

**Modeling Intraindividual Variability in Affect (MIVA): Formalized-theoretical Approach, Computational Model, and Parameter-recovery Study**

Maria Wirth<sup>1</sup>, Andreas Voss<sup>2</sup>, Stefan T. Radev<sup>3</sup> & Klaus Rothermund<sup>1</sup>

<sup>1</sup>Friedrich Schiller University Jena

<sup>2</sup>Heidelberg University

<sup>3</sup>Rensselaer Polytechnic Institute

*Behavior Research Methods*, in press

Author Note:

Maria Wirth and Klaus Rothermund, Department of Psychology, Friedrich Schiller University Jena, Jena, Germany. Andreas Voss, Department of Psychology, Heidelberg University, Heidelberg, Germany. Stefan T. Radev, Department of Cognitive Science and Center for Modeling, Simulation, and Imaging in Medicine, Rensselaer Polytechnic Institute, Troy, New York, USA.

Correspondence concerning this article should be addressed to Maria Wirth, Department of Psychology, Friedrich Schiller University Jena, Am Steiger 3/1, 07743 Jena, Germany.

Contact: +493641 945115 [maria.wirth@uni-jena.de](mailto:maria.wirth@uni-jena.de)

**Abstract**

No matter how angry, sad, or happy we are, eventually, we will feel different. Studying this ebb and flow of affective experience in daily life provides important insights into psychological functioning and well-being. We have developed a parsimonious formalized model of intraindividual variability in affect (MIVA), resting on the assumption that such affective changes reflect transactions between an individual and its proximal environment. We provide an outline of its theoretical background, scope, and mathematical formulation. We situate MIVA within the research field of affect dynamics and illustrate the models' behavior under realistic conditions using a simulation study. We use simulation-based inference to train a custom neural network on MIVA simulations, which we employ to rapidly estimate the model's parameters on a multitude of synthetic experiments with different configurations. Our simulation study demonstrates that the synthesis between a computational model and probabilistic neural networks results in an efficient and flexible tool for model-based inference of affect dynamics. Our simulation study also offers insights into the data requirements for a precise recovery of the model's parameters and recommendations for future data collection. The potential of MIVA for providing insights into affect dynamics is discussed.

*Keywords:* affect dynamics, simulation-based inference, emotional reactivity, emotion regulation, mathematical modeling, neuronal networks

## **Modeling Intraindividual Variability in Affect (MIVA): Formalized-theoretical Approach, Computational Model, and Parameter-recovery Study**

Affective experience is inherently dynamic. Within a day, an individual may experience all possible shades of feelings, from the ecstasy of joy to the depths of despair. Describing the principles and regularities underlying affective fluctuations, identifying the processes driving them, and understanding the implications they have for psychological functioning are the aims of affect dynamic research<sup>1</sup> (Koval & Kuppens, 2024; Kuppens, 2015; Kuppens et al., 2010). Past affect dynamic work has relied on two approaches for extracting relevant information concerning affect dynamics: simple summary statistics (such as the intraindividual standard deviation or first-order autocorrelation) and model-based approaches (Dejonckheere et al., 2019; Loossens et al., 2020). Model-based approaches try to capture the observed variability in a process of interest by fitting a computational model to empirical data (Ram et al., 2014). Inspired by contemporary affect dynamic models (Chow et al., 2005; Kuppens et al., 2010; Loossens et al., 2020), we developed the Model of Intraindividual Variability in Affect (MIVA; Wirth et al., 2022, 2023) with a specific focus on formalizing the relation between fluctuations in hedonic valence and individual differences in responding to affect-eliciting events, an issue that is still of high relevance in affect dynamic research (Ryan et al., 2025).

Although simple summary statistics have been criticized as descriptive and offering little information concerning underlying processes (Adolf et al., 2017; Deboeck et al., 2009; Jahng et al., 2008; Mestdagh et al., 2018; Trull et al., 2015), they still seem to dominate empirical research on affect dynamics (Vanhasbroeck et al., 2022). As model-based approaches, including our own, might be difficult to understand, implement, and communicate, in this article, we seek to achieve two goals: First, we elaborate on MIVA's theoretical background, its formalization, and embed MIVA within the research field of affect

---

<sup>1</sup> Affective fluctuations are also studied in affective computing research. However, affective computing usually focuses on developing artificial systems and devices that can recognize, interpret, process, and simulate human affect (Bosse et al., 2014; Ryan et al., 2025)

dynamics. Second, capitalizing on the amortized simulation-based inference (Cranmer et al., 2020; Radev et al., 2022), we introduce a new, efficient inference method for estimating MIVA model parameters. Using the results of a comprehensive simulation study, we provide specific recommendations for data collections aiming to apply the MIVA model (i.e., number of days and measurements per day). Concluding, we will discuss the potential of MIVA for providing insights into affect dynamics.

### **MIVA: Basic Concepts and their Relations**

We developed MIVA to integrate the translational process between external events and affective experience into a model-based approach, as this had been pointed out as an important step for better understanding affect dynamics in previous work (Adolf et al., 2017; Chow et al., 2005; Kuppens et al., 2010; Vanhasbroeck et al., 2021). Specifically, the MIVA model intends to provide a useful heuristic for characterizing fluctuations in hedonic valence and tries to capture interindividual differences in response to affective-eliciting events.

We will shortly outline the model's parameters and their relations; interested readers can find more details in the Supplemental Material (<https://osf.io/spxf8/files/2u35k>). Generally, the model assumes that fluctuations in affective experience reflect transactions between an organism and its proximal environment (Frijda, 2007; Larsen, 2000; Reich et al., 2003). Rather than objective characteristics, changes in affect are based on an individual's subjective appraisal of a situational change or event (Frijda, 1993; Roseman, 1996; Scherer, 2009). Following many appraisal theories (Frijda, 1993; Scherer, 1993; Smith & Ellsworth, 1985), the model includes a parameter for hedonic event quality, known as event pleasantness. If an event is appraised as pleasant, the elicited affect is positive. If the event is appraised as unpleasant, the resulting affect is negative.

How events and their appraisal can result in changes in affect across shorter and longer time scales is captured by two model parameters: *Affective reactivity* encompasses individual differences in the extremity of short-term changes in affective experience caused by

situational changes (Bolger & Zuckerman, 1995; Davidson, 1998; Mroczek & Almeida, 2004; Rosenberg, 1998). Individuals with high affective reactivity show substantial alterations in affective experience if they encounter situational changes, compared to their average affect level. Individuals with low reactivity only show minor changes in affect. The impact of valent events on affective experience, thus, depends on how a person generally reacts to events and on how the person appraises specific events. Given that affective reactivity is usually higher for events that are appraised as unpleasant compared to those appraised as pleasant (Bucks et al., 2005; Bylsma et al., 2011), our model differentiates between affective reactivity for pleasant and unpleasant events.

*Affect regulation* captures interindividual differences in longer-term changes in affective experience that originated from affect-eliciting events. Specifically, it encompasses any process that aims at changing the intensity, quality, or duration of affective experiences and/or expressions (Gross, 2013; Koole & Rothermund, 2011). Individuals with high affect regulation show a steep and fast decline of affective reactions. For individuals with low affect regulation, affect persists over longer periods. How fast the impact of events can be regulated also depends on the event's pleasantness. Affective reactions to unpleasant events are attenuated faster and more strongly than affective reactions to pleasant events (Hemenover, 2003; Volokhov & Demaree, 2010). Thus, our model also differentiates between affect regulation of pleasant and unpleasant event impact (for similar reasoning, see Ong et al., 2025).

Additionally, our model accounts for interindividual differences in an individual's typical affective experience in the absence of affect-eliciting events, called *anchoring*. This state may be predominantly positive or negative rather than neutral (Diener & Diener, 1996). Anchoring exhibits cross-situational stability and is an independent source of affective experience (Alessandri et al., 2014).

Taken together, we have identified the basic processes (i.e., event appraisal, reactivity, regulation, accumulation, and anchoring) underlying the variability and fluctuations in everyday affective experiences and have developed a conceptual model based on the central insights that have emerged from the research on affect generation and fluctuation.

### **Mathematical Formalization of MIVA**

So far, the presentation of MIVA has been restricted to a verbal account of its most important principles. To make the model accessible to empirical research and to allow rigorous testing of predictions derived from it, its principles were translated into a mathematical model (for an overview on formalizing verbal affect dynamic theories, see e.g., Haslbeck et al., 2022).

MIVA focuses on explaining variability in hedonic valence (positive–negative), which is not necessarily the only dimension of affect but is fundamental to its description (Barrett & Russell, 1999; Kuppens et al., 2013; Russell, 2003; Schiller et al., 2024; Watson et al., 1999). Although the question of whether valence is best conceptualized as one bipolar dimension or as two separate unipolar dimensions is debated (Green et al., 1993; Kuppens et al., 2013; Tellegen et al., 1999), we assume that affect is assessed on a bipolar scale. Truly unipolar scales (not interpreted as bipolar by participants) require a two-step procedure: First, asking if individuals felt happy/unhappy, and then asking to what extent they felt happy/unhappy (Schmukle & Egloff, 2009; Segura & González-Romá, 2003). Affective experience is rarely assessed using this two-step procedure, and many affect dynamic researchers often use single-item measures (Dejonckheere et al., 2022; Fritz et al., 2024; Laurenceau et al., 2025), which are an efficient, sufficient, and low-burden assessment tool for momentary affect (Cloos et al., 2023; Tuerlinckx et al., 2026). Moreover, a unidimensional bipolar approach as employed in the MIVA model is already sufficient to explain a wealth of affect dynamics (cf. Zavlis et al., 2025). Conceptualizing hedonic valence as two separate dimensions would severely restrict MIVA's empirical applicability. Moreover, many studies show rather low co-occurrence of

positive and negative affect (Riediger et al., 2014; Schmukle & Egloff, 2009; Schneider & Stone, 2015; Watson & Stanton, 2017). Similarly, event pleasantness is conceptualized as being assessed on a bipolar scale.

The following principles were considered during the formalization process: (1) The mathematical formulation needs to allow a non-linear mapping between event pleasantness and affective experience (for similar reasoning, see Zavlis et al., 2025), so that differences in event pleasantness result in larger differences in affect when event pleasantness is low (i.e., close to zero), compared to the same pleasantness difference for extreme events (Thagard & Nerb, 2002; Tong et al., 2009). (2) Affective predictions need to be confined to the scale range of the assessed affective outcome. A sigmoid function can meet these requirements, as it is constrained by a pair of horizontal asymptotes as  $x \rightarrow \pm \infty$ , and it is often used when a specific mathematical formulation is lacking (Gibbs & MacKay, 2000). To take advantage of the distributional properties of the sigmoid function (convex for values less than 0 and concave for values greater than 0), affect and event pleasantness need to be assessed on a bipolar scale (with unpleasant affect and events being indicated by values smaller than 0).<sup>2</sup> This also helps to communicate the bipolarity of the affect measure to participants (Schimmack, 2003).

The previously described ideas and principles are expressed in the following equation:

$$A_j = \frac{r}{1 + e^{-(a+Ru \cdot Iu_j+Rp \cdot Ip_j)}} - \frac{r}{2} \quad (1.1)$$

where  $A_j$  is the predicted affective experience at an observation  $j$ , measured on a bipolar scale (negative-positive affect) with range  $r$  and centered around zero.<sup>3</sup>  $a$  is the parameter for

<sup>2</sup> However, this is not necessarily true for event ratings. If event ratings are assessed on two unipolar scales with values  $> 0$ , the impact of unpleasant events can lead to a decrease in affect instead of adding the product of  $Ru_i \cdot Iu_{ij}$ , this product is subtracted such that

$$A_{ij} = \frac{r}{1 + e^{-(a_i - Ru_i \cdot Iu_{ij} + Rp_i \cdot Ip_{ij})}} - \frac{r}{2}$$

<sup>3</sup> Dependence on scale range can be avoided by standardizing. Centering the affect scale around zero is not strictly necessary for model estimation, but it provides meaningful advantages for interpretation. When affect is centered, the anchoring parameter directly reflects deviations from a neutral affective state:  $a = 0$  indicates neutral baseline affect, positive values indicate a positive affective disposition, and negative values indicate a

anchoring, which represents the mean affect level when no valent event is experienced.  $Ru$  and  $Rp$  are the reactivity parameters for unpleasant and pleasant events, respectively.  $Iu_j$  and  $Ip_j$  provide accumulated event pleasantness ( $p$ ) from all current and past unpleasant and pleasant events, respectively, and are computed as:

$$Ip_j = \sum_{E_p, t < t_j} P_{E_p} \cdot \frac{1}{1 + Dp \cdot \Delta t_E} \quad (1.2)$$

$$Iu_j = \sum_{E_u, t < t_j} P_{E_u} \cdot \frac{1}{1 + Du \cdot \Delta t_E} \quad (1.3)$$

In equations 1.2 and 1.3,  $E_p$  and  $E_u$  are pleasant and unpleasant affect-eliciting events, respectively, occurring before time  $t_j$  with a specific pleasantness  $P_E$ , and  $\Delta t_E$  is the time that has passed since the event ( $\Delta t = t_j - t_E$ ).  $Dp_i$  and  $Du_i$  represent the magnitude of regulation that dampens the impact of pleasant or unpleasant events, respectively. For a current event ( $\Delta t = 0$ ), the discounting function equals 1, meaning that a current event exerts its strongest influence on affect. Past events are discounted and have less impact on current affect. This discounting is relatively rapid shortly after the occurrence of an event but falls more slowly thereafter (Setodji et al., 2019; Verduyn et al., 2009, 2012).  $Ip_j$  is always positive (because  $P_{E_p}$  is positive) and leads to an increase in affect,  $Iu_j$  is always negative (because  $P_{E_u}$  is negative) and leads to a decrease in affect.

The behavior of MIVA is similar to previous models of affect dynamics (Chow et al., 2005; Kuppens et al., 2010). Once a valent event occurs and is appraised as such, affect ( $A_j$ ) will show a deviation from the average affect level ( $a$ ). The direction of the fluctuation depends on event pleasantness ( $P_E$ ), with pleasant events increasing affect such that it becomes more positive or less negative. Unpleasant events, in turn, decrease affective experience such that it becomes more negative or less positive. The amplitude of the fluctuation depends on the reactivity to different situational changes ( $Ru$  or  $Rp$ , respectively).

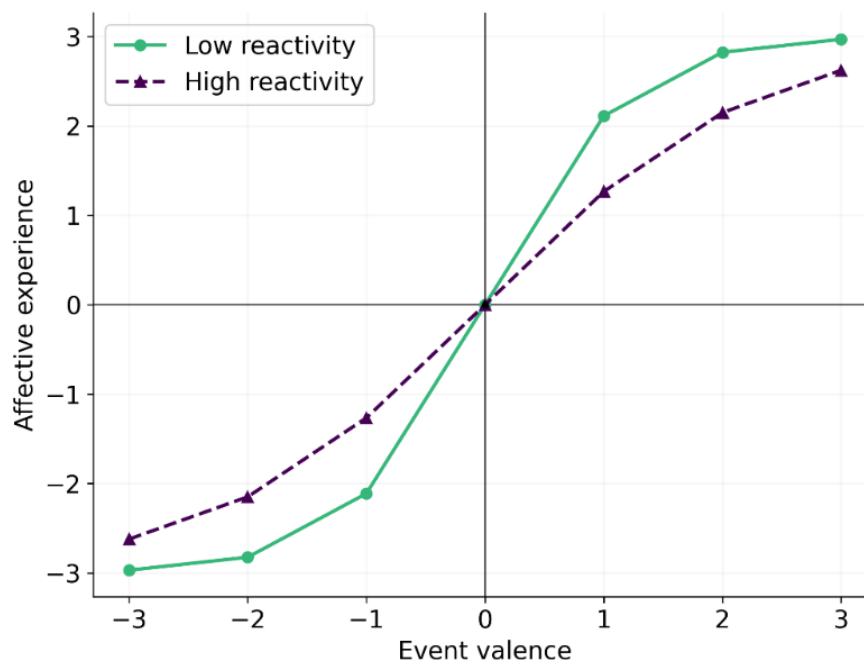
---

negative disposition. Additionally, centering ensures that the mathematical formulation of the decay functions naturally returns affect toward a psychologically meaningful neutral point rather than toward an arbitrary numerical value.

The effect of interindividual differences in reactivity on the relation between event pleasantness and resulting affective experience can be seen in Figure 1. To reach the same level of negative affect, an individual with high reactivity (dashed line) needs to experience a moderately unpleasant event, while someone with low reactivity (dotted line) needs to experience a highly unpleasant event.

**Figure 1**

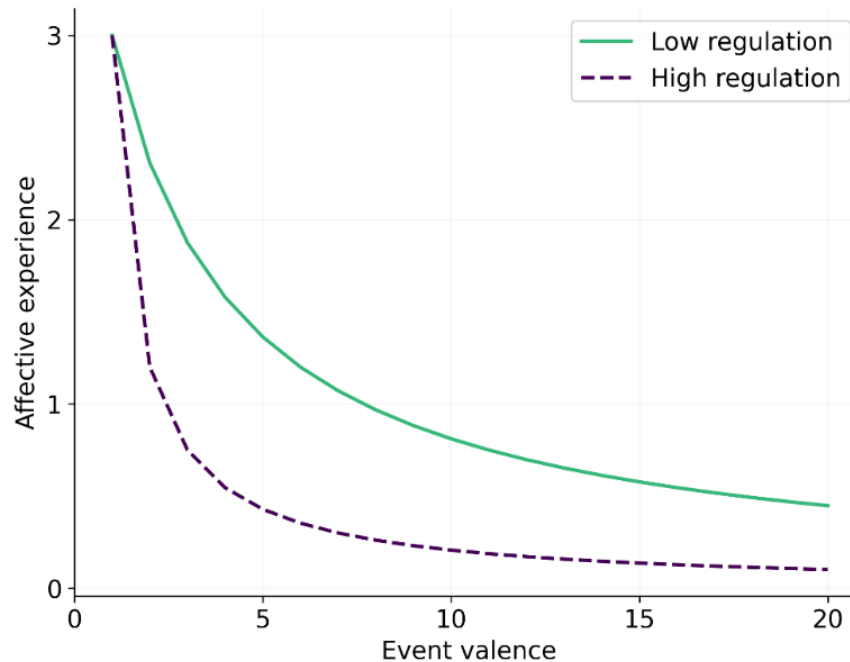
*Affective experience as a function of event valence and interindividual differences in reactivity.*



If no other affect-eliciting events are experienced, affective experience will return to its average level depending on the strength of the regulatory processes ( $D_u$  or  $D_p$ , respectively). The effect of interindividual differences in affect regulation on changes in affective experiences after an affect-eliciting event can be seen in Figure 2. Individuals with high affect regulation (dotted line) show a fast decline in affective experience across time compared to individuals with low affect regulation (solid line).

**Figure 2**

*Changes in affective experience as a function of time since the event (in minutes) and interindividual differences in regulation.*

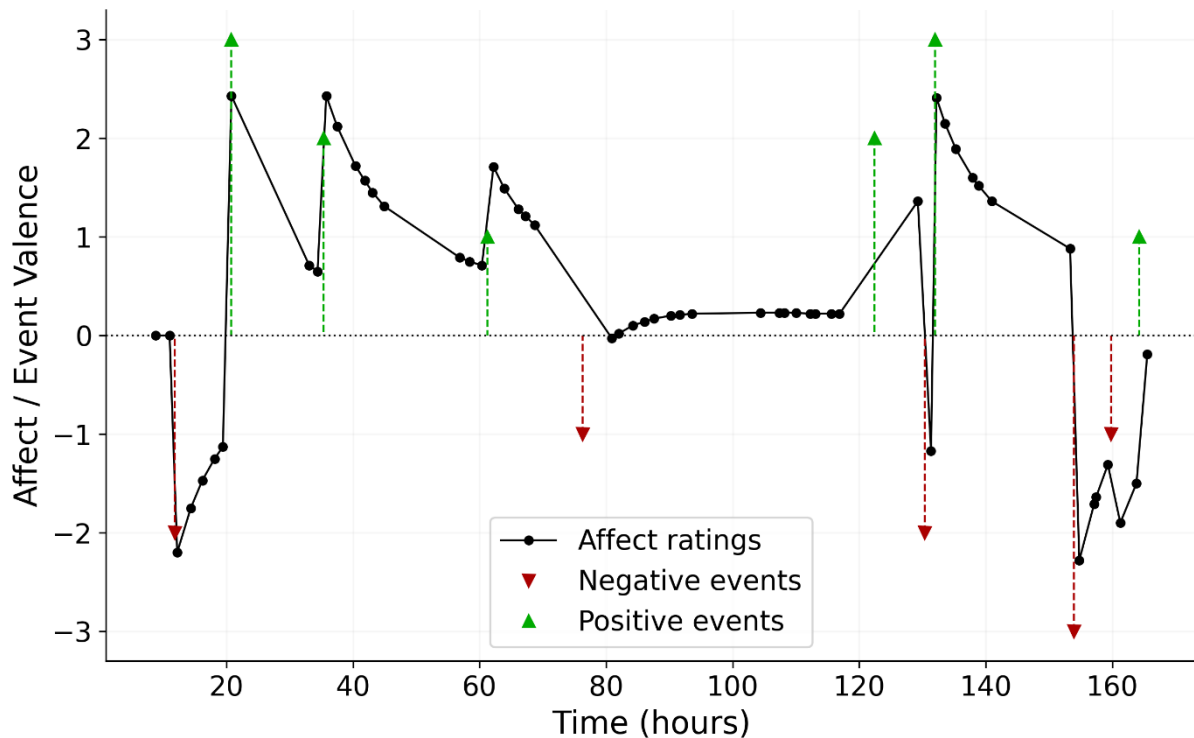


If other events are experienced, the course of the affective experience depends on the cumulated impact of all past events. An example of the impact of event pleasantness on affect dynamics can be seen in Figure 3. It shows a simulated data set for one hypothetical participant, reporting affective experience, event occurrence, and pleasantness for 7 days and 8 measurements per day. Affect and event ratings range between -3 (highly unpleasant/negative) to 3 (highly pleasant/positive). An average number of 2 events per day (following a Poisson distribution) and the following MIVA-parameter values were used for the simulation:  $a = 0.0$ ,  $Ru = 1.0$ ,  $Rp = 1.0$ ,  $Du = 0.2$ ,  $Dp = 0.2$ . Events occurred slightly before the corresponding measurement occasion. It can be seen that affect-eliciting events (vertical lines) have a strong initial impact on affective experience (black circles) that gradually declines thereafter. The first unpleasant event reported on the third measurement occasion decreased affect (it became more negative). On subsequent measurement occasions, the impact of this event is discounted, and affect becomes less negative. The occurrence of the

first pleasant event at the eighth measurement occasion increased affective experience (it became more positive). On subsequent occasions, the impact of this pleasant event is discounted, and affective experience becomes less positive.

**Figure 3**

*Simulated affective experience for one hypothetical participant.*



*Note:* Black dots represent the affect ratings, and vertical lines show valent events. Vertical lines above the midpoint indicate positive events, and vertical lines below the midpoint indicate negative events.

### Computational Models of Affect Dynamics

Before we turn to our simulation study, we situate the MIVA model within the landscape of computational models in affect dynamics. Computational models in affect dynamics are statistical or mathematical models that formalize how one believes the target system (i.e., affect) works (cf. Vanhasbroeck et al., 2021), which has several advantages over verbal accounts of affect dynamics (Haslbeck et al., 2022; Ryan et al., 2025; Vanhasbroeck et al., 2021). Computational models allow us to quantify how complex affect processes evolve dynamically over time within individuals and how this differs across individuals (Lafit et al.,

2022; Tuerlinckx et al., 2026; Vanhasbroeck et al., 2021). While many different computational models of affect dynamics exist (Ryan et al., 2025; Tuerlinckx et al., 2026; Vanhasbroeck et al., 2021), it has been reasoned that many researchers desire models that are sensitive to the serial dependence of the data, can quantify regulatory processes, and can take contextual factors into account (Adolf & Ceulemans, 2025; Kuppens & Verduyn, 2017). Especially the inclusion of contextual information into computational models should greatly enhance the understanding of why affect fluctuates (Vanhasbroeck et al., 2021).

As affect dynamic researchers can pick from a variety of computational models (Cui et al., 2023; Tuerlinckx et al., 2026; Vanhasbroeck et al., 2021), we want to discuss similarities and differences between MIVA and the most widely used models. The field's main reference point seems to be (vector) autoregressive models (Tuerlinckx et al., 2026; Vanhasbroeck et al., 2022). They are used to study the serial dependence of the affect system (also known as inertia; Koval et al., 2021; Koval & Kuppens, 2012, 2024) and, in terms of vector autoregressive models, also the cross-lagged dependencies of two or more affect variables. Thus, (vector) autoregressive models can be used to model affect dynamics for unidimensional approaches to affect (as the MIVA model<sup>4</sup>) but also discrete affect approaches, modeling, for example, the dependence of positive and negative affect. It has been conceptualized and empirically shown (Koval et al., 2015) that inertia can capture regulatory processes.

There are, however, several challenges of applying (vector) autoregressive models to the study of affect dynamics using experience sampling data (Berkhout et al., 2025; Cui et al., 2023; Lafit et al., 2022; Vanhasbroeck et al., 2021). For example, the prerequisite of equally spaced observations, which is usually not met by empirical data, leads to biased estimates

---

<sup>4</sup> In principle, it would be possible to extend the MIVA model to include multidimensional affect measures. Different reactivity and regulation parameters could be estimated for each dimension of affective experience and correlated errors could be used to bind these processes together. As the numbers of parameters would substantially increase in such a model, probably more data is required to reliably identify the interrelations of affect dimensions.

across different interval lengths (Berkhout et al., 2025). More importantly, unlike the MIVA model, (vector) autoregressive models do not explicitly model contextual information. Rather, it is assumed that dynamic errors or innovations account for context (Adolf & Ceulemans, 2025).

To circumvent problems of discrete time models, such as the (vector) autoregressive model, continuous time models have been proposed, the most popular ones in affect dynamic research being the Dynamic Affect model (DynAffect; Kuppens et al., 2010) and the Affective Ising Model (AIM; Loossens et al., 2020, 2022). Generally, MIVA, DynAffect, and AIM assume that affective processes unfold continuously across time; however, the facets of affective experience modeled differ across models.

MIVA focuses on affective valence (unidimensional, bipolar), whereas DynAffect models valence and arousal (bidimensional, bipolar), and the AIM focuses on positive and negative affect (bidimensional, unipolar). Accordingly, the data required for parameter estimation varies, with all models requiring an indicator of measurement time as well as one or two affect measures. As MIVA focuses on event-related changes in affect, it also requires the collection of event time and a rating of event pleasantness. This should be kept in mind when designing ESM studies or applying the MIVA model to existing data sets.

All models assume that fluctuations occur due to perturbations of the system, but how perturbations, that is, changes in context, are incorporated into the computational model, differs. In the DynAffect model and the AIM, context changes are only modeled by stochastic deflections (for AIM applications with input function linking context changes, see Vanhasbroeck et al., 2022, 2024), overlooking how these changes might shape affective experience (Prince et al., 2025). In contrast, MIVA incorporates context changes in terms of the occurrence of affect-eliciting events and the appraisal thereof explicitly into the mathematical model formulation.

As shown in Table 1, while the estimated parameters of MIVA and the DynAffect model can be interpreted in terms of psychological processes (e.g., regulation), the AIM does not attach strong interpretations to its parameters in terms of psychological processes but focuses on distributional and qualitative features of affective experience (Loossens et al., 2022). In comparison to MIVA and AIM, the DynAffect model is unbounded, thus potentially producing nonsensical predictions that fall outside of the data range (Vanhasbroeck et al., 2022, 2024). Also, the DynAffect model does not capture potential non-linearities of the affect system, while the MIVA assumes a non-linear mapping between event pleasantness and affective experience, and the AIM itself is a non-linear stochastic model. Interestingly, DynAffect and AIM assume that recovery is represented by exponential decay. However, it has been argued that the assumption of an exponential decay is unsubstantiated (Zavlis et al., 2025). The MIVA model, in turn, assumes hyperbolic decay but could, in principle, also handle other decay functions (e.g., linear).

All models have undergone empirical testing; for example, the 5-parameter MIVA model was the best-fitting model compared to alternative models in which one or more aspects of the MIVA model were absent (Wirth et al., 2022). While empirical studies of the DynAffect and MIVA parameters have shown external validity (e.g., in relation to well-being indicators), the empirical applications of the AIM are mostly focused on its performance compared to other computational models (see Table 1).

Empirical applications of the computational models hinge on how well researchers understand the model, its parameters, and scope, but it also seems important which software is required for parameter estimation, as well as understanding how many data points are needed to precisely estimate the parameters. DynAffect, as the earliest introduced model, has been implemented using different software and packages; the AIM has capitalized on a single package implemented in one software. While previous applications of MIVA have relied on customized code using the rstan package (Stan Development Team, 2025), our simulation

Table 1

*Comparison between MIVA, DynAffect, and AIM*

Model	Scope of model	Estimated Parameters	Data needed to estimate parameters	non-linearity of affect system	bounded model	decay function	Software	Empirical applications	Sample size recommendation
Model of Intraindividual Variability in Affect (MIVA)	Characterize interindividual differences in affective valence dynamics by incorporating event effects	<i>Anchoring</i> - person's average levels of experience when events are absent <i>Reactivity</i> – short-term changes in affective experience related to event <i>Regulation</i> – longer-term changes in affective experience following an event	time, affective experience, event pleasantness, event time	yes	yes	exponential, other options possible (e.g., linear)	RStan; BayesFlow; Python	Affect dynamics and well-being (Wirth et al., 2022)  Age differences in affect dynamics (Wirth et al., 2023)	yes
Dynamic Affect Model (DynAffect)	Characterize interindividual differences in core affect dynamics (valence and arousal)	“ <i>home base</i> ” - person's average levels of experience <i>variability</i> – fluctuation in affective experience across time “ <i>attractor strength</i> ” – stability of affective experience	time, valence, and arousal experience	no	no	exponential	RJags; Bayesian Hierarchical Ornstein–Uhlenbeck Modeling toolbox; MATLAB	Relation to (emotional) traits; (Kuppens et al., 2010; Oravecz & Brick, 2019) Age differences (Wood et al., 2018)  Psychopathology (Ebner-Priemer et al., 2015)	no

Affective Ising Model (AIM)	Capturing nonlinear phenomena in the dynamics of positive and negative affect	$A1, A2$ - positive feedback parameters $A12$ - mutual interaction between the positive and negative pools $\Theta1, \Theta2$ – thresholds; how easy it is to excite positive or negative pool $N1, N2$ – entropy of the system, that is, which affect states are more probable $D$ - diffusion constant, the smaller $D$ , the slower the affect state evolves	time, positive and negative affective experience	yes	yes	exponential	Julia software package	Model comparison (Loossens et al., 2020; Vanhasbroeck et al., 2022, 2024)  Relation between discrete affective experiences (Tongco-Rosario et al., 2024)	no
-----------------------------	---	---	--	-----	-----	-------------	------------------------	--	----

---

study capitalizes on simulation-based inference (SBI; Cranmer et al., 2020). SBI can reduce computational costs for large-scale parameter recovery studies by training neural networks on simulations. Once trained, these networks can perform inference almost instantaneously and across multiple datasets with varying configurations (e.g., numbers of measurements), a property known as *amortized inference*. Training a generative network for all existing and future data compatible with a model and sharing it with other researchers increases the reproducibility of findings and the estimation efficiency (no further training needed) in terms of new data. Here, we utilize SBI as implemented in BayesFlow 2 (Kühmichel et al., 2026), which allowed us to conduct a comprehensive, otherwise costly simulation study<sup>5</sup>. This simulation study provides insights into how design factors of ESM studies (i.e., measurement days and occasions per day) are related to a precise estimation of the MIVA parameters. This information is currently not available for DynAffect and AIM. More general advice on the design of ESM studies can be found, for example, in the handbook published by Myin-Germeys and Kuppens (2026).

### Simulation Study

The value of a computational model for its research field hinges mostly on its ease of use and the availability of guidance with regard to the empirical data collection in terms of, for instance, sample size. This simulation study had two aims. Firstly, a computational bottleneck of the extensive parameter estimation is addressed by using amortized SBI as implemented in BayesFlow 2 (Kühmichel et al., 2026) for the MIVA model. Secondly, we provide guidance on data collection and on the application of the MIVA model to already collected data. In a comprehensive simulation study in which our MIVA model was assumed

---

<sup>5</sup> To compare the speed of parameter estimation, MIVA parameters were estimated from 20 simulated datasets with 14 days of observations and 20 measurements per day using a Stan model. On a modern CPU (AMD Ryzen 7 at 3.80 GHz), fitting the model with 2000 MCMC samples per chain (4 chains run in parallel) took about 4\*24 seconds per dataset. Parameter estimation for 1000 data sets, 8 x 4 different combinations of days and measurements per day, 4 chains and assuming 24 seconds for each chain; would approximately take ~ 35.5 days.

to be the true data-generating process, we investigated the estimation precision across varying numbers of observations (i.e., study days and measurements per day).

Amortized SBI is designed to offset the cost of Bayesian inference over multiple models, datasets, and contextual factors (e.g., data set sizes) by learning approximate posteriors *globally* over the entire range of plausible parameters and datasets (Radev et al., 2022, Figure 1b).

Compared to MCMC-based estimation using Stan (Carpenter et al., 2017), which we applied for estimating MIVA parameters in previous publications (Wirth et al., 2022, 2023), amortized SBI distills posterior sampling into a neural network that becomes an expert in “inverting the model”<sup>6</sup>. Following simulation-based training, the network can obtain posterior estimates from any empirical or simulated data set compatible with the MIVA model at a tiny fraction of the time required for training. This idea has become a cornerstone of modern probabilistic modeling workflows (Bürkner et al., 2023, 2025; Deistler et al., 2025; Lavin et al., 2021; Zammit-Mangion et al., 2023) and can be viewed as a complementary approach to MCMC whenever repeated model (re)-fits are too costly or computationally infeasible (Li et al., 2024).

In BayesFlow, amortized posterior estimation involves two separate neural networks trained jointly. The summary network learns to embed datasets with varying numbers of measurements into vector representations of a fixed size. Specifically, the summary network takes raw data — which in MIVA's case includes affect ratings, event timings, and event pleasantness ratings for each participant — and compresses it into a fixed-length vector of numbers. This is necessary because different participants have different numbers of observations and events, but the inference network needs fixed-size input. The embeddings can be interpreted as learned sufficient summary statistics  $h(x)$ , which can be used in place of

---

<sup>6</sup> While the forward model is an explicit, sometimes complex, simulator ( $p(x|\theta)$ ) that generates data from parameters, the inverse mapping  $p(\theta | x)$  is generally intractable. Amortized SBI trains a conditional density estimator to directly approximate this inverse mapping.

the full data set  $x$  for conditioning the posterior  $p(\theta | x)$ , that is,  $p(\theta | h(x)) \approx p(\theta | x)$ . Due to joint training, these statistics are guaranteed to be maximally informative in principle (Radev et al., 2022). A sufficient statistic captures all the information in the data relevant to parameter estimation — once you have it, the raw data adds nothing more. For example, for estimating the mean of a normal distribution, the sample mean is sufficient. Approximately sufficient means that the summary network learns statistics that capture nearly all relevant information, without requiring us to derive them analytically. The network discovers through training which features of the data are most diagnostic of each parameter.

The inference network is realized as a generative network (e.g., a score-based diffusion model, see Arruda et al., 2025, for more details) that approximates the full joint posterior of model parameters given a dataset embedding. During training, true model parameters are drawn from prior distributions, and synthetic datasets are simulated from MIVA. Thus, the inference network takes those summary statistics and outputs the posterior distribution over the model parameters. It has learned (during training on thousands of simulated datasets) how patterns in the summaries relate to the underlying parameter values.

Then, the neural networks are trained via standard backpropagation, so that the networks learn to approximate the true posteriors. Following training, the networks' weights are “frozen”<sup>7</sup> and no further learning is required, amortizing the cost of upcoming evaluations.

## Study Design

The majority of studies interested in affect dynamic use data of intensive longitudinal designs, including experience sampling methods (ESM), in which participants are asked about affect and related experiences multiple times per day for several days or weeks (Cloos et al., 2024; Hamaker et al., 2015; Vogelsmeier et al., 2024). In this simulation study, we will cover the most common designs found in ESM studies (Pirla et al., 2022; Riediger & Rauters, 2018;

---

<sup>7</sup> This is done to create a static, deterministic model that can be used for inference. Parameters are locked, and no further learning (updating of weights via backpropagation) occurs when processing new data.

Seizer et al., 2024; Wrzus & Neubauer, 2023; Zawadzki et al., 2019) concerning the length of data collection, with 7 to 21 days and 5 to 20 measurement occasions per day. As data in ESM studies can follow different sampling schemes, we decided to mimic one that is most common, namely, interval-based or time-contingent sampling in which daily measurement occasions are scheduled at prespecified times within a time interval (Cloos et al., 2024; Conner & Mehl, 2015; Tuerlinckx et al., 2026). The exact timing of each measurement occasion per day was determined by using the following common procedure: For each day, a time window of 12 hours (between 9 AM and 9 PM) was divided equally by the number of notifications per day, and within the resulting time slots, one measurement was scheduled randomly. Although measurement occasions were sampled at discrete time points, time is treated as continuous when estimating the MIVA parameters.

Determining the number of affect-eliciting events per day was informed by our previous studies (Wirth et al., 2022, 2023), with an average rate of about 3 events per day. The number of events per participant was drawn from a Poisson distribution, with the added provision that there could be no more than one event between two consecutive measurement occasions. This restricts the maximum number of events to the product of days and measurements per day. Further, we only included data sets that entailed at least 2 events over the entire simulated data collection, as an estimation of reactivity and regulation is otherwise intractable.<sup>8</sup> For each simulated data set, MIVA parameters were drawn from the following prior distributions: For anchoring, a truncated normal prior ( $N(0, 1); [-3, 3]$ ) was used; for reactivity and regulation, truncated gamma priors ( $G(2, 2); [0.1, 3.0]$ ) were adopted. These distributions are based on previous results from the MIVA model (Wirth et al., 2022, 2023).

Event pleasantness ratings were simulated on a 7-point scale, as this is one of the most common response scales (Schimmack, 2003). The scales had a symmetrical response format

---

<sup>8</sup> While determining the optimal number of events for good or excellent recovery of MIVA reactivity and regulation parameters is beyond the scope of this manuscript, initial exploration has shown that on average, few (< 1 event per day) and too many events (> 3 events per day) may impede recovery.

to indicate a bipolar scale: -3 indicating highly unpleasant and +3 indicating highly pleasant events. Pleasantness for each event was sampled randomly from six possible values -3, -3, -1, 1, 2, 3 (i.e., excluding 0). Affect ratings were simulated on a scale from -50 to +50 to allow for the detection of more fine-grained changes.

Given that actual assessments of affective experience are prone to measurement error (Dejonckheere & Mestdagh, 2021), simulated affect ratings ( $A_{ij}$ ) were assumed to be influenced by random noise. For this purpose, Gaussian noise with standard deviation  $s$  was added to the exponent in Eq. (1.1):

$$R_j = \frac{r}{1 + e^{-(a+Ru \cdot Iu_j + Rp \cdot Ip_j + e)}} - \frac{r}{2}, \quad \text{with } e \sim N(0, s) \quad (2)$$

where  $s$  was drawn from a uniform distribution between 0 and 1 for each simulated participant.

### Network Architectures and Simulation-Based Training

Since simulations from the MIVA model are extremely fast, we used online training in which the networks never see the same simulation twice. Accordingly, we trained the networks for 1000 epochs, with 200 mini-batches per epoch and a batch size of 32 for a total of 16,000,000 simulations using the Adam optimizer with an initial learning rate of  $1 \times 10^{-4}$ . In each batch, we varied the number of days and number of measurements per day between 7 and 21 and 5 and 20, respectively. The inference network is implemented as a consistency model (Schmitt et al., 2024), which is a few-step neural sampler designed for fast inference. The summary network is implemented as a custom encoder-only transformer designed to reduce time series of varying lengths into fixed-size vectors. Training, validation, and inference were implemented in the BayesFlow Python library (version 2.0.8; Kühmichel et al., 2026). The networks were trained on a single GPU machine equipped with NVIDIA<sup>®</sup> A2000 graphics card.

### Performance Validation

To evaluate the trained inference model, we systematically varied the study design along two dimensions: the number of days and the number of measurements per day. Specifically, we considered 8 levels of study duration (7, 9, 11, 13, 15, 17, 19, 21 days) and 4 levels of measurements per day (5, 10, 15, 20), resulting in 32 distinct conditions in total. For each design, we generated 500 independent test simulations. For every simulated data set, the inference network then produced 1000 posterior samples for the five model parameters: Anchoring, Positive Reactivity, Negative Reactivity, Positive Regulation, and Negative Regulation. In total, the analysis comprised 16,000 simulated test cases ( $32 \times 500$ ) and 16,000,000 posterior samples ( $32 \times 500 \times 1000$ ).

Training of the networks took 70 minutes. Despite the scale of the evaluation, inference was computationally very efficient: generating all posterior samples across all designs and test simulations took less than 3 minutes in total for 16 million posterior samples (excluding simulation and metric computation).

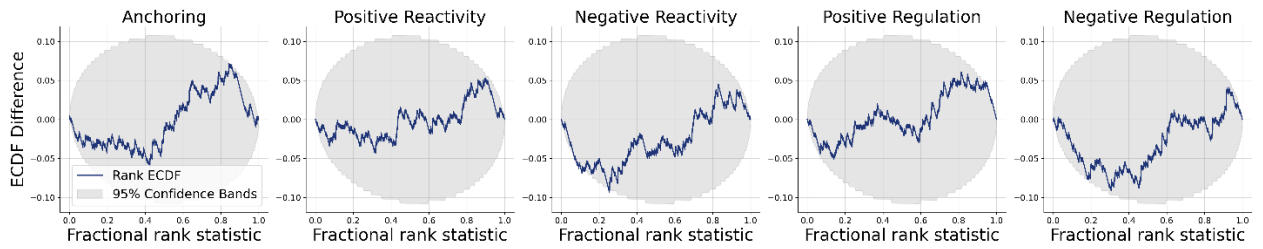
### ***Simulation-based Calibration Check***

The empirical cumulative distribution function (ECDF) of rank statistics is an indicator of a simulation-based calibration check (SBC; Säilynoja et al., 2022) that allows for visually detecting systematic biases in the approximate posteriors. Specifically, if posteriors are well-calibrated (i.e., unbiased), the rank statistics should be uniformly distributed. The ECDF difference plots were obtained by subtracting the values of the expected theoretical cumulative distribution functions from the observed ECDF values. SBC was based on 200 simulated participants. For each, a parameter vector was drawn from the prior, a time series was simulated under the present study design (13 days  $\times$  10 occasions), and 1,000 posterior samples were obtained from the trained network. The ECDF in Figure 4 is constructed from the resulting 200 rank values per parameter. For each of 200 simulated participants, we computed the fractional rank statistic of the true parameter (the proportion of the 1,000 posterior samples falling below the true value). Under a well-calibrated posterior, these ranks

are uniformly distributed across SBC iterations. Figure 4 plots the difference between the empirical and uniform cumulative distribution functions, so the expected curve is the horizontal line at zero, with the grey 95 % confidence band. Deviations outside the band would indicate miscalibration: A U-shape points to over-confident posteriors, an inverted U-shape to under-confident posteriors, and a one-sided deviation indicates a directional bias. As can be seen in Figure 4, the ECDF difference for all MIVA parameters across all conditions was within the 95% confidence bands, indicating no systematic biases in posterior estimation.

**Figure 4**

*Empirical cumulative distribution function (ECDF) of rank statistics for MIVA parameters across all simulated conditions.*



*Note.* Simulation-based calibration check indicating no systematic biases in posterior estimation (empirical cumulative distribution functions; ECDFs within confidence bands). The plots are based on 200 simulated participants and 1000 posterior samples per participant.

### **Calibration Error**

The Calibration Error evaluates whether posterior uncertainty correctly reflects the true uncertainty of the parameters. For each credibility level  $\alpha \in (0,1)$ , we compute the posterior interval

$$I_{\alpha,i,k} = \left[ Q_{\frac{1-\alpha}{2}}, Q_{1-\frac{1-\alpha}{2}} \right] \quad (3)$$

where  $Q_q$  denotes the  $q$ -th posterior quantile. Thus, the empirical coverage of each marginal posterior is given by:

$$\hat{C}_{\alpha,k} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(\theta_{i,k} \in I_{\alpha,i,k}) \quad (4)$$

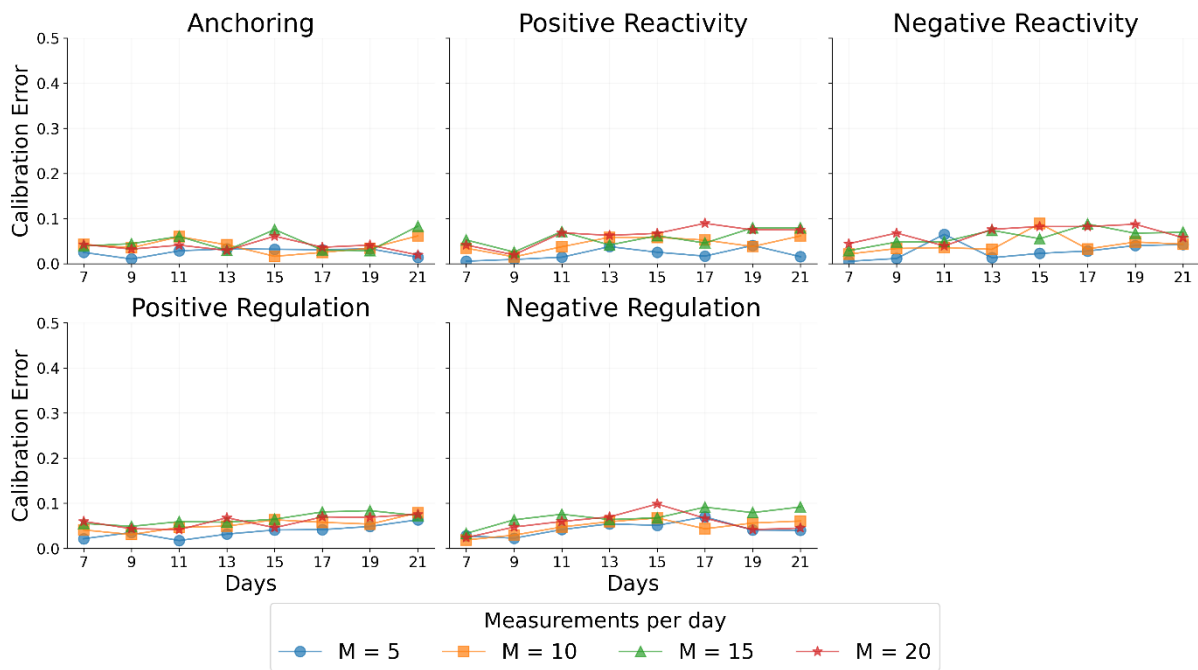
Calibration error summarizes the deviation between nominal and empirical coverage across credibility levels:

$$\text{CalibrationError}_k = \int_0^1 | \hat{C}_{\alpha,k} - \alpha | d\alpha \quad (5)$$

In practice, we approximate the integral using a finite grid of credibility levels between 0 and 1. Lower values indicate better calibrated posterior uncertainty. As can be seen in Figure 5, there was no indication of undesirable patterns, with calibration error remaining below 10% across all study design conditions.

**Figure 5**

*Calibration Error for MIVA parameters across days (7 – 21) and measurements per day (5 – 20).*



*Note.* Calibration error for 500 simulated data sets and 1000 posterior samples per condition.

***Normalized root-mean-squared error (NRMSE)***

The Bayes NRMSE evaluates absolute parameter recovery accuracy while accounting for posterior uncertainty. For each posterior draw  $s$  and parameter  $k$ , the root mean squared error (RMSE) across  $N$  test datasets is

$$\text{RMSE}_k^{(s)} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_{i,k}^{(s)} - \theta_{i,k})^2} \quad (6)$$

To make parameters comparable across scales, the RMSE is normalized by the empirical range of the true parameters

$$\text{NRMSE}_k^{(s)} = \frac{\text{RMSE}_k^{(s)}}{\text{Range}(k)} \quad (7)$$

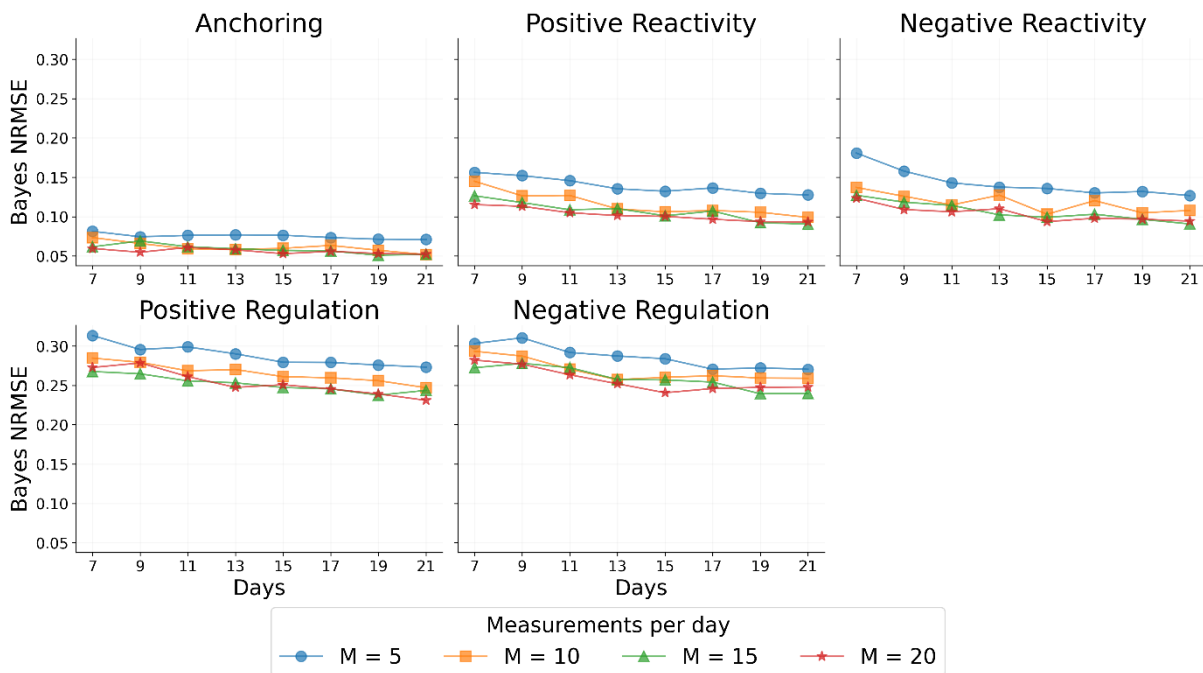
Finally, the metric aggregates across posterior draws, with lower values indicating more accurate parameter recovery.

$$\text{Bayes NRMSE}_k = \text{median}_{s=1}^S (\text{NRMSE}_k^{(s)}) \quad (8)$$

As can be seen in Figure 6, NRMSE decreased with increasing number of days, whereas the number of measurement occasions was less relevant for the NRMSE. Overall, parameter recovery was most accurate for anchoring, with regulation parameters showing somewhat lower accuracy.

**Figure 6**

*NRMSE for MIVA parameters across days (7 – 21) and measurements per day (5 – 20).*



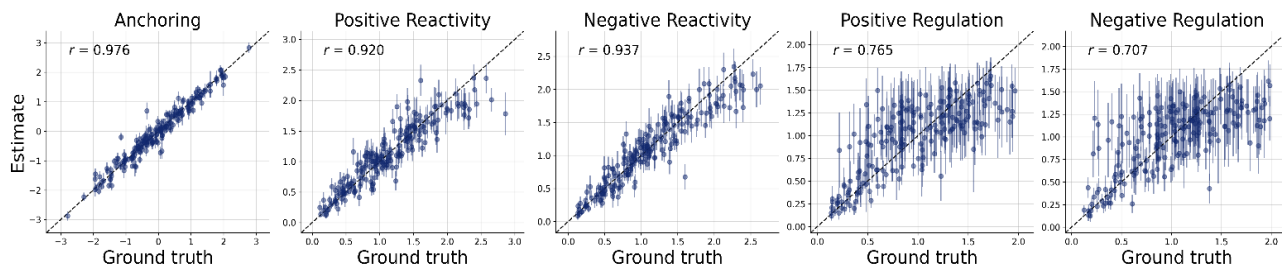
*Note.* NRMSE between posterior samples and ground truth parameters for 500 simulated data sets and 1000 posterior samples per condition.

## Correlation

The correlation indicates the linear relation between inferred and true parameter values, with higher values (closer to 1) indicating better recovery. When looking at the correlations for one simulation condition (13 days and 10 measurement points; Figure 7) across 200 simulated participants and 1000 posterior samples per participant, we see that there is a good match between ground truths (true parameter values) and point estimates.

### Figure 7

*Bayes correlation for MIVA parameter for one simulated condition (13 days and 10 measurement occasions)*



*Note.* Points and vertical lines indicate posterior medians  $\pm$  median absolute deviation.

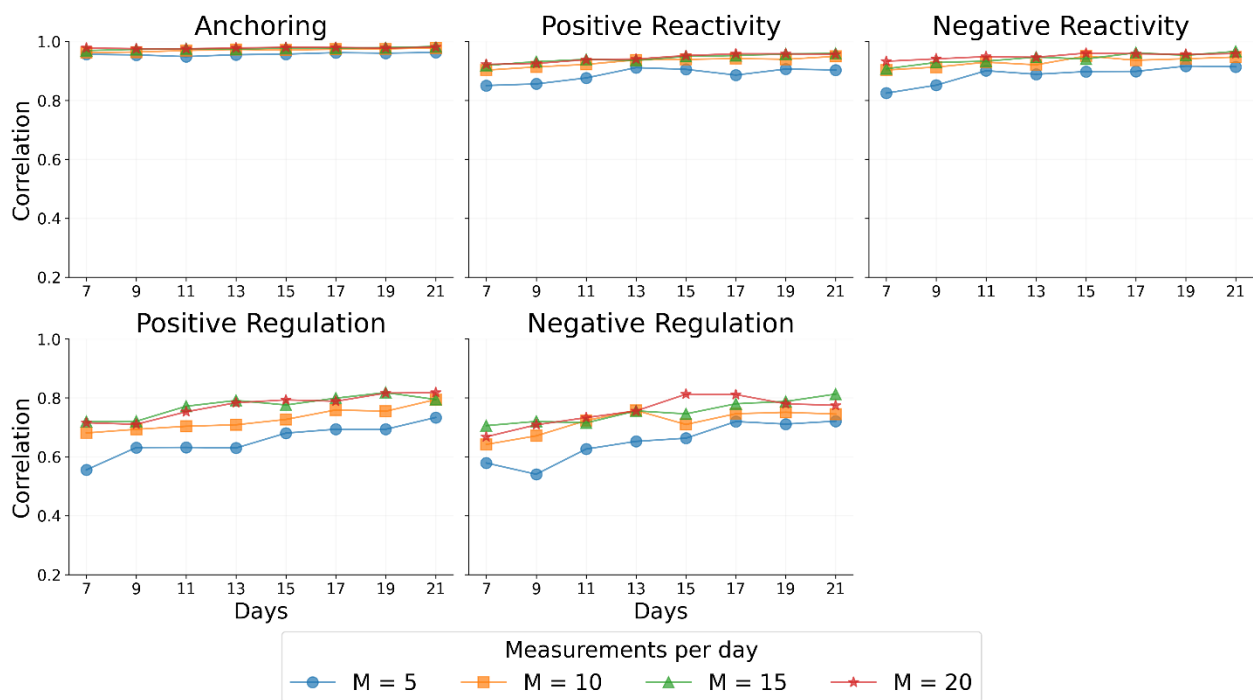
Correlations across all conditions were highest for the anchoring parameter and lowest for the two regulation parameters. Given the overall high recovery of the anchoring parameter, there were no substantial increases in correlation size if days or measurement occasions were increased. For all other MIVA parameters, increasing the number of days or the number of measurement occasions resulted in small to moderate increases in correlation size, with the regulation parameters showing the highest increase (Figure 8). Given that an increase in either the number of days or measurements per day did not necessarily result in meaningful increases in correlations between true and recovered values, we will review the correlation results with a focus on relatively few data points.

For the reactivity parameters, correlations above .80 and below .90 were found for 28 (reactivity to positive events) and 27 (reactivity for negative events) simulation conditions,

with as few as 35 measurement points (7 days and 5 occasions). Correlations equal or above .95 were found for 7 (reactivity to positive events) and 8 (reactivity for negative events) simulation conditions, with about 255 measurement points (17 days and 15 occasions) for reactivity to positive events and 150 measurement points (15 days and 10 occasions) for reactivity to negative events. The combination of the highest number of days (21) and measurement occasions (20) yielded high recovery ( $r = .96$ ).

### Figure 8

*Bayes Correlations for all five MIVA parameters across days (7 – 21) and measurements per day (5 – 20).*



*Note.* Correlations between posterior samples and ground truth parameters for 500 simulated data sets and 1000 posterior samples per configuration.

For the regulation parameters, for conditions with few days (7, 9) and few measurement occasions (5), correlations were between .54 and .63. Correlations between .70 and .80 were found for 20 (regulation of positive event impact) and 21 (regulation of negative event impact) simulation conditions, with 105 or 85 measurement points (7 days and 15 occasions

for regulation of positive event impact and 17 days and 5 occasions for regulation of negative event impact). Correlations above .80 and below .90 were found for 3 simulation conditions, with 285 or 300 measurement points (19 days and 15 occasions for regulation of positive event impact and 15 days and 20 occasions for regulation of negative event impact). Correlations equal to or above .90 were not achieved with the simulated conditions. The combination of the highest number of days (21) and measurement occasions (20) yielded correlations between .77 and .82.

To summarize, we found no indication of systematic biases in posterior estimation, the posterior uncertainty reflected the true uncertainty of the parameters well, the NRMSE decreased with increasing number of days, and correlations between simulated and estimated parameters were moderate to high, depending on the MIVA parameter in question. Given that the MIVA model was assumed to be the true data-generating process, we additionally ran a simulation study with the same design parameters but in which the translation between events and affective experience was different. Specifically, it was assumed that in some cases, events had no impact on affective experience, and in other cases, events were not reported but still had an effect on affective experience. Results of this simulation study can be found at <https://osf.io/spxf8/files/pr2xq>. Generally, performance validation indicators were robust concerning this misspecification (for robustness of amortized Bayesian inference in light of contaminated observations, see Wu et al., 2026).

## Discussion

In this manuscript, we presented an outline of the theoretical background of MIVA. The model captures three processes assumed to be fundamental for the elicitation and fluctuation of affective experience, namely anchoring, reactivity, and regulation (Hamaker et al., 2015; Krone et al., 2018; Kuppens et al., 2010; Larsen, 2000; Röcke & Brose, 2013). We further situated the MIVA model within the affect dynamics literature and compared it to other common computational models.

To elucidate the conditions under which it would produce reliable estimates, we conducted a simulation study that capitalized on recent advances in generative deep learning and neural network architectures (Kühmichel et al., 2026; Radev et al., 2022, 2023). Our simulation study demonstrated that the synthesis between a computational model and probabilistic neural networks results in an efficient and flexible tool for model-based inference of affect dynamics.

### **Practical Implications**

Our simulation study showed that the anchoring and reactivity parameters can be recovered with high precision with as few as 35 data points. To achieve a high precision in recovery of the regulation parameters, additional data points were needed, with at least 15 days and 10 measurement occasions. Around 150 measurement points might seem a lot, given the average ESM design with 7 days and 6 measurement occasions (Pirla et al., 2022; Wrzus & Neubauer, 2023). However, doubts have been raised about whether these designs provide sufficient data to adequately capture affect dynamic processes (Haslbeck & Ryan, 2022; Kuppens et al., 2022; Schuurman et al., 2015). Moreover, simulation studies for other model-based approaches also indicate that around 150 to 200 measurement points are necessary to assess affect dynamic indicators with high precision (Adolf et al., 2017; Deboeck et al., 2009; Kuppens et al., 2010; Wood et al., 2018). Other approaches used to extract indicators of regulatory processes from ESM data using lagged effects of stressful events (Scott et al., 2017; Smyth et al., 2018; Wrzus et al., 2015) might be equally disadvantaged when it comes to relatively short assessment periods.

An interesting question is why the estimation of regulation parameters requires more measurement points than the reactivity parameter. This is because the long-term effects of events on affective experience are more easily discernible from the short-term effects of subsequently occurring events (i.e., reactivity) if more data points are available. With few data points stretched out across several hours, it becomes more difficult to accurately model the

unobserved affective experience. Also, with few data points, the likelihood of events being reported on successive measurement occasions increases, again leading to problems in differentiating between the short- (reactivity) and long-term (regulation) effects of events on affective experience (Almeida et al., 2020; Scott et al., 2017; Smyth et al., 2018).

Although increasing the number of days or measurement occasions both lead to similar increases in recovery of regulation and reactivity parameters, if the subject of investigation is an individual's average affect fluctuations, the assessment period should include both weekdays and weekends as well as work and non-work times (Ram et al., 2014; Scott et al., 2020). When interested in affect dynamics related to a specific event (e.g., university exam; Dejonckheere et al., 2021), increasing the number of assessments per day before and after this event might be recommended, as there are more within-person affect fluctuations within than across days (Scott et al., 2020).

Generally, assessing affect dynamics at time scales that do not map onto those at which the underlying processes unfold may result in misleading or inaccurate inferences (Hamaker, 2025; Haslbeck & Ryan, 2022; Lazarus et al., 2021). Very frequent measurements with short time intervals can lead to a proportional increase in measurement error compared with true change (Boker & Nesselroade, 2002; Hopwood et al., 2022). Too sparse measurement could lead to missing important micro changes in affect (Hamaker & Wichers, 2017; Kalokerinos et al., 2020; Kuppens et al., 2022). Most importantly, the appropriate temporal resolution of measurements should be informed by theoretical work on affective change. However, only a few theoretical ideas concerning the time scale on which affective changes occur in the real world have been put forward (Fried, 2020; Hopwood et al., 2022; Lazarus et al., 2021). The Flex3 model (Hollenstein, 2015, 2021) proposes that real-time, temporally adjacent shifts from one affective state to another capture affect regulation. Other affect dynamic concepts assume a less intricate temporal resolution and propose cyclic changes in affective experience across weeks, months, or seasons (Chow et al., 2005; Larsen & Kasimatis, 1990; Liu & West,

2016; Ollero et al., 2023). However, current model-based approaches to affect dynamics, including MIVA, have not formulated specific assumptions about the exact time-dependent nature of changes in affective experience. Rather, the timing of affect fluctuations is implicitly assumed to happen on a minute-to-hourly basis. The specific focus of MIVA was to understand how events elicit affective fluctuations and how much and how long they influence the underlying affect processes. This focus was inspired by work calling attention to the neglected role of event appraisals in affect dynamic research (Vanhasbroeck et al., 2022; Wenzel et al., 2022). Going beyond this event-based focus and integrating an in-depth exploration of time-related processes will greatly enhance our understanding of affect dynamics.

### **Strength and Limitations**

To complement current methods and models of affect dynamics, we developed a theory-based approach that specifically focuses on the effects of events on the processes assumed to underlie affect dynamics. The MIVA parameters anchoring, reactivity, and regulation capture affect characteristics that differ between individuals, are relatively stable across time, and are relatively consistent over situations. Its parameters reflect essential processes that account for observed changes in affective experience (Wirth et al., 2022).

### ***Event-dependence***

While focusing on the short- and long-term effects of events affords a closer look at how situational appraisals are translated into affective fluctuations, it comes at the expense of having to rely on participants reporting such events. The MIVA parameters regulation and reactivity are conceptualized as event-dependent, meaning that they cannot be estimated for participants who do not report any events. Thus, in our simulation study, we also had to constrain the used datasets to those containing at least two events. The event-based nature of MIVA limits its applicability for empirical datasets of everyday affective experience. Research has just started exploring the characteristics of samples in ESM studies that do not

report events in daily life (Charles et al., 2021). How well their affect dynamics can be investigated under controlled laboratory conditions remains an open question. Adopting MIVA to datasets from controlled laboratory studies is an important future avenue because it allows to (a) determine affect dynamics of individuals who do not experience or report daily events, (b) examine affect dynamics with a high temporal resolution on a second-to-second basis, and (c) shed light on the yet unresolved issue of overlap between affect dynamics assessed in laboratory versus daily life (Davidson, 2015; Klein et al., 2023; Koval et al., 2015; Kuppens et al., 2022).

### *Cyclic changes and changes in model parameters over time*

To keep MIVA as parsimonious as possible, it captures only certain parts of the complex reality. It currently does not include parameters for cyclic changes in affective experience. Such intricacies have not been incorporated because the model then may become harder to manage both conceptually and computationally, requiring large sample sizes and more measurement occasions to estimate the model parameters with reasonable precision (Bosse et al., 2014; Krone et al., 2018). Additionally, MIVA currently rests on the assumption that its parameters are stable over time. However, due to critical life events, developmental processes, or interventions, the model parameters may exhibit changes that are more or less transient (Bringmann et al., 2016; Koval & Kuppens, 2012). Statistical advances have accounted for changes in affect dynamic indicators, such as the continuous time dynamic modeling approach (Driver & Voelkle, 2018) or the time-varying change point autoregressive model (Albers & Bringmann, 2020). These statistical approaches, however, are descriptive and exploratory rather than driven by specific theoretical ideas of time-dependent changes in affect dynamics. What is needed is a deeper understanding of the time course at which affect fluctuates and how this time-dependent change maps onto the frequency with which studies investigate affect dynamics. This will help to develop a more comprehensive theory of affect dynamics that explains the underlying causal mechanisms and makes accurate predictions

about the magnitude, shape, and direction of affect fluctuation and the time scales and contextual conditions under which they occur (Fried, 2020; Lazarus et al., 2021). Otherwise, the field is left with an abundance of exploratory and pluralistic approaches and difficulties in determining which indicators are theoretically meaningful (Kuppens et al., 2022).

### ***Range of applicability of the model***

Our simulation study demonstrated the utility of the amortized Bayesian inference as implemented in the BayesFlow toolkit for assessing affect dynamics. We have shown that the same trained network can be utilized across different study designs. This makes the method especially effective when inferences are drawn across different scenarios with a large number of data sets and/or varying numbers of participants or assessment occasions. The re-use of trained networks for multiple datasets and across multiple researchers will increase the comparability and reproducibility of findings in affect dynamic research. Going beyond the current simulation study, the posteriors estimated via the BayesFlow method can be used to test hypotheses about particular parameter values, compute individual differences, or compare means between conditions in a fully Bayesian way. When focusing especially on the comparison of affect dynamics between different groups (e.g., age groups or patient groups), using hierarchical BayesFlow models including hyperpriors would be advisable. To increase the applicability and usability of the novel simulation-based BayesFlow method, developing and maintaining user-friendly software is crucial. So far, BayesFlow offers a user-friendly API, which encapsulates the details of neural network architectures and training procedures. Concerning MIVA, our simulator, approximation scripts, and results are fully open and reproducible. Thus, we hope that amortized Bayesian inference will enhance model-based inference in affect dynamic research.

### **Conclusion**

Affective experience is inherently dynamic as it continuously fluctuates as a result of internal or external events. We applied a formalized theoretical approach to develop a

parsimonious model of affect dynamics. MIVA allows estimating parameters for anchoring, reactivity, and regulation based on affective states in combination with daily events. MIVA was empirically evaluated by conducting a simulation study using amortized Bayesian inference, a novel approach capitalizing on the latest developments in machine learning. Our simulation demonstrates the applicability of amortized methods to studying affect dynamics and offers insight into the conditions under which the MIVA model parameters can be estimated with high precision.

## **Declarations**

### **Funding**

This work was supported by an early-career grant from Friedrich Schiller University Jena awarded to MW.

### **Conflicts of Interest/ Competing Interests**

The authors have no known conflicts of interest or competing interests to disclose.

### **Ethics Approval**

Not applicable, data was simulated.

### **Consent to participate**

Not applicable, data was simulated.

### **Consent for publication**

Not applicable, data was simulated.

### **Availability of data and materials**

Data included in this manuscript are available at <https://osf.io/spxf8/>

### **Code Availability**

Scripts included in this manuscript are available at <https://osf.io/spxf8/>

### **Authors' contributions**

M. Wirth, A. Voss, S. Radev, and K. Rothermund conceptualized and designed the study. S. Radev and A. Voss simulated and analyzed the data. M. Wirth and K. Rothermund assisted in interpreting the data. M. Wirth wrote the first draft, and all authors critically edited the manuscript.

### References

- Adolf, J. K., & Ceulemans, E. (2025). Improved estimation of autoregressive models through contextual impulses and robust modeling. *Psychological Methods*.  
<https://doi.org/10.1037/met0000761>
- Adolf, J. K., Voelkle, M. C., Brose, A., & Schmiedek, F. (2017). Capturing Context-Related Change in Emotional Dynamics via Fixed Moderated Time Series Analysis. *Multivariate Behavioral Research*, 52(4), 499–531.  
<https://doi.org/10.1080/00273171.2017.1321978>
- Albers, C. J., & Bringmann, L. F. (2020). Inspecting Gradual and Abrupt Changes in Emotion Dynamics With the Time-Varying Change Point Autoregressive Model. *European Journal of Psychological Assessment*, 36(3), 492–499. <https://doi.org/10.1027/1015-5759/a000589>
- Alessandri, G., Zuffianò, A., Fabes, R., Vecchione, M., & Martin, C. (2014). Linking Positive Affect and Positive Self-beliefs in Daily Life. *Journal of Happiness Studies*, 15(6), 1479–1493. <https://doi.org/10.1007/s10902-013-9487-y>
- Almeida, D. M., Marcusson-Clavertz, D., Conroy, D. E., Kim, J., Zawadzki, M. J., Sliwinski, M. J., & Smyth, J. M. (2020). Everyday stress components and physical activity: Examining reactivity, recovery and pileup. *Journal of Behavioral Medicine*, 43(1), 108–120. <https://doi.org/10.1007/s10865-019-00062-z>
- Arruda, J., Bracher, N., Köthe, U., Hasenauer, J., & Radev, S. T. (2025). *Diffusion Models in Simulation-Based Inference: A Tutorial Review (Version 2)*. arXiv.  
<https://doi.org/10.48550/ARXIV.2512.20685>
- Backs, R. W., da Silva, S. P., & Han, K. (2005). A Comparison of Younger and Older Adults' Self-Assessment Manikin Ratings of Affective Pictures. *Experimental Aging Research*, 31(4), 421–440. <https://doi.org/10.1080/03610730500206808>

- Barrett, L. F., & Russell, J. A. (1999). The Structure of Current Affect: Controversies and Emerging Consensus. *Current Directions in Psychological Science*, 8(1), 10–14.  
<https://doi.org/10.1111/1467-8721.00003>
- 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>
- Boker, S. M., & Nesselroade, J. R. (2002). A Method for Modeling the Intrinsic Dynamics of Intraindividual Variability: Recovering the Parameters of Simulated Oscillators in Multi-Wave Panel Data. *Multivariate Behavioral Research*, 37(1), 127–160.  
[https://doi.org/10.1207/S15327906MBR3701\\_06](https://doi.org/10.1207/S15327906MBR3701_06)
- Bolger, N., & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of Personality and Social Psychology*, 69(5), 890–902.  
<https://doi.org/10.1037/0022-3514.69.5.890>
- Bosse, T., Broekens, J., Dias, J., & Van Der Zwaan, J. (Eds.). (2014). *Emotion Modeling: Towards Pragmatic Computational Models of Affective Processes* (Vol. 8750). Springer International Publishing. <https://doi.org/10.1007/978-3-319-12973-0>
- Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., Tuerlinckx, F., & Kuppens, P. (2016). Assessing Temporal Emotion Dynamics Using Networks. *Assessment*, 23(4), 425–435. <https://doi.org/10.1177/1073191116645909>
- Bürkner, P.-C., Schmitt, M., & Radev, S. T. (2025). *Simulations in Statistical Workflows* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2503.24011>
- Bürkner, P.-C., Scholz, M., & Radev, S. T. (2023). Some models are useful, but how do we know which ones? Towards a unified Bayesian model taxonomy. *Statistics Surveys*, 17, 216–310.

- Bylsma, L. M., Taylor-Clift, A., & Rottenberg, J. (2011). Emotional reactivity to daily events in major and minor depression. *Journal of Abnormal Psychology, 120*(1), 155–167. <https://doi.org/10.1037/a0021662>
- 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). <https://doi.org/10.18637/jss.v076.i01>
- Charles, S. T., Mogle, J., Chai, H. W., & Almeida, D. M. (2021). The mixed benefits of a stressor-free life. *Emotion, 21*(5), 962–971. <https://doi.org/10.1037/emo0000958>
- Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as a Thermostat: Representing Emotion Regulation Using a Damped Oscillator Model. *Emotion, 5*(2), 208–225. <https://doi.org/10.1037/1528-3542.5.2.208>
- Cloos, L., Ceulemans, E., & Kuppens, P. (2023). Development, validation, and comparison of self-report measures for positive and negative affect in intensive longitudinal research. *Psychological Assessment, 35*(3), 189–204. <https://doi.org/10.1037/pas0001200>
- Cloos, L., Mestdagh, M., Vanpaemel, W., Ceulemans, E., & Kuppens, P. (2024). Measuring Continuous Affect in Daily Life With Intensity Profile Drawings. *Assessment, 10731911241266286*. <https://doi.org/10.1177/10731911241266286>
- Conner, T. S., & Mehl, M. R. (2015). Ambulatory Assessment: Methods for Studying Everyday Life. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (1st ed., pp. 1–15). Wiley. <https://doi.org/10.1002/9781118900772.etrds0010>
- Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences, 117*(48), 30055–30062. <https://doi.org/10.1073/pnas.1912789117>

- Cui, J., Hasselman, F., & Lichtwarck-Aschoff, A. (2023). Unlocking nonlinear dynamics and multistability from intensive longitudinal data: A novel method. *Psychological Methods*. <https://doi.org/10.1037/met0000623>
- Davidson, R. J. (1998). Affective Style and Affective Disorders: Perspectives from Affective Neuroscience. *Cognition & Emotion*, *12*(3), 307–330.  
<https://doi.org/10.1080/026999398379628>
- Davidson, R. J. (2015). Comment: Affective Chronometry Has Come of Age. *Emotion Review*, *7*(4), 368–370. <https://doi.org/10.1177/1754073915590844>
- Deboeck, P. R., Montpetit, M. A., Bergeman, C. S., & Boker, S. M. (2009). Using derivative estimates to describe intraindividual variability at multiple time scales. *Psychological Methods*, *14*(4), 367–386. <https://doi.org/10.1037/a0016622>
- Deistler, M., Boelts, J., Steinbach, P., Moss, G., Moreau, T., Gloeckler, M., Rodrigues, P. L. C., Linhart, J., Lappalainen, J. K., Miller, B. K., Gonçalves, P. J., Lueckmann, J.-M., Schröder, C., & Macke, J. H. (2025). *Simulation-Based Inference: A Practical Guide* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2508.12939>
- Dejonckheere, E., Demeyer, F., Geusens, B., Piot, M., Tuerlinckx, F., Verdonck, S., & Mestdagh, M. (2022). Assessing the reliability of single-item momentary affective measurements in experience sampling. *Psychological Assessment*, *34*(12), 1138–1154.  
<https://doi.org/10.1037/pas0001178>
- Dejonckheere, E., & Mestdagh, M. (2021). On the Signal-to-Noise Ratio in Real-Life Emotional Time Series. In C. E. Waugh & P. Kuppens (Eds.), *Affect Dynamics* (pp. 131–152). Springer International Publishing. [https://doi.org/10.1007/978-3-030-82965-0\\_7](https://doi.org/10.1007/978-3-030-82965-0_7)
- Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of

- psychological well-being. *Nature Human Behaviour*, 3(5), 478–491.  
<https://doi.org/10.1038/s41562-019-0555-0>
- Dejonckheere, E., Mestdagh, M., Verdonck, S., Lafit, G., Ceulemans, E., Bastian, B., & Kalokerinos, E. K. (2021). The relation between positive and negative affect becomes more negative in response to personally relevant events. *Emotion*, 21(2), 326–336.  
<https://doi.org/10.1037/emo0000697>
- Diener, E., & Diener, C. (1996). Most people are happy. *Psychological Science*, 7(3), 181–185. <https://doi.org/10.1111/j.1467-9280.1996.tb00354.x>
- Driver, C. C., & Voelkle, M. C. (2018). Understanding the Time Course of Interventions with Continuous Time Dynamic Models. In K. Van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous Time Modeling in the Behavioral and Related Sciences* (pp. 79–109). Springer International Publishing. [https://doi.org/10.1007/978-3-319-77219-6\\_4](https://doi.org/10.1007/978-3-319-77219-6_4)
- Ebner-Priemer, U. W., Houben, M., Santangelo, P., Kleindienst, N., Tuerlinckx, F., Oravecz, Z., Verleysen, G., Van Deun, K., Bohus, M., & Kuppens, P. (2015). Unraveling affective dysregulation in borderline personality disorder: A theoretical model and empirical evidence. *Journal of Abnormal Psychology*, 124(1), 186–198.  
<https://doi.org/10.1037/abn0000021>
- Fried, E. I. (2020). Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature. *Psychological Inquiry*, 31(4), 271–288.  
<https://doi.org/10.1080/1047840X.2020.1853461>
- Frijda, N. H. (1993). The place of appraisal in emotion. *Cognition & Emotion*, 7(3–4), 357–387. <https://doi.org/10.1080/02699939308409193>
- Frijda, N. H. (2007). What might emotions be? Comments on the Comments. *Social Science Information*, 46(3), 433–443. <https://doi.org/10.1177/05390184070460030112>

- Fritz, J., Piccirillo, M. L., Cohen, Z. D., Frumkin, M., Kirtley, O., Moeller, J., Neubauer, A. B., Norris, L. A., Schuurman, N. K., Snippe, E., & Bringmann, L. F. (2024). So You Want to Do ESM? 10 Essential Topics for Implementing the Experience-Sampling Method. *Advances in Methods and Practices in Psychological Science*, 7(3), 25152459241267912. <https://doi.org/10.1177/25152459241267912>
- Gibbs, M. N., & Mackay, D. J. C. (2000). Variational Gaussian process classifiers. *IEEE Transactions on Neural Networks*, 11(6), 1458–1464. <https://doi.org/10.1109/72.883477>
- Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, 64(6), 1029–1041. <https://doi.org/10.1037/0022-3514.64.6.1029>
- Gross, J. J. (2013). Emotion regulation: Taking stock and moving forward. *Emotion*, 13(3), 359–365. <https://doi.org/10.1037/a0032135>
- Hamaker, E. L. (2025). Analysis of Intensive Longitudinal Data: Putting Psychological Processes in Perspective. *Annual Review of Clinical Psychology*, 21(1), 379–405. <https://doi.org/10.1146/annurev-clinpsy-081423-022947>
- Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling Affect Dynamics: State of the Art and Future Challenges. *Emotion Review*, 7(4), 316–322. <https://doi.org/10.1177/1754073915590619>
- Hamaker, E. L., & Wichers, M. (2017). No Time Like the Present: Discovering the Hidden Dynamics in Intensive Longitudinal Data. *Current Directions in Psychological Science*, 26(1), 10–15. <https://doi.org/10.1177/0963721416666518>
- Haslbeck, J. M. B., & Ryan, O. (2022). Recovering Within-Person Dynamics from Psychological Time Series. *Multivariate Behavioral Research*, 57(5), 735–766. <https://doi.org/10.1080/00273171.2021.1896353>

- Haslbeck, J. M. B., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2022). Modeling psychopathology: From data models to formal theories. *Psychological Methods, 27*(6), 930–957. <https://doi.org/10.1037/met0000303>
- Hemenover, S. H. (2003). Individual differences in rate of affect change: Studies in affective chronometry. *Journal of Personality and Social Psychology, 85*(1), 121–131. <https://doi.org/10.1037/0022-3514.85.1.121>
- Hollenstein, T. (2015). This Time, It's Real: Affective Flexibility, Time Scales, Feedback Loops, and the Regulation of Emotion. *Emotion Review, 7*(4), 308–315. <https://doi.org/10.1177/1754073915590621>
- Hollenstein, T. (2021). Affect Dynamics and Time Scales: Pictures of Movies. In C. E. Waugh & P. Kuppens (Eds.), *Affect Dynamics* (pp. 117–129). Springer International Publishing. [https://doi.org/10.1007/978-3-030-82965-0\\_6](https://doi.org/10.1007/978-3-030-82965-0_6)
- Hopwood, C. J., Bleidorn, W., & Wright, A. G. C. (2022). Connecting Theory to Methods in Longitudinal Research. *Perspectives on Psychological Science, 17*(3), 884–894. <https://doi.org/10.1177/17456916211008407>
- Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods, 13*(4), 354–375. <https://doi.org/10.1037/a0014173>
- Kalokerinos, E. K., Murphy, S. C., Koval, P., Bailen, N. H., Crombez, G., Hollenstein, T., Gleeson, J., Thompson, R. J., Van Ryckeghem, D. M. L., Kuppens, P., & Bastian, B. (2020). Neuroticism may not reflect emotional variability. *Proceedings of the National Academy of Sciences, 117*(17), 9270–9276. <https://doi.org/10.1073/pnas.1919934117>
- Klein, R. J., Rapaport, R., Gyorda, J. A., Jacobson, N. C., & Robinson, M. D. (2023). Second-to-Second Affective Responses to Images Correspond With Affective Reactivity,

- Variability, and Instability in Daily Life. *Experimental Psychology*, 70(1), 14–31.  
<https://doi.org/10.1027/1618-3169/a000564>
- Koole, S. L., & Rothermund, K. (2011). “I feel better but I don’t know why”: The psychology of implicit emotion regulation. *Cognition & Emotion*, 25(3), 389–399.  
<https://doi.org/10.1080/02699931.2010.550505>
- Koval, P., Brose, A., Pe, M. L., Houben, M., Erbas, Y., Champagne, D., & Kuppens, P. (2015). Emotional inertia and external events: The roles of exposure, reactivity, and recovery. *Emotion*, 15(5), 625–636. <https://doi.org/10.1037/emo0000059>
- Koval, P., Burnett, P. T., & Zheng, Y. (2021). Emotional Inertia: On the Conservation of Emotional Momentum. In C. E. Waugh & P. Kuppens (Eds.), *Affect Dynamics* (pp. 63–94). Springer International Publishing. [https://doi.org/10.1007/978-3-030-82965-0\\_4](https://doi.org/10.1007/978-3-030-82965-0_4)
- Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, 12(2), 256–267.  
<https://doi.org/10.1037/a0024756>
- Koval, P., & Kuppens, P. (2024). Changing feelings: Individual differences in emotional inertia. In *Change in Emotion and Mental Health* (pp. 3–21). Elsevier.  
<https://doi.org/10.1016/B978-0-323-95604-8.00007-1>
- Krone, T., Albers, C. J., Kuppens, P., & Timmerman, M. E. (2018). A multivariate statistical model for emotion dynamics. *Emotion*, 18(5), 739–754.  
<https://doi.org/10.1037/emo0000384>
- Kühmichel, L., Huang, J. M., Pratz, V., Arruda, J., Olischläger, H., Habermann, D., Kucharsky, S., Elsemüller, L., Mishra, A., Bracher, N., Jedhoff, S., Schmitt, M., Bürkner, P.-C., & Radev, S. T. (2026). *BayesFlow 2.0: Multi-Backend Amortized Bayesian Inference in Python* (Version 1). arXiv.  
<https://doi.org/10.48550/ARXIV.2602.07098>

- Kuppens, P. (2015). It's About Time: A Special Section on Affect Dynamics. *Emotion Review*, 7(4), 297–300. <https://doi.org/10.1177/1754073915590947>
- Kuppens, P., Dejonckheere, E., Kalokerinos, E. K., & Koval, P. (2022). Some Recommendations on the Use of Daily Life Methods in Affective Science. *Affective Science*, 3(2), 505–515. <https://doi.org/10.1007/s42761-022-00101-0>
- Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, 99(6), 1042–1060. <https://doi.org/10.1037/a0020962>
- Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, 139(4), 917–940. <https://doi.org/10.1037/a0030811>
- Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology*, 17, 22–26. <https://doi.org/10.1016/j.copsyc.2017.06.004>
- Lafit, G., Meers, K., & Ceulemans, E. (2022). A Systematic Study into the Factors that Affect the Predictive Accuracy of Multilevel VAR(1) Models. *Psychometrika*, 87(2), 432–476. <https://doi.org/10.1007/s11336-021-09803-z>
- Larsen, R. J. (2000). Toward a Science of Mood Regulation. *Psychological Inquiry*, 11(3), 129–141. [https://doi.org/10.1207/S15327965PLI1103\\_01](https://doi.org/10.1207/S15327965PLI1103_01)
- Larsen, R. J., & Kasimatis, M. (1990). Individual differences in entrainment of mood to the weekly calendar. *Journal of Personality and Social Psychology*, 58(1), 164–171. <https://doi.org/10.1037/0022-3514.58.1.164>
- Laurenceau, J.-P., DiGiovanni, A. M., & Bolger, N. (2025). Intensive Longitudinal Methods: Toward a Psychological Science of Daily Life. *Annual Review of Psychology*. <https://doi.org/10.1146/annurev-psych-040325-025418>
- Lavin, A., Krakauer, D., Zenil, H., Gottschlich, J., Mattson, T., Brehmer, J., Anandkumar, A., Choudry, S., Rocki, K., Baydin, A. G., Prunkl, C., Paige, B., Isayev, O., Peterson, E.,

- McMahon, P. L., Macke, J., Cranmer, K., Zhang, J., Wainwright, H., ... Pfeffer, A. (2021). *Simulation Intelligence: Towards a New Generation of Scientific Methods* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2112.03235>
- Lazarus, G., Song, J., Crawford, C. M., & Fisher, A. J. (2021). A Close Look at the Role of Time in Affect Dynamics Research. In C. E. Waugh & P. Kuppens (Eds.), *Affect Dynamics* (pp. 95–116). Springer International Publishing. [https://doi.org/10.1007/978-3-030-82965-0\\_5](https://doi.org/10.1007/978-3-030-82965-0_5)
- Li, C., Vehtari, A., Bürkner, P.-C., Radev, S. T., Acerbi, L., & Schmitt, M. (2024). *Amortized Bayesian Workflow* (Version 3). arXiv. <https://doi.org/10.48550/ARXIV.2409.04332>
- Liu, Y., & West, S. G. (2016). Weekly Cycles in Daily Report Data: An Overlooked Issue. *Journal of Personality*, *84*(5), 560–579. <https://doi.org/10.1111/jopy.12182>
- Loossens, T., Meers, K., Vanhasbroeck, N., Anarat, N., Verdonck, S., & Tuerlinckx, F. (2022). Efficient estimation of bounded gradient-drift diffusion models for affect on CPU and GPU. *Behavior Research Methods*, *54*(3), 1428–1443. <https://doi.org/10.3758/s13428-021-01674-7>
- Loossens, T., Mestdagh, M., Dejonckheere, E., Kuppens, P., Tuerlinckx, F., & Verdonck, S. (2020). The Affective Ising Model: A computational account of human affect dynamics. *PLOS Computational Biology*, *16*(5), e1007860. <https://doi.org/10.1371/journal.pcbi.1007860>
- Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelineing 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>
- Mroczek, D. K., & Almeida, D. M. (2004). The Effect of Daily Stress, Personality, and Age on Daily Negative Affect. *Journal of Personality*, *72*(2), 355–378. <https://doi.org/10.1111/j.0022-3506.2004.00265.x>

- Myin-Germeys, I., & Kuppens, P. (Eds.). (2026). *The Open Handbook of Experience Sampling Methodology – Third edition*. Leuven University Press.  
<https://doi.org/10.11116/9789461666932>
- Ollero, M. J. F., Estrada, E., Hunter, M. D., & Cáncer, P. F. (2023). Characterizing affect dynamics with a damped linear oscillator model: Theoretical considerations and recommendations for individual-level applications. *Psychological Methods*.  
<https://doi.org/10.1037/met0000615>
- Ong, A. D., Dejonckheere, E., & Ram, N. (2025). Positive Affect Dynamics. *Current Directions in Psychological Science*, 34(5), 301–308.  
<https://doi.org/10.1177/09637214251339454>
- Oravecz, Z., & Brick, T. R. (2019). Associations Between Slow- and Fast-Timescale Indicators of Emotional Functioning. *Social Psychological and Personality Science*, 10(7), 864–873. <https://doi.org/10.1177/1948550618797128>
- Pirla, S., Taquet, M., & Quidbach, J. (2022). Measuring affect dynamics: An empirical framework. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-022-01829-0>
- Prince, C., Demaree, H., Goldenberg, A., & LaFollette, K. J. (2025). *Bridging Internal and External Forces in Affect Dynamics*. PsyArXiv.  
[https://doi.org/10.31234/osf.io/jvxza\\_v1](https://doi.org/10.31234/osf.io/jvxza_v1)
- Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L., & Kothe, U. (2022). BayesFlow: Learning Complex Stochastic Models With Invertible Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(4), 1452–1466.  
<https://doi.org/10.1109/TNNLS.2020.3042395>
- Radev, S. T., Schmitt, M., Schumacher, L., Elsemüller, L., Pratz, V., Schälte, Y., Köthe, U., & Bürkner, P.-C. (2023). *BayesFlow: Amortized Bayesian Workflows With Neural Networks*. <https://doi.org/10.48550/ARXIV.2306.16015>

- Ram, N., Conroy, D. E., Pincus, A. L., Lorek, A., Rebar, A., Roche, M. J., Coccia, M., Morack, J., Feldman, J., & Gerstorf, D. (2014). Examining the Interplay of Processes Across Multiple Time-Scales: Illustration With the Intraindividual Study of Affect, Health, and Interpersonal Behavior (iSAHIB). *Research in Human Development, 11*(2), 142–160. <https://doi.org/10.1080/15427609.2014.906739>
- Reich, J. W., Zautra, A. J., & Davis, M. (2003). Dimensions of Affect Relationships: Models and Their Integrative Implications. *Review of General Psychology, 7*(1), 66–83. <https://doi.org/10.1037/1089-2680.7.1.66>
- Riediger, M., & Rauters, A. (2018). Experience Sampling in Lifespan Developmental Methodology. In M. Riediger & A. Rauters, *Oxford Research Encyclopedia of Psychology*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190236557.013.346>
- Riediger, M., Wrzus, C., & Wagner, G. G. (2014). Happiness is pleasant, or is it? Implicit representations of affect valence are associated with contrahedonic motivation and mixed affect in daily life. *Emotion, 14*(5), 950–961. <https://doi.org/10.1037/a0037711>
- Röcke, C., & Brose, A. (2013). Intraindividual Variability and Stability of Affect and Well-Being: Short-Term and Long-Term Change and Stabilization Processes. *GeroPsych, 26*(3), 185–199. <https://doi.org/10.1024/1662-9647/a000094>
- Roseman, I. J. (1996). Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. *Cognition & Emotion, 10*(3), 241–278. <https://doi.org/10.1080/026999396380240>
- Rosenberg, E. L. (1998). Levels of Analysis and the Organization of Affect. *Review of General Psychology, 2*(3), 247–270. <https://doi.org/10.1037/1089-2680.2.3.247>
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172. <https://doi.org/10.1037/0033-295X.110.1.145>

- Ryan, O., Dablander, F., & Haslbeck, J. M. B. (2025). Toward a generative model for emotion dynamics. *Psychological Review*, *132*(2), 416–441.  
<https://doi.org/10.1037/rev0000513>
- Säilynoja, T., Bürkner, P.-C., & Vehtari, A. (2022). Graphical test for discrete uniformity and its applications in goodness-of-fit evaluation and multiple sample comparison. *Statistics and Computing*, *32*(2), 32. <https://doi.org/10.1007/s11222-022-10090-6>
- Scherer, K. R. (1993). Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition & Emotion*, *7*(3–4), 325–355.  
<https://doi.org/10.1080/02699939308409192>
- Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion*, *23*(7), 1307–1351.  
<https://doi.org/10.1080/02699930902928969>
- Schiller, D., Yu, A. N. C., Alia-Klein, N., Becker, S., Cromwell, H. C., Dolcos, F., Eslinger, P. J., Frewen, P., Kemp, A. H., Pace-Schott, E. F., Raber, J., Siltan, R. L., Stefanova, E., Williams, J. H. G., Abe, N., Aghajani, M., Albrecht, F., Alexander, R., Anders, S., ... Lowe, L. (2024). The Human Affectome. *Neuroscience & Biobehavioral Reviews*, *158*, 105450. <https://doi.org/10.1016/j.neubiorev.2023.105450>
- Schimmack, U. (2003). Affect Measurement in Experience Sampling Research. *Journal of Happiness Studies*, *4*(1), 79–106. <https://doi.org/10.1023/A:1023661322862>
- Schmitt, M., Pratz, V., Köthe, U., Bürkner, P.-C., & Radev, S. T. (2024). Consistency Models for Scalable and Fast Simulation-Based Inference. *Advances in Neural Information Processing Systems*. 38th Conference on Neural Information Processing Systems.  
<https://doi.org/10.52202/079017-4030>
- Schmukle, S. C., & Egloff, B. (2009). Exploring bipolarity of affect ratings by using polychoric correlations. *Cognition & Emotion*, *23*(2), 272–295.  
<https://doi.org/10.1080/02699930801987330>

- Schneider, S., & Stone, A. A. (2015). Mixed emotions across the adult life span in the United States. *Psychology and Aging, 30*(2), 369–382. <https://doi.org/10.1037/pag0000018>
- Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in  $n = 1$  psychological autoregressive modeling. *Frontiers in Psychology, 6*. <https://doi.org/10.3389/fpsyg.2015.01038>
- Scott, S., Ram, N., Smyth, J. M., Almeida, D. M., & Sliwinski, M. J. (2017). Age differences in negative emotional responses to daily stressors depend on time since event. *Developmental Psychology, 53*(1), 177–190. <https://doi.org/10.1037/dev0000257>
- Scott, S., 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>
- Segura, S. L., & González-Romá, V. (2003). How Do Respondents Construe Ambiguous Response Formats of Affect Items? *Journal of Personality and Social Psychology, 85*(5), 956–968. <https://doi.org/10.1037/0022-3514.85.5.956>
- Seizer, L., Schiepek, G., Cornelissen, G., & Löchner, J. (2024). A primer on sampling rates of ambulatory assessments. *Psychological Methods*. <https://doi.org/10.1037/met0000656>
- Setodji, C. M., Martino, S. C., Dunbar, M. S., & Shadel, W. G. (2019). An exponential effect persistence model for intensive longitudinal data. *Psychological Methods, 24*(5), 622–636. <https://doi.org/10.1037/met0000211>
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*(4), 813–838. <https://doi.org/10.1037/0022-3514.48.4.813>
- Smyth, J. M., Sliwinski, M. J., Zawadzki, M. J., Scott, S., Conroy, D. E., Lanza, S. T., Marcusson-Clavertz, D., Kim, J., Stawski, R. S., Stoney, C. M., Buxton, O. M., Sciamanna, C. N., Green, P. M., & Almeida, D. M. (2018). Everyday stress response

- targets in the science of behavior change. *Behaviour Research and Therapy*, *101*, 20–29. <https://doi.org/10.1016/j.brat.2017.09.009>
- Stan Development Team. (2025). *RStan: The R interface to Stan*. [Computer software]. <https://mc-stan.org/>
- Tellegen, A., Watson, D., & Clark, L. A. (1999). On the Dimensional and Hierarchical Structure of Affect. *Psychological Science*, *10*(4), 297–303. <https://doi.org/10.1111/1467-9280.00157>
- Thagard, P., & Nerb, J. (2002). Emotional Gestalts: Appraisal, Change, and the Dynamics of Affect. *Personality and Social Psychology Review*, *6*(4), 274–282. [https://doi.org/10.1207/S15327957PSPR0604\\_02](https://doi.org/10.1207/S15327957PSPR0604_02)
- Tong, E. M. W., Ellsworth, P. C., & Bishop, G. D. (2009). An S-shaped relationship between changes in appraisals and changes in emotions. *Emotion*, *9*(6), 821–837. <https://doi.org/10.1037/a0017812>
- Tongco-Rosario, G., Sio, C., & Soriano, J. (2024). Investigations on Anger Experience with Other Basic Emotions Using Affective Ising Model: *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies*, 738–745. <https://doi.org/10.5220/0012470300003657>
- Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective Dynamics in Psychopathology. *Emotion Review*, *7*(4), 355–361. <https://doi.org/10.1177/1754073915590617>
- Tuerlinckx, F., Kuppens, P., Ariens, S., Cloos, L., Dejonckheere, E., Lafit, G., Niemeijer, K., Revol, J., Schat, E., Schreuder, M., Vanhasbroeck, N., & Ceulemans, E. (2026). New developments in experience sampling methodology. *British Journal of Mathematical and Statistical Psychology*, *79*, 46–65. <https://doi.org/10.1111/bmsp.12398>
- Vanhasbroeck, N., Ariens, S., Tuerlinckx, F., & Loossens, T. (2021). Computational Models for Affect Dynamics. In C. E. Waugh & P. Kuppens (Eds.), *Affect Dynamics* (pp. 213–

- 260). Springer International Publishing. [https://doi.org/10.1007/978-3-030-82965-0\\_10](https://doi.org/10.1007/978-3-030-82965-0_10)
- Vanhasbroeck, N., Loossens, T., Anarat, N., Ariens, S., Vanpaemel, W., Moors, A., & Tuerlinckx, F. (2022). Stimulus-Driven Affective Change: Evaluating Computational Models of Affect Dynamics in Conjunction with Input. *Affective Science*. <https://doi.org/10.1007/s42761-022-00118-5>
- Vanhasbroeck, N., Niemeijer, K., & Tuerlinckx, F. (2024). Nonlinearity in affect dynamics persists after accounting for the valence of daily-life events. *Emotion*, *24*(5), 1206–1223. <https://doi.org/10.1037/emo0001336>
- Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Van Mechelen, I. (2009). Predicting the duration of emotional experience: Two experience sampling studies. *Emotion*, *9*(1), 83–91. <https://doi.org/10.1037/a0014610>
- Verduyn, P., Van Mechelen, I., Kross, E., Chezzi, C., & Van Bever, F. (2012). The relationship between self-distancing and the duration of negative and positive emotional experiences in daily life. *Emotion*, *12*(6), 1248–1263. <https://doi.org/10.1037/a0028289>
- Vogelsmeier, L. V. D. E., Cloos, L., Kuppens, P., & Ceulemans, E. (2024). Evaluating dynamics in affect structure with latent Markov factor analysis. *Emotion*, *24*(3), 782–794. <https://doi.org/10.1037/emo0001307>
- Volokhov, R. N., & Demaree, H. A. (2010). Spontaneous emotion regulation to positive and negative stimuli. *Brain and Cognition*, *73*(1), 1–6. <https://doi.org/10.1016/j.bandc.2009.10.015>
- Watson, D., & Stanton, K. (2017). Emotion Blends and Mixed Emotions in the Hierarchical Structure of Affect. *Emotion Review*, *9*(2), 99–104. <https://doi.org/10.1177/1754073916639659>

- Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*(5), 820–838.  
<https://doi.org/10.1037/0022-3514.76.5.820>
- Wenzel, M., Rowland, Z., & Kubiak, T. (2022). How much variance can event intensity and emotion regulation strategies explain in momentary affect in daily life? *Emotion, 22*(8), 1969–1979. <https://doi.org/10.1037/emo0000816>
- Wirth, M., Voss, A., & Rothermund, K. (2023). Age Differences in Everyday Emotional Experience: Testing Core Predictions of Socioemotional Selectivity Theory With the MIVA Model. *The Journals of Gerontology: Series B, 78*(7), 1152–1162.  
<https://doi.org/10.1093/geronb/gbad033>
- Wirth, M., Voss, A., Wirth, S., & Rothermund, K. (2022). Affect dynamics and well-being: Explanatory power of the model of intraindividual variability in affect (MIVA). *Cognition and Emotion, 36*(2), 188–210.  
<https://doi.org/10.1080/02699931.2021.1993148>
- Wood, J., Oravecz, Z., Vogel, N., Benson, L., Chow, S.-M., Cole, P., Conroy, D. E., Pincus, A. L., & Ram, N. (2018). Modeling Intraindividual Dynamics Using Stochastic Differential Equations: Age Differences in Affect Regulation. *The Journals of Gerontology: Series B, 73*(1), 171–184. <https://doi.org/10.1093/geronb/gbx013>
- Wrzus, C., Luong, G., Wagner, G. G., & Riediger, M. (2015). Can't get it out of my head: Age differences in affective responsiveness vary with preoccupation and elapsed time after daily hassles. *Emotion, 15*(2), 257–269. <https://doi.org/10.1037/emo0000019>
- Wrzus, C., & Neubauer, A. B. (2023). Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. *Assessment, 30*(3), 825–846. <https://doi.org/10.1177/10731911211067538>

- Wu, Y., Radev, S. T., & Tuerlinckx, F. (2026). Testing and improving the robustness of amortized bayesian inference for cognitive models. *Psychological Methods*.  
<https://doi.org/10.1037/met0000814>
- Zammit-Mangion, A., Sainsbury-Dale, M., & Huse, R. (2023). Neural Methods for Amortized Inference. *Annual Review of Statistics and Its Application*, 12, 311–335.
- Zavlis, O., Bentall, R. P., Fonagy, P., & Rigoli, F. (2025). A Formal Theory of Mood Instability. *Clinical Psychological Science*, 21677026251363862.  
<https://doi.org/10.1177/21677026251363862>
- Zawadzki, M. J., Scott, S., Almeida, D. M., Lanza, S. T., Conroy, D. E., Sliwinski, M. J., Kim, J., Marcusson-Clavertz, D., Stawski, R. S., Green, P. M., Sciamanna, C. N., Johnson, J. A., & Smyth, J. M. (2019). Understanding stress reports in daily life: A coordinated analysis of factors associated with the frequency of reporting stress. *Journal of Behavioral Medicine*, 42(3), 545–560. <https://doi.org/10.1007/s10865-018-00008-x>