# Supplementary Information

**State Self-Compassion Dynamics: Partial Evidence for the Bipolar Continuum Hypothesis**

## Table of contents

# 1 Study 1

## 1.1 Baseline Measures

To exclude the possibility of emotional disorders or psychological conditions, we administered a series of validated psychological assessment questionnaires.

*Depression Anxiety Stress Scale-21* (DASS-21). The DASS-21 (Lovibond and Lovibond 1995) is a 21-item self-report questionnaire to assess depression (*i.e.*, "I felt down-hearted and blue"), anxiety (*i.e.*, "I felt I was close to panic") and stress (*i.e.*, "I found it difficult to relax") symptoms, over the past week. Items are rated on a 4-point Likert scale ranging from 0 (*It never happened to me*) to 3 (*It happened to me almost always*). Both the original and its Italian version (Bottesi et al. 2015) demonstrate adequate reliability.

*Rosenberg Self - Esteem Scale* (RSES). The RSES (Rosenberg 1965) assesses individual self-esteem levels through a 10-item scale (*e.g.,* "I feel that I'm a person of worth, at least on an equal basis with others"). Respondents rate each statement on a 4-point Likert scale, ranging from 4 (*Strongly Agree*) to 1 (*Strongly Disagree*). Higher total scores indicate higher self-esteem.

*Self-Compassion Scale* (SCS). The SCS (Neff 2003), is a self-report questionnaire comprising 26 items designed to assess individuals' enduring Self-Compassion traits in their daily lives. The SCS encompasses six subscales: Self-Kindness, Common Humanity, Mindfulness, Self-Judgment, Isolation, and Over-Identification. Respondents were instructed to rate the frequency of their Self-Compassionate attitudes using a 5-point scale, ranging from 1 (*Almost Never*) to 5 (*Almost Always*). To ensure consistency in scoring, negative items were reverse coded, with higher scores indicating a greater absence of negative Self-Compassion traits. The psychometric properties of the SCS have been found to be robust. For the total score, the Cronbach's alpha coefficient was reported as 0.96 in the original work by Neff (2003). Additionally, the test-retest reliability demonstrated adequate results, with a correlation coefficient ($r$) of 0.93 for the total score and ranging from 0.80 to 0.88 for the subscales.

*Difficulties in Emotion Regulation Scale* (DERS). The DERS is a 36-item self-report measure developed to assess the complexities in emotion regulation processes among individuals (Gratz and Roemer 2004). This scale is divided into six subscales, each targeting a specific dimension of emotion regulation difficulties: Nonacceptance of Emotional Responses (NER), Difficulties Engaging in Goal-Directed Behavior (DEGB), Impulse Control Difficulties (ICD), Lack of Emotional Awareness (LEA), Limited Access to Emotion Regulation Strategies (LAERS), and Lack of Emotional Clarity (LEC). Respondents are asked to rate each item on a 5-point Likert scale that ranges from 0 (*Never*) to 4 (*Always*), reflecting the frequency with which they experience each emotion regulation difficulty. The higher the score, the greater the difficulties in emotion regulation an individual is likely to have. Both the original version and its Italian adaptation (Sighinolfi et al. 2010), have demonstrated strong psychometric properties, including adequate reliability and validity.

The following table presents the descriptive statistics for all the aforementioned measures. It includes the estimated posterior mean (Estimate), Standard Error, and the 95% credibility interval, computed using a Bayesian model. Bayesian modeling was employed to account for deviations from Gaussianity.

| Variable | Estimate | Std. Error | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|
| $DASS-21_{Stress}$ | 7.58 | 0.45 | 6.92 | 8.14 |
| $DASS-21_{Anxiety}$ | 1.04 | 0.19 | 0.89 | 1.93 |
| $DASS-21_{Depression}$ | 2.97 | 0.29 | 2.02 | 3.63 |
| $RSES$ | 28.05 | 0.26 | 27.53 | 28.55 |
| $SCS_{total-score}$ | 17.45 | 0.21 | 17.05 | 17.87 |
| $SCS_{Self-Kindness}$ | 2.98 | 0.04 | 2.89 | 3.06 |
| $SCS_{Common-Humanity}$ | 3.11 | 0.04 | 3.02 | 3.20 |
| $SCS_{Mindfulness}$ | 3.11 | 0.04 | 3.03 | 3.20 |
| $SCS_{Self-Judgment}$ | 2.74 | 0.04 | 2.65 | 2.82 |
| $SCS_{Isolation}$ | 2.74 | 0.05 | 2.65 | 2.84 |
| $SCS_{Over-Identification}$ | 2.78 | 0.05 | 2.68 | 2.88 |
| $DERS_{totalscore}$ | 67.89 | 2.07 | 63.79 | 71.64 |
| $DERS_{NER}$ | 6.23 | 0.43 | 5.95 | 7.02 |
| $DERS_{DEGB}$ | 13.41 | 0.78 | 12.46 | 15.78 |
| $DERS_{ICD}$ | 4.99 | 0.02 | 4.95 | 5.00 |
| $DERS_{LEA}$ | 11.52 | 0.82 | 9.91 | 12.98 |
| $DERS_{LAERS}$ | 13.01 | 0.67 | 11.52 | 14.11 |
| $DERS_{LEC}$ | 9.91 | 0.65 | 8.36 | 11.14 |

There is no evidence of emotional disorders among participants. The obtained scores are consistent with those reported in other studies using the same measures within community samples (Bottesi et al. 2015; Sighinolfi et al. 2010; Neff, Whittaker, and Karl 2017; Sica et al. 2021).

## 1.2 EMA Survey Questions

For each notification in the Ecological Momentary Assessment (EMA) protocol, participants were prompted to answer the following questions.

1. Think about the most notable event that has occurred since you last received a notification. If this is your first notification of the day, consider the most significant event from the start of the day. How would you evaluate this event?
   1) Very unpleasant
   2) Unpleasant
   3) Neither unpleasant nor pleasant
   4) Pleasant
   5) Very pleasant
2. At this moment I feel NERVOUS.
   1) Not at all
   2) A little
   3) Moderately
   4) Quite a bit
   5) Very much
3. At this moment I feel UPSET.
   1) Not at all
   2) A little
   3) Moderately
   4) Quite a bit
   5) Very much
4. At this moment I feel SATISFIED.
   1) Not at all
   2) A little
   3) Moderately
   4) Quite a bit
   5) Very much
5. At this moment I feel CHEERFUL.
   1) Not at all
   2) A little
   3) Moderately
   4) Quite a bit
   5) Very much
6. I'm giving myself the caring and tenderness I need.
   1) Completely false
   2) Moderately false
   3) Slightly false
   4) Slightly true

5) Moderately true
6) Completely true

7. I'm obsessing and fixating on everything that's wrong.

   1) Completely false
   2) Moderately false
   3) Slightly false
   4) Slightly true
   5) Moderately true
   6) Completely true

8. I'm remembering that there are lots of others in the world feeling like I am.

   1) Completely false
   2) Moderately false
   3) Slightly false
   4) Slightly true
   5) Moderately true
   6) Completely true

9. I feel like I'm struggling more than others right now.

   1) Completely false
   2) Moderately false
   3) Slightly false
   4) Slightly true
   5) Moderately true
   6) Completely true

10. I feel intolerant and impatient toward myself.

    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true
    5) Moderately true
    6) Completely true

11. I'm keeping things in perspective.

    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true
    5) Moderately true
    6) Completely true

12. At this moment I am able to accept my flaws and weaknesses.

    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true

5) Moderately true
6) Completely true

13. At this moment I let myself be carried away by my emotions.

    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true
    5) Moderately true
    6) Completely true

## 1.3 Data Quality Management and Subject Selection

To ensure data quality, we employed a comprehensive set of behavioral indicators to detect inattentive or insufficient-effort responding, including the compliance rate and metrics designed to identify aberrant response patterns (Ulitzsch et al. 2024; Hasselhorn, Ottenstein, and Lischetzke 2023). These indicators evaluated both response patterns and supplementary data (e.g., timing metrics, engagement levels), flagging anomalies that might suggest inattentive or low-effort responding (Meade and Craig 2012). Common anomalies included response invariability, inconsistencies across responses, and multivariate outliers.

Notably, our data-quality assessment was conducted post hoc, rather than by embedding directed prompts (e.g., instructed-response items) in the survey. Although such items can effectively detect careless responding, they present several drawbacks (Ulitzsch et al. 2024). First, adding items to detect inattentiveness can lengthen the questionnaire, paradoxically increasing the chance of respondent fatigue—particularly in studies using Ecological Momentary Assessment (EMA). Second, in repeated-measures contexts, these items may confuse participants who must maintain clarity about task expectations across multiple assessment occasions. By employing a post-hoc analysis using behavioral indicators, we minimized participant burden and preserved the integrity of the EMA design.

Compliance rate, defined as the ratio of completed assessments to the total number of expected assessments, served as a key criterion for gauging insufficient engagement. Participants failing to meet a prespecified compliance threshold were excluded from analyses to maintain the reliability of the dataset.

### 1.3.1 Compliance Rate

Non-compliance introduces systematic missing data, which can bias statistical inference if the data are not missing at random. To mitigate this risk, participants with a compliance rate below 50% were excluded. Based on this criterion, seven participants were removed from the dataset.

### 1.3.2 Time to Complete

The **Time to Complete** (TTC) metric was calculated from the moment participants opened an EMA survey to the time they submitted it. Due to the right-skewed nature of survey completion times, we calculated each participant's median TTC using log-transformed values. We then applied the lower fence of the interquartile range (IQR) rule to these log-transformed medians to identify unusually short completion times. No participant fell below this lower-fence threshold, indicating no evidence of careless responding based on survey duration.

### 1.3.3 Careless responding on the State Self-Compassion Scale

We assessed careless responding on the State Self-Compassion Scale items by using four metrics provided by the **careless** R package: the Longstring Index, Intra-Individual Response Variability (IRV), the Even-Odd Inconsistency Index, and Mahalanobis Distance. These metrics, designed to evaluate response patterns, offer a robust framework for identifying inattentive or inconsistent behavior.

- **Longstring Index**: Measures the longest sequence of identical responses within a survey. High values suggest "straightlining," a common marker of inattentiveness.
- **Intra-Individual Response Variability (IRV)**: Reflects the variability in responses across consecutive items. Low IRV may indicate lower engagement or attention.
- **Even-Odd Inconsistency Index**: Assesses internal consistency by correlating mean scores of even- and odd-numbered items. Lower correlations suggest greater inconsistency.
- **Mahalanobis Distance ($D^2$)**: Identifies multivariate outliers by quantifying deviations from an expected response pattern. High $D^2$ values indicate unusual response profiles.

We performed these analyses at two levels: person and occasion.

#### 1.3.3.1 Person-Level Analysis of Behavioral Indices

Participants were flagged as potential careless responders if their index scores exceeded the 95th percentile on any of the four metrics. Participants flagged on more than two indices were classified as careless responders.

The following R script illustrates how we counted participants who exceeded thresholds for various combinations of indices:

```r
vectors <- list(mahad_bad, longstring_bad, irv_bad, even_odd_bad)

# Count shared elements across index combinations
shared_counts <- map(2:4, ~ {
  combos <- combn(vectors, .x, simplify = FALSE)
  shared <- map(combos, ~ Reduce(intersect, .x)) %>% unlist() %>% unique()
  length(shared)
}) %>% set_names(paste0("shared_by_", 2:4))

# Results
shared_counts
# $shared_by_2
# [1] 3
#
# $shared_by_3
# [1] 0
#
# $shared_by_4
# [1] 0
```

No participant exceeded thresholds on more than two indices during the EMA phase, so no exclusions were made based on these metrics.

#### 1.3.3.2 Occasion-Level Analysis of Behavioral Indices

To complement the participant-level analysis, we conducted an occasion-level analysis to capture instances of momentary inattentiveness. This approach allowed us to detect subtle, transient lapses in engagement that might otherwise remain undetected in participant-level summaries.

By evaluating responses at the occasion level:

- momentary lapses in engagement could be flagged without excluding an entire participant;
- the analysis gained granularity, enabling targeted handling of problematic data points;
- the design aligns with recent work emphasizing the importance of evaluating variability over time in EMA research (Hasselhorn, Ottenstein, and Lischetzke 2023).

We calculated the four careless-responding indices (Longstring, IRV, Even-Odd Inconsistency, and Mahalanobis Distance) separately for each occasion on the State Self-Compassion Scale (SCS). We then flagged occasions exceeding **adjusted fences** derived from bootstrapped IQRs. Finally, we aggregated flagged occasions to compute a proportion of flagged data points per participant.

##### 1.3.3.2.1 Results

- **Distribution of Flagged Occasions**:
  - **80.4%** of occasions were never flagged by any metric.
  - **17.7%** of occasions were flagged on one metric, indicating mild evidence of momentary inattention.
  - **1.77%** of occasions were flagged on two metrics, suggesting stronger evidence of momentary lapses.
  - **<0.2%** of occasions were flagged on three or more metrics, representing very rare instances of substantial carelessness.
- **Proportion of Flagged Occasions per Participant**:
  - the median proportion of flagged occasions per participant was **<5%**, reflecting low overall incidence of momentary inattentiveness;
  - no participant showed persistently high flagging rates across occasions, underscoring the reliability of the dataset.

#### 1.3.4 Final Assessment

Both the person-level and occasion-level analyses confirmed a low incidence of inattentive or careless responding. These findings attest to the dataset's robustness and highlight the strengths of this study's design.

Compared to other study designs—particularly those involving paid participants, such as Amazon Mechanical Turk—our sample exhibited notably fewer inattentive responses. This discrepancy is likely attributable to varying levels of intrinsic motivation (Aruguete et al. 2019). For instance, participants on Amazon Mechanical Turk may exhibit lower intrinsic motivation due to minimal monetary compensation, while our volunteers, who committed to completing weekly surveys over a two-month period, were likely driven by higher intrinsic motivation.

Recent research supports these observations in the context of EMA studies. For example, Hasselhorn, Ottenstein, and Lischetzke (2023) used multilevel latent class analysis (ML-LCA) with volunteer participants to identify profiles of momentary careless responding at the occasion level and latent classes of individuals who varied in their distribution of careless responses across occasions. They identified four latent classes: "careful," "frequently careless," and two categories of "infrequently careless" respondents. Notably, the "frequently careless" class accounted for only 2% of participants, even in a more intensive EMA design involving multiple daily notifications over seven consecutive days.

Study design plays a critical role in influencing participant compliance and data quality. Intensive EMA designs, characterized by frequent and intrusive assessments, are associated with participant fatigue, reduced response accuracy, and increased inattentiveness (Shiffman, Stone, and Hufford 2008). In contrast, our study employed a less intensive approach, with participants receiving only one notification per week rather than multiple daily prompts.

Other additional factors likely contributed to the minimization of inattentive responses in our study:

- **Brevity of EMA questionnaires:** Participants faced minimal cognitive burden, encouraging thoughtful responses.
- **Flexibility to discontinue participation:** Participants retained the option to withdraw at any time, reducing potential disengagement due to fatigue or other external pressures.

These features collectively reduced the likelihood of inattentive responses compared to traditional cross-sectional surveys (Welling, Fischer, and Schinkel-Bielefeld 2021) and more intensive EMA studies (Hasselhorn, Ottenstein, and Lischetzke 2023).

#### 1.3.5 Data Reanalysis

To evaluate whether excluding flagged occasions—defined as those identified on $geq$2 metrics—would affect the primary outcomes, we conducted a secondary analysis of the Study 1 data. This analysis focused on the nomothetic model assessing the relationships between the Uncompassionate Self (UCS) and Compassionate Self (CS) components of state self-compassion, after controlling for Negative Affect and Context Evaluation. Results from the full dataset are provided in Figure 1.

The reanalysis produced results that were effectively identical to those derived from the full dataset, underscoring the robustness of the findings. Hence, we retained all participants who met the compliance requirements in our final analyses. Below is a table summarizing the posterior estimates from the reanalysis, alongside their standard deviations and 89% credibility intervals:

| Variable | Mean | SD | Q5.5 | Q94.5 |
|---|---|---|---|---|
| alpha_ucs | -0.0005 | 0.0403 | -0.0644 | 0.0643 |

| Variable | Mean | SD | Q5.5 | Q94.5 |
|---|---|---|---|---|
| beta_cs | -0.4470 | 0.0183 | -0.4770 | -0.4180 |
| beta_negative_affect | 0.1160 | 0.0059 | 0.1070 | 0.1250 |
| beta_context_valence | 0.0101 | 0.0055 | 0.0013 | 0.0189 |

These results confirm the reliability of the final dataset and the validity of our conclusions, demonstrating a high level of data quality in this long-term EMA study design.

## 1.4 Model Convergence Diagnostics

To ensure that the model parameters accurately reflected participants' behavior in our task, we conducted posterior predictive checks, which were adequate for all examined models.

To confirm satisfactory convergence of the Markov chain Monte Carlo (MCMC) procedure for parameter estimation, we employed the convergence statistic $\hat{R}$ and monitored the number of divergent transitions during sampling. We verified that all population-level parameters had $\hat{R} < 1.01$. Additionally, posterior sampling of the models did not result in any divergent transitions.

## 1.5 Multilevel Correlations between CS and UCS components of State Self-Compassion

We conducted a multilevel analysis to compute the correlation between the Compassionate Self (CS) and Uncompassionate Self (UCS) components of state self-compassion, and each of the six subscales of the Self-Compassion Scale (Neff, 2003).

Multilevel correlations provide a more accurate estimate of the relationships between variables in hierarchical or clustered data compared to simple bivariate correlations. In our study, participants were assessed multiple times across different moments and days, creating a nested data structure where measurements are clustered within individuals. A multilevel approach accounts for this clustering by separating within-person variation from between-person variation, which leads to more precise estimates of the associations. In contrast, a simple bivariate correlation would not distinguish between these sources of variation, potentially conflating individual-level effects with population-level patterns and resulting in biased estimates.

By using multilevel correlations, we estimated an average effect across participants, adjusting for the random intercepts attributable to individual differences in the hierarchical model. For both components of state self-compassion, the correlations with the subscales of the Self-Compassion Scale were modest. For CS, the largest multilevel correlation was with self-kindness, $r = 0.24$, 89% CI [0.15, 0.32]; for UCS, the largest multilevel correlation was with isolation, $r = 0.21$, 89% CI [0.10, 0.31].

### 1.5.1 CS Component of State Self-Compassion

```r
fit_cs <- brm(
  state_cs ~ self_kindness + common_humanity + mindfulness +
    self_judgment + isolation + over_identification +
    (1 | user_id),
  data = d,
  family = student(),
  backend = "cmdstanr",
  iter = 10000,
  chains = 4
)
```

| Variable | Estimate | Std. Error | 89% CI Lower | 89% CI Upper |
|---|---|---|---|---|
| Intercept | 0.06 | 0.04 | 0.00 | 0.12 |
| Self-Kindness | 0.23 | 0.06 | 0.13 | 0.32 |
| Common Humanity | 0.18 | 0.05 | 0.10 | 0.26 |
| Mindfulness | 0.06 | 0.05 | -0.03 | 0.14 |
| Self-Judgment | -0.04 | 0.05 | -0.13 | 0.04 |
| Isolation | -0.11 | 0.05 | -0.19 | -0.02 |
| Over-Identification | -0.00 | 0.06 | -0.10 | 0.09 |

### 1.5.2 UCS Component of State Self-Compassion

```
fit_ucs <- brm(
  state_ucs ~ self_kindness + common_humanity + mindfulness +
    self_judgment + isolation + over_identification +
    (1 | user_id),
  data = d,
  family = student(),
  backend = "cmdstanr",
  iter = 10000,
  chains = 4
)
```

| Variable | Estimate | Std. Error | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|
| Intercept | -0.02 | 0.04 | -0.10 | 0.06 |
| Self-Kindness | -0.14 | 0.06 | -0.24 | -0.02 |
| Common Humanity | -0.03 | 0.05 | -0.14 | 0.07 |
| Mindfulness | -0.01 | 0.06 | -0.13 | 0.10 |
| Self-Judgment | 0.10 | 0.06 | -0.01 | 0.23 |
| Isolation | 0.21 | 0.05 | 0.10 | 0.31 |
| Over-Identification | 0.12 | 0.06 | 0.00 | 0.25 |

## 1.6 The Impact of Negative Affect and Event Unpleasantness on State Self-Compassion

In this section, we describe the statistical analysis used to assess the impact of contextual variables on state self-compassion. Our analysis specifically focuses on three key areas: differences across individuals, variations between days within the same individual, and fluctuations within a single day for each individual.

Prior to implementing the final Bayesian hierarchical models, we performed a model selection process for determining the most fitting structures for both random and fixed effects within our dataset. In the final models, we examined the CS and UCS aspects of SSC as a function of six predictors: negative affect and context evaluation. Each of these predictors was uniquely centered to distinctly capture and differentiate the three dimensions of variance we were interested in – namely, inter-individual differences, between-day variations within individuals, and within-day fluctuations for each individual.

We started by exploring the full fixed-effect structure and proceeded with a systematic comparison of models featuring varying degrees of random-effect complexity. Once we determined the optimal random-effect structure, we turned our attention to assessing models with different fixed-effects configurations. To facilitate model comparison, we used the Leave-One-Out (LOO) method, a robust Bayesian model selection technique implemented within Stan.

This technique evaluates out-of-sample prediction accuracy by sequentially excluding individual observations from the dataset and assessing the model's performance on these excluded points. Models demonstrating lower LOO values were interpreted as having superior fit and enhanced predictive accuracy. In our modeling process, we integrated regularizing priors, which served to mitigate overfitting by applying constraints that direct the model towards more plausible outcomes. Additionally, we employed partial pooling to boost the accuracy of estimations across various groups. We fitted the models using the cmdstan interface and with the brms package, which leverages the computational power of Stan for Bayesian inference.

### 1.6.1 CS Component

#### 1.6.1.1 Random effects

| Model | elpd_diff | se_diff | elpd_loo | se_elpd_loo | p_loo | se_p_loo | looic | se_looic |
|---|---|---|---|---|---|---|---|---|
| Model 1: Basic Model | 0.00 | 0.00 | -8383.79 | 114.38 | 2760.33 | 23.56 | 16767.57 | 228.77 |
| Model 2: Add Random Effect user_id | -281.13 | 55.04 | -8664.92 | 123.20 | 24642.01 | 87.90 | 17329.84 | 246.39 |
| Model 3: Add Random Effects for user_id and user_id:day | -551.34 | 45.91 | -8935.13 | 110.27 | 2312.90 | 19.53 | 17870.25 | 220.54 |
| Model 4: Add Random Slopes for na_moment, na_day on user_id | -1666.63 | 71.51 | -10050.41 | 105.37 | 369.61 | 3.49 | 20100.83 | 210.75 |
| Model 5: Complex Random Effects Structure | -6024.44 | 114.52 | -14408.22 | 85.91 | 9.62 | 0.16 | 28816.44 | 171.81 |

The model comparison, utilizing the LOO method, indicates that there is no valid justification for employing a random-effect structure more complex than participant-level clustering. In other words, the simpler approach of clustering at the participant level provides an adequate representation for our data, as evidenced by the minimal improvements gained from more intricate random-effect structures.

#### 1.6.1.2 Fixed-Effects

| Model | elpd_diff | se_diff | elpd_loo | se_elpd_loo | p_loo | se_p_loo | looic | se_looic |
|---|---|---|---|---|---|---|---|---|
| Model 1: Full Fixed Effects | 0.00 | 0.00 | -10050.04 | 105.34 | 368.99 | 3.49 | 20100.07 | 210.69 |
| Model 2: Only na | -21.57 | 7.81 | -10071.61 | 105.26 | 366.92 | 3.48 | 20143.21 | 210.52 |
| Model 3: Only con | -1580.99 | 64.59 | -11631.03 | 110.22 | 378.08 | 3.39 | 23262.05 | 220.43 |

Based on the model comparison using the LOO method, the best-fitting model is "Model 1: Full Fixed Effects." It exhibits the highest estimated log pointwise predictive density (`elpd_loo`) and the lowest Leave-One-Out Information Criterion (`looic`), suggesting superior predictive performance compared to the other models.

The final model for predicting the compassionate responding component of state self-compassion is as follows:

$$
\begin{aligned}
sc &\sim \text{Student-t}(\mu, \sigma, \nu) \\
\mu &= \beta_0 + \\
&\quad \beta_{\text{na\_moment}} \times \text{na\_moment} + \beta_{\text{na\_day}} \times \text{na\_day} + \beta_{\text{na\_person}} \times \text{na\_person} + \\
&\quad \beta_{\text{context\_moment}} \times \text{context\_moment} + \beta_{\text{context\_day}} \times \text{context\_day} + \\
&\quad \beta_{\text{context\_person}} \times \text{context\_person} + \\
&\quad b_{\text{user\_id}}[j] + b_{\text{bysubj\_day}}[k] \\
b_{\text{user\_id}}[j] &\sim \mathcal{N}(0, \Sigma_{\text{user\_id}}) \\
b_{\text{bysubj\_day}}[k] &\sim \mathcal{N}(0, \sigma^2_{\text{bysubj\_day}}) \\
\beta &\sim \text{priors1} \\
\sigma &\sim \text{Half-Cauchy}(0, \text{scale}) \\
\nu &\sim \text{Exponential}(\text{rate})
\end{aligned}
$$

$$
\begin{aligned}
&\text{Where:} \\
\Sigma_{\text{user\_id}} &= \text{Full covariance matrix for random effects within user\_id} \\
\sigma^2_{\text{bysubj\_day}} &= \text{Var}(b_{\text{bysubj\_day}}[k])
\end{aligned}
$$

### 1.6.2 USC Component

In parallel with our analysis of the SC component, we conducted a model comparison for the USC component.

#### 1.6.2.1 Random-Effects

| Model | elpd_diff | se_diff | elpd_loo | se_elpd_loo | p_loo | se_p_loo | looic | se_looic |
|---|---|---|---|---|---|---|---|---|
| Model 1: Basic Model | 0.00 | 0.00 | -8325.73 | 103.75 | 2507.22 | 22.25 | 16651.46 | 207.50 |
| Model 2: Add Random Effect user_id | -393.72 | 36.24 | -8719.45 | 101.44 | 2181.38 | 18.63 | 17438.91 | 202.89 |
| Model 3: Add Random Effects for user_id and user_id:day | -539.99 | 93.24 | -8865.73 | 140.28 | 37464.65 | 116.55 | 17731.46 | 280.56 |
| Model 4: Add Random Slopes for na_moment, na_day on user_id | -1646.63 | 64.84 | -9972.36 | 96.36 | 345.11 | 3.36 | 19944.72 | 192.72 |
| Model 5: Complex Random Effects Structure | -5363.17 | 98.42 | -13688.90 | 77.90 | 8.61 | 0.15 | 27377.80 | 155.79 |

Our evaluation of random-effect structures using the LOO method yielded results consistent with those observed for the SC component. The model comparison for USC indicates that there is no compelling justification for employing a random-effect structure more complex than clustering at the participant level. This echoes the findings from the SC component analysis, where participant-level clustering proved sufficient to adequately represent our data.

#### 1.6.2.2 Fixed-Effects

| Model | elpd_diff | se_diff | elpd_loo | se_elpd_loo | p_loo | se_p_loo | looic | se_looic |
|---|---|---|---|---|---|---|---|---|
| Model 1: Full Fixed Effects | 0.00 | 0.00 | -9973.73 | 96.34 | 346.60 | 3.37 | 19947.46 | 192.67 |
| Model 2: Only na | -18.65 | 6.72 | -9992.38 | 96.39 | 346.37 | 3.40 | 19984.75 | 192.79 |
| Model 3: Only con | -1938.50 | 69.27 | -11912.23 | 101.58 | 363.87 | 3.46 | 23824.47 | 203.15 |

When considering fixed-effect structures for the USC component, our analysis identified ‘Model 1: Full Fixed Effects’ as the best-fitting model. This model exhibited the highest estimated log pointwise predictive density (`elpd_loo`) and the lowest Leave-One-Out Information Criterion (`looic`) among the options. These results closely mirror the findings from the SC component analysis, where ‘Model 1: Full Fixed Effects’ also emerged as the preferred model.

The congruence in results between the SC and USC components underscores the consistency and reliability of our modeling approach. For both SC and USC, we have selected ‘Model 1: Full Fixed Effects’ as the optimal model, demonstrating superior predictive performance compared to more complex alternatives.

The final model for predicting the uncompassionate responding component of state self-compassion is as follows:

$$
\begin{aligned}
usc &\sim \text{Student-t}(\mu, \sigma, \nu) \\
\mu &= \beta_0 + \\
&\quad \beta_{\text{na\_moment}} \times \text{na\_moment} + \beta_{\text{na\_day}} \times \text{na\_day} + \beta_{\text{na\_person}} \times \text{na\_person} + \\
&\quad \beta_{\text{context\_moment}} \times \text{context\_moment} + \beta_{\text{context\_day}} \times \text{context\_day} + \\
&\beta_{\text{context\_person}} \times \text{context\_person} + \\
&\quad b_{\text{user\_id}}[j] + b_{\text{bysubj\_day}}[k] \\
b_{\text{user\_id}}[j] &\sim \mathcal{N}(0, \Sigma_{\text{user\_id}}) \\
b_{\text{bysubj\_day}}[k] &\sim \mathcal{N}(0, \sigma^2_{\text{bysubj\_day}}) \\
\beta &\sim \text{priors1} \\
\sigma &\sim \text{Half-Cauchy}(0, \text{scale}) \\
\nu &\sim \text{Exponential}(\text{rate})
\end{aligned}
$$

$$
\begin{aligned}
&\text{Where:} \\
\Sigma_{\text{user\_id}} &= \text{Full covariance matrix for random effects within user\_id} \\
\sigma^2_{\text{bysubj\_day}} &= \text{Var}(b_{\text{bysubj\_day}}[k])
\end{aligned}
$$

The two models were estimated using a Student's t-distribution with identity links for the mean ($\mu$), scale ($\sigma$), and degrees of freedom ($\nu$). The analysis was based on 12621 observations, 326 participants, with the posterior distribution derived from 12000 post-warmup draws across four chains.

In both cases, the model diagnostics indicate satisfactory convergence with Rhat values close to 1 for all parameters. The Bulk_ESS and Tail_ESS values suggest adequate effective sample sizes for reliable estimation and inference.

## 1.7 Nomothetic Analysis of the Relationship Between UCS and CS

In this section, we present a nomothetic analysis of the relationship between the two components of state self-compassion, UCS and CS. This analysis was conducted using a Bayesian hierarchical model, where UCS was predicted by CS to assess the association between these components. The model also incorporated covariates, such as within-day centered negative affect and context evaluation, to capture the influence of these factors on the UCS-CS relationship. Random effects were included at the participant, day, and measurement levels to account for individual and temporal variability, critical for understanding fluctuations in state self-compassion.

Although the results of this nomothetic analysis are not discussed in the main manuscript, they are included here for completeness. The idionomic analysis described in the main manuscript, which combines data from two experiments, provides a more comprehensive view of the UCS-CS association. In contrast, this section presents a mixed-effects model separately for each experiment to address the same research question.

The outcomes of this nomothetic model offer a direct test of the Bipolar Continuum Hypothesis. A credible negative fixed-effect slope for CS would support the Bipolar Continuum Hypothesis, indicating that higher levels of CS are associated with lower levels of UCS, thus suggesting a bipolar relationship. Conversely, a non-credible or positive slope would challenge the Bipolar Continuum Hypothesis, implying that CS and UCS may function independently or synergistically rather than as opposing components.

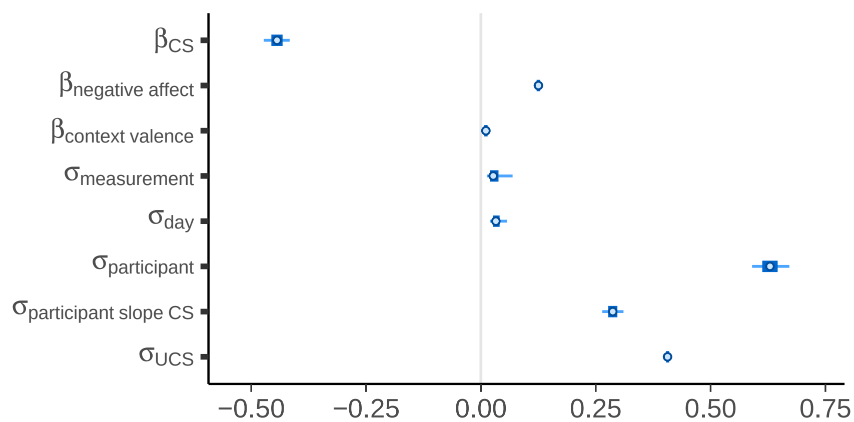


Figure 1: The estimates are expressed in terms of inter-individual differences and intra-individual variations, both within a day and across different days. The bars represent 89% credibility intervals. $\beta_{CS}$ represents the beta coefficients for CS; $\beta_{\text{negative affect}}$ and $\beta_{\text{context valence}}$ denote the beta coefficients for negative affect and context valence, respectively; $\sigma_{\text{measurement}}$ is the standard deviation of the distribution of random effects coefficients for the 5 daily observations; $\sigma_{\text{day}}$ represents the standard deviation of the distribution of random effects coefficients across 10 days; $\sigma_{\text{participant}}$ indicates the standard deviation of the distribution of random effects coefficients across participants (subjects); $\sigma_{\text{participant slope CS}}$ is the standard deviation of the slopes describing the effect of CS on UCS for each participant; $\sigma_{UCS}$ is the estimated standard deviation of the population residuals distribution.

The central finding of this analysis was a robust negative association between CS and UCS, with a median slope estimate of -0.44 (89% CI [-0.47, -0.42]). This result supports the Bipolar Continuum Hypothesis by demonstrating that increases in CS are associated with decreases in UCS.

After controlling for CS, negative affect had a modest positive effect on UCS ($\beta$ = 0.12, 89% CI [0.116, 0.13]), while the evaluation of context unpleasantness showed a minimal impact ($\beta$ = 0.01, 89% CI [0.002, 0.02]). These findings suggest that contextual factors and negative affect play nuanced roles in shaping state self-compassion.

The random-effect slopes for CS varied substantially across participants (median estimate: 0.29, 89% CI [0.26, 0.32]), reflecting significant individual differences in how CS affects UCS. This variability indicates that the relationship between CS and UCS is influenced by factors such as personality and contextual circumstances.

The analysis also revealed considerable individual differences in UCS levels ($\beta$ = 0.63, 89% CI [0.59, 0.67]) and moderate day-to-day variability ($\beta$ = 0.04, 89% CI [0.02, 0.06]), though the latter was less pronounced. This variability emphasizes the state-dependent nature of self-compassion and the influence of daily emotional states and contextual stressors.

Measurement error was minimal, with a median estimate of 0.01 (89% CI [0.001, 0.039]), indicating the reliability of the assessment tools used.

In summary, this nomothetic analysis shows a strong negative association between CS and UCS, supporting the Bipolar Continuum Hypothesis The substantial variability in random slopes across participants highlights the complexity of the CS-UCS relationship, which is influenced by individual differences and context. Despite the robust findings, unexplained variability in UCS (median estimate: 0.41, 89% CI [0.40, 0.414]) suggests that additional factors, possibly related to unmeasured psychological variables, influence UCS levels.

### 1.7.1 Model Implementation

The hierarchical Bayesian model included random effects for participants, days, and measurements and used a t-distribution to account for potential outliers in UCS. The following priors were employed to regularize the estimates:

$$
\begin{aligned}
\alpha_{ucs} &\sim \text{Normal}(0, 1) \\
\beta_{cs}, \beta_{covariates} &\sim \text{Normal}(0, 1) \\
z_{participant}, z_{day}, z_{measurement}, z_{participant\_slope\_cs} &\sim \text{Normal}(0, 1) \\
\sigma_{participant}, \sigma_{day}, \sigma_{measurement}, \sigma_{participant\_slope\_cs}, \sigma_{ucs} &\sim \text{Exponential}(1) \\
\nu &\sim \text{Gamma}(2, 0.1)
\end{aligned}
$$

The Stan code for the model is presented below:

```
data {
  int<lower=0> N; // Total observations
  int<lower=0> P; // Participants
  int<lower=0> D; // Days
```

```
  int<lower=0> M; // Measurements per day
  array[N] int<lower=1, upper=P> participant;
  array[N] int<lower=1, upper=D> day;
  array[N] int<lower=1, upper=M> measurement;
  array[N] real CS;
  array[N] real UCS;
  array[N] real neg_affect;
  array[N] real decentering;
  array[N] real context_eval;
}

parameters {
  real alpha_ucs;
  real beta_cs;
  array[3] real beta_covariates;
  vector[P] z_participant;
  vector[D] z_day;
  vector[M] z_measurement;
  vector[P] z_participant_slope_cs;
  real<lower=0> sigma_participant;
  real<lower=0> sigma_day;
  real<lower=0> sigma_measurement;
  real<lower=0> sigma_participant_slope_cs;
  real<lower=0> sigma_ucs;
  real<lower=0> nu;
}

model {
  alpha_ucs ~ normal(0, 1);
  beta_cs ~ normal(0, 1);
  beta_covariates ~ normal(0, 1);
  z_participant ~ normal(0, 1);
  z_day ~ normal(0, 1);
  z_measurement ~ normal(0, 1);
  z_participant_slope_cs ~ normal(0, 1);
  sigma_participant ~ exponential(1);
  sigma_day ~ exponential(1);
  sigma_measurement ~ exponential(1);
  sigma_participant_slope_cs ~ exponential(1);
  sigma_ucs ~ exponential(1);
  nu ~ gamma(2, 0.1);

  for (n in 1:N) {
    UCS[n] ~ student_t(nu,
                  alpha_ucs +
                  (beta_cs + sigma_participant_slope_cs * z_participant_slope_cs[partici
                  beta_covariates[1] * neg_affect[n] +
```

```
                      beta_covariates[2] * decentering[n] +
                      beta_covariates[3] * context_eval[n] +
                      sigma_participant * z_participant[participant[n]] +
                      sigma_day * z_day[day[n]] +
                      sigma_measurement * z_measurement[measurement[n]],
                      sigma_ucs);
  }
}
```

## 1.8 Data Quality

To ensure data quality, we employed a comprehensive set of behavioral indicators to detect inattentive or insufficient-effort responding, including the compliance rate and metrics designed to identify aberrant response patterns (Ulitzsch et al. 2024; Hasselhorn, Ottenstein, and Lischetzke 2023). These indicators evaluated both response patterns and supplementary data (e.g., timing metrics, engagement levels), flagging anomalies that might suggest inattentive or low-effort responding (Meade and Craig 2012). Common anomalies included response invariability, inconsistencies across responses, and multivariate outliers.

Notably, our data-quality assessment was conducted post hoc, rather than by embedding directed prompts (e.g., instructed-response items) in the survey. Although such items can effectively detect careless responding, they present several drawbacks (Ulitzsch et al. 2024). First, adding items to detect inattentiveness can lengthen the questionnaire, paradoxically increasing the chance of respondent fatigue—particularly in studies using Ecological Momentary Assessment (EMA). Second, in repeated-measures contexts, these items may confuse participants who must maintain clarity about task expectations across multiple assessment occasions. By employing a post-hoc analysis using behavioral indicators, we minimized participant burden and preserved the integrity of the EMA design.

Compliance rate, defined as the ratio of completed assessments to the total number of expected assessments, served as a key criterion for gauging insufficient engagement. Participants failing to meet a prespecified compliance threshold were excluded from analyses to maintain the reliability of the dataset.

### 1.8.1 Compliance Rate

Non-compliance introduces systematic missing data, which can bias statistical inference if the data are not missing at random. To mitigate this risk, participants with a compliance rate below 50% were excluded. Based on this criterion, seven participants were removed from the dataset.

### 1.8.2 Time to Complete

The **Time to Complete** (TTC) metric was calculated from the moment participants opened an EMA survey to the time they submitted it. Due to the right-skewed nature of survey completion times, we calculated each participant's median TTC using log-transformed values. We then applied the lower fence of the interquartile range (IQR) rule to these log-transformed medians to identify unusually short completion times. No participant fell below this lower-fence threshold, indicating no evidence of careless responding based on survey duration.

### 1.8.3 Careless responding on the State Self-Compassion Scale

We assessed careless responding on the State Self-Compassion Scale items by using four metrics provided by the **careless** R package: the Longstring Index, Intra-Individual Response Variability (IRV), the Even-Odd Inconsistency Index, and Mahalanobis Distance. These metrics, designed to evaluate response patterns, offer a robust framework for identifying inattentive or inconsistent behavior.

- **Longstring Index**: Measures the longest sequence of identical responses within a survey. High values suggest "straightlining," a common marker of inattentiveness.
- **Intra-Individual Response Variability (IRV)**: Reflects the variability in responses across consecutive items. Low IRV may indicate lower engagement or attention.
- **Even-Odd Inconsistency Index**: Assesses internal consistency by correlating mean scores of even- and odd-numbered items. Lower correlations suggest greater inconsistency.
- **Mahalanobis Distance** ($D^2$): Identifies multivariate outliers by quantifying deviations from an expected response pattern. High $D^2$ values indicate unusual response profiles.

We performed these analyses at two levels: person and occasion.

#### 1.8.3.1 Person-Level Analysis of Behavioral Indices

Participants were flagged as potential careless responders if their index scores exceeded the 95th percentile on any of the four metrics. Participants flagged on more than two indices were classified as careless responders.

The following R script illustrates how we counted participants who exceeded thresholds for various combinations of indices:

```r
vectors <- list(mahad_bad, longstring_bad, irv_bad, even_odd_bad)

# Count shared elements across index combinations
shared_counts <- map(2:4, ~ {
  combos <- combn(vectors, .x, simplify = FALSE)
  shared <- map(combos, ~ Reduce(intersect, .x)) %>% unlist() %>% unique()
  length(shared)
}) %>% set_names(paste0("shared_by_", 2:4))

# Results
shared_counts
# $shared_by_2
# [1] 3
#
# $shared_by_3
# [1] 0
#
# $shared_by_4
# [1] 0
```

No participant exceeded thresholds on more than two indices during the EMA phase, so no exclusions were made based on these metrics.

#### 1.8.3.2 Occasion-Level Analysis of Behavioral Indices

To complement the participant-level analysis, we conducted an occasion-level analysis to capture instances of momentary inattentiveness. This approach allowed us to detect subtle, transient lapses in engagement that might otherwise remain undetected in participant-level summaries.

By evaluating responses at the occasion level:

- momentary lapses in engagement could be flagged without excluding an entire participant;
- the analysis gained granularity, enabling targeted handling of problematic data points;
- the design aligns with recent work emphasizing the importance of evaluating variability over time in EMA research (Hasselhorn, Ottenstein, and Lischetzke 2023).

We calculated the four careless-responding indices (Longstring, IRV, Even-Odd Inconsistency, and Mahalanobis Distance) separately for each occasion on the State Self-Compassion Scale (SCS). We then flagged occasions exceeding **adjusted fences** derived from bootstrapped IQRs. Finally, we aggregated flagged occasions to compute a proportion of flagged data points per participant.

##### 1.8.3.2.1 Results

- **Distribution of Flagged Occasions**:
  - **80.4%** of occasions were never flagged by any metric.
  - **17.7%** of occasions were flagged on one metric, indicating mild evidence of momentary inattention.
  - **1.77%** of occasions were flagged on two metrics, suggesting stronger evidence of momentary lapses.
  - **<0.2%** of occasions were flagged on three or more metrics, representing very rare instances of substantial carelessness.
- **Proportion of Flagged Occasions per Participant**:
  - the median proportion of flagged occasions per participant was **<5%**, reflecting low overall incidence of momentary inattentiveness;
  - no participant showed persistently high flagging rates across occasions, underscoring the reliability of the dataset.

### 1.8.4 Final Assessment

Both the person-level and occasion-level analyses confirmed a low incidence of inattentive or careless responding. These findings attest to the dataset's robustness and highlight the strengths of this study's design.

Compared to other study designs—particularly those involving paid participants, such as Amazon Mechanical Turk—our sample exhibited notably fewer inattentive responses. This discrepancy is likely attributable to varying levels of intrinsic motivation (Aruguete et al. 2019). For instance, participants on Amazon Mechanical Turk may exhibit lower intrinsic motivation due to minimal monetary compensation, while our volunteers, who committed to completing weekly surveys over a two-month period, were likely driven by higher intrinsic motivation.

Recent research supports these observations in the context of EMA studies. For example, Hasselhorn, Ottenstein, and Lischetzke (2023) used multilevel latent class analysis (ML-LCA) with volunteer participants to identify profiles of momentary careless responding at the occasion level and latent classes of individuals who varied in their distribution of careless responses across occasions. They identified four latent classes: "careful," "frequently careless," and two categories of "infrequently careless" respondents. Notably, the "frequently careless" class accounted for only 2% of participants, even in a more intensive EMA design involving multiple daily notifications over seven consecutive days.

Study design plays a critical role in influencing participant compliance and data quality. Intensive EMA designs, characterized by frequent and intrusive assessments, are associated with participant fatigue, reduced response accuracy, and increased inattentiveness (Shiffman, Stone, and Hufford 2008). In contrast, our study employed a less intensive approach, with participants receiving only one notification per week rather than multiple daily prompts.

Other additional factors likely contributed to the minimization of inattentive responses in our study:

- **Brevity of EMA questionnaires:** Participants faced minimal cognitive burden, encouraging thoughtful responses.
- **Flexibility to discontinue participation:** Participants retained the option to withdraw at any time, reducing potential disengagement due to fatigue or other external pressures.

These features collectively reduced the likelihood of inattentive responses compared to traditional cross-sectional surveys (Welling, Fischer, and Schinkel-Bielefeld 2021) and more intensive EMA studies (Hasselhorn, Ottenstein, and Lischetzke 2023).

### 1.8.5 Data Reanalysis

To evaluate whether excluding flagged occasions—defined as those identified on $geq$2 metrics—would affect the primary outcomes, we conducted a secondary analysis of the Study 1 data. This analysis focused on the nomothetic model assessing the relationships between the Uncompassionate Self (UCS) and Compassionate Self (CS) components of state self-compassion, after controlling for Negative Affect and Context Evaluation. Results from the full dataset are provided in Figure 1.

The reanalysis produced results that were effectively identical to those derived from the full dataset, underscoring the robustness of the findings. Hence, we retained all participants who met the compliance requirements in our final analyses. Below is a table summarizing the posterior estimates from the reanalysis, alongside their standard deviations and 89% credibility intervals:

| Variable | Mean | SD | Q5.5 | Q94.5 |
|---|---|---|---|---|
| alpha_ucs | -0.0005 | 0.0403 | -0.0644 | 0.0643 |

| Variable | Mean | SD | Q5.5 | Q94.5 |
|---|---|---|---|---|
| beta_cs | -0.4470 | 0.0183 | -0.4770 | -0.4180 |
| beta_negative_affect | 0.1160 | 0.0059 | 0.1070 | 0.1250 |
| beta_context_valence | 0.0101 | 0.0055 | 0.0013 | 0.0189 |

These results confirm the reliability of the final dataset and the validity of our conclusions, demonstrating a high level of data quality in this long-term EMA study design.

# 2 Study 2

## 2.1 Baseline Measures

| Variable | Estimate | Std. Error | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|
| $DASS-21_{Stress}$ | 7.24 | 0.73 | 6.10 | 9.37 |
| $DASS-21_{Anxiety}$ | 3.43 | 1.55 | 0.95 | 6.32 |
| $DASS-21_{Depression}$ | 3.00 | 0.36 | 2.17 | 3.92 |
| $RSES$ | 22.88 | 0.52 | 21.84 | 23.86 |
| $SCS_{total-score}$ | 17.00 | 0.42 | 16.19 | 17.84 |
| $SCS_{Self-Kindness}$ | 2.88 | 0.10 | 2.70 | 3.07 |
| $SCS_{Common-Humanity}$ | 3.00 | 0.09 | 2.83 | 3.17 |
| $SCS_{Mindfulness}$ | 3.07 | 0.09 | 2.89 | 3.23 |
| $SCS_{Self-Judgment}$ | 2.69 | 0.09 | 2.51 | 2.87 |
| $SCS_{Isolation}$ | 2.69 | 0.09 | 2.50 | 2.87 |
| $SCS_{Over-Identification}$ | 2.67 | 0.09 | 2.50 | 2.85 |

There was no indication of emotional disorders among the participants. The obtained scores aligned with those found in previous studies that used the same assessment tools in community samples (Bottesi et al. 2015; Neff, Whittaker, and Karl 2017; Sica et al. 2021).

## 2.2 EMA Survey Questions

For each notification in the Ecological Momentary Assessment (EMA) protocol, participants were prompted to answer the following questions.

1. Think about the most notable event that has occurred since you last received a notification. If this is your first notification of the day, consider the most significant event from the start of the day. How would you evaluate this event?
    1) Very unpleasant
    2) Unpleasant
    3) Neither unpleasant nor pleasant
    4) Pleasant
    5) Very pleasant
2. At this moment I feel NERVOUS.
    1) Not at all
    2) A little
    3) Moderately
    4) Quite a bit
    5) Very much
3. At this moment I feel UPSET.
    1) Not at all
    2) A little
    3) Moderately
    4) Quite a bit
    5) Very much
4. At this moment I feel SATISFIED.
    1) Not at all
    2) A little
    3) Moderately
    4) Quite a bit
    5) Very much
5. At this moment I feel CHEERFUL.
    1) Not at all
    2) A little
    3) Moderately
    4) Quite a bit
    5) Very much
6. I'm giving myself the caring and tenderness I need.
    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true

   5) Moderately true
   6) Completely true

7. I'm obsessing and fixating on everything that's wrong.
   1) Completely false
   2) Moderately false
   3) Slightly false
   4) Slightly true
   5) Moderately true
   6) Completely true

8. I'm remembering that there are lots of others in the world feeling like I am.
   1) Completely false
   2) Moderately false
   3) Slightly false
   4) Slightly true
   5) Moderately true
   6) Completely true

9. I feel like I'm struggling more than others right now.
   1) Completely false
   2) Moderately false
   3) Slightly false
   4) Slightly true
   5) Moderately true
   6) Completely true

10. I feel intolerant and impatient toward myself.
    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true
    5) Moderately true
    6) Completely true

11. I'm keeping things in perspective.
    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true
    5) Moderately true
    6) Completely true

12. At this moment I am able to accept my flaws and weaknesses.
    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true

5) Moderately true
6) Completely true

13. At this moment I let myself be carried away by my emotions.
    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true
    5) Moderately true
    6) Completely true
14. Recently, I have been able to observe my thoughts and feelings without being drawn in.
    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true
    5) Moderately true
    6) Completely true
15. Recently, I have struggled with my thoughts and feelings.
    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true
    5) Moderately true
    6) Completely true
16. Recently, I have experienced my thoughts and feelings as separate from myself.
    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true
    5) Moderately true
    6) Completely true
17. Recently, I have been caught up in my thoughts.
    1) Completely false
    2) Moderately false
    3) Slightly false
    4) Slightly true
    5) Moderately true
    6) Completely true

### 2.3 Data Quality Management and Subject Selection

To ensure consistency and comparability with Study 1, we applied the same procedures to identify and exclude participants with inattentive or careless responding in Study 2.

#### 2.3.1 Compliance Rate

Participants with a compliance rate below 50% were excluded from the analysis, resulting in the exclusion of two participants.

#### 2.3.2 Time to Completion (TTC)

Following the approach used in Study 1, we applied the lower fence of the interquartile range (IQR) rule to identify participants with unusually short time-to-completion (TTC) values. No participants had TTC values below this threshold, suggesting no evidence of careless responding based on average response times.

#### 2.3.3 Careless Responding on the State Self-Compassion Scale

To assess inattentive responding, we calculated four indices—Longstring Index, Intra-Individual Response Variability (IRV), Even-Odd Inconsistency Index, and Mahalanobis Distance—based on responses to the State Self-Compassion Scale. Participants were flagged as potential careless responders if their scores exceeded the 95th percentile on any index. Participants flagged on more than two indices were classified as careless responders.

##### 2.3.3.1 Integration of Metrics to Identify Careless Responding

To evaluate the overlap of flagged participants across indices, we used the following R script to calculate shared participants flagged by two, three, or four indices:

```r
vectors <- list(mahad_bad, longstring_bad, irv_bad, even_odd_bad)

# Identify shared flagged participants across combinations of indices
shared_counts <- map(2:4, ~ {
  combos <- combn(vectors, .x, simplify = FALSE)
  shared <- map(combos, ~ Reduce(intersect, .x)) %>% unlist() %>% unique()
  length(shared)
}) %>% set_names(paste0("shared_by_", 2:4))

# Output results
shared_counts
# $shared_by_2
# [1] 4
#
# $shared_by_3
```

```
# [1] 0
#
# $shared_by_4
# [1] 0
```

#### 2.3.3.2 Final Assessment

No participants exceeded the threshold on more than two indices during the EMA phase of Study 2. Consequently, all participants who met the response frequency criterion were retained for the final statistical analyses.

### 2.4 Model Convergence Diagnostics

Models diagnostics produced similar results as in Study 1.

## 2.5 Impact of Academic Exam on State Self-Compassion

We employ a hierarchical Bayesian model to compare the CS or UCS components of state self-compassion across two time points: the day before and the day after an academic exam. This model accounts for the hierarchical structure of the EMA data, which includes repeated measures across multiple days and times. Specifically, we compare 12 administration distributed over separate days, each containing 5 notifications at different times, against a single notification on the evening following the exam.

Our model incorporates random intercepts for subjects, days, and measurement times, as well as random slopes for the exam day effects:

$$sc_n \sim \text{SkewNormal}\left(\mu_n, \sigma, \text{skewness}\right),$$

where

$$\mu_n = \alpha + \alpha_j[\text{subj}_n] + \alpha_d[\text{day}_n] + \alpha_m[\text{meas}_n] + (\beta_{\text{pre}} + \beta_{j,\text{pre}}[\text{subj}_n]) \cdot \text{exam\_day\_pre}_n + (\beta_{\text{post}} + \beta_{j,\text{post}}[\text{subj}_n]) \cdot \text{exam\_day\_post}_n.$$

For the model's parameters, we used regularization priors:

$$
\begin{aligned}
\alpha &\sim \mathcal{N}(0, 2.5) \\
\alpha_j &\sim \mathcal{N}(0, \sigma_j) \\
\alpha_d &\sim \mathcal{N}(0, \sigma_d) \\
\alpha_m &\sim \mathcal{N}(0, \sigma_m) \\
\beta_{\text{pre}} &\sim \mathcal{N}(0, 1) \\
\beta_{\text{post}} &\sim \mathcal{N}(0, 1) \\
\beta_{j,\text{pre}} &\sim \mathcal{N}(0, \sigma_{\beta_{j,\text{pre}}}) \\
\beta_{j,\text{post}} &\sim \mathcal{N}(0, \sigma_{\beta_{j,\text{post}}}) \\
\sigma &\sim \text{Exponential}(1) \\
\sigma_j &\sim \text{Exponential}(1) \\
\sigma_d &\sim \text{Exponential}(1) \\
\sigma_m &\sim \text{Exponential}(1) \\
\sigma_{\beta_{j,\text{pre}}} &\sim \text{Exponential}(1) \\
\sigma_{\beta_{j,\text{post}}} &\sim \text{Exponential}(1) \\
\text{skewness} &\sim \mathcal{N}(0, 1)
\end{aligned}
$$

Belows is shown the Stan implementation of the model:

```
data {
  int<lower=1> N; // Number of observations
  int<lower=1> J; // Number of subjects
  int<lower=1> D; // Number of days
```

```
  int<lower=1> M; // Number of measurements per day
  array[N] int<lower=1, upper=J> subj; // Subject index
  array[N] int<lower=1, upper=D> day; // Day index
  array[N] int<lower=1, upper=M> meas; // Moment index
  array[N] real sc; // Dependent variable
  array[N] real exam_day_pre; // 1 if exam day is 'pre', 0 otherwise
  array[N] real exam_day_post; // 1 if exam day is 'post', 0 otherwise
}

parameters {
  real alpha; // Global intercept
  array[J] real alpha_j; // Random intercepts for subjects
  array[D] real alpha_d; // Random intercepts for days
  array[M] real alpha_m; // Random intercepts for measurements
  real beta_pre; // Main effect of exam day 'pre'
  real beta_post; // Main effect of exam day 'post'
  array[J] real beta_j_pre; // Random slopes for exam_day_pre
  array[J] real beta_j_post; // Random slopes for exam_day_post
  real<lower=0> sigma; // Standard deviation for psc
  real<lower=0> sigma_j; // SD for subject random intercepts
  real<lower=0> sigma_d; // SD for day random intercepts
  real<lower=0> sigma_m; // SD for measurement random intercepts
  real<lower=0> sigma_beta_j_pre; // SD for random slopes (pre)
  real<lower=0> sigma_beta_j_post; // SD for random slopes (post)
  real skewness; // Skewness parameter for the skew normal distribution
}

model {
  // Priors
  alpha ~ normal(0, 2.5);
  alpha_j ~ normal(0, sigma_j);
  alpha_d ~ normal(0, sigma_d);
  alpha_m ~ normal(0, sigma_m);
  beta_pre ~ normal(0, 1);
  beta_post ~ normal(0, 1);
  beta_j_pre ~ normal(0, sigma_beta_j_pre);
  beta_j_post ~ normal(0, sigma_beta_j_post);
  sigma ~ exponential(1);
  sigma_j ~ exponential(1);
  sigma_d ~ exponential(1);
  sigma_m ~ exponential(1);
  sigma_beta_j_pre ~ exponential(1);
  sigma_beta_j_post ~ exponential(1);
  skewness ~ normal(0, 1);

  // Likelihood
  for (n in 1:N) {
```

```
    sc[n] ~ skew_normal(
      alpha + alpha_j[subj[n]] + alpha_d[day[n]] + alpha_m[meas[n]] +
      (beta_pre + beta_j_pre[subj[n]]) * exam_day_pre[n] +
      (beta_post + beta_j_post[subj[n]]) * exam_day_post[n],
      sigma, skewness
    );
  }
}

generated quantities {
  array[N] real y_rep;
  array[N] real log_lik;

  for (n in 1:N) {
    y_rep[n] = skew_normal_rng(
      alpha + alpha_j[subj[n]] + alpha_d[day[n]] + alpha_m[meas[n]] +
      (beta_pre + beta_j_pre[subj[n]]) * exam_day_pre[n] +
      (beta_post + beta_j_post[subj[n]]) * exam_day_post[n],
      sigma, skewness
    );

    log_lik[n] = skew_normal_lpdf(
      sc[n] |
      alpha + alpha_j[subj[n]] + alpha_d[day[n]] + alpha_m[meas[n]] +
      (beta_pre + beta_j_pre[subj[n]]) * exam_day_pre[n] +
      (beta_post + beta_j_post[subj[n]]) * exam_day_post[n],
      sigma, skewness
    );
  }
}
```

This hierarchical Bayesian model was also applied to the UCS component of state self-compassion, following the same structure and parameterization.

### 2.5.1 Impact of Academic Exam on Negative Affect

In a separate analysis, we found a large effect of academic examinations on students' negative affect. Specifically, we observed a pronounced decrease in negative affect from the day before to the day after the exams. This pattern was consistent across two separate examinations. For the first exam, we found a substantial standardized decrease in negative affect on the day following the exam, relative to the day prior. The magnitude of this decrease was -0.92 (Standard Error, $SE = 0.10$), translating to a Cohen's $d$ value of -0.98, with an 89% CI of [-1.23, -0.78]. A parallel trend was observed for the second exam, where the standardized decrease in negative affect mirrored that of the first exam, being -0.39 ($SE = 0.08$). This yielded a Cohen's $d$ value of -0.54, with the 89% CI of [-0.75, -0.36].

### 2.5.2 Testing the Bipolar Continuum Hypothesis Through Contextual Influences of the State Self-Compassion.

For our analyses, we employed the identical models utilized in Study 1.

### 2.5.3 Nomothetic Analysis of the Relationship Between UCS and CS

As observed in Study 1, both negative affect and contextual valence are key indicators of an individual's position on the bipolar continuum of state self-compassion, which ranges from CS to UCS. High levels of negative affect were associated with proximity to the UCS pole, while low levels suggested alignment with the CS pole. Similarly, we expected that higher levels of contextual valence (i.e., more positive evaluations of context) would be associated with the CS pole. Additionally, decentering, an essential aspect of mindfulness, was anticipated to demonstrate an inverse relationship with UCS, with higher levels of decentering linked to the CS pole and lower levels indicating proximity to the UCS pole.

To replicate the analysis from Study 1, we tested the Bipolar Continuum Hypothesis with a nomothetic Bayesian hierarchical analysis. In this model, UCS was predicted linearly from CS, while also accounting for within-day centered negative affect, decentering, and the unpleasantness of the event as covariates. Random effects for participants, days, and the event unpleasantness were incorporated to address individual variability and measurement precision (see Figure 2).

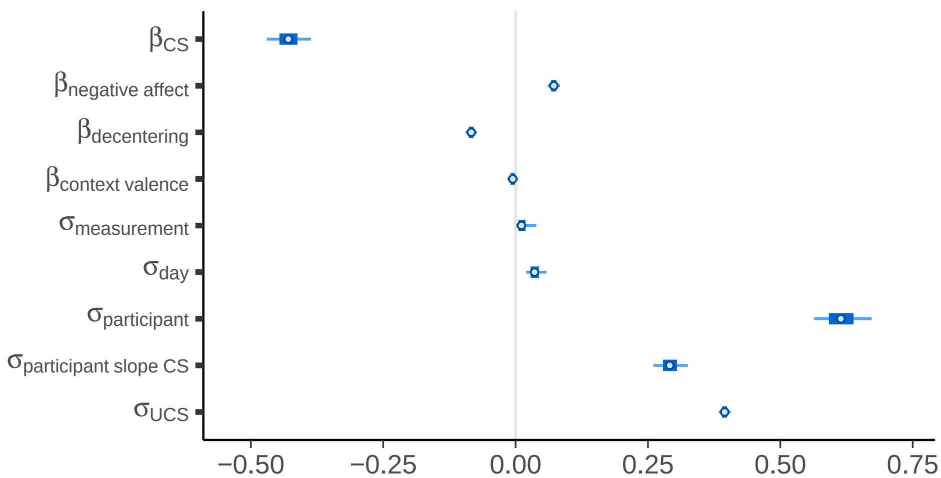


Figure 2: The estimates are expressed in terms of inter-individual differences and intra-individual variations, both within a day and across different days. The bars represent 89% credibility intervals. $\beta_{CS}$ represents the beta coefficients for CS; $\beta_{\text{negative affect}}$, $\beta_{\text{context valence}}$ and $\beta_{\text{decentering}}$ denote the beta coefficients for negative affect, context valence and decentering respectively; $\sigma_{\text{measurement}}$ is the standard deviation of the distribution of random effects coefficients for the 5 daily observations; $\sigma_{\text{day}}$ represents the standard deviation of the distribution of random effects coefficients across 10 days; $\sigma_{\text{participant}}$ indicates the standard deviation of the distribution of random effects coefficients across participants (N subjects); $\sigma_{\text{participant slope CS}}$ is the standard deviation of the slopes describing the effect of CS on UCS for each participant; $\sigma_{UCS}$ is the estimated standard deviation of the population residuals distribution.

The analysis revealed several significant fixed effects. A strong negative relationship between

CS and UCS was observed ($\beta$ = -0.43, 89% CI [-0.47, -0.39]), confirming that higher levels of CS correspond to lower levels of UCS and thus supporting the Bipolar Continuum Hypothesis. Higher levels of negative affect were associated with increased UCS ($\beta$ = 0.07, 89% CI [0.06, 0.08]), indicating a modest positive influence. Decentering showed an inverse relationship with UCS ($\beta$ = -0.08, 89% CI [-0.09, -0.07]), suggesting that higher mindfulness (decentering) is associated with lower levels of UCS. The impact of event unpleasantness was slight and marginally negative ($\beta$ = -0.005, 89% CI [-0.016, 0.005]).

The random effects analysis revealed substantial variability in UCS levels among participants ($\beta$ = 0.61, 89% CI [0.56, 0.67]), indicating individual differences in state self-compassion responses. Daily fluctuations were minor, with a median estimate of 0.04 (89% CI [0.02, 0.06]). The model also demonstrated high reliability, as indicated by minimal measurement error (median estimate: 0.01, 89% CI [0.001, 0.039]).

The analysis further highlighted individual differences in how CS influences UCS, with a random slope median estimate of 0.29 (89% CI [0.26, 0.32]). There was moderate unexplained variability in UCS (median estimate: 0.39, 89% CI [0.38, 0.41]), suggesting that additional factors beyond the scope of this model contribute to UCS levels.

In summary, the results from Study 2 echo those of Study 1, demonstrating a strong inverse relationship between CS and UCS, reinforcing the validity of the Bipolar Continuum Hypothesis. This robust negative correlation, observed even when accounting for momentary negative affect and other contextual factors, supports the hypothesis that self-compassion operates as a bipolar construct rather than as independent or synergistic elements. The influence of negative affect and decentering on UCS, coupled with the consistent individual differences, highlights the dynamic nature of state self-compassion and the nuanced roles of personal and contextual factors.

These findings underscore the applicability of the Bipolar Continuum Hypothesis across diverse contexts and individual differences. The minimal daily fluctuations and substantial between-person variability further emphasize the role of personal traits and situational factors in shaping state self-compassion responses.

For our analysis, we employed the identical Stan model utilized in Study 1.

## 2.6 Levels of Personal Concern

This section details the script used to test the hypothesis that the correlation between the SC and USC components of state self-compassion is influenced by the level of personal concern. The same statistical model was applied to three distinct datasets, each representing different time points relative to an examination period:

1. **Pre-Exam Dataset**: Data collected one day before the exam.
2. **Post-Exam Dataset**: Data collected one day after the exam.
3. **Baseline Dataset**: Data collected during a period unrelated to any exam.

```
bform <-
  bf(mvbind(psc, nsc) ~ 1 + (1 | user_id)) +
  set_rescor(TRUE)

mod <- brm(
  bform,
  data = d,
  backend = "cmdstanr",
  chains = 4
)
```

## 2.7 Decentering and SC and USC Correlation

Decentering (Bennett et al. 2021; Bernstein et al. 2015) and self-compassion are related constructs. Both constructs contribute to psychological well-being through slightly different mechanisms. Decentering aids in diminishing identification with negative thoughts and feelings, enabling individuals to observe them without judgment or automatic reactions. This detachment can create space for self-compassion to emerge, as individuals can acknowledge their challenging experiences with greater kindness and understanding. Self-compassion, in turn, can facilitate the decentering process by offering a gentler and more welcoming approach to difficult emotions and thoughts. When individuals can treat themselves with compassion, they may find it easier to observe their internal experiences without being overwhelmed. Self-compassion allows individuals to confront failures, mistakes, and self-criticism with kindness and understanding rather than judgment or reproach. This approach can facilitate decentering, as individuals learn to view such experiences as common to all humans, rather than reflections of their personal worth. Both decentering and self-compassion contribute to psychological resilience. Decentering helps maintain a more balanced and objective perspective, while self-compassion provides emotional support and a more positive response to challenges. Mindfulness plays a key role in connecting decentering and self-compassion. Mindfulness practice encourages both decentering and self-compassion by promoting non-judgmental observation of internal experiences and greater self-kindness.

# 3 Dimensionality Test

Previous research has extensively explored the bipolar continuum hypothesis by investigating the dimensionality of *trait* self-compassion using various psychometric approaches (Brenner et al. 2017; Cleare et al. 2018; Neff 2016; Neff, Whittaker, and Karl 2017; Neff et al. 2019; Petrocchi, Ottaviani, and Couyoumdjian 2014). These studies consistently show that models including a general factor yield superior fit indices, offering strong empirical support for Neff's hypothesis.

The current study seeks to build upon this foundation by conducting a novel psychometric evaluation of the dimensionality of *state* self-compassion. Our analysis is unique in that it examines repeated measurements of *state* self-compassion within the same participants. To address the multilevel nature of the data, we applied a series of multilevel Confirmatory Factor Analysis (CFA) models to the combined datasets from both studies. This approach allows for a precise representation of the hierarchical structure, where repeated measurements are nested within days, which are further nested within individuals.

We specified and compared three distinct models to determine the underlying structure of state self-compassion:

- The **One-Factor Model** hypothesizes that state self-compassion can be represented by a single latent factor, capturing variance at both the within-person and between-person levels.
- The **Two-Factor Model** conceptualizes state self-compassion as comprising two correlated but distinct latent factors—CS (compassionate self-responding) and UCS (uncompassionate self-responding)—functioning at both the within-person and between-person levels.
- The **Bifactor Model** examines whether state self-compassion is best characterized by a general self-compassion factor along with orthogonal specific factors (CS and UCS), isolating the variance explained by the general factor from that explained by the specific factors.

### 3.0.1 Mplus script for the One-Factor Model

```
TITLE: Multilevel One-Factor Model with covariates at moment, day, and person level;

DATA:
  FILE = "data.dat";

VARIABLE:
  NAMES = user_id day time_window scs_pos_1 scs_neg_2 scs_pos_3 scs_neg_4
    scs_neg_5 scs_pos_6 scs_pos_7 scs_neg_8 neg_aff_Moment neg_aff_Day neg_aff_Person;
  USEVARIABLES = scs_pos_1 scs_neg_2 scs_pos_3 scs_neg_4 scs_neg_5 scs_pos_6
    scs_pos_7 scs_neg_8 neg_aff_Moment neg_aff_Day neg_aff_Person;
  CLUSTER = user_id;  ! Clustering at the subject level
  WITHIN = neg_aff_Moment neg_aff_Day;  ! Covariates at the moment and day level
```

```
  BETWEEN = neg_aff_Person;  ! Covariate at the person level

MISSING = .;

ANALYSIS:
  TYPE = TWOLEVEL;  ! Multilevel model with user_id as the cluster
  ESTIMATOR = ML;  ! Maximum likelihood estimation

MODEL:
  %WITHIN%  ! Level 1: Within-subject and within-day variation
    SelfCompassion_w BY scs_pos_1 scs_neg_2 scs_pos_3 scs_neg_4 scs_neg_5
      scs_pos_6 scs_pos_7 scs_neg_8;
    SelfCompassion_w ON neg_aff_Moment neg_aff_Day;
    ! Regress self-compassion factor on momentary and daily negative affect

  %BETWEEN%  ! Level 2: Between-subject variation
    SelfCompassion_b BY scs_pos_1 scs_neg_2 scs_pos_3 scs_neg_4 scs_neg_5
      scs_pos_6 scs_pos_7 scs_neg_8;
    SelfCompassion_b ON neg_aff_Person;
    ! Regress between-level self-compassion factor on person-level negative affect

OUTPUT:
  TECH1 TECH8 STANDARDIZED MODINDICES;
```

### 3.0.2 Table: Standardized Factor Loadings for Multilevel One-Factor Model

| Item | Within-Level Standardized Loadings | Between-Level Standardized Loadings |
|---|---|---|
| **SCS_POS_1** | 0.579 | 0.778 |
| **SCS_NEG_2** | 0.673 | 0.925 |
| **SCS_POS_3** | 0.274 | 0.363 |
| **SCS_NEG_4** | 0.567 | 0.795 |
| **SCS_NEG_5** | 0.652 | 0.934 |
| **SCS_POS_6** | 0.550 | 0.738 |
| **SCS_POS_7** | 0.587 | 0.835 |
| **SCS_NEG_8** | 0.279 | 0.262 |

### 3.0.3 Additional Summary Statistics

- **Within-Level Explained Variance for SELFCOMP**: 19.6%
- **Between-Level Explained Variance for SELFCOMP**: 21.7%
- **Effect of Negative Affect on SELFCOMP (within level)**:
  - **NEG_AFF_MOMENT**: -0.272
  - **NEG_AFF_DAY**: -0.350

- **Effect of Negative Affect on SELFCOMP (between level)**:
  - **NEG_AFF_PERSON**: -0.466

### 3.0.4 Mplus script for the Two-Factor Model

```
TITLE: Multilevel Two-Factor Model with covariates at moment, day, and person level;

DATA:
  FILE = "data.dat";

VARIABLE:
  NAMES = user_id day time_window scs_pos_1 scs_neg_2 scs_pos_3 scs_neg_4
    scs_neg_5 scs_pos_6 scs_pos_7 scs_neg_8 neg_aff_Moment neg_aff_Day neg_aff_Person;
  USEVARIABLES = scs_pos_1 scs_neg_2 scs_pos_3 scs_neg_4 scs_neg_5 scs_pos_6
    scs_pos_7 scs_neg_8 neg_aff_Moment neg_aff_Day neg_aff_Person;
  CLUSTER = user_id;  ! Clustering at the subject level
  WITHIN = neg_aff_Moment neg_aff_Day;  ! Covariates at the moment and day level
  BETWEEN = neg_aff_Person;  ! Covariate at the person level

MISSING = .;

ANALYSIS:
  TYPE = TWOLEVEL;  ! Multilevel model with user_id as the cluster
  ESTIMATOR = ML;  ! Maximum likelihood estimation

MODEL:
  %WITHIN%  ! Level 1: Within-subject and within-day variation
    PositiveSelfCompassion_w BY scs_pos_1 scs_pos_3 scs_pos_6 scs_pos_7;
    ! Positive self-compassion factor on within level
    NegativeSelfCompassion_w BY scs_neg_2 scs_neg_4 scs_neg_5 scs_neg_8;
    ! Negative self-compassion factor on within level
    PositiveSelfCompassion_w ON neg_aff_Moment neg_aff_Day;
    ! Regress positive factor on momentary and daily negative affect
    NegativeSelfCompassion_w ON neg_aff_Moment neg_aff_Day;
    ! Regress negative factor on momentary and daily negative affect

  %BETWEEN%  ! Level 2: Between-subject variation
    PositiveSelfCompassion_b BY scs_pos_1 scs_pos_3 scs_pos_6 scs_pos_7;
    ! Positive self-compassion factor on between level
    NegativeSelfCompassion_b BY scs_neg_2 scs_neg_4 scs_neg_5 scs_neg_8;
    ! Negative self-compassion factor on between level
    PositiveSelfCompassion_b ON neg_aff_Person;
    ! Regress positive factor on person-level negative affect
    NegativeSelfCompassion_b ON neg_aff_Person;
    ! Regress negative factor on person-level negative affect

OUTPUT:
  TECH1 TECH8 STANDARDIZED MODINDICES;
```

### 3.0.5 Table: Standardized Factor Loadings for Multilevel Two-Factor Model

| Item | Within-Level Standardized Loadings | Between-Level Standardized Loadings |
|---|---|---|
| **CS Factor** | | |
| **SCS_POS_1** | 0.632 | 0.902 |
| **SCS_POS_3** | 0.270 | 0.321 |
| **SCS_POS_6** | 0.581 | 0.749 |
| **SCS_POS_7** | 0.639 | 0.959 |
| **UCS Factor** | | |
| **SCS_NEG_2** | 0.710 | 0.949 |
| **SCS_NEG_4** | 0.582 | 0.796 |
| **SCS_NEG_5** | 0.682 | 0.945 |
| **SCS_NEG_8** | 0.279 | 0.278 |

### 3.0.6 Additional Summary Statistics

- **Within-Level Correlation between POSITIVE and NEGATIVE factors**: 0.825
- **Between-Level Correlation between POSITIVE and NEGATIVE factors**: 0.817
- **Residual Variances (Within Level)**:
  - **SCS_POS_1**: 0.601, **SCS_NEG_2**: 0.497, **SCS_POS_3**: 0.927, **SCS_NEG_4**: 0.661, **SCS_NEG_5**: 0.535, **SCS_POS_6**: 0.662, **SCS_POS_7**: 0.591, **SCS_NEG_8**: 0.922
- **Residual Variances (Between Level)**:
  - **SCS_POS_1**: 0.187, **SCS_NEG_2**: 0.100, **SCS_POS_3**: 0.897, **SCS_NEG_4**: 0.367, **SCS_NEG_5**: 0.106, **SCS_POS_6**: 0.439, **SCS_POS_7**: 0.080, **SCS_NEG_8**: 0.923

### 3.0.7 Mplus script for the Bifactor Model

```
TITLE: Multilevel Bifactor Model with covariates at moment, day, and person level;

DATA:
  FILE = "data.dat";

VARIABLE:
  NAMES = uid day tw scp1 scn2 scp3 scn4 scn5 scp6 scp7 scn8 na_mom na_day na_per;
  USEVARIABLES = scp1 scn2 scp3 scn4 scn5 scp6 scp7 scn8 na_mom na_day na_per;
  CLUSTER = uid;
  WITHIN = na_mom na_day;
  BETWEEN = na_per;

  MISSING = .;

ANALYSIS:
  TYPE = TWOLEVEL;
  ESTIMATOR = ML;

MODEL:
  %WITHIN%
    Gen_w BY scp1* scn2 scp3 scn4 scn5 scp6 scp7 scn8;
    Pos_w BY scp1* scp3 scp6 scp7;
    Neg_w BY scn2* scn4 scn5 scn8;

    ! Fix variances to help with model identification
    Gen_w@1;  ! Fix variance of the general factor
    Pos_w@1;  ! Constrain variance of the positive factor
    Neg_w@1;  ! Constrain variance of the negative factor

    ! Ensure no covariance between factors
    Pos_w WITH Gen_w@0;
    Neg_w WITH Gen_w@0;
    Pos_w WITH Neg_w@0;

    ! Regress factors on covariates
    Gen_w ON na_mom na_day;
    Pos_w ON na_mom na_day;
    Neg_w ON na_mom na_day;

  %BETWEEN%
    Gen_b BY scp1* scn2 scp3 scn4 scn5 scp6 scp7 scn8;
    Pos_b BY scp1* scp3 scp6 scp7;
    Neg_b BY scn2* scn4 scn5 scn8;

    ! Fix variances on the between level as well
```

```
    Gen_b@1;  ! Fix variance of the general factor
    Pos_b@1;  ! Constrain variance of the positive factor
    Neg_b@1;  ! Constrain variance of the negative factor

    ! Ensure no covariance between factors
    Pos_b WITH Gen_b@0;
    Neg_b WITH Gen_b@0;
    Pos_b WITH Neg_b@0;

    ! Regress factors on covariates
    Gen_b ON na_per;
    Pos_b ON na_per;
    Neg_b ON na_per;

OUTPUT:
  TECH1 TECH8 STANDARDIZED MODINDICES(ALL);
```

#### 3.0.8 Table: Standardized Factor Loadings for Multilevel Bifactor Model

| Item | Within-Level Standardized Loadings | Between-Level Standardized Loadings |
|---|---|---|
| **General Factor (GEN__W / GEN__B)** | | |
| **SCP1** | 0.614 | 0.860 |
| **SCN2** | 0.579 | 0.774 |
| **SCP3** | 0.327 | 0.410 |
| **SCN4** | 0.490 | 0.697 |
| **SCN5** | 0.585 | 0.819 |
| **SCP6** | 0.583 | 0.781 |
| **SCP7** | 0.623 | 0.935 |
| **SCN8** | 0.210 | 0.143 |
| **Positive Factor (POS__W / POS__B)** | | |
| **SCP1** | 0.219 | 0.215 |
| **SCP3** | -0.313 | -0.329 |
| **SCP6** | -0.009 | -0.073 |
| **SCP7** | 0.246 | 0.322 |
| **Negative Factor (NEG__W / NEG__B)** | | |
| **SCN2** | 0.473 | 0.626 |

| Item | Within-Level Standardized Loadings | Between-Level Standardized Loadings |
|---|---|---|
| **SCN4** | 0.291 | 0.378 |
| **SCN5** | 0.310 | 0.422 |
| **SCN8** | 0.208 | 0.288 |

The table presents standardized loadings for both the general factor and the specific positive and negative factors at the within and between levels. These loadings illustrate the bifactor structure and how each item loads onto the general and specific factors.

### 3.1 Goodness-of-fit indices

Goodness-of-fit indices for each model are presented in Table 1.

**Table 1. Goodness-of-Fit Indices for the One-Factor, Two-Factor, and Bifactor Models**

| Model | CFI | TLI | RMSEA | SRMR (Within) | SRMR (Between) | AIC | BIC |
|---|---|---|---|---|---|---|---|
| One-Factor | 0.935 | 0.910 | 0.050 | 0.033 | 0.058 | 475881.5 | 476195.3 |
| Two-Factor | 0.972 | 0.959 | 0.034 | 0.024 | 0.050 | 474810.2 | 475139.6 |
| Bifactor | 0.987 | 0.971 | 0.029 | 0.016 | 0.031 | 474376.7 | 474816.0 |

As shown in Table 1, the goodness-of-fit indices progressively improve from the One-Factor Model to the Two-Factor Model, with the Bifactor Model demonstrating the best overall fit. The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) both exceed the commonly accepted threshold of 0.95 for the Two-Factor and Bifactor Models, indicating strong model fit. Additionally, the Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) values decrease across models, further supporting improved fit with more complex structures.

We conducted a series of Likelihood Ratio Tests (LRTs) to formally assess improvements in model fit:

- Comparing the One-Factor Model with the Two-Factor Model yielded an LRT statistic of 1068.792 with 5 degrees of freedom ($p < 0.001$), indicating that the Two-Factor Model significantly improves fit. This suggests that conceptualizing state self-compassion as two distinct components (CS and UCS) better represents the data.
- The comparison between the Two-Factor Model and the Bifactor Model resulted in an LRT statistic of 506.656 with 17 degrees of freedom ($p < 0.001$), demonstrating that the Bifactor Model offers further improvement in fit. This suggests that, beyond distinct compassionate and uncompassionate components, a general self-compassion factor provides additional explanatory power.

In summary, from a psychometric standpoint, both the Two-Factor and Bifactor Models provide better fit than the One-Factor Model, with the Bifactor Model yielding the best statistical fit. These results suggest that self-compassion is best conceptualized as comprising both a general factor and specific components (CS and UCS). However, while the Bifactor Model offers superior statistical fit, its added complexity raises questions about the practical significance of modeling both general and specific factors.

### 3.2 Practical Importance

Dimensionality assessments should not rely solely on statistical fit indices but should also incorporate the substantive significance of each factor (Rodriguez, Reise, and Haviland 2016; Pere J. Ferrando and Lorenzo-Seva 2018). Although more complex models often exhibit better statistical fit, they may overfit the data, especially when an excessive number of factors are included (Pere Joan Ferrando and Lorenzo-Seva 2019). Therefore, achieving a balance between model complexity and interpretability is crucial for drawing meaningful conclusions.

To determine whether the data support a unidimensional or multifactorial structure for *state* self-compassion, we applied several criteria:

1. **Factor Correlation**:
   In the Two-Factor Model, the within-level correlation between the Compassionate Self (CS) and Uncompassionate Self (UCS) factors was 0.825, while the between-level correlation was 0.817. These high correlations suggest a strong overlap between the two factors, which might indicate a common underlying construct at both the within- and between-person levels.

2. **Explained Common Variance (ECV)**:
   ECV quantifies how much of the shared variance among items is explained by the general versus specific factors. In the Bifactor Model, the general factor ($\mathrm{Gen}_w$) exhibited high standardized loadings on items (e.g., SCP1 = 0.614, SCN2 = 0.579 at the within-person level), indicating that it accounts for the majority of the common variance. By comparison, the loadings of the specific factors ($\mathrm{POS}_w$ and $\mathrm{NEG}_w$) were notably lower, suggesting that they explain less variance. This finding underscores the dominance of the general factor in accounting for self-compassion.

3. **Omega Hierarchical ($\omega_H$)**:
   $\omega_H$ estimates how much of the variance in total scores is attributable to the general factor. In the Bifactor Model, the general factor explained a substantial proportion of the variance (e.g., SCP1 = 0.614, SCN2 = 0.579), while the specific factors contributed relatively little. This suggests that the general factor plays a predominant role in explaining individual differences in self-compassion, further supporting the interpretation of a primarily unidimensional construct.

4. **Bifactor Model Analysis**:
   The Bifactor Model highlights the predominance of the general factor, as the loadings of the specific factors were consistently lower (e.g., $\mathrm{POS}_w$ on SCP3 = -0.313, $\mathrm{NEG}_b$ on SCN8 = 0.288). This indicates that the specific factors (CS and UCS) provide limited explanatory power beyond the general self-compassion factor. Consequently, the Bifactor Model supports a conceptualization where a single general factor underlies the variance in self-compassion, with the specific factors adding relatively small additional value.

5. **Change in R-squared**:
   The Bifactor Model shows that the general factor accounts for a large proportion of the variance (e.g., R-squared for SCP1 = 0.425 at the within-person level, 0.786 at the between-person level). The specific factors contribute only marginally to the overall variance explained, suggesting that a unidimensional structure may be more parsimonious without significantly compromising the model's explanatory power.

In summary, the results of the multilevel CFA analyses lend support to Neff's hypothesis that, also for *state* self-compassion, CS and UCS are related at the global level. The strong correlations between CS and UCS at the latent level indicate that individuals who exhibit high levels of CS tend to have lower levels of UCS, and vice versa, reinforcing the notion that these two aspects are inversely related at a general level.

However, it is important to emphasize that multilevel CFA focuses on *latent, person-level relationships*, which capture broader, long-term patterns. The present findings show that, for *state* self-compassion, CS and UCS are correlated at a global, trait-like level across individuals. In contrast, analyses targeting *momentary dynamics* may reveal that, for *state* self-compassion, CS and UCS operate more independently within individuals during short-term, context-specific situations. This distinction highlights that while Neff's hypothesis may hold at the global level, further investigation is necessary to explore the relationship between CS and UCS at the momentary, within-person level.

# 4 Idionomic Analysis

The analyses described in the previous section provided evidence supporting the "essential unidimensionality" (Reise, Bonifay, and Haviland 2013) of state self-compassion. Our multilevel framework, which accounted for repeated measurements within individuals over a three-month period, demonstrated that compassionate and uncompassionate self-responding (CS and UCS) are inversely related at the *nomothetic level* – that is, when examined across the entire sample.

Our psychometric evaluations, using purely *internal criteria* (fit indices for item scores), revealed that more complex models improved statistical fit. However, these improvements do not fundamentally challenge Neff's hypothesis of a bipolar continuum. While minor deviations and better fit indices were observed with more complex models, their practical importance was limited, and the overall pattern of findings remained aligned with the notion of CS and UCS as opposing dimensions.

This conclusion is further corroborated when considering *external criteria*, including covariates such as momentary negative affect and context evaluation, as described in our two studies. These situational factors did not substantially disrupt the core relationship between CS and UCS proposed by the bipolar continuum hypothesis.

It is important to note, however, that all these previous analyses were conducted at a *nomothetic* level, focusing on group-level patterns. This approach, while valuable, may obscure important differences in how self-compassion operates in the everyday life of individuals. The variability in how individuals experience and express self-compassion across different contexts might not be fully captured in these broad, averaged analyses.

To address this limitation and gain deeper insights into person-specific dynamics, we conducted an *idionomic analysis* to examine the relationship between UCS and CS at the individual level (Ciarrochi et al. 2024; Ferrari et al. 2022; Sahdra et al. 2024). This approach allows us to explore potential heterogeneity in self-compassion processes that may be masked by nomothetic analyses, providing a more nuanced understanding of how the CS and UCS constructs interact within individuals over time.

## 4.1 Idionomic Analysis of the Relationship Between UCS and CS

For the present purposes, an idionomic analysis was conducted in two stages. The first stage involved a strictly idiographic approach, focusing on the individual-level patterns. In the second stage, we applied a nomothetic approach to examine and describe group-level regularities as well as the variability in these effects across individuals (e.g., Ciarrochi et al. 2024; Sahdra et al. 2024).

**Step 1: Fitting Individual-Level Hierarchical Bayesian Models.** For each participant separately, we implemented a hierarchical Bayesian model using Stan to estimate the relationship between UCS and CS. The model incorporated additional covariates, including negative affect and context evaluation, as well as lagged effects of CS from the previous measurement within the same day. The model for each participant was specified as follows:

$$\begin{aligned}\mathrm{UCS}_n \sim t_\nu\Big(&\alpha + \gamma_{\mathrm{CS}} \cdot \mathrm{CS}_n + \gamma_{\mathrm{neg\_aff}} \cdot \mathrm{neg\_aff}_n \\ &+ \gamma_{\mathrm{context}} \cdot \mathrm{context}_n + \phi \cdot \mathrm{lag\_CS}_n \\ &+ \gamma_{\mathrm{interaction}} \cdot \mathrm{CS}_n \cdot \mathrm{neg\_aff}_n, \sigma\Big),\end{aligned}$$

where:

- $\mathrm{UCS}_n$ represents the CS score for observation $n$,
- $\alpha$ denotes the intercept,
- $\gamma_{\mathrm{CS}}$ is the coefficient for the primary predictor, CS ($\mathrm{CS}_n$),
- $\gamma_{\mathrm{neg\_aff}}$ and $\gamma_{\mathrm{context}}$ are coefficients for negative affect ($\mathrm{neg\_aff}_n$) and context evaluation ($\mathrm{context}_n$), respectively,
- $\phi$ represents the autoregressive coefficient for the lagged CS within the same day ($\mathrm{lag\_CS}_n$),
- $\sigma$ is the scale parameter (standard deviation) of the distribution,
- $\nu$ denotes the degrees of freedom of the Student's $t$-distribution,
- $\gamma_{\mathrm{interaction}}$ is the coefficient for the interaction term between CS and negative affect.

This model allows for the examination of the association between UCS and CS while controlling for the effects of negative affect, context evaluation, lagged CS effects within the same day, and the CS × negative-affect interaction. We employed a Student's $t$-distribution ($t_\nu$) to account for potential outliers or heavy-tailed distributions in the data.

For each participant, we examined the posterior distribution of the `gamma_CS` coefficient to estimate the proportion of posterior draws that were negative. This enabled us to evaluate whether there was evidence of a negative association between UCS and CS at the individual level, in line with Neff's bipolar continuum hypothesis. Additionally, we computed the mean posterior estimates for the `gamma_neg_aff`, `gamma_context`, and `gamma_interaction` parameters, which represent the effects of negative affect, context evaluation, and the interaction between CS and negative affect, respectively, on UCS for each participant.

**Step 2: Aggregate Analysis Using a Hierarchical Model.** After the idiographic analysis, we used a hierarchical model (e.g., Ciarrochi et al. 2024) in the `brms` package (`R`) to summarize the proportion of negative estimates for the `gamma_CS` parameter across participants. This model employed a binomial distribution, with the total number of posterior samples as the denominator and the proportion of negative estimates as the response. A random intercept for participants was included to account for individual variability.

Hierarchical models were also applied to the mean posterior estimates of the `gamma_neg_aff`, `gamma_context`, and `gamma_interaction` parameters, representing the influences of negative affect, context evaluation, and the interaction between CS and negative affect on UCS. Each model included a fixed effect (intercept) and random intercepts for participants, capturing individual differences. A Student-$t$ likelihood was used to account for potential outliers and accommodate the heavy-tailed nature of the effect distributions, providing robust aggregate estimates of each parameter and the heterogeneity of their relationships with UCS.

#### 4.1.1 Results

The analysis of the `gamma_CS` parameter across participants indicated that 81.0% (89% CI [0.796, 0.823]) of the posterior estimates for the association between uncompassionate self-responding (UCS) and compassionate self-responding (CS) were negative. This supports Neff's bipolar continuum hypothesis, which posits an inverse relationship between UCS and CS. However, the variability in individual effects (`sd(Intercept)` = 1.25, corresponding to 0.196 on the probability scale) suggests moderate heterogeneity in the strength of this relationship across participants.

For the influence of negative affect on UCS, represented by the `gamma_neg_aff` parameter, the analysis revealed a positive overall effect (intercept = 0.38, 89% CI [0.34, 0.42]), indicating that higher levels of negative affect are generally associated with increased UCS. The variability between individuals (`sd(Intercept)` = 0.19, corresponding to 0.547 on the probability scale) suggests substantial heterogeneity in how strongly negative affect influences UCS.

The analysis of context evaluation showed a small but credible negative effect on UCS (mean = -0.04, 95% CI [-0.07, -0.01]), suggesting that higher context evaluation scores are associated with a slight reduction in UCS. The variability in individual responses (`sd(Intercept)` = 0.17, corresponding to 0.543 on the probability scale) also indicates substantial heterogeneity in the relationship between context evaluation and UCS across participants.

Finally, the interaction between CS and negative affect was negligible, with the 89% credible interval spanning zero (-0.03, 0.00), suggesting no meaningful interaction effect between these variables on UCS.

#### 4.1.2 Discussion

Our idionomic analysis provides insight into the complex, momentary relationship between CS and UCS. Although the group-level results generally support Neff's bipolar continuum hypothesis, substantial heterogeneity emerged in the strength and direction of the CS-UCS association across individuals. While most participants displayed the anticipated inverse relationship between CS and UCS, a notable proportion showed no clear association, and a smaller subset exhibited positive correlations—indicating that, for some individuals, CS and UCS may vary together rather than oppositely, which contradicts the bipolar continuum hypothesis (Ferrari et al., 2023; Ullrich et al., 2020). This variability suggests that Neff's hypothesis may not be universally applicable and underscores the importance of idiographic approaches in revealing the nuanced patterns often obscured in aggregate, group-level analyses.

Additionally, our analysis revealed that negative affect consistently increased UCS across participants, indicating that elevated negative emotional states tend to intensify uncompassionate self-responding. However, the variability observed across individuals highlights that the influence of negative affect on UCS is not uniform. Contrary to expectations, we found no credible evidence that negative affect moderated the CS-UCS relationship, suggesting that momentary emotional states do not substantially alter how CS influences UCS within individuals (Dejonckheere et al., 2021).

## 4.2 Idiographic Analysis

This section details the specification of the model fitted separately to individual data from both experiments, constituting the first step of our idiographic analysis.

### 4.2.1 Model Structure

We employed a hierarchical Bayesian model with the following characteristics:

1. **Response Variable**: Uncompassionate Self-Responding (UCS).
2. **Primary Predictor**: Compassionate Self-Responding (CS).
3. **Covariates**:
    - `neg_aff_Moment`: Momentary negative affect.
    - `context_Moment`: Context evaluation.
    - `lag_CS_same_day`: Lagged CS from the same day.
4. **Distribution**: Student's t-distribution to account for potential outliers or heavy-tailed response distributions
5. **Temporal Dependency**: Incorporated an AR(1) component to account for autoregressive effects of UCS.

### 4.2.2 Model Parameters

- `gamma_CS`: Coefficient representing the association between CS and UCS.
- `gamma_neg_aff`: Slope for negative affect (`neg_aff_Moment`).
- `gamma_context`: Slope for context evaluation (`context_Moment`).
- `phi`: Autoregressive coefficient for the lagged CS effect.
- `nu`: Degrees of freedom for the Student's *t*-distribution (estimated by the model).
- `alpha`: Intercept term.
- `sigma`: Scale parameter of the Student's *t*-distribution.

### 4.2.3 Stan Model Specification

```
data {
  int<lower=1> N;
  vector[N] CS;
  vector[N] UCS;
  vector[N] neg_aff_Moment;
  vector[N] context_Moment;
  vector[N] lag_CS_same_day;
}

parameters {
  real alpha_raw;
  real gamma_CS_raw;
```

```
  real gamma_neg_aff_raw;
  real gamma_context_raw;
  real phi_raw;
  real<lower=0> sigma;
  real<lower=2> nu;
}

transformed parameters {
  real alpha = alpha_raw;
  real gamma_CS = gamma_CS_raw;
  real gamma_neg_aff = gamma_neg_aff_raw;
  real gamma_context = gamma_context_raw;
  real phi = phi_raw * 0.5;  // Constraining phi to [-0.5, 0.5] for stability
}

model {
  // Priors
  alpha_raw ~ normal(0, 1);
  gamma_CS_raw ~ normal(0, 1);
  gamma_neg_aff_raw ~ normal(0, 1);
  gamma_context_raw ~ normal(0, 1);
  phi_raw ~ normal(0, 1);
  sigma ~ cauchy(0, 2.5);
  nu ~ gamma(2, 0.1);

  // Likelihood
  UCS ~ student_t(nu, alpha + gamma_CS * CS + gamma_neg_aff * neg_aff_Moment +
                  gamma_context * context_Moment + phi * lag_CS_same_day, sigma);
}
```

This Stan code implements the hierarchical Bayesian model described above. The model uses weakly informative priors for all parameters and constrains the autoregressive coefficient `phi` to the interval [-0.5, 0.5] to ensure model stability. The degrees of freedom parameter `nu` is constrained to be greater than 2 to ensure finite variance of the Student's $t$-distribution.

### 4.3 Hierarchical Model Analysis

In the second step of our idiographic analysis, we employed a hierarchical model approach, analogous to a meta-analysis (Ciarrochi et al. 2024). This method allowed us to synthesize the individual-level results across all participants, providing both an aggregate estimate of the effects and a measure of heterogeneity among participants.

#### 4.3.1 Model Specification

We fitted a hierarchical binomial model to the proportion of negative posterior draws for the `gamma_CS` parameter. In this framework, each participant was treated as a unique "study," enabling us to estimate:

1. The aggregate proportion of negative effects across all participants.
2. The degree of heterogeneity in these effects among participants.

The model was implemented using the `brms` package in `R`.

#### 4.3.2 R Code for Model Fitting

```
fit_gamma_cs_binom <- brm(
  bf(n_negative | trials(n_total) ~ 1 + (1 | id), family = "binomial"),
  data = beta_cs_df,
  prior = c(
    prior(normal(0, 1), class = "Intercept"),
    prior(normal(0, 1), class = "sd")
  ),
  iter = 100000, warmup = 2000, chains = 4, cores = 4,
  backend = "cmdstanr",
  control = list(adapt_delta = 0.99, max_treedepth = 15)
)
```

#### 4.3.3 Model Components

- **Formula**: `n_negative | trials(n_total) ~ 1 + (1 | id)`
  - `n_negative`: Count of negative posterior draws.
  - `n_total`: Total number of posterior draws
  - `(1 | id)`: Random intercept for each participant.
- **Family**: Binomial, appropriate for modeling proportions.
- **Priors**:
  - `normal(0, 1)` for the intercept and standard deviation of random effects.
  - These priors are weakly informative, allowing the data to drive the results while providing some regularization.

- **MCMC Settings**:
  - 100,000 iterations with 2,000 warmup.
  - 4 chains run in parallel.
  - Increased `adapt_delta` and `max_treedepth` for improved MCMC convergence.

#### 4.3.4 Extension to Other Parameters

We applied a similar hierarchical modeling approach to analyze the posterior distributions of other key parameters from the idiographic analyses:

- `gamma_neg_aff`: Coefficient for negative affect.
- `gamma_context`: Coefficient for context evaluation.
- `gamma_interaction`: Coefficient for potential interaction effects.

## 4.4 Posterior Predictive Checks

In the following sections, we present the results of the Posterior Predictive Checks for the four models used to assess the aggregate effects and heterogeneity in our idiographic analysis. These models evaluate the linear association between UCS and CS, the effect of negative affect, the influence of contextual evaluation, and the interaction between CS and negative affect.

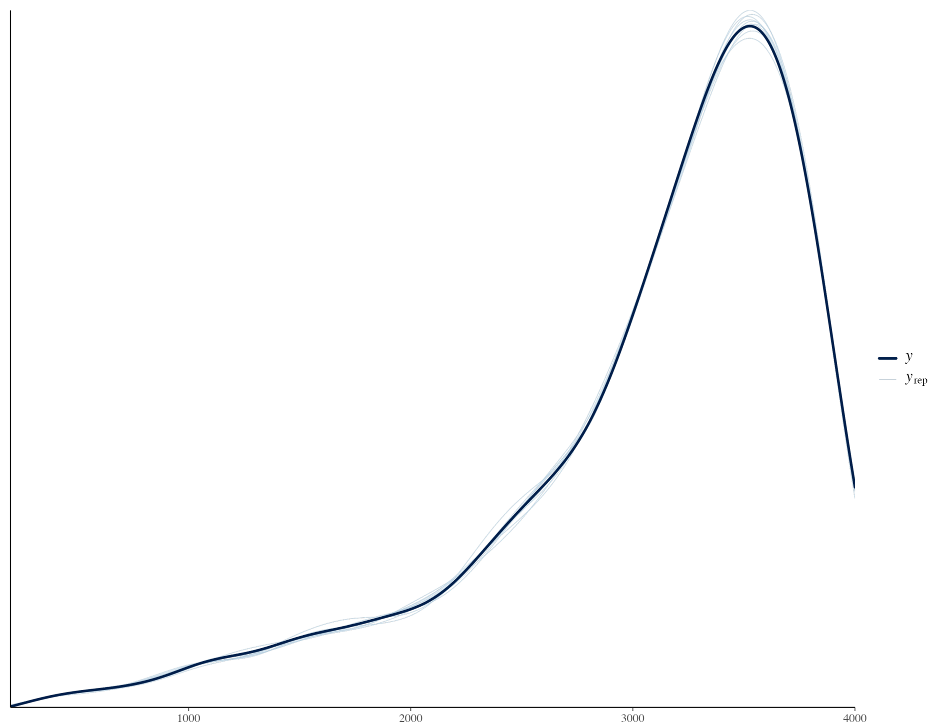


Figure 3: Idiographic analysis. Posterior Predictive Check for the hierarchical model fitted to the proportion of posterior draws of the `gamma_CS`parameter that were negative.

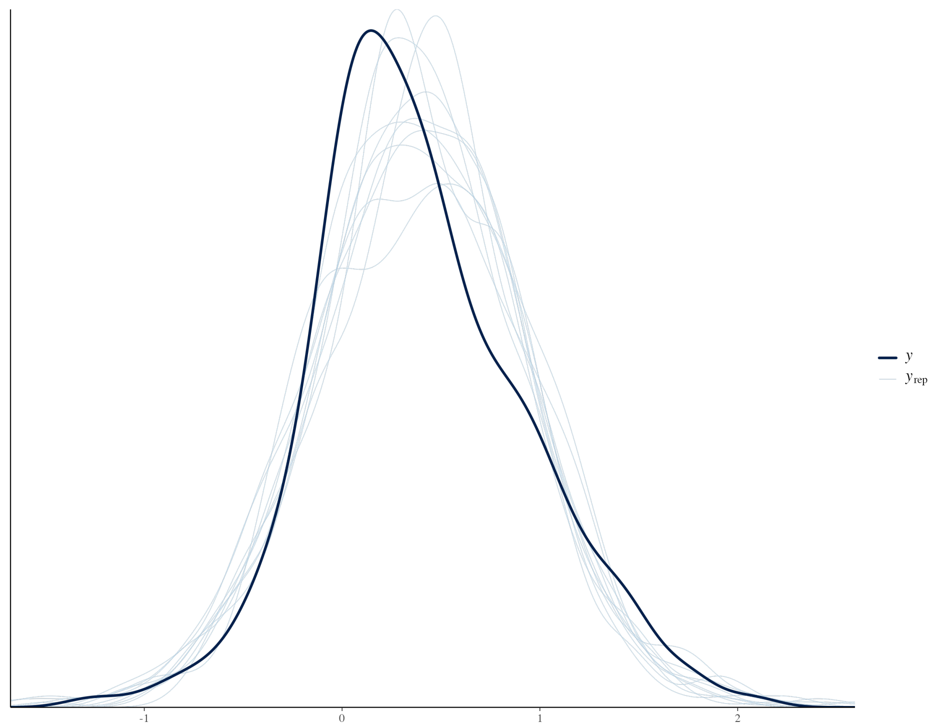


Figure 4: Idiographic analysis. Posterior Predictive Check for the hierarchical model fitted to the mean of the posterior draws of the `gamma_neg_aff` parameter.

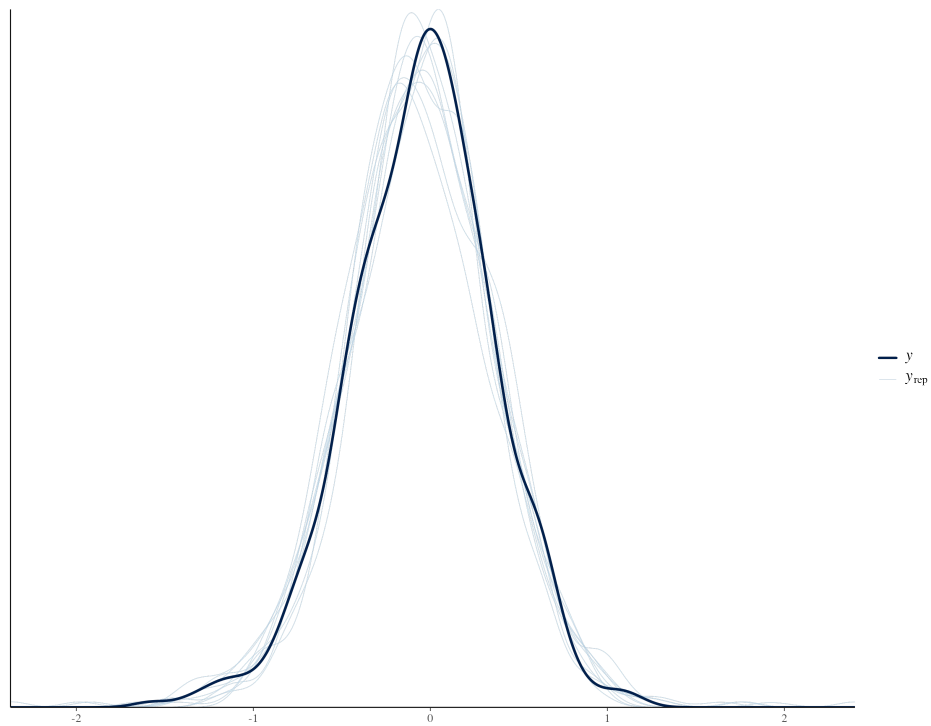


Figure 5: Idiographic analysis. Posterior Predictive Check for the hierarchical model fitted to the mean of the posterior draws of the `gamma_context` parameter.

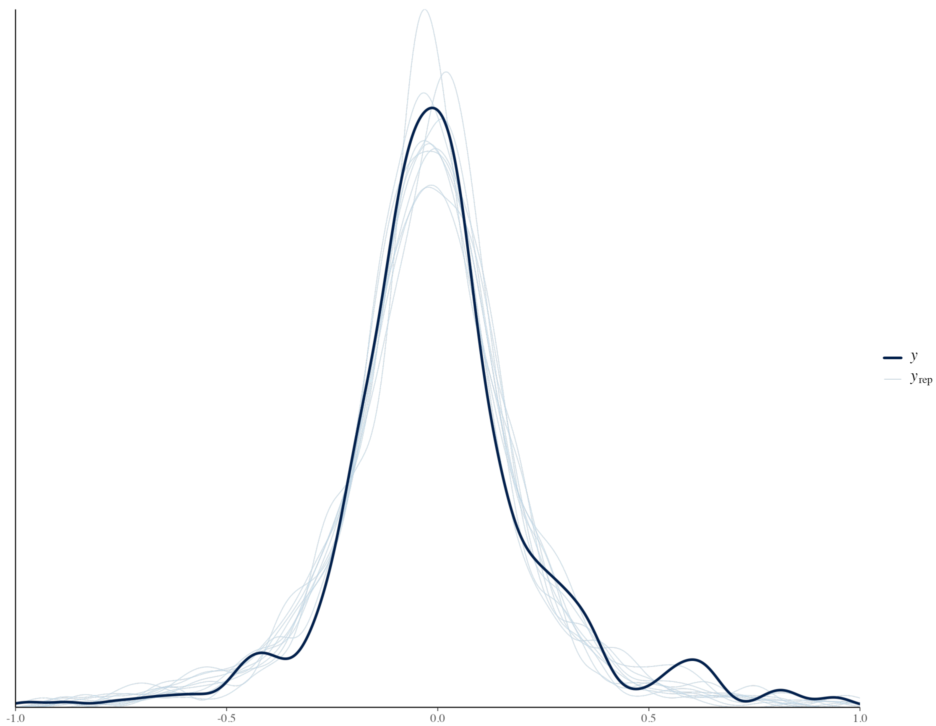


Figure 6: Idiographic analysis. Posterior Predictive Check for the hierarchical model fitted to the mean of the posterior draws of the `gamma_interaction` parameter.

# 5 Examining Response Bias

In this study, we identified a subset of participants who demonstrated a positive association between the Compassionate Self (CS) and Uncompassionate Self (UCS) components of the State Self-Compassion Scale. This positive association, characterized by higher levels of CS coinciding with higher levels of UCS, stands in direct opposition to the Bipolar Continuum Hypothesis, which posits an inverse relationship between these components. A Reviewer suggested that this finding might be attributed to response biases, such as "careless responding" or "insufficient effort." These biases could manifest as participants consistently selecting the same response option, irrespective of their actual subjective state.

To evaluate this alternative explanation, we analyzed four indices of careless responding—Longstring Index, Intra-Individual Response Variability (IRV), Even-Odd Inconsistency Index, and Mahalanobis Distance. We compared these indices between two groups: participants exhibiting a positive UCS-CS association and all other participants. To maximize statistical power, this analysis was conducted using the combined dataset from both studies ($N = 495$).

## 5.1 Analysis Steps

### 5.1.1 Step 1: Identifying Positive UCS-CS Associations

A hierarchical Bayesian model was fitted individually for each participant to estimate the slope between UCS and CS. This model accounted for the hierarchical structure of the data by incorporating random intercepts for participants, days, and measurements within days, as well as random slopes for CS and negative affect (NA) at the participant level. Interaction terms between CS and NA, as well as contextual factors (momentary, daily, and person-level evaluations), were included to account for potential modulation of UCS. A t-distribution was employed for the error term to address potential non-normality in the UCS data. Weakly informative priors ensured plausible parameter estimation, and posterior predictive checks confirmed that the model provided a good fit to the observed data.

Participants were classified as exhibiting a positive UCS-CS association if the lower bound of the 89% credible interval for the slope coefficient was positive. Based on this criterion, 52 participants (10.5% of the sample) were identified as belonging to the positive-slope group.

### 5.1.2 Step 2: Comparing Careless Responding Indices

To assess whether the positive UCS-CS associations were driven by response biases, a Bayesian multilevel analysis was conducted to compare the four indices of careless responding between the positive-slope group and the remaining participants. Separate regression models were fitted for each index, accounting for the nested structure of the data with random effects specified for participants, days, and time windows within days. These models allowed us to robustly estimate differences in systematic response behavior between the two groups.

## 5.2 Results and Interpretation

The Bayesian analyses provided posterior estimates of differences in careless responding indices between participants with positive UCS-CS associations and those without. These estimates allowed us to evaluate whether the observed positive associations reflected genuine deviations from the Bipolar Continuum Hypothesis or were artifacts of systematic response biases. By incorporating hierarchical modeling and Bayesian estimation, this analysis offered a nuanced approach to understanding the potential role of response behavior in shaping the observed patterns, contributing to the reliability and interpretability of the study's findings.

### 5.2.1 Longstring Index

The Longstring index measures the longest sequence of consecutive identical responses provided by a participant during a single measurement occasion. A high Longstring index may indicate "straightlining," a response pattern often associated with careless or inattentive behavior. If the posterior estimate of the average Longstring index is credibly higher for participants with a positive UCS-CS association, it would suggest that this group exhibits a greater tendency toward careless responding.

To investigate the relationship between group membership and the Longstring index, we employed a hierarchical Bayesian cumulative ordinal regression model using the `brm` function from the `brms` R package. The model specification is as follows:

```
mod_longstring <- brm(
  formula = longstring_val ~ is_pos_slope_group +
    (1 | user_id / day / time_window),
  family = cumulative(link = "logit"),  # Cumulative ordinal regression
  data = df_result,                     # Input data
  backend = "cmdstanr"                  # Backend for computation
)
```

**Model Details:**

- **Outcome Variable**: `longstring_val`, representing the Longstring index, is treated as an ordinal variable.
- **Predictor**: `is_pos_slope_group`, a binary variable indicating group membership:
  - `1`: Participants with a positive UCS-CS slope, where the 89% credible interval does not include zero.
  - `0`: All other participants.
- **Random Effects**: The model accounts for hierarchical clustering at three levels:
  - **User ID** (`user_id`): Captures between-subject variability.
  - **Day** (`day`): Accounts for within-subject, day-to-day variability.
  - **Time Window** (`time_window`): Addresses measurement occasion-specific variability within each day.
- **Link Function**: A cumulative logit link is used, appropriate for ordinal outcomes.

This model structure allows us to estimate the impact of group membership on the Longstring index while appropriately accounting for the nested structure of the data.

Table 15: Posterior mean, standard error, 95% credible interval and $\hat{R}$ statistic for each parameter of the constant effect model bmod1.

| Parameter | Mean | SE | Lower bound | Upper bound | Rhat |
|---|---|---|---|---|---|
| Intercept[1] | -3.205 | 0.118 | -3.448 | -2.987 | 1.016 |
| Intercept[2] | 1.316 | 0.078 | 1.164 | 1.470 | 1.006 |
| Intercept[3] | 3.660 | 0.131 | 3.422 | 3.927 | 1.014 |
| Intercept[4] | 5.105 | 0.166 | 4.802 | 5.441 | 1.015 |
| Intercept[5] | 6.372 | 0.198 | 6.009 | 6.779 | 1.014 |
| Intercept[6] | 7.118 | 0.221 | 6.712 | 7.567 | 1.013 |
| Intercept[7] | 7.444 | 0.232 | 7.006 | 7.919 | 1.014 |
| is_pos_slope_group1 | 0.047 | 0.203 | -0.355 | 0.441 | 1.000 |

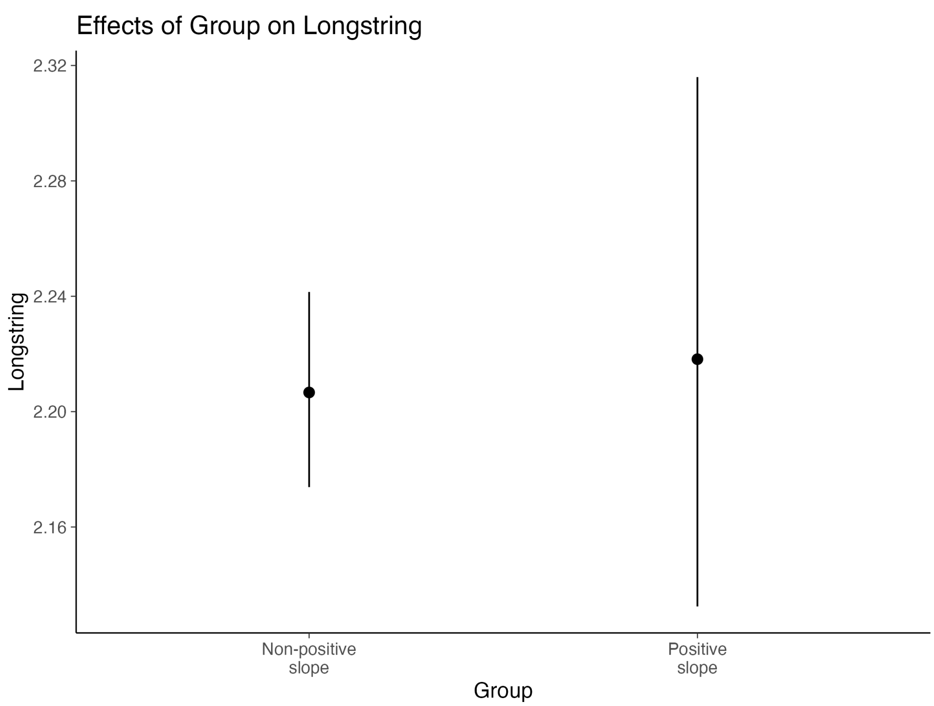


Figure 7: Conditional effect of Group on Longstring Index. Vertical bars represent 89% credibility intervals.

### 5.2.2 Intra-Individual Response Variability

Intra-Individual Response Variability (IRV) is defined as the standard deviation of responses across a set of consecutive items for a given individual (Dunn et al., 2018). A lower IRV is often interpreted as an indicator of a greater tendency toward careless responding.

### 5.2.3 Even-Odd Inconsistency Index

The Even-Odd Inconsistency Index measures response inconsistency by dividing a unidimensional scale into two halves—an even-odd split—and analyzing the relationship between the two subsets of items. The procedure is as follows:

1. **Split the scale**: Items are divided into two subsets, with even-numbered items forming one subset and odd-numbered items forming the other.

2. **Compute subscale scores**: The average response for each subset is calculated, resulting in two scores—one for the even subscale and one for the odd subscale.

3. **Calculate within-person correlation**: A correlation is computed between the two subscale scores for each participant, treating the even scores as variable $x$ and the odd scores as variable $y$, resulting in a correlation coefficient $r(x, y)$.

4. **Correct for scale length**: The correlation is adjusted for the reduced length of the scale using the Spearman–Brown prophecy formula.

5. **Compute the inconsistency score**: The final score is calculated as $0 - r(x, y)$, where higher values indicate a greater likelihood of careless responding.

For the present analysis, the Compassionate Self (CS) items were treated as odd, and the Uncompassionate Self (UCS) items were treated as even. Higher scores on this index indicate greater response inconsistency, suggesting a higher tendency toward careless responding.

### 5.2.4 Mahalanobis Distance

Mahalanobis Distance ($D^2$) is a multivariate measure used to detect outliers, including patterns that may indicate careless responding. It evaluates the distance of an individual's response set from the overall distribution of responses in the dataset. Higher $D^2$ values suggest that a participant's responses deviate markedly from the typical response pattern, potentially reflecting careless responding. Lower $D^2$ values indicate responses closer to the average, consistent with careful engagement.

## References


Aruguete, Mara S, Ho Huynh, Blaine L Browne, Bethany Jurs, Emilia Flint, and Lynn E McCutcheon. 2019. "How Serious Is the 'Carelessness' Problem on Mechanical Turk?" *International Journal of Social Research Methodology* 22 (5): 441–49.

Bennett, Marc P, Rachel Knight, Shivam Patel, Tierney So, Darren Dunning, Thorsten Barnhofer, Patrick Smith, Willem Kuyken, Tamsin Ford, and Tim Dalgleish. 2021. "Decentering as a Core Component in the Psychological Treatment and Prevention of Youth Anxiety and Depression: A Narrative Review and Insight Report." *Translational Psychiatry* 11 (1): 288.

Bernstein, Amit, Yuval Hadash, Yael Lichtash, Galia Tanay, Kathrine Shepherd, and David M Fresco. 2015. "Decentering and Related Constructs: A Critical Review and Metacognitive Processes Model." *Perspectives on Psychological Science* 10 (5): 599–617.

Bottesi, G., M. Ghisi, G. Altoè, E. Conforti, G. Melli, and C. Sica. 2015. "The Italian version of the Depression Anxiety Stress Scales-21: Factor structure and psychometric properties on community and clinical samples." *Comprehensive Psychiatry* 60: 170–81.

Brenner, Rachel E, Patrick J Heath, David L Vogel, and Marcus Credé. 2017. "Two Is More Valid Than One: Examining the Factor Structure of the Self-Compassion Scale (SCS)." *Journal of Counseling Psychology* 64 (6): 696–707.

Ciarrochi, Joseph, Baljinder Sahdra, Madeleine I Fraser, Steven C Hayes, Keong Yap, and Andrew T Gloster. 2024. "The Compassion Connection: Experience Sampling Insights into Romantic Attraction." *Journal of Contextual Behavioral Science* 32: 100749.

Cleare, Seonaid, Andrew Gumley, Chris J Cleare, and Rory C O'Connor. 2018. "An Investigation of the Factor Structure of the Self-Compassion Scale." *Mindfulness* 9: 618–28.

Ferrando, Pere J, and Urbano Lorenzo-Seva. 2018. "Assessing the Quality and Appropriateness of Factor Solutions and Factor Score Estimates in Exploratory Item Factor Analysis." *Educational and Osychological Measurement* 78 (5): 762–80.

Ferrando, Pere Joan, and Urbano Lorenzo-Seva. 2019. "An External Validity Approach for Assessing Essential Unidimensionality in Correlated-Factor Models." *Educational and Psychological Measurement* 79 (3): 437–61.

Ferrari, Madeleine, Joseph Ciarrochi, Keong Yap, Baljinder Sahdra, and Steven C Hayes. 2022. "Embracing the Complexity of Our Inner Worlds: Understanding the Dynamics of Self-Compassion and Self-Criticism." *Mindfulness* 13 (7): 1652–61.

Gratz, KL, and L Roemer. 2004. "Difficulties in Emotion Regulation Scale (DERS)." *Journal of Psychopathology and Behavioral Assessment* 26: 41–54.

Hasselhorn, Kilian, Charlotte Ottenstein, and Tanja Lischetzke. 2023. "Modeling Careless Responding in Ambulatory Assessment Studies Using Multilevel Latent Class Analysis: Factors Influencing Careless Responding." *Psychological Methods.*

Lovibond, P. F., and S. H. Lovibond. 1995. "The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories." *Behaviour Research and Therapy* 33 (3): 335–43.

Meade, Adam W, and S Bartholomew Craig. 2012. "Identifying Careless Responses in Survey Data." *Psychological Methods* 17 (3): 437.

Neff, Kristin D. 2003. "The Development and Validation of a Scale to Measure Self-Compassion." *Self and Identity* 2 (3): 223–50.

———. 2016. "The Self-Compassion Scale Is a Valid and Theoretically Coherent Measure of Self-Compassion." *Mindfulness* 7: 264–74.

Neff, Kristin D, István Tóth-Király, Lisa M Yarnell, Kohki Arimitsu, Paula Castilho, Nima

Ghorbani, Hailan Xiaoxia Guo, et al. 2019. "Examining the Factor Structure of the Self-Compassion Scale in 20 Diverse Samples: Support for Use of a Total Score and Six Subscale Scores." *Psychological Assessment* 31 (1): 27–45.

Neff, Kristin D, Tiffany A Whittaker, and Anke Karl. 2017. "Examining the Factor Structure of the Self-Compassion Scale in Four Distinct Populations: Is the Use of a Total Scale Score Justified?" *Journal of Personality Assessment* 99 (6): 596–607.

Petrocchi, Nicola, Cristina Ottaviani, and Alessandro Couyoumdjian. 2014. "Dimensionality of Self-Compassion: Translation and Construct Validation of the Self-Compassion Scale in an Italian Sample." *Journal of Mental Health* 23 (2): 72–77.

Reise, Steven P, Wes E Bonifay, and Mark G Haviland. 2013. "Scoring and Modeling Psychological Measures in the Presence of Multidimensionality." *Journal of Personality Assessment* 95 (2): 129–40.

Rodriguez, Anthony, Steven P Reise, and Mark G Haviland. 2016. "Evaluating Bifactor Models: Calculating and Interpreting Statistical Indices." *Psychological Methods* 21 (2): 137–50.

Rosenberg, M. 1965. *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Sahdra, Baljinder K, Joseph Ciarrochi, Korena S Klimczak, Jennifer Krafft, Steven C Hayes, and Michael Levin. 2024. "Testing the Applicability of Idionomic Statistics in Longitudinal Studies: The Example of 'Doing What Matters'." *Journal of Contextual Behavioral Science*, 100728.

Shiffman, Saul, Arthur A Stone, and Michael R Hufford. 2008. "Ecological Momentary Assessment." *Annual Review of Clinical Psychology* 4 (1): 1–32.

Sica, Claudio, Corrado Caudek, Silvia Cerea, Ilaria Colpizzi, Maria Caruso, Paolo Giulini, and Gioia Bottesi. 2021. "Health Anxiety Predicts the Perceived Dangerousness of Covid-19 over and Above Intrusive Illness-Related Thoughts, Contamination Symptoms, and State and Trait Negative Affect." *International Journal of Environmental Research and Public Health* 18 (4): 1933.

Sighinolfi, Cecilia, Andrea Norcini Pala, Luigi Rocco Chiri, Igor Marchetti, and Claudio Sica. 2010. "Difficulties in Emotion Regulation Scale (DERS): The Italian Translation and Adaptation." *Psicoterapia Cognitiva Comportamentale* 16 (2): 141–70.

Ulitzsch, Esther, Steffen Nestler, Oliver Lüdtke, and Gabriel Nagy. 2024. "A Screen-Time-Based Mixture Model for Identifying and Monitoring Careless and Insufficient Effort Responding in Ecological Momentary Assessment Data." *Psychological Methods*.

Welling, Jana, Rosa-Linde Fischer, and Nadja Schinkel-Bielefeld. 2021. "Is It Possible to Identify Careless Responses with Post-Hoc Analysis in EMA Studies?" In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 150–56.