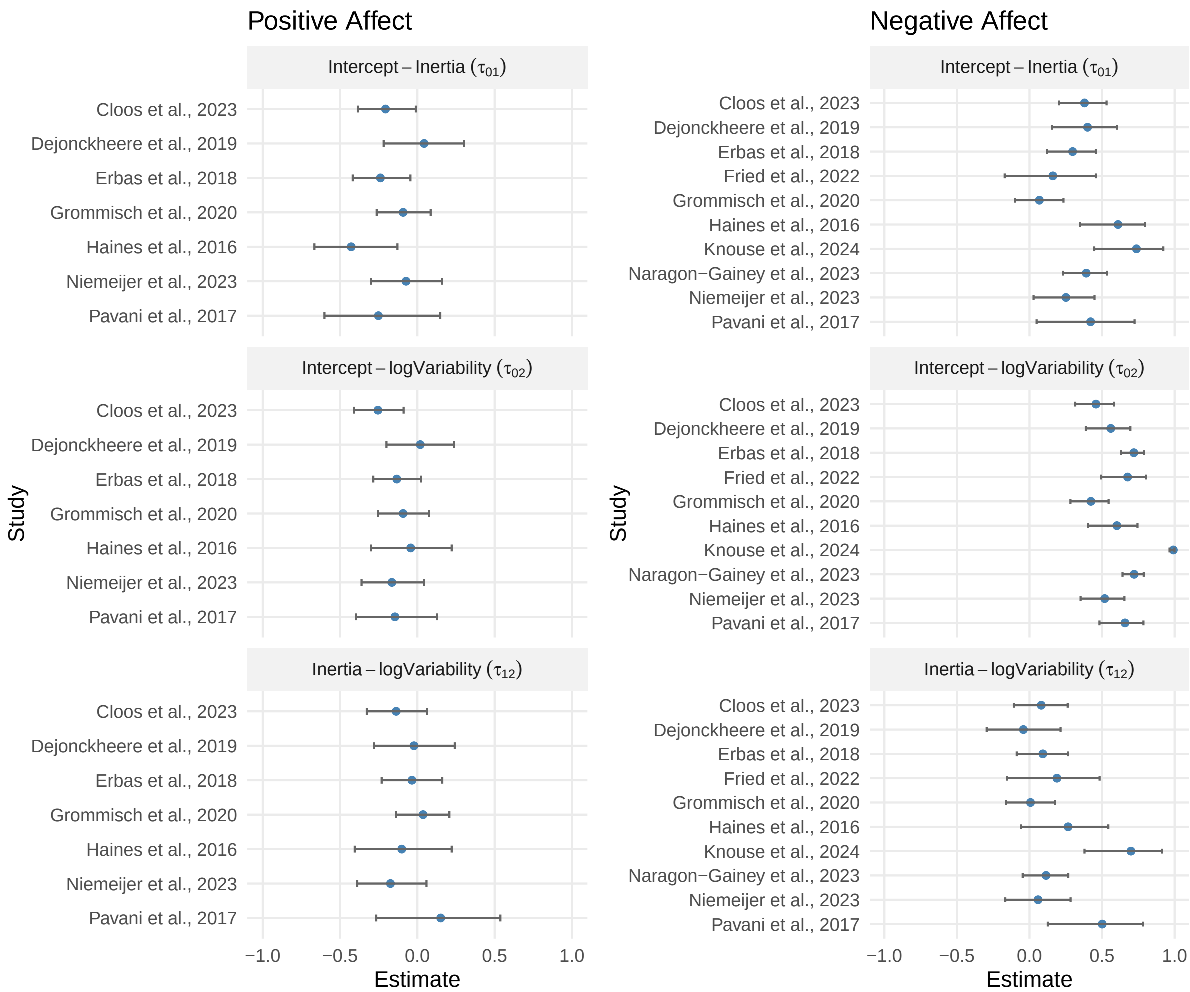


Note. Intraindividual standard deviations are computed as $\sqrt{\exp(\omega_0)}$. The random-effect variance of log variability (τ_{22}) represents the between-person heterogeneity in intraindividual variance on the log scale. To interpret this, $\exp(\sqrt{\tau_{22}})$ gives the proportional change in variability: a person 1SD over the mean in the random effect for variability would have $\exp(\sqrt{\tau_{22}})$ times higher with-person variability than a typical person.

shown in Appendix D. For positive affect, we observe small correlations across datasets. In contrast, for negative affect, correlations are positive, in particular the correlation between intercept and variability which are moderate to strong (r ranging from .4 to .6). These correlations indicate that individuals with higher mean levels of affect also tended to show greater intraindividual variability, a typical pattern in affect dynamic research (e.g., Bos

Figure 4

Estimates of correlations between random effects for intercept, log variability, and inertia for positive and negative affect. Error bars represent 95% *credible intervals in the Bayesian location-scale models.*



et al., 2019; Mestdagh et al., 2018).[5]

In one dataset (Knouse et al., 2024), however, the correlation between intercept and variability approached unity, reflecting strong floor effects in the measures of negative emotions, and thus in the aggregated measure for NA (as can also be seen in Figure B1 and by the spikes at zero in Figure B2 in Appendix B, meaning no variability was observed for these emotions through the study). These floor effects in the dataset are potentially due to the lower number of measurement occasions (6 days, 3x/day; see Table 1), which may have limited the potential occasions at which negative emotions could be observed. Because these floor effects may

[5] For bounded rating scales, the association between mean levels and variability is constrained by the scale limits and typically follows an inverted-U–shaped relation, such that variability is highest at intermediate mean levels and necessarily decreases as means approach the lower or upper boundaries. In the present datasets, mean affect levels were well below the upper scale boundary, placing the observations on the ascending (left) portion of this function.

have compromised the identifiability of random effects in the location-scale model, we report the synthesized effects excluding the dataset by Knouse et al. (2024) in the main text; results including it are provided in Appendix E for completeness and transparency.

Bayesian model evaluation: Hierarchical meta-analytic models

We examined traceplots, PSR, and ESS values. Traceplots indicated good mixing and convergence (see supplementary material on OSF). All PSR values were $\hat{R} = 1.00$, indicating excellent convergence. All bulk and tail ESS were $\geq 4{,}000$, indicating adequate Monte Carlo precision for posterior inference.

Primary hypothesis: Association between ERS and intraindividual variability across datasets

Figure 5 shows the estimated associations between ERS and intraindividual variability affect across datasets from the location-scale models. Figures 6 and 7 present the results for the specific positive and negative emotions, respectively. Points represent posterior regression coefficients from Bayesian models and the error bars indicate the 95% credible intervals.[6]

Positive and negative affect

For both PA and NA, most datasets show positive associations between ERS and variability. Exceptions include Haines et al. (2016) for PA, and Naragon-Gainey et al. (2023) for NA, which show near-zero associations. The pooled estimate across datasets and affect valences is positive ($\hat{\omega} = 0.24$, $SD = 0.08$, 95% CI $[0.09, 0.42]$).

In addition, we fitted separate subgroup models for PA and NA to examine whether affective valence influenced the results. These models likewise yielded positive pooled estimates within each subgroup (PA: $\hat{\omega} = 0.28$, $SD = 0.09$, 95% CI $[0.09, 0.46]$; NA: $\hat{\omega} = 0.21$, $SD = 0.09$, 95% CI $[0.04, 0.41]$). These results support our main hypothesis that higher ERS is associated with greater intraindividual variability in affect.

It is important to note that the composition of PA and NA scores differs across datasets, as each study assessed slightly different sets of positive and negative emotions from which PA and NA scores are derived, and some studies (e.g., Fried et al., 2022; Naragon-Gainey et al., 2023) only assessed negative emotions.

Positive and negative emotions

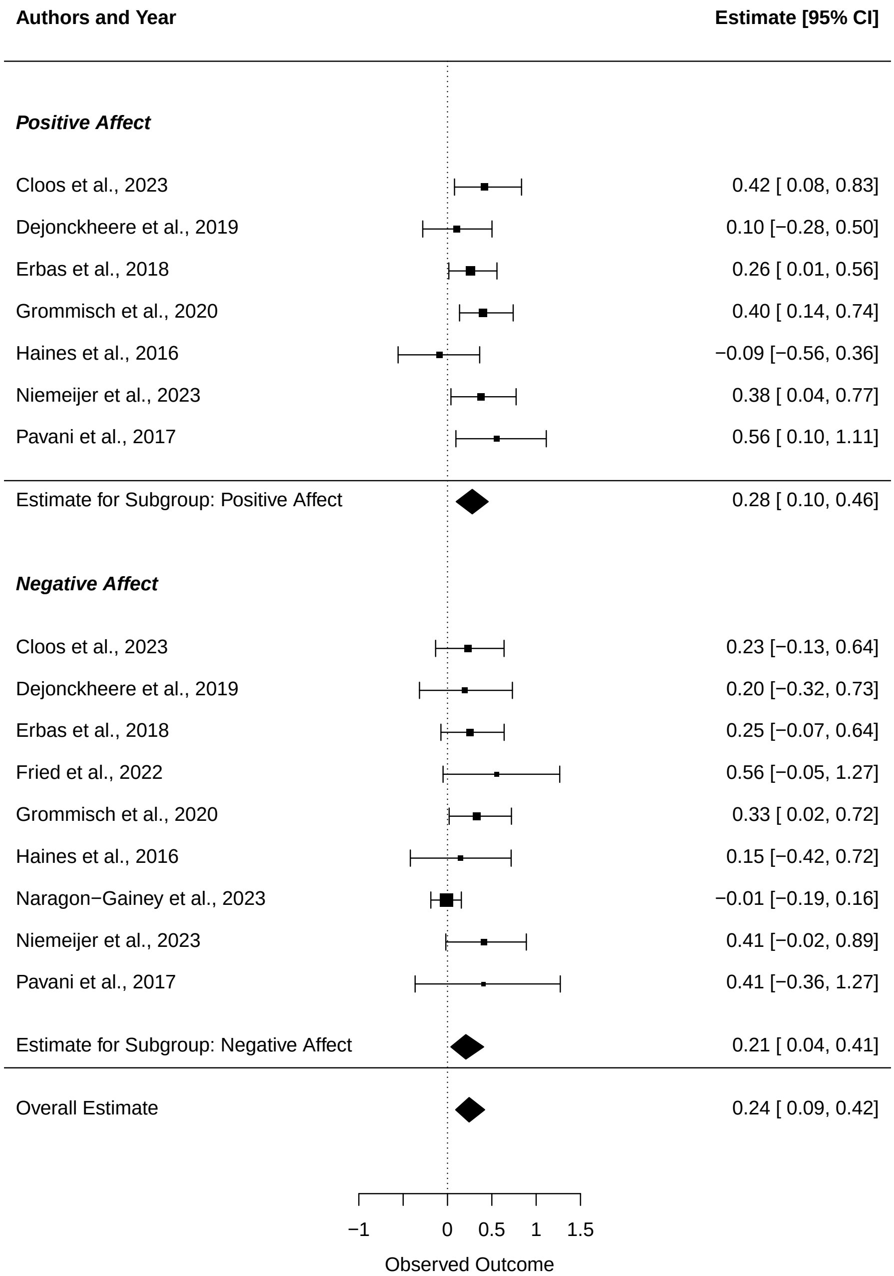
For positive and negative emotions, we observed patterns similar to those for aggregated PA and NA scores, with positive associations between ERS and variability in most datasets. The pooled estimate across all emotions is positive ($\hat{\omega} = 0.25$, $SD = 0.08$, 95% CI $[0.09, 0.42]$).

For positive emotions (Figure 6), results were largely consistent across datasets, again with the exception of the dataset by Haines et al. (2016), which showed a negative associations between ERS and variability of *happy* and *confident*. The pooled estimate across datasets for the subgroups of positive emotions is positive ($\hat{\omega} = 0.29$, $SD = 0.09$, 95% CI $[0.12, 0.47]$).

[6] For robustness and transparency, we report results including the dataset by Knouse et al. (2024) as well as results based on frequentist meta-analysis in Appendix E. Neither including this dataset nor using a frequentist approach did alter the results.

Figure 5

Estimated associations between ERS and intraindividual variability for positive and negative affect across datasets in the location-scale models.

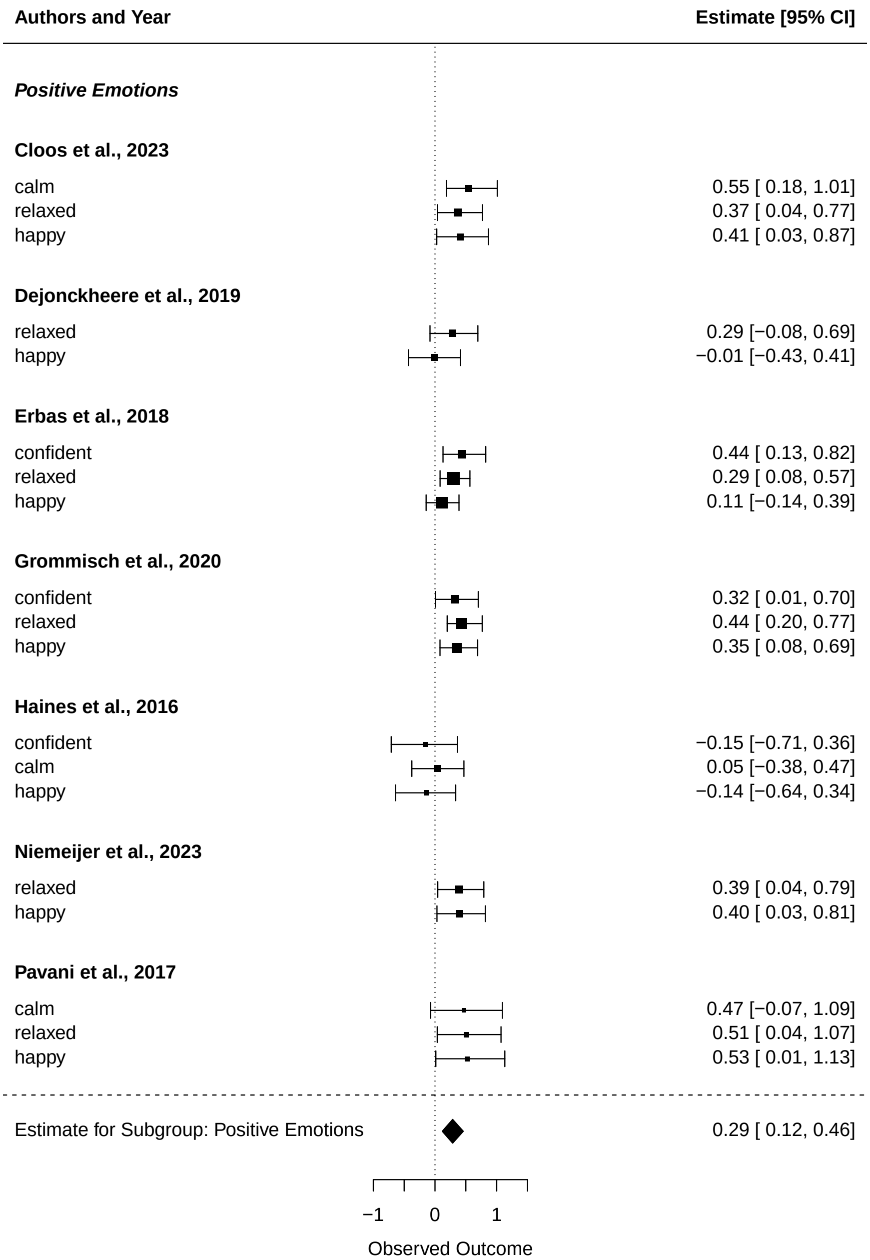


Note: Points represent posterior regression coefficients from Bayesian models; error bars indicate 95% credible intervals.

We observed that the associations between ERS and variability in negative emotions were relatively consistent within each dataset but slightly more heterogeneous between datasets (Figure 7). Associations tended to be positive for the datasets of Cloos et al. (2023), Dejonckheere et al. (2019a), Erbas et al. (2018), Fried et al. (2022), Grommisch et al. (2020), and Niemeijer et al. (2023), whereas those from Haines et al. (2016) and

Figure 6

Estimated associations between ERS and intraindividual variability for positive emotions across datasets in the location-scale models.

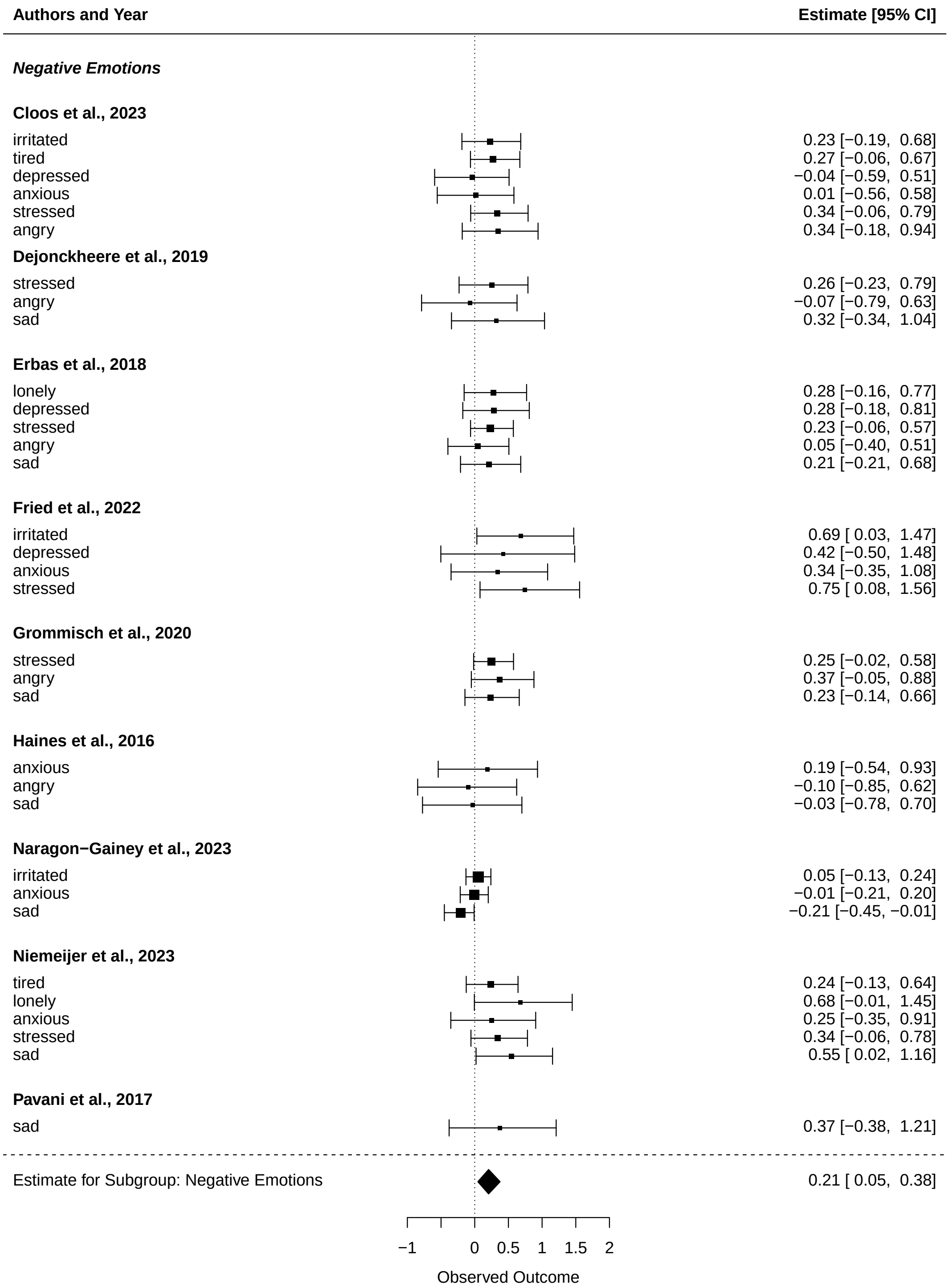


Note: Points represent posterior regression coefficients from Bayesian models; error bars indicate 95% credible intervals.

Naragon-Gainey et al. (2023) were closer to zero or negative. The pooled estimate across datasets for the negative emotion subgroup was positive ($\hat{\omega} = 0.21$, $SD = 0.09$, 95% CI $[0.04, 0.39]$). These results again corroborate our main hypothesis that higher ERS is associated with greater intraindividual variability in individual emotions.

Figure 7

Estimated associations between ERS and intraindividual variability for negative emotions across datasets in the location-scale models.



Note: Points represent posterior regression coefficients from Bayesian models; error bars indicate 95% credible intervals.

Interpretation of effect sizes

Because variability is modeled on the log scale (Equation 4), the regression coefficient ω_1 represents the change in the *log variance* per unit increase in ERS. Taking the exponential of ω_1 gives the multiplicative change in variance. To interpret this in terms of the standard deviation, we take the square root of the exponentiated coefficient:

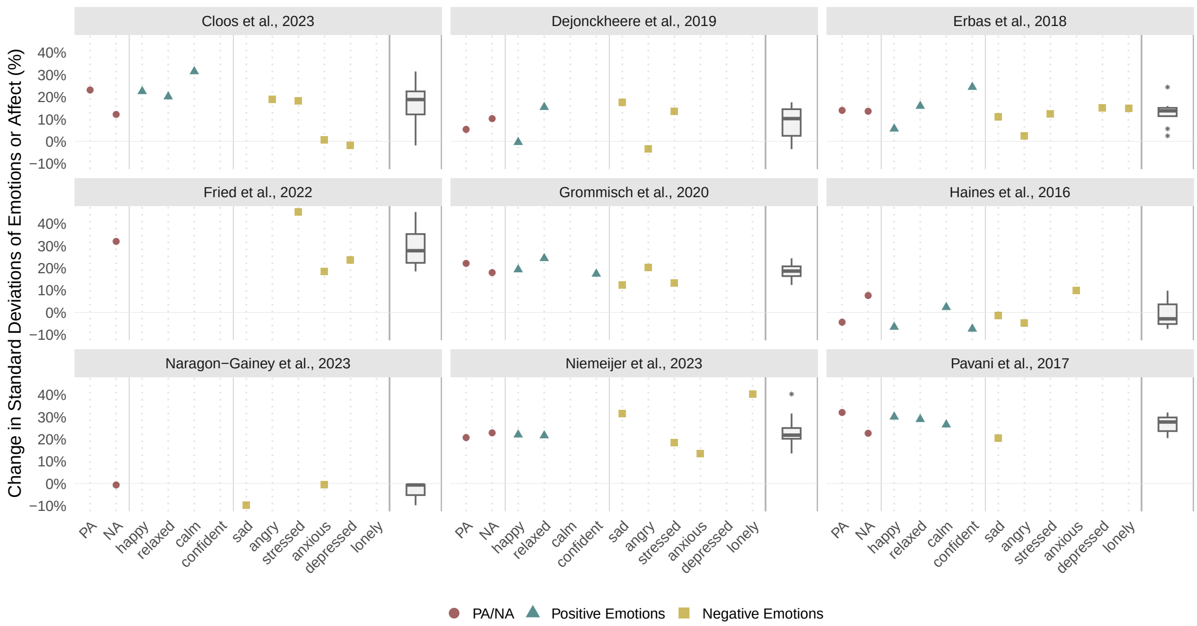
$$\text{Proportional change in SD} = \sqrt{\exp(\omega_1)}$$

Using the pooled meta-analytic estimate of the PA/NA model $\hat{\omega}_1 = 0.24$, we obtain $\sqrt{\exp(0.24)} \approx 1.128$. This corresponds to an approximate 12.8% increase in the standard deviation of affect or emotion for a one-unit increase in ERS, providing an intuitive and interpretable effect size. Note that posterior means of ERS scores ranged from -0.7 to 1.0 averaged across affect/emotions and datasets. Thus, a one-unit increase in ERS represents a substantial, but plausible change in response style.

Figure 8 shows the estimated percent change in intraindividual standard deviations associated with a one-unit increase in ERS across datasets and aggregated PA/NA scores or emotion items. For most datasets, higher ERS was consistently associated with increased variability, with expected changes ranging approximately from 5% to 35%. Overall, the figure underscores that ERS is systematically associated with larger estimates of variability across most affective measures, while also highlighting between-study and between-emotion differences that contribute to heterogeneity in the meta-analytic estimates.

Figure 8

Percent change in intraindividual standard deviations of affect or emotions for one unit increase in ERS per dataset for PA, NA, as well as positive and negative emotions.



Note: ERS: Extreme Response Style; PA: Positive Affect; NA: Negative Affect.

Secondary hypothesis: Association between ERS and inertia across datasets

Across the aggregated affect scores (PA/NA) and across all individual emotion measures, the regression coefficients in the individual datasets are close to zero. Analogously, the pooled estimate across datasets is nearly zero, indicating no association between ERS and inertia (affect: $\hat{\omega} = -0.01$, $SD = 0.02$, 95% CI $[-0.04, 0.02]$; emotions: $\hat{\omega} < 0.01$, $SD = 0.02$, 95% CI $[-0.03, 0.05]$; with similar results for the subgroups analyses). Appendix F presents the meta-analytic associations between ERS and inertia for PA and NA, as well as for positive and negative emotions. These results do not support our secondary hypothesis of a negative association between ERS and affect or emotional inertia.

Discussion

In the present study, we examined the association between ERS and affect dynamics−specifically intraindividual variability and inertia−across multiple ESM datasets. For both, aggregated PA and NA scores as well as individual emotion items, higher levels of ERS were consistently associated with greater intraindividual variability. In contrast, we found no evidence for an association between ERS and affect inertia. This null finding appears to be driven, at least in part, by the very small amount of between-person variance in inertia observed across datasets, limiting the potential for stable response-style differences to explain variation in inertia (see Figure 3 and Appendix D). These findings were robust across aggregated PA/NA scores and individual emotion items, across subgroup analyses taking the valence of affect and emotions into account, as well as alternative modeling approaches (see Appendix E). Furthermore, though the association between ERS and variability was largely consistent and positive, there remained heterogeneity in effect sizes across datasets.

Conceptually, these findings suggest that commonly used indicators of intraindividual variability may partly reflect systematic response behavior rather than purely psychological fluctuations in affect or emotion. That is, variability in repeated self-reports does not exclusively capture genuine affective dynamics but may also be influenced by stable individual differences in how respondents use rating scales.

In contrast, the absence of evidence regarding ERS effects on inertia may suggest that temporal dependencies in affect are less susceptible to response style biases, at least in the datasets examined here. More generally, the small between-person variance in inertia indicates that individual differences in affective carry-over may be limited in naturalistic ESM data, reducing the scope for response styles to explain such differences, specifically their tendency to use extreme categories.

From a methodological perspective, these results have important implications for multivariate and longitudinal modeling of ILD. If ERS systematically inflates within-person variance, failing to account for response styles may bias estimates obtained from location–scale models, multilevel dynamic models, or related approaches that explicitly model intraindividual variability, particularly when variability parameters are interpreted as stable person characteristics.

By jointly modeling psychometric response tendencies and affect dynamics, as done here, researchers can potentially obtain more valid estimates of person-specific parameters and improve the interpretability of model-based inferences.

Strengths and limitations

This study has several notable strengths. First, by synthesizing results across multiple ESM datasets, we increased effective between-person sample size, allowing for more precise estimation of Level-2 associations than would be possible within any single dataset. Second, the joint IRTree–location-scale modeling approach enabled uncertainty in the estimation of ERS trait levels to be propagated into the analysis of affect dynamics, reducing bias that could arise from treating ERS estimates as error-free predictors. Third, the consistency of results across diverse datasets supports the robustness and generalizability of the main findings. Finally, similar results for aggregated PA/NA scores and individual emotion items indicate that the association between ERS and variability is not driven by a particular level of aggregation.

Several limitations should be acknowledged. The included datasets were not originally designed to address the present research questions and therefore varied with respect to emotion measures, scale formats, sampling designs, and the constructs assessed in background questionnaires. These differences introduced researcher degrees of freedom, for example, in selecting background scales for modeling ERS, which we addressed through preregistration.

In particular, ESM emotion items, measurement scales, and sampling schemes differed across datasets, and not all datasets provided information for all emotion items and valence. As a result, pooled estimates for PA and NA were based on only partially overlapping sets of emotions. At the same time, such heterogeneity is characteristic of ESM research and therefore enhances the ecological validity and generalizability of our findings (see Brose et al., 2020).

Additionally, differences in sampling intensity (e.g., number of days and measurement occasions per day) may affect the reliability of affect dynamics estimates and contributed to non-convergence in some models. Excluding non-converged models further reduced the number of datasets included in the final analyses.

The selection of datasets represents a convenience sample drawn from open data repositories and our research network. Although the included studies varied in design and sample characteristics, they may not be fully representative of the broader population of ESM studies. To explore the potential impact of study design heterogeneity, we examined the impact of the number of ESM days, the number of measurement occasions per day, and the type of ESM response scale (slider vs. rating). We found that none of these key design characteristics moderated the estimated associations between ERS and intraindividual variability. The absence of significant moderator effects may again reflect the comparably small number of included studies, which limits statistical power to detect potential moderation by study design characteristics.

A further limitation is that negative emotion items were right-skewed, leading to positive correlations between person means and standard deviations. Although such mean–variance associations are typical in ESM data (e.g., Hamaker et al., 2018; Mestdagh et al., 2018), they may limit the validity and interpretability of variability estimates. At the same time, the observed between-person differences in variability and their associations with ERS were substantial and psychologically plausible (except for one dataset that was excluded from further analyses), suggesting that the results are not driven solely by distributional artifacts. We did not apply transformations or alternative distributional models (e.g., by Lesaffre et al., 2007), as the joint IRTree–location-scale model is already complex and challenging to estimate; adding further modeling layers would have increased the risk of non-convergence while reducing the interpretability of results.

Lastly, we did not explicitly account for unequal sampling intervals or the night gap between measurement days when modeling inertia (e.g., Berkhout et al., 2025; Hamaker et al., 2018). As a result, we treated inertia as having the same interpretation across different sampling schedules, and implicitly assumed that affective and emotional processes pause overnight and resume the following day. This assumption may not hold universally and may be more problematic for some datasets than others, given that studies differed in the number and timing of daily assessments.

However, given the very small between-person variance in inertia observed across datasets, it seems unlikely that explicitly modeling unequal intervals or night gaps would substantially change the present conclusions regarding the absence of systematic ERS effects on inertia. Nonetheless, future research may benefit from incorporating continuous-time or interval-sensitive models to further disentangle substantive temporal dynamics from measurement-related influences.

Future directions

Several promising avenues for future research follow from the present findings. First, an important next step is to examine whether ERS biases substantive conclusions about the role of affect dynamics in predicting psychological outcomes. For example, intraindividual variability has been linked to depressive symptoms and other forms of psychopathology. Future studies could investigate whether including ERS as a covariate attenuates or alters the estimated associations between affect variability and such outcomes, thereby clarifying the extent to which previously reported effects may partly reflect systematic response behavior rather than genuine affective instability.

Second, although the joint IRTree–location-scale model provides a principled way to account for ERS, its complexity and computational demands may limit its feasibility for applied and clinical researchers. Future work could therefore explore simpler approaches to controlling for ERS, such as using manifest ERS indices derived from background questionnaires (e.g., proportions of extreme responses instead of the latent ERS estimates from the IRTree model) as covariates in longitudinal multilevel models. Simulation studies could compare the performance of different strategies, such as latent versus manifest ERS controls, under varying response-style prevalences.

Third, despite accounting for ERS, substantial unexplained between-person variance in intraindividual variability remained. Future research should focus on identifying substantive determinants of variability, such as contextual sensitivity or emotion regulation strategies, while explicitly separating these processes from response-style effects.

Fourth, the present study focused on affective experiences, but the association between ERS and intraindividual variability is likely not limited to this domain. Similar mechanisms may operate for other constructs commonly assessed using intensive longitudinal methods, such as goal pursuit, cognitive states, or daily behaviors. Extending this line of research to additional domains would help clarify the generality of ERS-related biases in ILD modeling.

Finally, the present findings raise broader questions about person-specific data quality in ILD. ERS may be largely inconsequential for some individuals but highly influential for those with stronger extreme responding tendencies. Future research could explore person-specific indicators of reliability and validity that help identify when variability estimates are likely to be contaminated by response styles and when they can be interpreted as reflecting genuine psychological dynamics.

Conclusion

Taken together, the present findings demonstrate that ERS constitutes a meaningful psychometric confound in the estimation of intraindividual variability in ESM data, while leaving affect inertia largely unaffected. These results underscore the importance of accounting for response styles when interpreting variability parameters in multilevel and dynamic models of ILD. By bridging psychometric modeling and intensive longitudinal analysis, the present study contributes to more valid and interpretable modeling of affect dynamics in daily life.

Declarations

AI statement: We used used Generative Artificial Intelligence (AI; GPT-5.2) for language improvement, generating R code (e.g., for visualizations or data cleaning), and interactive online search. We carefully reviewed and edited all AI-generated content.

Supplemental online material: Preregistration, data preprocessing scripts, analysis scripts, Mplus output files, and additional materials are available on OSF.

Ethics approval: This study involved secondary analyses of previously collected datasets. All original studies received ethical approval from their respective institutional review boards, with the exception of the dataset reported by Pavani et al. (2017), which was exempt from ethics review in accordance with French legislation permitting non-interventional research to proceed without formal ethics committee approval.

References

Adams, D. J., Bolt, D. M., Deng, S., Smith, S. S., & Baker, T. B. (2019). Using multidimensional item response theory to evaluate how response styles impact measurement. *British Journal of Mathematical and Statistical Psychology*, *72*(3), 1–20. https://doi.org/10.1111/bmsp.12169

Asparouhov, T., & Muthen, B. (2023). *Bayesian analysis using Mplus: Technical implementation* (tech. rep.). http://www.statmodel.com/download/Bayes2.pdf

Baird, B. M., Lucas, R. E., & Donnellan, M. B. (2017). The role of response styles in the assessment of intraindividual personality variability. *Journal of Research in Personality*, *69*, 170–179. https://doi.org/10.1016/j.jrp.2016.06.015

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156. https://doi.org/10.1509/jmkr.38.2.143.18840

Berkhout, S. W., Schuurman, N. K., & Hamaker, E. L. (2025). Let sleeping dogs lie? How to deal with the night gap problem in experience sampling method data. *Psychological Methods*. https://doi.org/10.1037/met0000762

Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, *17*, 665–678. https://doi.org/10.1037/a0028111

Böckenholt, U., & Meiser, T. (2017). Response style analysis with threshold and multi-process IRT models: A review and tutorial. *British Journal of Mathematical and Statistical Psychology*, *70*, 159–181. https://doi.org/10.1111/bmsp.12086

Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement*, *33*, 335–352. https://doi.org/10.1177/0146621608329891

Bolt, D. M., Lu, Y., & Kim, J.-S. (2014). Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods*, *19*, 528–541. https://doi.org/10.1037/met0000016

Bos, E. H., de Jonge, P., & Cox, R. F. A. (2019). Affective variability in depression: Revisiting the inertia–instability paradox. *British Journal of Psychology*, *110*(4), 814–827. https://doi.org/10.1111/bjop.12372

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455. https://doi.org/10.1080/10618600.1998.10474787

Brose, A., Schmiedek, F., Gerstorf, D., & Voelkle, M. C. (2020). The measurement of within-person affect variation. *Emotion*, *20*(4), 677–699. https://doi.org/10.1037/emo0000583

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*. https://doi.org/10.18637/jss.v080.i01

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1–32. https://doi.org/10.18637/jss.v076.i01

Carver, C. S. (2015). Control processes, priority management, and affective dynamics. *Emotion Review*, *7*(4), 301–307. https://doi.org/10.1177/1754073915590616

Cloos, L., Ceulemans, E., & Kuppens, P. (2023). Development, validation, and comparison of self-report measures for positive and negative affect in intensive longitudinal research. *Psychological Assessment*, *55*(3), 189–204. https://doi.org/10.1037/pas0001200

Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *The Journal of Abnormal and Social Psychology*, *60*, 151–174. https://doi.org/10.1037/h0040372

Czyz, E. K., Yap, J. R., King, C. A., & Nahum-Shani, I. (2021). Using intensive longitudinal data to identify early predictors of suicide-related outcomes in high-risk adolescents: Practical and conceptual considerations. *Assessment*, *28*(8), 1949–1959. https://doi.org/10.1177/1073191120939168

De Boeck, P., & Partchev, I. (2012). IRTrees: Tree-based item response models of the GLMM family. *Journal of Statistical Software*, *48*, 1–28. https://doi.org/10.18637/jss.v048.c01

De Jong, M. G., Steenkamp, J.-B. E., Fox, J.-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, *45*, 104–115. https://doi.org/10.1509/jmkr.45.1.104

Dejonckheere, E., Kalokerinos, E. K., Bastian, B., & Kuppens, P. (2019a). Poor emotion regulation ability mediates the link between depressive symptoms and affective bipolarity. *Cognition and Emotion*, *33*(5), 1076–1083. https://doi.org/10.1080/02699931.2018.1524747

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

Deng, S., McCarthy, D. E., Piper, M. E., Baker, T. B., & Bolt, D. M. (2018). Extreme response style and the measurement of intra-individual variability in affect. *Multivariate Behavioral Research*, *53*(2), 199–218. https://doi.org/10.1080/00273171.2017.1413636

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, *76*(4), 662–676. https://doi.org/10.1037/0022-3514.76.4.662

Erbas, Y., Ceulemans, E., Kalokerinos, E. K., Houben, M., Koval, P., Pe, M. L., & Kuppens, P. (2018). Why I don't always know what I'm feeling: The role of stress in within-person fluctuations in emotion differentiation. *Journal of Personality and Social Psychology*, *115*(2), 179–191. https://doi.org/10.1037/pspa0000126

Ernst, A. F., Timmerman, M. E., Jeronimus, B. F., & Albers, C. J. (2021). Insights into individual differences in emotion dynamics with clustering. *Assessment*, *28*(4), 1186–1206. https://doi.org/10.1177/1073191119873714

Falk, C. F., & Cai, L. (2016). A flexible full-information approach to the modeling of response styles. *Psychological Methods*, *21*, 328–347. https://doi.org/10.1037/met0000059

Ferrando, P. J. (2014). A factor-analytic model for assessing individual differences in response scale usage. *Multivariate Behavioral Research*, *49*, 390–405. https://doi.org/10.1080/00273171.2014.911074

Franck, E., & De Raedt, R. (2007). Self-esteem reconsidered: Unstable self-esteem outperforms level of self-esteem as vulnerability marker for depression. *Behaviour Research and Therapy*, *45*(7), 1531–1541. https://doi.org/10.1016/j.brat.2007.01.003

Fried, E. I., Papanikolaou, F., & Epskamp, S. (2022). Mental health and social contact during the COVID-19 pandemic: An ecological momentary assessment study. *Clinical Psychological Science*, *10*(2), 340–354. https://doi.org/10.1177/21677026211017839

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Grommisch, G., Koval, P., Hinton, J. D. X., Gleeson, J., Hollenstein, T., Kuppens, P., & Lischetzke, T. (2020). Modeling individual differences in emotion regulation repertoire in daily life with multilevel latent profile analysis. *Emotion*, *20*(8), 1462–1474. https://doi.org/10.1037/emo0000669

Haines, S. J., Gleeson, J., Kuppens, P., Hollenstein, T., Ciarrochi, J., Labuschagne, I., Grace, C., & Koval, P. (2016). The wisdom to know the difference: Strategy-situation fit in emotion regulation in daily life is associated with well-being. *Psychological Science*, *27*(12), 1651–1659. https://doi.org/10.1177/0956797616669086

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling*, 1–18. https://doi.org/10.1080/10705511.2017.1402334

Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. *Multivariate Behavioral Research*, *53*(6), 820–841. https://doi.org/10.1080/00273171.2018.1446819

Hamaker, E. L., Grasman, R. P. P. P., & Kamphuis, J. H. (2016). Modeling BAS dysregulation in bipolar disorder: Illustrating the potential of time series analysis. *Assessment*, *23*(4), 436–446. https://doi.org/10.1177/1073191116632339

Hamilton, D. L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, *69*, 192–203. https://doi.org/10.1037/h0025606

Hedeker, D., Demirtas, H., & Mermelstein, R. J. (2009a). A mixed ordinal location scale model for analysis of ecological momentary assessment (EMA) data. *Statistics and its Interface*, *2*(4), 391–401. https://doi.org/10.4310/sii.2009.v2.n4.a1

Hedeker, D., & Mermelstein, R. J. (2020). Modeling variation in intensive longitudinal data. In *Multilevel Modeling Methods with Introductory and Advanced Applications*. Information Age Publishing. https://doi.org/10.1108/978-1-64802-873-120251015

Hedeker, D., Mermelstein, R. J., Berbaum, M. L., & Campbell, R. T. (2009b). Modeling mood variation associated with smoking: An application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. *Addiction*, *104*(2), 297–307. https://doi.org/10.1111/j.1360-0443.2008.02435.x

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location/scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, *64*(2), 627–634. https://doi.org/10.1111/j.1541-0420.2007.00924.x

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Henninger, M. (2021). A novel Partial Credit extension using varying thresholds to account for response styles. *Journal of Educational Measurement*, *58*, 104–129. https://doi.org/10.1111/jedm.12268

Henninger, M., & Meiser, T. (2020a). Different approaches to modeling response styles in Divide-by-Total IRT models (Part I): A model integration. *Psychological Methods*, *25*, 560–576. https://doi.org/10.1037/met0000249

Henninger, M., & Meiser, T. (2020b). Different approaches to modeling response styles in Divide-by-Total IRT models (Part II): Applications and novel extensions. *Psychological Methods*, *25*, 577–595. https://doi.org/10.1037/met0000268

Henninger, M., & Meiser, T. (2022). Quality control: Response style modeling. In D. McCaffrey (Ed.), *International Encyclopedia of Education*. https://doi.org/10.1016/B978-0-12-818630-5.10041-7

Henninger, M., Vanhasbroeck, N., & Tuerlinckx, F. (2025). Affect dynamics or response bias? The relationship between extreme response style and affect dynamics in a controlled experiment. *Psychological Assessment*. https://doi.org/10.1037/pas0001370

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. https://doi.org/10.1037/a0038822

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment. *Psychological Methods*, *13*, 354–375. https://doi.org/10.1037/a0014173

Jangraw, D. C., Keren, H., Sun, H., Bedder, R. L., Rutledge, R. B., Pereira, F., Thomas, A. G., Pine, D. S., Zheng, C., Nielson, D. M., & Stringaris, A. (2023). A highly replicable decline in mood during rest and simple tasks. *Nature Human Behaviour*, *7*(4), 596–610. https://doi.org/10.1038/s41562-023-01519-7

Jenkins, B. N., Ong, L. Q., Ong, A. D., Lee, H. Y., & Boehm, J. K. (2024). Mean affect moderates the association between affect variability and mental health. *Affective Science*, *5*(2), 99–114. https://doi.org/10.1007/s42761-024-00238-0

Jongerling, J., Laurenceau, J. P., & Hamaker, E. L. (2015). A multilevel AR(1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, *50*(3), 334–349. https://doi.org/10.1080/00273171.2014.1003772

Kalokerinos, E., Russo-Batterham, D., Grewal, K., Moeck, E., Wilson, E., Smith, I., Greenaway, K., Garrett, P., Shrestha, K., Kuppens, P., & Koval, P. (2023). The EMOTE Database: An open, searchable database of experience sampling data mapping everyday life. *Manuscript in preparation*. Retrieved June 25, 2025, from https://emotedatabase.com

Khorramdel, L., & von Davier, M. (2014). Measuring response styles across the Big Five: A multiscale extension of an approach using multinomial processing trees. *Multivariate Behavioral Research*, *49*, 161–177. https://doi.org/10.1080/00273171.2013.866536

Knouse, L. E., Ziegler, M., Lavine, I., Zhang, J., Cheng, Y., & Ul Ain, H. (2024). Avoidant automatic thoughts are associated with task avoidance and inattention in the moment. *Cognitive Therapy and Research*, *48*(5), 866–879. https://doi.org/10.1007/s10608-023-10410-8

Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, *13*(6), 1132–1141. https://doi.org/10.1037/a0033579

Koval, P., Sütterlin, S., & Kuppens, P. (2016). Emotional inertia is associated with lower well-being when controlling for differences in emotional context. *Frontiers in Psychology*, *6:1997*. https://doi.org/10.3389/fpsyg.2015.01997

Kuppens, P., Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. *Emotion*, *7*(2), 262–274. https://doi.org/10.1037/1528-3542.7.2.262

Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology*, *17*, 22–26. https://doi.org/10.1016/j.copsyc.2017.06.004

Larsen, R. J., & Diener, E. (1987). Affect intensity as an individual difference characteristic: A review. *Journal of Research in Personality*, *21*(1), 1–39. https://doi.org/10.1016/0092-6566(87)90023-7

Leckie, G., French, R., Charlton, C., & Browne, W. (2014). Modeling heterogeneous variance-covariance components in two-level models. *Journal of Educational and Behavioral Statistics*, *39*(5), 307–332. https://doi.org/10.3102/1076998614546494

Lesaffre, E., Rizopoulos, D., & Tsonaka, R. (2007). The logistic transform for bounded outcome scores. *Biostatistics*, *8*(1), 72–85. https://doi.org/10.1093/biostatistics/kxj034

Maher, J. P., Dzubur, E., Nordgren, R., Huh, J., Chou, C. P., Hedeker, D., & Dunton, G. F. (2019). Do fluctuations in positive affective and physical feeling states predict physical activity and sedentary time? *Psychology of Sport and Exercise*, *41*, 153–161. https://doi.org/10.1016/j.psychsport.2018.01.011

McKone, K. M. P., & Silk, J. S. (2022). The emotion dynamics conundrum in developmental psychpathology: Similarities, distinctions, and adaptiveness of affective variability and socioaffective flexibility. *Clinical Child and Family Psychology Review*, *25*, 44–74. https://doi.org/10.1007/s10567-022-00382-8

McNeish, D. (2021). Specifying location-scale models for heterogeneous variances as multilevel SEMs. *Organizational Research Methods*, *24*(3), 630–653. https://doi.org/10.1177/1094428120913083

Meiser, T., Plieninger, H., & Henninger, M. (2019). IRTree models with ordinal and multidimensional decision nodes for response styles and trait-based rating responses. *British Journal of Mathematical and Statistical Psychology*, *72*, 501–516. https://doi.org/10.1111/bmsp.12158

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, *23*(4), 690–707. https://doi.org/10.1037/met0000153

Mohr, C. D., Arpin, S., & McCabe, C. T. (2015). Daily affect variability and context-specific alcohol consumption. *Drug and Alcohol Review*, *34*(6), 581–587. https://doi.org/10.1111/dar.12253

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335. https://doi.org/10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

Naragon-Gainey, K., DeMarree, K. G., Kyron, M. J., McMahon, T. P., Park, J., & Biehler, K. M. (2023). Decentering from emotions in daily life: Dynamic associations with affect, symptoms, and well-being. *Clinical Psychological Science*, *11*(5), 841–862. https://doi.org/10.1177/21677026221147262

Niemeijer, K. (2024). Quantifying the robustness of mobile sensing for predicting momentary emotions: A multiverse study. https://doi.org/10.17605/OSF.IO/HCKSQ

Niemeijer, K., Mestdagh, M., Verdonck, S., Meers, K., & Kuppens, P. (2023). Combining experience sampling and mobile sensing for digital phenotyping with m-Path Sens: Performance study. *JMIR Formative Research*, *7*(1), e43296. https://doi.org/10.2196/43296

Palmier-Claus, J. E., Taylor, P. J., Gooding, P., Dunn, G., & Lewis, S. (2012). Affective variability predicts suicidal ideation in individuals at ultra-high risk of developing psychosis: An experience sampling study. *British Journal of Clinical Psychology*, *51*(1), 72–83. https://doi.org/10.1111/j.2044-8260.2011.02013.x

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp. 17–59). Academic Press. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

Pavani, J.-B., Le Vigouroux, S., Kop, J.-L., Congard, A., Dauvier, B., & Denissen, J. (2017). A network approach to affect regulation dynamics and personality trait-induced variations: Extraversion and neuroticism moderate reciprocal influences between affect and affect regulation strategies. *European Journal of Personality*, *31*(4), 329–346. https://doi.org/10.1002/per.2109

Piasecki, T. M., Hedeker, D., Dierker, L. C., & Mermelstein, R. J. (2016). Progression of nicotine dependence, mood level, and mood variability in adolescent smokers. *Psychology of Addictive Behaviors*, *32*(8), 859–860. https://doi.org/10.1037/adb0000429_2

Plieninger, H. (2017). Mountain or molehill? A simulation study on the impact of response styles. *Educational and Psychological Measurement*, *77*, 32–53. https://doi.org/10.1177/0013164416636655

Plieninger, H. (2020). Developing and applying IR-Tree models : Guidelines, caveats, and an extension to multiple groups. *Organizational Research Methods*, 1–17. https://doi.org/10.1177/1094428120911096

Plieninger, H., & Heck, D. W. (2018). A new model for acquiescence at the interface of psychometrics and cognitive psychology. *Multivariate Behavioral Research*, *53*, 633–654. https://doi.org/10.1080/00273171.2018.1469966

Plieninger, H., & Meiser, T. (2014). Validity of multiprocess IRT models for separating content and response styles. *Educational and Psychological Measurement*, *74*, 875–899. https://doi.org/10.1177/0013164413514998

Rast, P., Hofer, S. M., & Sparks, C. (2012). Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. *Multivariate Behavioral Research*, *47*(2), 177–200. https://doi.org/10.1080/00273171.2012.658328

RCoreTeam. (2025). R: A langauge and environment for statistical computing. https://www.R-project.org/

Röcke, C., & Brose, A. (2013). Intraindividual variability and stability of affect and well-being. *The Journal of Gerontopsychology and Geriatric Psychiatry*, *26*(3), 185–199. https://doi.org/10.1024/1662-9647/a000094

Russell, J. J., Moskowitz, D. S., Zuroff, D. C., Sookman, D., & Paris, J. (2007). Stability and variability of affective experience and interpersonal behavior in borderline personality disorder. *Journal of Abnormal Psychology*, *116*(3), 578–588. https://doi.org/10.1037/0021-843X.116.3.578

Scott, S. B., Sliwinski, M. J., Zawadzki, M., Stawski, R. S., Kim, J., Marcusson-Clavertz, D., Lanza, S. T., Conroy, D. E., Buxton, O., Almeida, D. M., & Smyth, J. M. (2020). A coordinated analysis of variance in affect in daily life. *Assessment*, *27*(8), 1683–1698. https://doi.org/10.1177/1073191118799460

Smit, A. C., Schat, E., & Ceulemans, E. (2023). The exponentially weighted moving average procedure for detecting changes in intensive longitudinal data in psychological research in real-time: A tutorial showcasing potential applications. *Assessment*, *30*(5), 1354–1368. https://doi.org/10.1177/10731911221086985

Soland, J., & Kuhfeld, M. (2020). Do response styles affect estimates of growth on social-emotional constructs? Evidence from four years of longitudinal survey scores. *Multivariate Behavioral Research*, *56*(6), 1–21. https://doi.org/10.1080/00273171.2020.1778440

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective dynamics in psychopathology. *Emotion Review*, 1–7. https://doi.org/10.1177/1754073915590617

Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology*, *117*(3), 647–661. https://doi.org/10.1037/a0012532

Ulitzsch, E., Henninger, M., & Meiser, T. (2024). Differences in response-scale usage are ubiquitous in cross-country comparisons and a potential driver of elusive relationships. *Scientific Reports*, *14*(1), 1–4. https://doi.org/10.1038/s41598-024-60465-0

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*, 195–217. https://doi.org/10.1093/ijpor/eds021

Vanhasbroeck, N., Vanbelle, S., Moors, A., Vanpaemel, W., & Tuerlinckx, F. (2024). Chasing consistency: On the measurement error in self-reported affect in experiments. *Behavior Research Methods*, *56*(4), 3009–3022. https://doi.org/10.3758/s13428-023-02290-3

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48. https://doi.org/10.18637/jss.v036.i03

Wang, L., Hamaker, E., & Bergeman, C. (2012). Investigating inter-individual differences in short-term intra- individual variability. *Psychological Methods*, *17*, 19–44. https://doi.org/10.1037/a0029317

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, *34*, 105–121. https://doi.org/10.1177/0146621609338593

Wendt, L. P., Wright, A. G., Pilkonis, P. A., Woods, W. C., Denissen, J. J., Kühnel, A., & Zimmermann, J. (2020). Indicators of affect dynamics: Structure, reliability, and personality correlates. *European Journal of Personality*, *34*(6), 1060–1072. https://doi.org/10.1002/per.2277

Wetzel, E., & Carstensen, C. H. (2017). Multidimensional modeling of traits and response styles. *European Journal of Psychological Assessment*, *33*, 352–364. https://doi.org/10.1027/1015-5759/a000291

Wetzel, E., Carstensen, C. H., & Böhnke, J. R. (2013). Consistency of extreme response style and non-extreme response style across traits. *Journal of Research in Personality*, *47*, 178–189. https://doi.org/10.1016/j.jrp.2012.10.010

Wichers, M., Wigman, J. T. W., & Myin-Germeys, I. (2015). Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review*, *7*(4), 362–367. https://doi.org/10.1177/1754073915590623

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org/

Wickham, H. (2025a). Forcats: Tools for Working with Categorical Variables (Factors). *R package version 1.0.1*. https://CRAN.R-project.org/package=forcats

Wickham, H. (2025b). Stringr: Simple, Consistent Wrappers for Common String Operations. *R package version 1.6.0*. https://CRAN.R-project.org/package=stringr

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). Dplyr: A Grammar of Data Manipulation. *R package version 1.1.4*. https://cran.r-project.org/package=dplyr

Wickham, H., Vaughan, D., & Girlich, M. (2024). Tidyr: Tidy Messy Data. *R package version 1.3.1*. https://CRAN.R-project.org/package=tidyr

Yanagida, R. (2025). Misty: Miscellaneous Functions 'T. Yanagida'. https://CRAN.R-project.org/package=misty

Zhu, H. (2024). kableExtra: Construct complex table with 'kable' and pipe syntax. *R package version 1.4.0*. https://doi.org/10.32614/CRAN.package.kableExtra

Appendix A

Deviations from Preregistration

Here, we document any deviations from our preregistered study design and analysis plan.

- On page 6 of the preregistration, we stated that results would be synthesized across studies using the `metafor` package in R (Viechtbauer, 2010). However, to remain within a Bayesian estimation framework, we instead conducted a multilevel meta-analysis using the `brms` package (Bürkner, 2017). The `metafor` package was subsequently used to create forest plots. Meta-analytic results based on a frequentist are provided in Appendix E.
- Table 1 of the preregistration erroneously lists the emotion *excited* for the dataset by Naragon-Gainey et al. (2023). This item was not included in the analyses, consistent with our preregistered criterion that emotions must be assessed in at least three datasets to be included (page 7 of the preregistration).
- On page 10 of the preregistration, we stated that we would use the default uninformative prior distributions implemented in Mplus and described the default inverse-Wishart priors for the IRTree model as $IW(1,4)$ and $IW(0,4)$ for variances and covariances, respectively. However, the IRTree specification included two latent agreement dimensions per construct, resulting in a total of four (rather than three) latent dimensions. Consequently, the default degrees of freedom for the inverse-Wishart priors are df $= 4+1 = 5$, such that the actual priors used were $IW(1,5)$ and $IW(0,5)$. These priors reflect the default Mplus settings and are consistent with our preregistered analysis plan.
- On page 11 of the preregistration, we stated that we would use 4 chains with 5 000 iterations each (including 2 500 warmup iterations). However, to ensure model convergence and address potential autocorrelation, we increased the number of iterations to 30 000 (including 15 000 warmup iterations) for all location-scale models.
- The models for *angry*, *lonely*, and *tired* in the data by Fried et al. (2022) did not converge even after increasing the number of iterations to 40 000. These models were therefore excluded from subsequent analyses.
- On page 11 of the preregistration, we stated that if very strong correlations (e.g., $r \approx .70$) were observed between random effects for the intercept, inertia, and variability (u_{0p}, u_{1p}, u_{2p}) in the location-scale model, we would explore potential remedies such as transformations. We observed such strong correlations between the random intercept and variability for negative emotions and negative affect in the dataset by Knouse et al. (2024), whereas correlations in the remaining datasets were moderate. The dataset by Knouse et al. (2024) included a relatively small number of measurement occasions (6 days, 3x/day), which may have limited observability of negative emotions and potentially compromised the identifiability of random effects. Consequently, we excluded this dataset from the primary analyses, but report results including the dataset for completeness in Appendix E.

Appendix B

Descriptive Results

Figure B1

Person means with 20% and 80% quantiles for positive and negative affect (top), positive emotions (middle), and negative emotions (bottom) across studies.

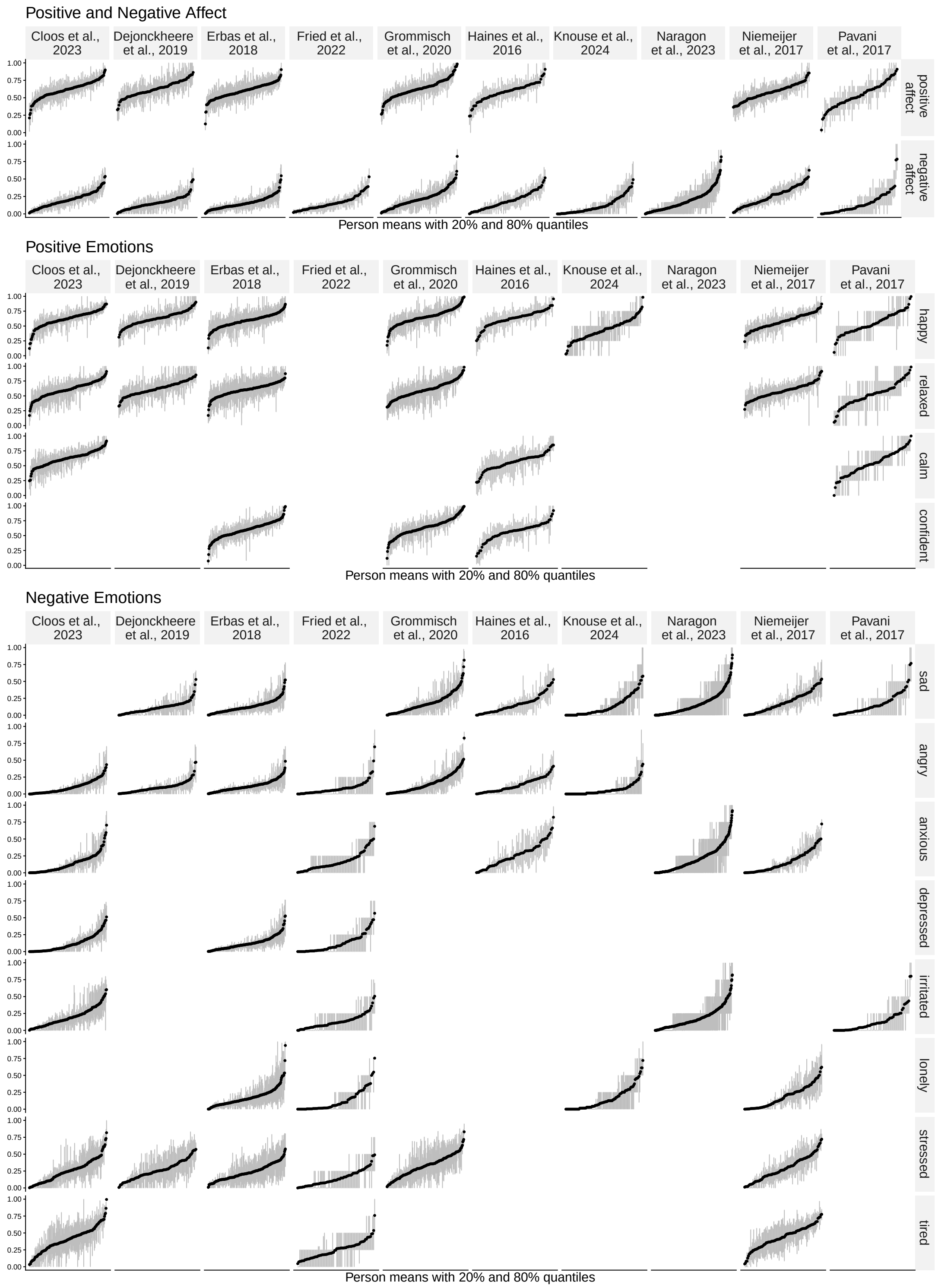
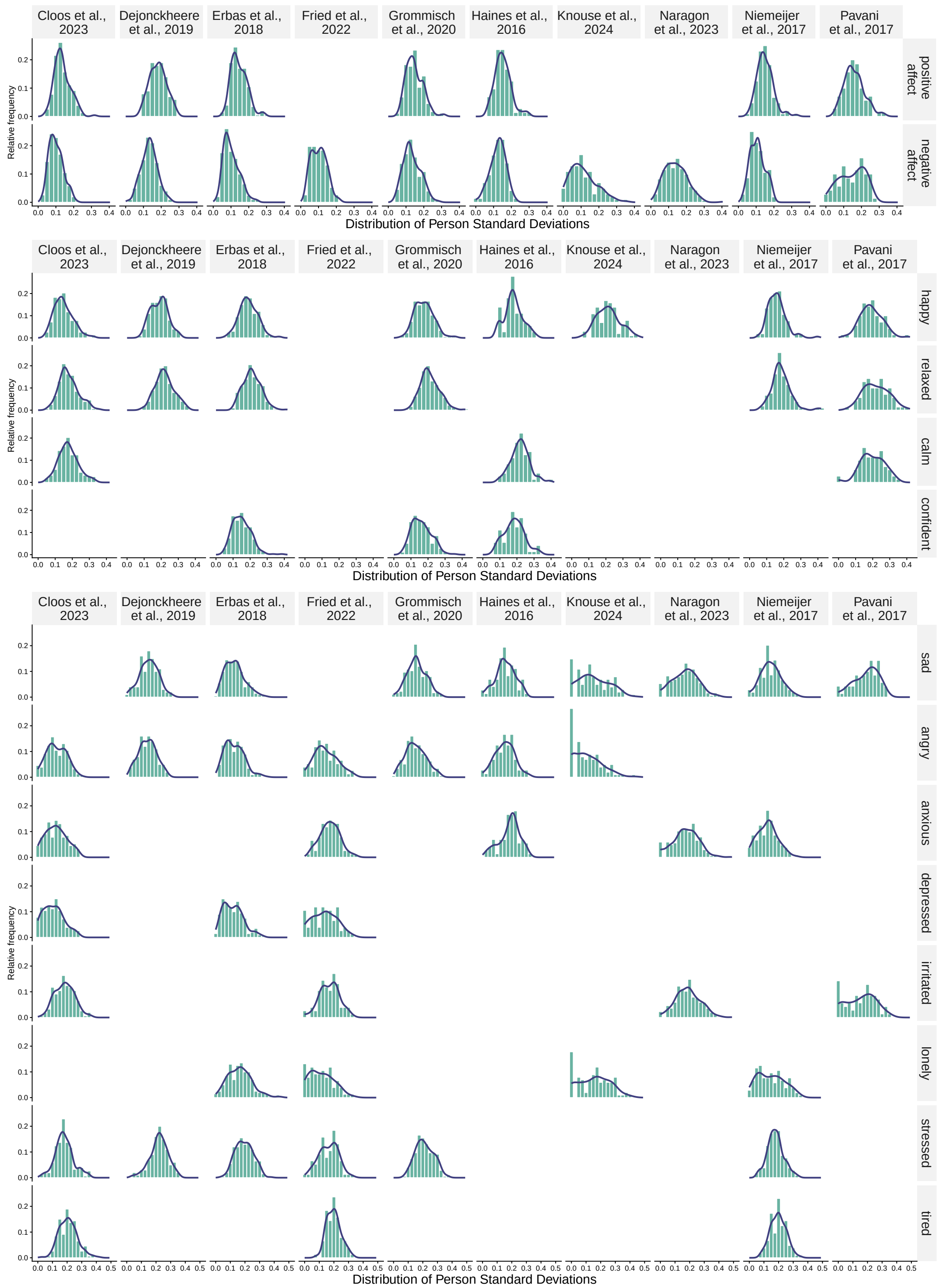


Figure B2

Distribution of person standard deviations for positive and negative affect (top), positive emotions (middle), and negative emotions (bottom) across studies.



Appendix C

Convergence Diagnostics

Table C1

Potential Scale Reduction (PSR) and Effective Sample Size (ESS) for Regression Parameters

	PSR γ_{11}	PSR ω_1	Bulk ESS γ_{11}	Bulk ESS ω_1	Tail ESS γ_{11}	Tail ESS ω_1
Cloos et al., 2023						
Positive Affect	1.00	1.00	10777.92	3631.22	19876.55	10026.80
Negative Affect	1.00	1.00	12995.49	7962.25	23273.70	16166.94
Happy	1.00	1.00	11872.56	4591.02	21567.40	11819.98
Calm	1.00	1.00	5893.50	2846.30	12612.39	8386.22
Relaxed	1.00	1.00	8345.95	4002.86	16139.66	9985.10
Angry	1.00	1.00	9793.78	8020.69	18853.90	15024.34
Stressed	1.00	1.00	10134.04	5529.27	20326.65	14794.18
Irritated	1.00	1.00	9984.14	10137.37	18202.72	18868.26
Anxious	1.00	1.00	10567.72	15688.36	19270.30	27621.23
Depressed	1.00	1.00	10605.25	14505.79	21410.49	23960.18
Tired	1.00	1.00	8778.38	6000.80	16930.19	14252.82
Dejonckheere et al., 2019						
Positive Affect	1.00	1.00	7745.45	10412.94	17503.63	26563.01
Negative Affect	1.00	1.00	8646.79	12410.54	20693.93	29025.57
Happy	1.00	1.00	8611.82	10732.55	18699.53	25423.55
Relaxed	1.00	1.00	7542.53	8100.49	17231.48	19902.88
Angry	1.00	1.00	6947.85	14032.22	15029.48	26606.00
Stressed	1.00	1.00	6238.30	10907.87	14983.90	23032.85
Sad	1.00	1.00	9523.96	12576.01	22042.89	27738.28
Erbas et al., 2018						
Positive Affect	1.00	1.00	7045.29	3534.62	11302.87	10044.36
Negative Affect	1.00	1.00	4386.98	4618.11	12198.00	10718.41
Happy	1.00	1.00	7423.52	9933.54	14354.10	14854.13
Confident	1.00	1.00	5313.58	2342.76	10681.01	7688.52
Relaxed	1.00	1.00	7448.44	2279.60	15052.39	8308.83
Angry	1.00	1.00	5059.53	13210.14	9747.70	20175.91
Stressed	1.00	1.00	5088.69	5363.38	10726.74	11761.54
Depressed	1.00	1.00	5283.60	4221.50	10633.94	10222.51

Continued on next page

Table continued

	PSR γ_{11}	PSR ω_1	Bulk ESS γ_{11}	Bulk ESS ω_1	Tail ESS γ_{11}	Tail ESS ω_1
Sad	1.00	1.00	6117.80	8131.33	14150.77	15264.44
Lonely	1.00	1.00	4543.75	6671.21	11041.21	13489.87
Fried et al., 2022						
Negative Affect	1.00	1.00	3335.05	3595.85	9767.76	11294.76
Stressed	1.00	1.00	2516.96	3818.84	8282.09	11289.92
Irritated	1.00	1.00	2523.57	3059.03	8189.30	9692.51
Anxious	1.00	1.00	2773.66	6428.65	7142.13	16390.13
Depressed	1.00	1.00	5016.59	4464.82	11121.60	12649.22
Grommisch et al., 2020						
Positive Affect	1.00	1.00	13159.08	2800.26	21736.04	9587.39
Negative Affect	1.00	1.00	11703.23	4481.48	21131.47	11555.43
Happy	1.00	1.00	12161.94	3351.84	19239.28	9273.70
Confident	1.00	1.00	9221.08	5129.75	16625.15	13130.75
Relaxed	1.00	1.00	14739.21	2250.95	18940.29	7786.13
Angry	1.00	1.00	7214.92	5589.70	14459.36	13153.30
Stressed	1.00	1.00	13804.56	5402.00	25009.27	15318.60
Sad	1.00	1.00	12539.55	8456.54	22426.34	12700.10
Haines et al., 2016						
Positive Affect	1.00	1.00	9498.45	15219.19	21545.09	29433.71
Negative Affect	1.00	1.00	8194.09	12414.58	22653.37	24475.29
Happy	1.00	1.00	7061.36	12589.41	16062.85	24113.09
Calm	1.00	1.00	9552.78	15258.14	21437.34	29243.38
Confident	1.00	1.00	6169.65	13554.54	15398.07	25615.61
Angry	1.00	1.00	7344.64	15215.46	18294.81	29598.69
Anxious	1.00	1.00	7615.18	14204.18	17708.97	30545.17
Sad	1.00	1.00	9790.94	15579.93	26205.01	31483.39
Knouse et al., 2024						
Negative Affect	1.00	1.00	2937.37	7688.98	11545.92	17911.32
Happy	1.00	1.00	5764.61	6865.48	14540.02	19408.83
Angry	1.00	1.00	2440.30	9150.12	8524.69	25838.25
Sad	1.01	1.00	582.68	10396.19	7006.98	26681.08
Lonely	1.01	1.00	627.83	11260.77	11321.12	26392.49

Continued on next page

Table continued

	PSR γ_{11}	PSR ω_1	Bulk ESS γ_{11}	Bulk ESS ω_1	Tail ESS γ_{11}	Tail ESS ω_1
Naragon-Gainey et al., 2023						
Negative Affect	1.00	1.00	9575.02	18222.71	11067.10	15998.53
Irritated	1.00	1.00	4247.59	16493.11	6663.27	16954.44
Anxious	1.00	1.00	5919.79	18451.21	9182.97	16693.06
Sad	1.00	1.00	2591.54	2519.58	5657.54	5798.63
Niemeijer et al., 2023						
Positive Affect	1.00	1.00	8110.38	5985.39	18298.14	14170.85
Negative Affect	1.00	1.00	3958.31	6211.55	12138.81	18114.15
Happy	1.00	1.00	9111.77	5477.69	19573.76	17093.25
Relaxed	1.00	1.00	4948.40	5675.26	16697.87	15472.35
Stressed	1.00	1.00	3042.52	7778.05	10573.67	20007.73
Anxious	1.00	1.00	3189.50	12976.14	11790.59	22487.60
Sad	1.00	1.00	8761.77	5289.61	18790.54	16648.12
Lonely	1.00	1.00	11313.26	6310.06	23838.93	17770.30
Tired	1.00	1.00	6928.67	8106.31	17121.13	19132.07
Pavani et al., 2017						
Positive Affect	1.00	1.00	4988.56	4622.16	12529.42	12756.18
Negative Affect	1.00	1.00	4491.01	8447.61	11972.15	17745.28
Happy	1.00	1.00	6040.71	4351.43	14097.57	11316.06
Calm	1.00	1.00	4213.63	5960.65	11214.45	12952.37
Relaxed	1.00	1.00	4639.40	5484.68	12601.76	14543.31
Sad	1.00	1.00	4352.09	7889.95	11555.49	17805.06

Appendix D

Model parameters: Intercepts, variability, inertia, and their variances and covariances for emotion items

Figure D1

Estimates of intercepts, standard deviations, inertia, variance of the standard deviation, and variance of inertia for positive emotions across datasets. Error bars represent 95% credible intervals in the Bayesian location-scale models.

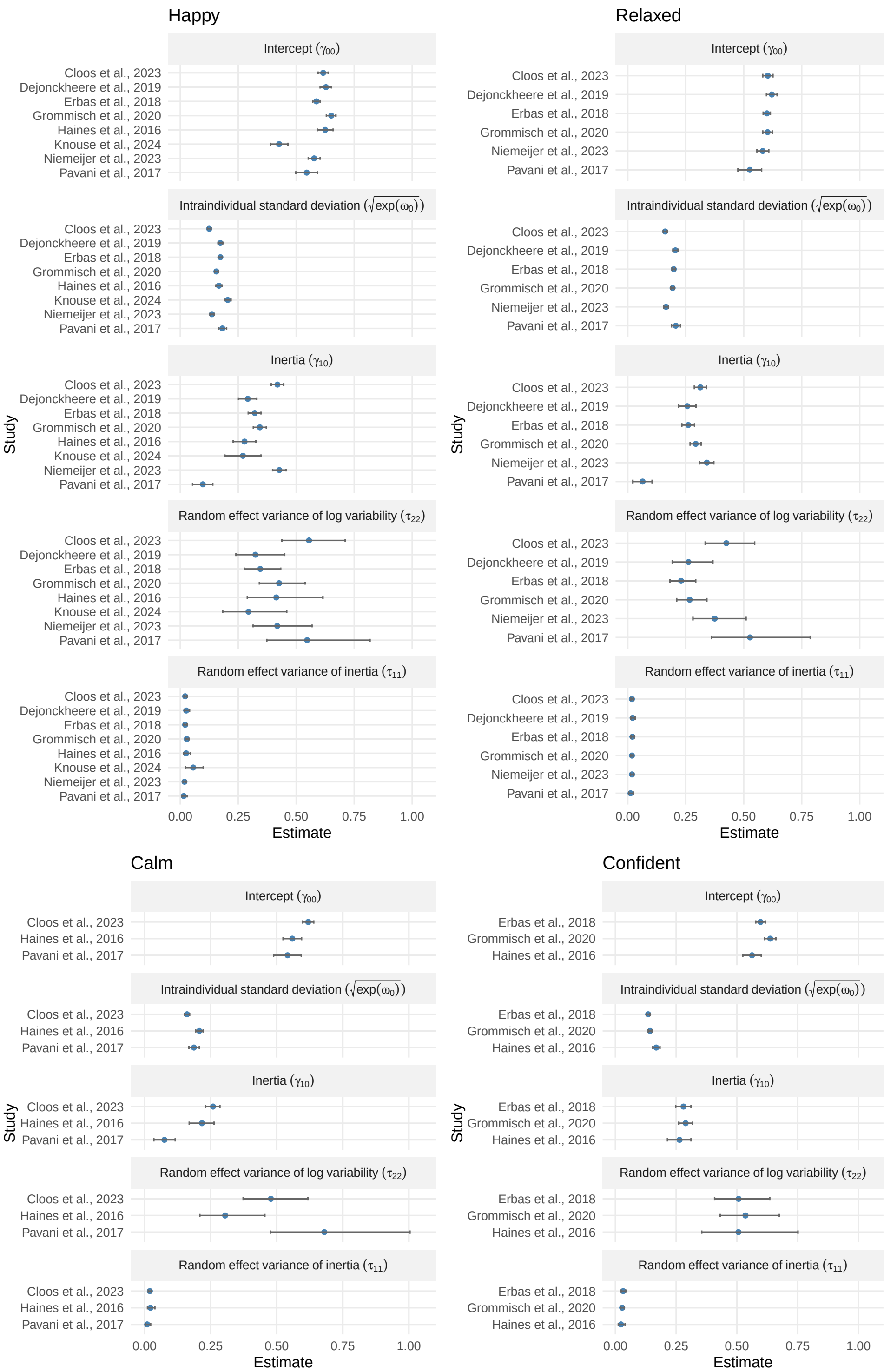


Figure D2

Estimates of intercepts, standard deviations, inertia, variance of the standard deviation, and variance of inertia for negative emotions across datasets. Error bars represent 95% credible intervals in the Bayesian location-scale models.

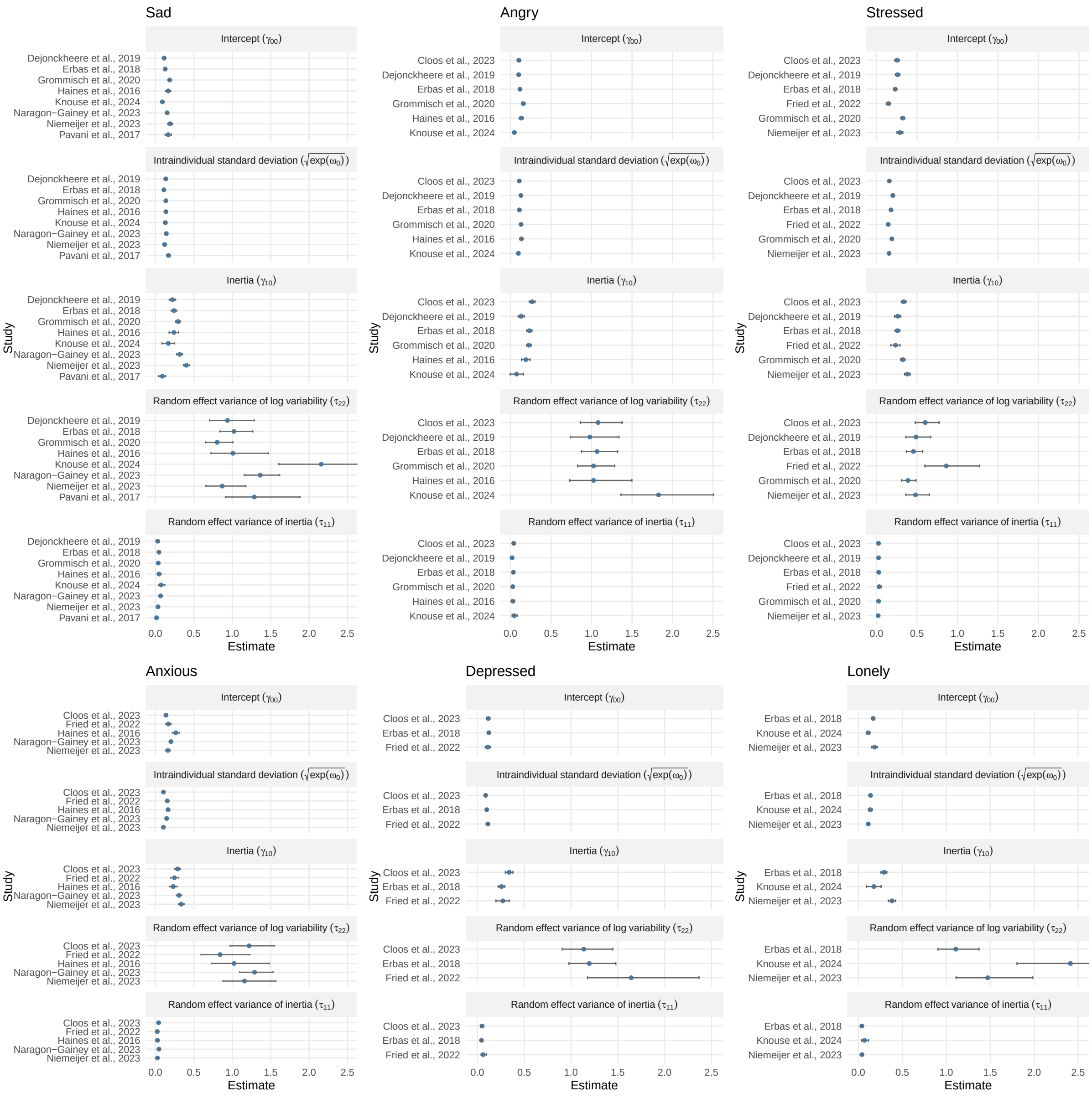


Figure D3

Estimates of correlations between random effects for intercept, variability, and inertia for positive emotions. Error bars represent 95% *credible intervals in the Bayesian location-scale models.*

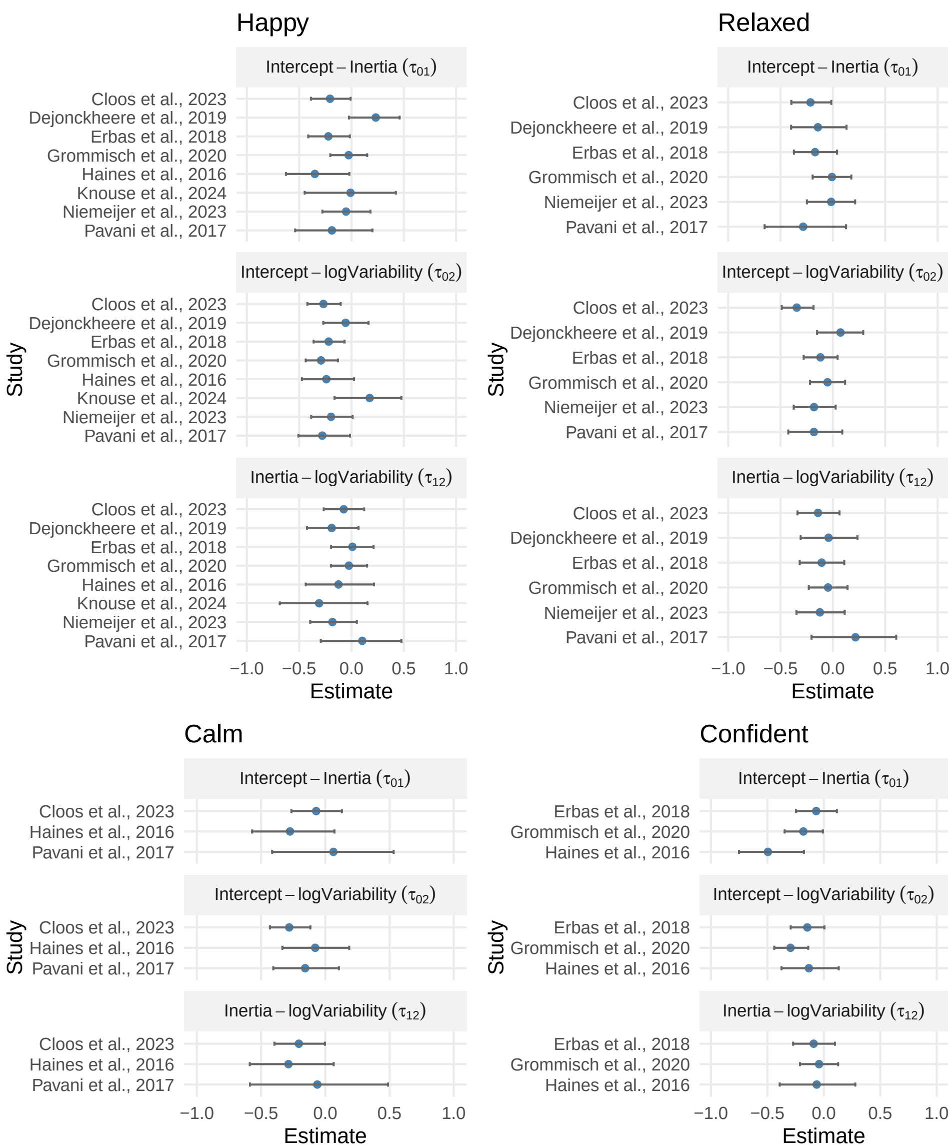
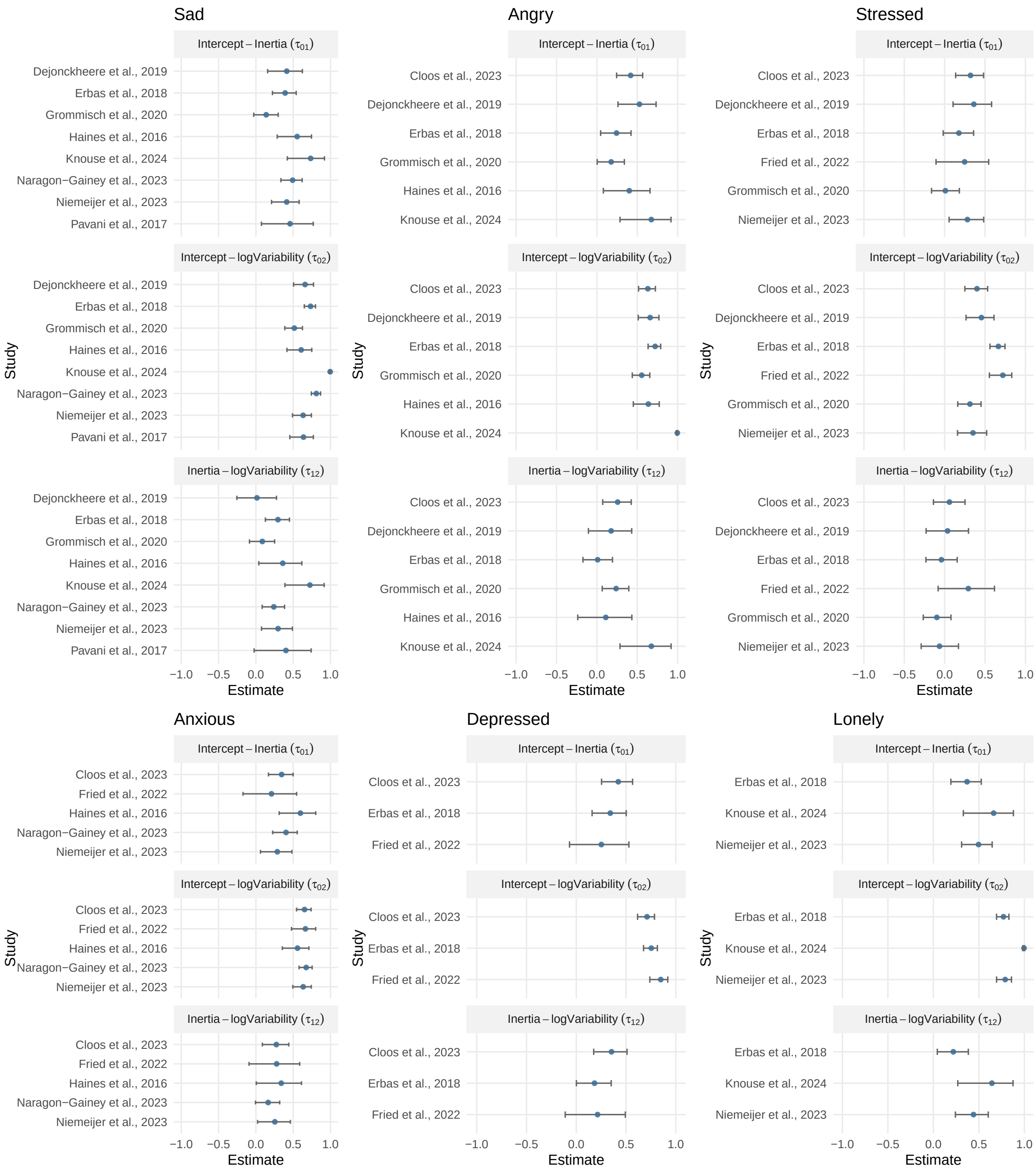


Figure D4

Estimates of correlations between random effects for intercept, variability, and inertia for negative emotions. Error bars represent 95% credible intervals in the Bayesian location-scale models.



Appendix E

Robustness checks

Results including the dataset by Knouse et al., 2024

The associations between ERS and variability of emotions in the dataset by Knouse et al. (2024) were mostly negative, with regression weights ranging from −0.71 to 0.06 for different emotions. Including the dataset into the Bayesian random-effects meta-analytic model did neither substantially affect the results nor did it change the overall conclusions.

Table E1

Results including the dataset by Knouse et al., 2024

		Estimate	SD	CI Lower	CI Upper
Variability					
Affect	Overall	0.22	0.09	0.05	0.39
	Positive	0.27	0.09	0.07	0.44
	Negative	0.18	0.09	0.01	0.37
Emotions	Overall	0.21	0.08	0.05	0.38
	Positive	0.26	0.09	0.09	0.44
	Negative	0.17	0.09	-0.01	0.35
Inertia					
Affect	Overall	-0.01	0.02	-0.04	0.02
	Positive	-0.01	0.02	-0.05	0.03
	Negative	-0.01	0.02	-0.05	0.03
Emotions	Overall	0.00	0.02	-0.03	0.05
	Positive	0.26	0.09	0.09	0.44
	Negative	0.17	0.09	-0.01	0.35

Note: SD = posterior standard deviation; CI = credible interval.

Results using frequentist estimation

Here, we provide the results of synthesized study-specific parameter estimates and their associated uncertainty using frequentist random-effects meta-analytic model from the `metafor` package in R (Viechtbauer, 2010) excluding the dataset by Knouse et al. (2024). Herein, each study was treated as a cluster providing multiple regression estimates (e.g., for PA and NA or for different emotions). We used cluster-robust variance estimation with a small-sample correction to obtain confidence intervals and *p*-values (i.e., CR2 adjustment; Hedges et al., 2010; Tipton, 2015). Posterior study-level effect estimates for the association between ERS and both intraindividual variability and inertia, obtained from the Bayesian location–scale models, were used as point estimates in the meta-analytic models. The corresponding posterior standard deviations served as the associated standard errors.

Table E2

Results using a frequentist random-effects meta-analytic model

		Estimate	SE	*z*-value	CI Lower	CI Upper	*p*-value
Variability							
Affect	Overall	0.22	0.07	3.30	0.06	0.37	0.006
	Positive	0.28	0.06	4.53	0.13	0.43	0.002
	Negative	0.17	0.07	2.27	0.00	0.34	0.027
Emotions	Overall	0.21	0.07	3.09	0.06	0.36	0.007
	Positive	0.28	0.06	4.53	0.13	0.43	0.002
	Negative	0.17	0.07	2.27	0.00	0.34	0.027
Inertia							
Affect	Overall	-0.01	0.01	-0.68	-0.04	0.02	0.262
	Positive	-0.01	0.01	-1.03	-0.04	0.02	0.175
	Negative	-0.01	0.02	-0.33	-0.05	0.04	0.377
Emotions	Overall	0.01	0.02	0.36	-0.03	0.04	0.364
	Positive	-0.01	0.01	-0.93	-0.03	0.02	0.198
	Negative	0.01	0.02	0.50	-0.04	0.05	0.317

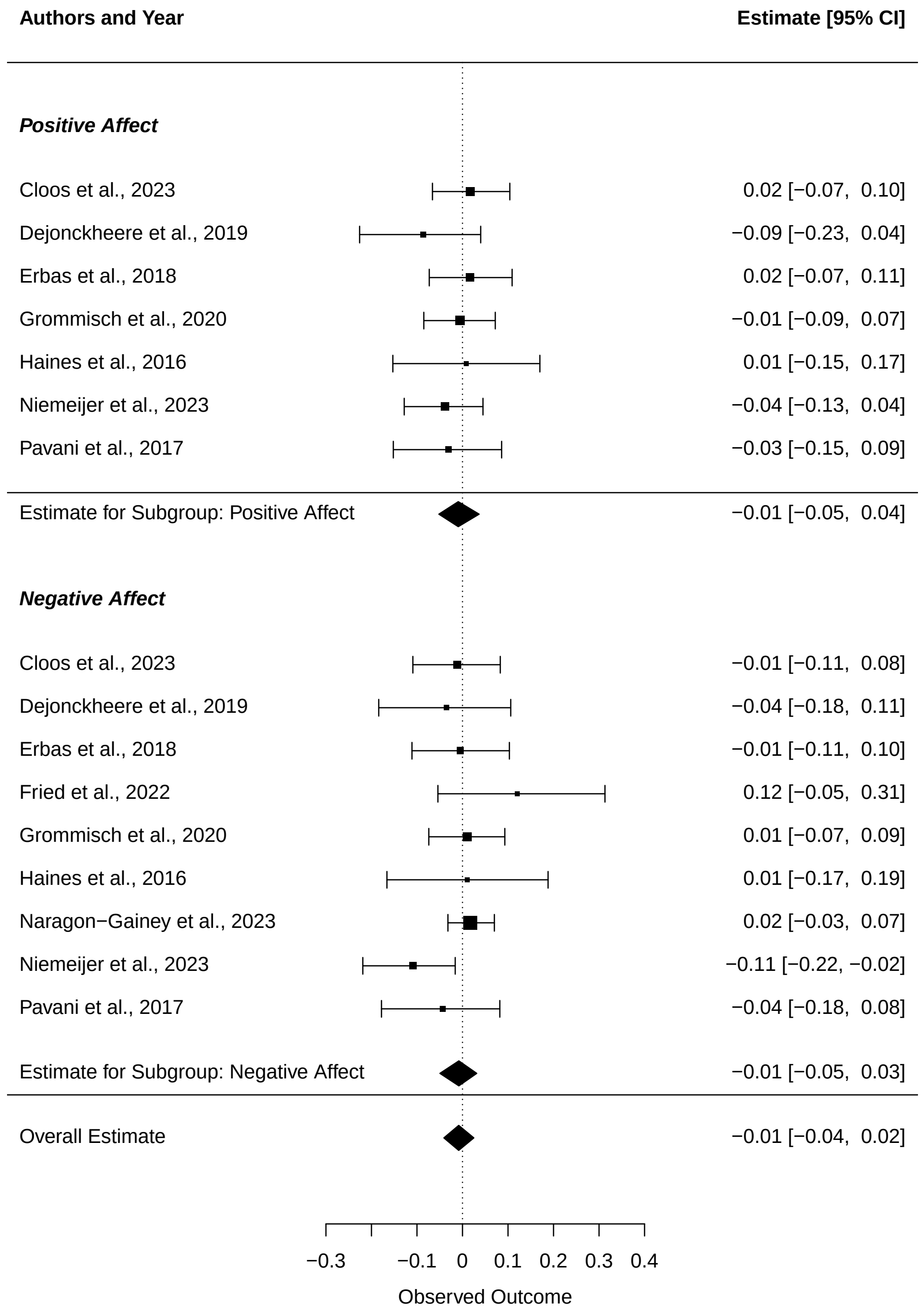
Note: Results excluding the dataset by Knouse et al., (2024). *p*-values are one-sided, as we had a directional hypothesis that higher ERS would be associated with greater variability and lower inertia. SE = standard error; CI = confidence interval.

Appendix F

Association between ERS and inertia across studies

Figure F1

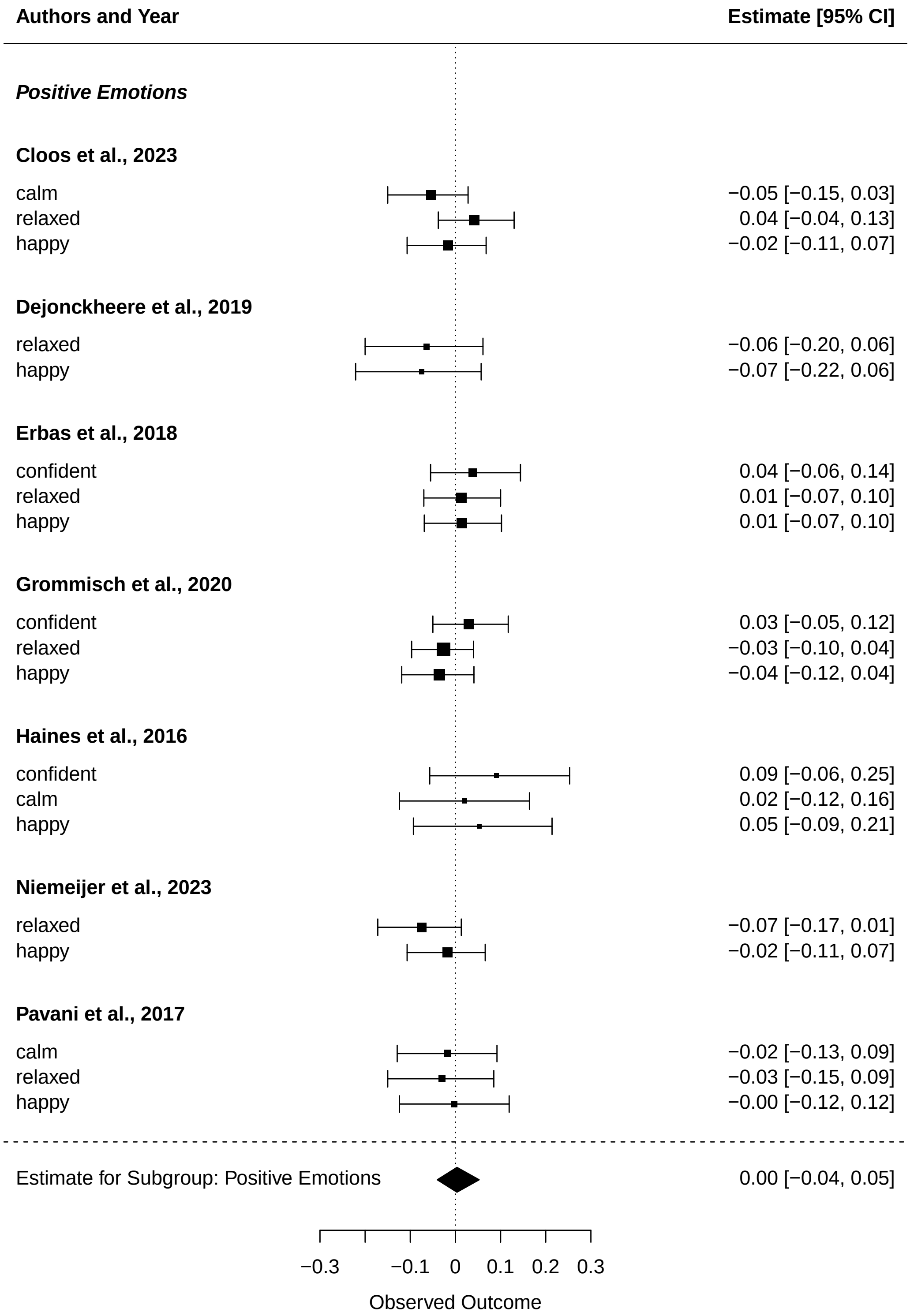
Estimated associations between ERS and inertia for positive and negative affect across studies in the location-scale models.



Note: Points represent posterior regression coefficients from Bayesian models; error bars indicate 95% credible intervals.

Figure F2

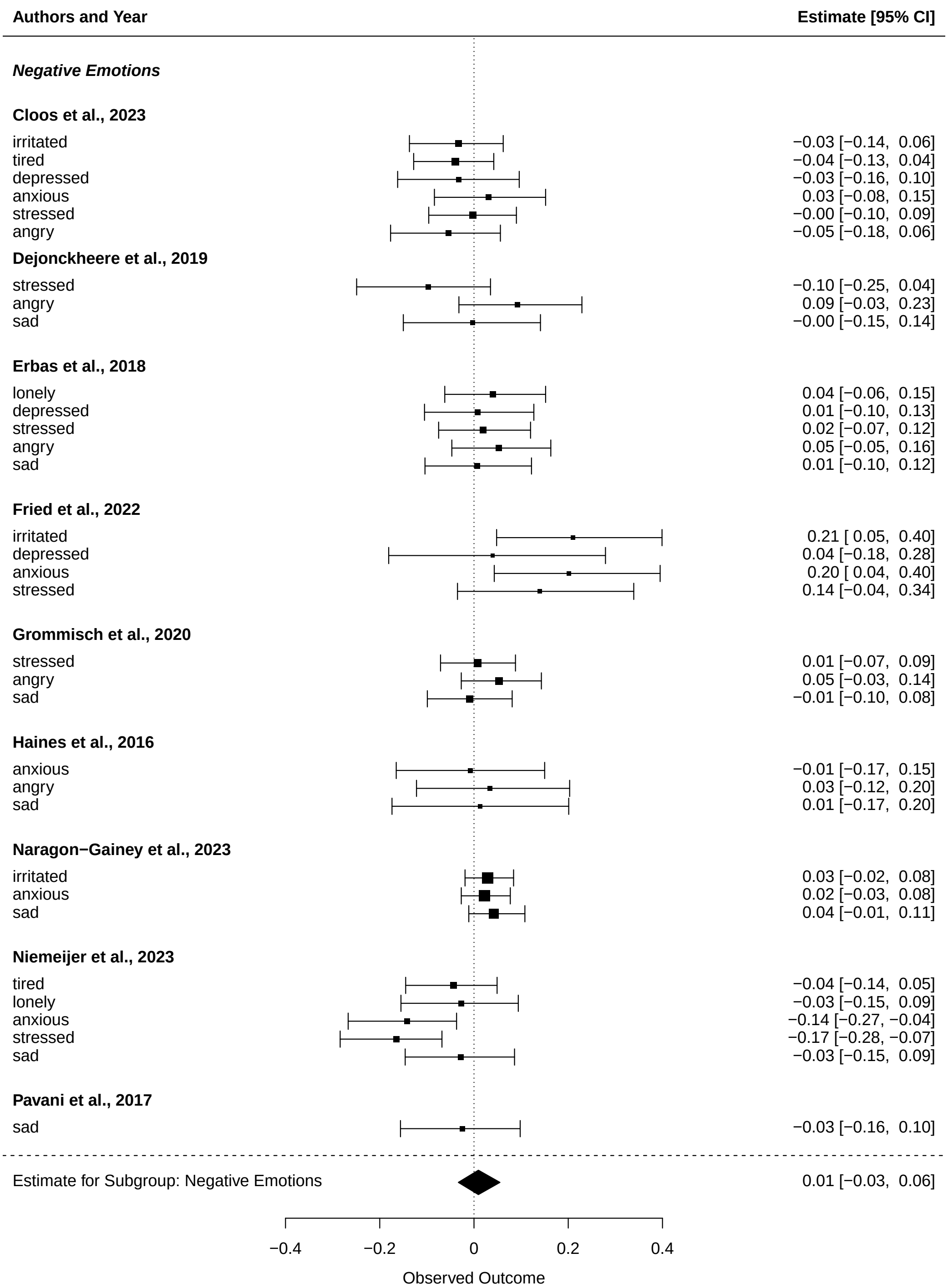
Estimated associations between ERS and inertia for positive emotions across studies in the location-scale models.



Note: Points represent posterior regression coefficients from Bayesian models; error bars indicate 95% credible intervals.

Figure F3

Estimated associations between ERS and inertia for negative emotions across studies in the location-scale models.



Note: Points represent posterior regression coefficients from Bayesian models; error bars indicate 95% credible intervals.