# Supplementary Information

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

## Table of contents

# 1 Study 1

## 1.1 Baseline Measures

To rule out 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 designed to assess symptoms of depression (*e.g.*, "I felt down-hearted and blue"), anxiety (*e.g.*, "I felt I was close to panic"), and stress (*e.g.*, "I found it difficult to relax") over the past week. Items are rated on a 4-point Likert scale ranging from 0 (*Did not apply to me at all*) to 3 (*Applied to me very much, or most of the time*). Both the original and Italian versions (Bottesi et al. 2015) demonstrate strong psychometric properties, including adequate reliability and validity.

**Rosenberg Self-Esteem Scale (RSES).**
The RSES (Rosenberg 1965) evaluates individual self-esteem levels using a 10-item scale (*e.g.*, "I feel that I am a person of worth, at least on an equal basis with others"). Respondents rate each statement on a 4-point Likert scale from 1 (*Strongly Disagree*) to 4 (*Strongly Agree*), with higher scores indicating greater self-esteem. The scale has been widely validated and demonstrates strong reliability in both its original and translated versions.

**Self-Compassion Scale (SCS).**
The SCS (Neff 2003) is a 26-item self-report measure designed to assess self-compassion in daily life. The SCS comprises six subscales: Self-Kindness, Common Humanity, Mindfulness, Self-Judgment, Isolation, and Over-Identification. Participants rate the frequency of self-compassionate behaviors on a 5-point Likert scale ranging from 1 (*Almost Never*) to 5 (*Almost Always*). Negative items are reverse-coded, and higher scores indicate greater self-compassion. The SCS demonstrates excellent psychometric properties, with a Cronbach's alpha of 0.96 for the total score in the original validation study (Neff 2003). Test-retest reliability is also strong, with a total score correlation of $r = 0.93$, and subscale correlations ranging from $r = 0.80$ to $r = 0.88$.

**Difficulties in Emotion Regulation Scale (DERS).**
The DERS (Gratz and Roemer 2004) is a 36-item self-report measure designed to assess challenges in emotion regulation. It includes six subscales: 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 rate items on a 5-point Likert scale ranging from 0 (*Almost Never*) to 4 (*Almost Always*), with higher scores indicating greater difficulties in emotion regulation. Both the original scale and its Italian adaptation (Sighinolfi et al. 2010) have shown robust psychometric properties, including strong reliability and validity.

### 1.1.1 Descriptive Statistics

Descriptive statistics for all administered measures are presented below, including the estimated posterior mean, standard error, and 95% credibility intervals computed using a

Bayesian model. Bayesian modeling was employed to accommodate deviations from Gaussianity and provide robust parameter estimates.

| Measure | Estimate | Std. Error | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|
| **DASS-21** | | | | |
| Stress | 7.58 | 0.45 | 6.92 | 8.14 |
| Anxiety | 1.04 | 0.19 | 0.89 | 1.93 |
| 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 |
| Self-Kindness | 2.98 | 0.04 | 2.89 | 3.06 |
| Common Humanity | 3.11 | 0.04 | 3.02 | 3.20 |
| Mindfulness | 3.11 | 0.04 | 3.03 | 3.20 |
| Self-Judgment | 2.74 | 0.04 | 2.65 | 2.82 |
| Isolation | 2.74 | 0.05 | 2.65 | 2.84 |
| Over-Identification | 2.78 | 0.05 | 2.68 | 2.88 |
| **DERS** | | | | |
| Total Score | 67.89 | 2.07 | 63.79 | 71.64 |
| Nonacceptance (NER) | 6.23 | 0.43 | 5.95 | 7.02 |
| Goal Behavior (DEGB) | 13.41 | 0.78 | 12.46 | 15.78 |
| Impulse Control (ICD) | 4.99 | 0.02 | 4.95 | 5.00 |
| Emotional Awareness (LEA) | 11.52 | 0.82 | 9.91 | 12.98 |
| Regulation Strategies | 13.01 | 0.67 | 11.52 | 14.11 |
| Emotional Clarity (LEC) | 9.91 | 0.65 | 8.36 | 11.14 |

#### 1.1.2 Summary

Scores obtained from these measures are consistent with community norms reported in previous studies (Bottesi et al. 2015; Sighinolfi et al. 2010; Neff, Whittaker, and Karl 2017; Sica et al. 2021), providing no evidence of notable emotional disorders among participants.

## 1.2 EMA Survey Questions

For each notification in the Ecological Momentary Assessment (EMA) protocol, participants were prompted to respond to the following questions. These items were designed to capture participants' momentary emotional states, cognitive appraisals, and self-compassionate attitudes.

---

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

---

### 1.2.1 Self-Compassionate Attitudes

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 mean proportion of occasions flagged on two or more metrics per participant was **<2%**, 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 supplementary analysis described in Section 1.7. 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 Analysis 1: Multilevel Reliability

To compute the multilevel reliability indices presented in the manuscript, we applied a multilevel confirmatory factor analysis (MCFA) approach as described by Lai (2021). This method allows for the separation of variability at different levels—within and between subjects—and the computation of reliability indices using factor loadings and variance components.

#### 1.4.1 Reliability Indices

We computed three types of reliability measures for the CS and UCS components of state self-compassion:

1. **Between-subject reliability** ($\tilde{\omega}^b$): Reflects the scale's ability to distinguish stable individual differences across participants.
2. **Within-subject reliability** ($\tilde{\omega}^w$): Indicates the consistency of scores across repeated measures for the same participant, accounting for dynamic changes over time.
3. **Overall composite reliability** ($\omega^{2L}$): Combines within- and between-subject variances to provide a measure of the scale's overall reliability.

These indices were derived by fitting an MCFA model to the data.

#### 1.4.2 Code Implementation

The analysis was implemented in R using the `lavaan` package for structural equation modeling and the `semTools` package to compute reliability indices.

```r
get_ssc_reliabilities <- function(input_path, output_path) {

  suppressPackageStartupMessages({
    library("dplyr")
    library("lavaan")
    library("semTools")
  })

  both_df <- readRDS(input_path)

  process_ssc_data <- function(df, mcfa_model) {
    suppressWarnings({
      fit <- cfa(mcfa_model, data = df, cluster = "person")
      comp_rel <- compRelSEM(
        fit,
        obs.var = FALSE, config = c("f1"), shared = "f1"
      )
    })
    return(comp_rel)
  }
```

```
# Define the multilevel SEM model
mcfa_model <-
  "
  level: 1
  f1 =~ NA * i1 + l1 * i1 + l2 * i2 + l3 * i3 + l4 * i4
  i1 ~~ ev1w * i1
  i2 ~~ ev2w * i2
  i3 ~~ ev3w * i3
  i4 ~~ ev4w * i4
  f1 ~~ 1 * f1

  level: 2
  f1 =~ NA * i1 + l1 * i1 + l2 * i2 + l3 * i3 + l4 * i4
  # fixed residual variances
  i1 ~~ ev1b * i1
  i2 ~~    0 * i2
  i3 ~~ ev3b * i3
  i4 ~~ ev4b * i4
  f1 ~~ vf1b * f1

  # tilde omega values:
  tilomgb := (l1 + l2 + l3 + l4)^2 * vf1b /
    ((l1 + l2 + l3 + l4)^2 * vf1b + ev1b + 0 + ev3b + ev4b)
  tilomgw := (l1 + l2 + l3 + l4)^2 * 1 /
    ((l1 + l2 + l3 + l4)^2 * 1 + ev1w + ev2w + ev3w + ev4w)
  # score reliability:
  omg2l := (l1 + l2 + l3 + l4)^2 * (1 + vf1b) /
    ((l1 + l2 + l3 + l4)^2 * (1 + vf1b) +
       ev1b + 0 + ev3b + ev4b + ev1w + ev2w + ev3w + ev4w)
  omgb := (l1 + l2 + l3 + l4)^2 * vf1b /
    ((l1 + l2 + l3 + l4)^2 * vf1b + ev1b + 0 + ev3b + ev4b +
       (ev1w + ev2w + ev3w + ev4w + (l1 + l2 + l3 + l4)^2) / 25.1)
"

# Prepare data
pos_ssc_df <- both_df %>% select(user_id, starts_with("scs_pos"))
neg_ssc_df <- both_df %>% select(user_id, starts_with("scs_neg"))
colnames(pos_ssc_df) <- c("person", "i1", "i2", "i3", "i4")
colnames(neg_ssc_df) <- c("person", "i1", "i2", "i3", "i4")

# Calculate reliabilities
pos_reliabilities <- process_ssc_data(pos_ssc_df, mcfa_model)
neg_reliabilities <- process_ssc_data(neg_ssc_df, mcfa_model)

# Format and return results
results <- list(
  positive = pos_reliabilities,
```

```
    negative = neg_reliabilities
  )

  saveRDS(results, output_path)
}
```

Below is a summary of the main steps in the R script:

1. **Model Specification**:
   The MCFA model defines a single latent factor (`f1`) at both levels:
   - **Level 1 (within-subject)**: Captures variability across repeated measures within participants.
   - **Level 2 (between-subject)**: Captures stable individual differences across participants. Residual variances and factor loadings were constrained as appropriate, and reliability indices ($\tilde{\omega}^b$, $\tilde{\omega}^w$, $\omega^{2L}$) were explicitly defined within the model.
2. **Data Preparation**:
   The input dataset contains repeated measures of four items (`i1` to `i4`) for each participant. Two subsets were created: one for the CS and one for the UCS components of state self-compassion.

### 1.4.3 Results

Using this method, we obtained the following reliability estimates:

- **CS**: Between-subject ($\tilde{\omega}^b$) = 0.82, Within-subject ($\tilde{\omega}^w$) = 0.63, Composite ($\omega^{2L}$) = 0.79.
- **UCS**: Between-subject ($\tilde{\omega}^b$) = 0.88, Within-subject ($\tilde{\omega}^w$) = 0.68, Composite ($\omega^{2L}$) = 0.83.

These results demonstrate that the state self-compassion scale effectively captures both stable individual differences and dynamic within-subject variability over time.

### 1.5 Analysis 2: Correlations Between the CS and UCS Components

To estimate the multilevel correlation between the components of state self-compassion (CS and UCS), we fitted a Bayesian multilevel model using the `brms` package. The model accounts for the nested structure of the data (measurements nested within days, and days nested within participants). This approach ensures that both within- and between-person variability in the relationship between CS and UCS are accurately modeled.

#### 1.5.1 Model Specification

The model was specified as follows:

```
f_joint <- brm(
  data = d1,
  family = student,
  bf(
    mvbind(SC, USC) ~ 1 + # Include intercepts for both components
      (1 | user_id) + # Random intercept for participants
      (1 | user_id:day) + # Random intercept for days within participants
      (1 | user_id:day:time_window)
      # Random intercept for time windows within days
  ) +
    set_rescor(TRUE), # Estimate residual correlations
  iter = 8000, # Total iterations
  warmup = 2000, # Warmup iterations
  chains = 4, # Number of chains
  cores = 8, # Number of cores for parallel computation
  seed = 210191, # Seed for reproducibility
  backend = "cmdstanr" # Backend for faster computations
)
```

#### 1.5.2 Extracting Residual Correlations

The residual correlation between CS and UCS at the momentary level was extracted from the posterior draws using the following function:

```
extract_overall_correlation <- function(model) {
  draws <- as_draws_df(model)
  rescor_cols <- grep("^rescor__SC__USC", names(draws), value = TRUE)
  rescor_samples <- draws[[rescor_cols]]
  overall_correlation <- median(rescor_samples, na.rm = TRUE)
  ci_lower <- quantile(rescor_samples, 0.055, na.rm = TRUE)
  ci_upper <- quantile(rescor_samples, 0.945, na.rm = TRUE)
  data.frame(
  correlation = overall_correlation, ci_lower = ci_lower, ci_upper = ci_upper
```

```
  )
}
```

### 1.5.3 Results

The overall within-person correlation between CS and UCS was estimated to be $r = -0.48$, with an 89% credible interval (CI) of $[-0.49, -0.47]$. This moderate negative correlation reflects the inverse relationship between CS and UCS at the state level, accounting for the nested data structure.

Additionally, the lagged correlation between state CS at a given time point and state UCS at the immediately preceding time point was weaker, estimated at $r = -0.10$ with an 89% CI of $[-0.12, -0.08]$. This highlights the temporal dynamics and suggests that immediate emotional and contextual factors influence the momentary relationship between CS and UCS.

## 1.6 Analysis 3: Impact of Contextual Influences on CS and UCS

To evaluate the impact of contextual variables on state self-compassion, we developed two Bayesian hierarchical models, one for each dependent variable: UCS and CS. Both models shared identical fixed-effect and random-effect structures, with predictors including negative affect (person, day, moment) and context evaluation (person, day, moment). This approach allowed us to examine state self-compassion dynamics across individuals, days, and moments.

### 1.6.1 Model Selection Process

A systematic model selection process was undertaken to identify the optimal random- and fixed-effect structures. Initially, we tested random-effect configurations ranging from simple participant-level clustering to more complex structures incorporating nested effects and random slopes. Using the Leave-One-Out Information Criterion (LOOIC) for model comparison, we identified that random intercepts for participants and days provided the best balance between model complexity and predictive accuracy.

For the fixed effects, we tested configurations that included all predictors (negative affect and context) at all levels (person, day, moment). The best-fitting model included all fixed effects, demonstrating superior predictive performance.

### 1.6.2 Final Models

Both models included fixed effects for negative affect and context at the person, day, and moment levels, along with random intercepts for participants and days.

### 1.6.3 Model Diagnostics

We checked convergence using R-hat values, which were close to 1 for all parameters, indicating proper convergence. Posterior predictive checks demonstrated good model fit for both CS and UCS, with simulated data closely resembling observed data distributions.

### 1.6.4 brms Code

The final models were specified as follows, using the `brms` package:

```
# Model for UCS
model_ucs <- brm(
  bf(usc ~ na_moment + na_day + na_person +
         context_moment + context_day + context_person +
         (1 | user_id) + (1 | user_id:day),
     sigma ~ 1),
  data = dataset,
  family = student(),
  chains = 4,
```

```
  cores = 8,
  iter = 4000,
  warmup = 1000,
  control = list(adapt_delta = 0.95, max_treedepth = 12),
  backend = "cmdstanr"
)

# Model for CS
model_cs <- brm(
  bf(cs ~ na_moment + na_day + na_person +
         context_moment + context_day + context_person +
         (1 | user_id) + (1 | user_id:day),
     sigma ~ 1),
  data = dataset,
  family = student(),
  chains = 4,
  cores = 8,
  iter = 4000,
  warmup = 1000,
  control = list(adapt_delta = 0.95, max_treedepth = 12),
  backend = "cmdstanr"
)
```

## 1.7 Supplementary Analysis: Nomothetic Evaluation of the CS-UCS Relationship

One objective of this study was to test Hypothesis **H4**, examining the association between the CS and UCS components of state self-compassion within the multilevel structure of our EMA data. While the idionomic approach used in the manuscript provides a detailed individual-level analysis, this supplementary section presents a traditional nomothetic analysis to align with prior literature.

### 1.7.1 Rationale

Multilevel models are commonly used to address questions like H4, but they often assume psychological homogeneity and aggregate data toward group-level trends, potentially obscuring individual variability and temporal dynamics (Sahdra et al., 2024; Gelman, Hill, & Vehtari, 2020). To address these concerns, the manuscript used an idionomic approach combining data from both studies. However, for completeness, this section employs Bayesian hierarchical models separately for each study to assess the UCS-CS relationship under a nomothetic framework.

### 1.7.2 Method

The Bayesian hierarchical model predicted UCS from CS while including the following covariates: negative affect and context evaluation. Random intercepts at the participant, day, and measurement levels captured individual and temporal variability. The model also incorporated random slopes for CS to reflect individual differences in the UCS-CS relationship. Predictors were scaled to enable consistent interpretation, and a Student's t-distribution was used for UCS to handle potential outliers.

The model's fixed effects were tested to evaluate the Bipolar Continuum Hypothesis. A credible negative slope for CS would support the hypothesis, indicating a bipolar relationship, while a positive or non-credible slope would challenge it.

### 1.7.3 Results

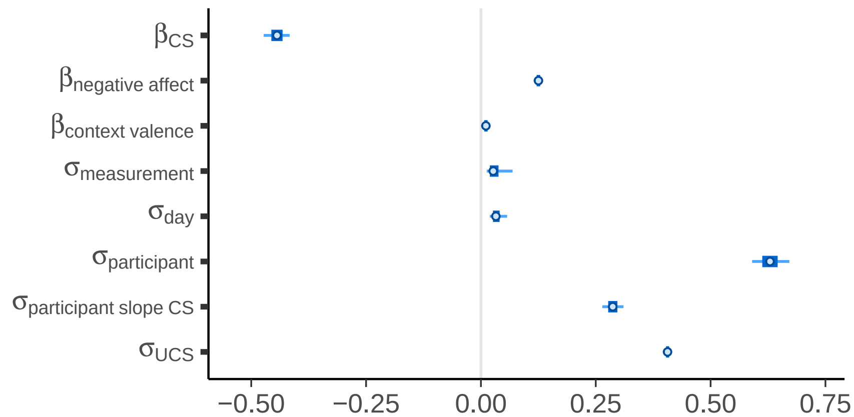


Figure 1: The estimates are expressed in terms of inter-individual differences and intra-individual variations, both within a day and across different days. The bars represent 89% credibility intervals. $\beta_{CS}$ represents the beta coefficients for CS; $\beta_{\text{negative affect}}$ and $\beta_{\text{context valence}}$ 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 analysis revealed a robust negative association between CS and UCS, with a median slope of -0.44 (89% CI [-0.47, -0.42]), consistent with the Bipolar Continuum Hypothesis.

- **Covariates**: Negative affect positively influenced UCS ($\beta = 0.12$, 89% CI [0.116, 0.13]), while context evaluation had minimal impact ($\beta = 0.01$, 89% CI [0.002, 0.02]).

- **Random Slopes**: The relationship between CS and UCS varied significantly across participants (median slope variability: 0.29, 89% CI [0.26, 0.32]), indicating substantial individual differences.

- **Random Effects**: UCS levels showed considerable variability across participants ($\beta = 0.63$, 89% CI [0.59, 0.67]) and moderate day-to-day fluctuations ($\beta = 0.04$, 89% CI [0.02, 0.06]).

- **Measurement Error**: Minimal error (median estimate: 0.01, 89% CI [0.001, 0.039]) demonstrated the reliability of the data.

These findings align with the Bipolar Continuum Hypothesis, showing that higher CS levels correspond to lower UCS levels. The variability in slopes highlights the complexity of the CS-UCS relationship, influenced by personality and contextual factors.

### 1.7.4 Model Specification

We implemented a hierarchical Bayesian model to account for participant-level, day-level, and measurement-level variability. The model included random effects and used the following prior distributions:

#### 1.7.4.1 Priors

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

#### 1.7.4.2 Likelihood

The model assumes that the observed values of ( UCS ) follow a Student-( t ) distribution. The parameters account for fixed effects (e.g., ( *{ucs}* ), ( {cs} )), random effects, and measurement error:

$$\text{UCS}[n] \sim \text{Student-}t\left(\nu, \mu_n, \sigma_{ucs}\right),$$

where:

$$\begin{aligned}
\mu_n = \alpha_{ucs} & \\
&+ \beta_{cs} \cdot \text{CS}[n] \\
&+ \beta_{neg_affect} \cdot \text{neg_affect}[n] \\
&+ \beta_{context_eval} \cdot \text{context_eval}[n] \\
&+ \sigma_{participant} \cdot z_{participant}[\text{participant}[n]] \\
&+ \sigma_{day} \cdot z_{day}[\text{day}[n]] \\
&+ \sigma_{measurement} \cdot z_{measurement}[\text{measurement}[n]] \\
&+ \sigma_{participant_slope_cs} \cdot z_{participant_slope_cs}[\text{participant}[n]].
\end{aligned}$$

#### 1.7.4.3 Stan Code

The full implementation in Stan is shown below:

```
data {
  int<lower=0> N; // Total observations
  int<lower=0> P; // Number of participants
  int<lower=0> D; // Number of days
  int<lower=0> M; // Number of measurements
  array[N] int<lower=1, upper=P> participant; // Participant index
```

```
  array[N] int<lower=1, upper=D> day;          // Day index
  array[N] int<lower=1, upper=M> measurement; // Measurement index
  array[N] real CS;                            // Predictor: CS
  array[N] real UCS;                           // Response: UCS
  array[N] real neg_affect;                    // Covariate: Negative affect
  array[N] real context_eval;                  // Covariate: Context evaluation
}

parameters {
  real alpha_ucs;                          // Intercept
  real beta_cs;                            // Effect of CS
  real beta_neg_affect;                    // Effect of negative affect
  real beta_context_eval;                  // Effect of context evaluation
  vector[P] z_participant;                 // Participant random effects
  vector[D] z_day;                         // Day random effects
  vector[M] z_measurement;                 // Measurement random effects
  vector[P] z_participant_slope_cs;        // Random slopes for CS by participant
  real<lower=0> sigma_participant;         // SD for participant random effects
  real<lower=0> sigma_day;                 // SD for day random effects
  real<lower=0> sigma_measurement;         // SD for measurement random effects
  real<lower=0> sigma_participant_slope_cs; // SD for CS slopes
  real<lower=0> sigma_ucs;                 // Measurement noise
  real<lower=0> nu;                        // Degrees of freedom for Student-t
}

model {
  // Priors
  alpha_ucs ~ normal(0, 1);
  beta_cs ~ normal(0, 1);
  beta_neg_affect ~ normal(0, 1);
  beta_context_eval ~ 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);

  // Likelihood
  for (n in 1:N) {
    UCS[n] ~ student_t(
      nu,
      alpha_ucs +
```

```
        beta_cs * CS[n] +
        beta_neg_affect * neg_affect[n] +
        beta_context_eval * context_eval[n] +
        sigma_participant * z_participant[participant[n]] +
        sigma_day * z_day[day[n]] +
        sigma_measurement * z_measurement[measurement[n]] +
        sigma_participant_slope_cs * z_participant_slope_cs[participant[n]],
      sigma_ucs
    );
  }
}
```

# 2 Study 2

## 2.1 Baseline Measures

The table below presents the descriptive statistics for the psychological assessment measures administered in Study 2. The results include the estimated posterior mean, standard error, and 95% credibility intervals, computed using a Bayesian model to account for potential deviations from Gaussian assumptions.

| Measure | Estimate | Std. Error | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|
| **DASS-21** | | | | |
| Stress | 7.24 | 0.73 | 6.10 | 9.37 |
| Anxiety | 3.43 | 1.55 | 0.95 | 6.32 |
| 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 |
| Self-Kindness | 2.88 | 0.10 | 2.70 | 3.07 |
| Common Humanity | 3.00 | 0.09 | 2.83 | 3.17 |
| Mindfulness | 3.07 | 0.09 | 2.89 | 3.23 |
| Self-Judgment | 2.69 | 0.09 | 2.51 | 2.87 |
| Isolation | 2.69 | 0.09 | 2.50 | 2.87 |
| Over-Identification | 2.67 | 0.09 | 2.50 | 2.85 |

## 2.2 Summary

The results indicate no evidence of emotional disorders among participants. The scores align closely with those reported in prior studies that utilized these assessment tools in community samples (Bottesi et al. 2015; Neff, Whittaker, and Karl 2017; Sica et al. 2021).

### 2.3 EMA Survey Questions

In the Ecological Momentary Assessment (EMA) protocol for Study 2, participants were prompted to answer the following questions at each notification. These items were designed to capture momentary emotional states, cognitive appraisals, self-compassion, and mindfulness-related processes.

---

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

---

### 2.3.1 Self-Compassion and Mindfulness-Related Processes

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

### 2.4 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.

#### 2.4.1 Compliance Rate

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

#### 2.4.2 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.

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

##### 2.4.3.1 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.

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

One participant exceeded the thresholds on more than two indices during the EMA phase. Consequently, this participant was excluded from the reanalysis of the data, as detailed in Section 2.4.4.

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

As in Study 1, to complement the participant-level analysis, we conducted an occasion-level analysis to capture instances of momentary inattentiveness.

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.

##### 2.4.3.2.1 Results

- **Distribution of Flagged Occasions**:
  - **79.3%** of occasions were never flagged by any metric.
  - **18.9%** of occasions were flagged on one metric, indicating mild evidence of momentary inattention.
  - **1.19%** 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 mean proportion of occasions flagged on two or more metrics per participant was **<2%**, reflecting low overall incidence of momentary inattentiveness;
  - no participant showed persistently high flagging rates across occasions, underscoring the reliability of the dataset.

#### 2.4.3.3 Final Assessment

Also in Study 2, both the person-level and occasion-level analyses showed a low incidence of inattentive or careless responding.

### 2.4.4 Data Reanalysis

To evaluate whether excluding data flagged at the person-level and at the occasion-level, we conducted a secondary analysis of the Study 2 data. As for Study 1, 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 2.

Also for Study 2, the reanalysis produced results that were effectively identical to those derived from the full dataset. 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 | 0.107 | 0.0526 | 0.0226 | 0.192 |
| beta_cs | -0.429 | 0.0250 | -0.469 | -0.389 |
| beta_negative_affect | 0.0562 | 0.00742 | 0.0444 | 0.0682 |
| beta_context_valence | -0.00365 | 0.00652 | -0.0141 | 0.00683 |
| beta_decentering | -0.108 | 0.00718 | -0.120 | -0.0969 |

### 2.5 Analysis 1: Impact of Academic Exam on State Self-Compassion

We employed a hierarchical Bayesian model to compare the CS and UCS components of state self-compassion across three time periods: the day before an academic exam (pre-exam), the day after the exam (post-exam), and a baseline period consisting of notifications collected on non-exam days distant from 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
    );
  }
}
```

The hierarchical Bayesian model was implemented separately for each component of state self-compassion: once with CS as the dependent variable and once with UCS as the dependent variable (`sc`).

## 2.6 Supplementary Analysis: Impact of Academic Exam on Negative Affect

We examined how stress levels, measured as negative affect, changed based on the timing of assessments relative to exams. Assessments were categorized into three time points: immediately before an exam, immediately after an exam, and during periods unrelated to exams. This approach allowed us to compare average levels of negative affect across these three periods.

Students participated in two exams during the study. The results showed a clear and consistent pattern: negative affect decreased significantly from the day before to the day after each exam.

**Exam 1.** For the first exam, we observed a substantial reduction in negative affect the day following the exam compared to the day prior. The standardized decrease was -0.92 ($SE = 0.10$), corresponding to a Cohen's $d$ of -0.98. The 89% credible interval (CI) ranged from -1.23 to -0.78, indicating a strong and reliable effect.

**Exam 2.** Similarly, for the second exam, the standardized decrease in negative affect was smaller but still notable, at -0.39 ($SE = 0.08$). This corresponds to a Cohen's $d$ of -0.54, with an 89% CI of [-0.75, -0.36].

These findings indicate that the timing of assessments relative to exam periods influenced participants' stress levels, as reflected by changes in negative affect. Specifically, assessments conducted immediately after exams consistently showed lower stress levels compared to those conducted before exams. This supports the effectiveness of the study design in capturing the expected temporal variation in stress levels among students.

### 2.7 Analysis 2: Impact of Contextual Influences on CS and UCS

In Study 2, we evaluated the impact of contextual variables on state self-compassion using two Bayesian hierarchical models, one for each dependent variable: UCS and CS. Both models incorporated fixed effects for negative affect, context, and decentering, each measured at the person, day, and moment levels, as well as random intercepts to account for variability across participants and days.

#### 2.7.1 brms Code

The final models were specified as follows, using the `brms` package:

```
# Model for UCS
model_ucs <- brm(
  bf(usc ~ na_moment + na_day + na_person +
         context_moment + context_day + context_person +
         dec_moment + dec_day + dec_person +
         (1 | user_id) + (1 | user_id:day),
     sigma ~ 1),
  data = dataset,
  family = student(),
  chains = 4,
  cores = 8,
  iter = 4000,
  warmup = 1000,
  control = list(adapt_delta = 0.95, max_treedepth = 12),
  backend = "cmdstanr"
)

# Model for CS
model_cs <- brm(
  bf(cs ~ na_moment + na_day + na_person +
         context_moment + context_day + context_person +
         dec_moment + dec_day + dec_person +
         (1 | user_id) + (1 | user_id:day),
     sigma ~ 1),
  data = dataset,
  family = student(),
  chains = 4,
  cores = 8,
  iter = 4000,
  warmup = 1000,
  control = list(adapt_delta = 0.95, max_treedepth = 12),
  backend = "cmdstanr"
)
```

## 2.8 Supplementary Analysis: Nomothetic Evaluation of the CS-UCS Relationship

We extended the main analyses by conducting a supplementary test of the Bipolar Continuum Hypothesis using a nomothetic Bayesian hierarchical model. This approach replicated the framework described in Section 1.7, allowing direct comparisons across studies.

In Study 2, UCS was modeled as a linear function of CS, with momentary negative affect, decentering, and event unpleasantness included as covariates. Random effects at the participant, day, and measurement levels accounted for individual variability and temporal fluctuations. This hierarchical structure enabled the simultaneous examination of both inter-individual differences and intra-individual changes over time.

### 2.8.1 Results

The findings supported the Bipolar Continuum Hypothesis, mirroring those observed in Study 1 (see Figure 2).

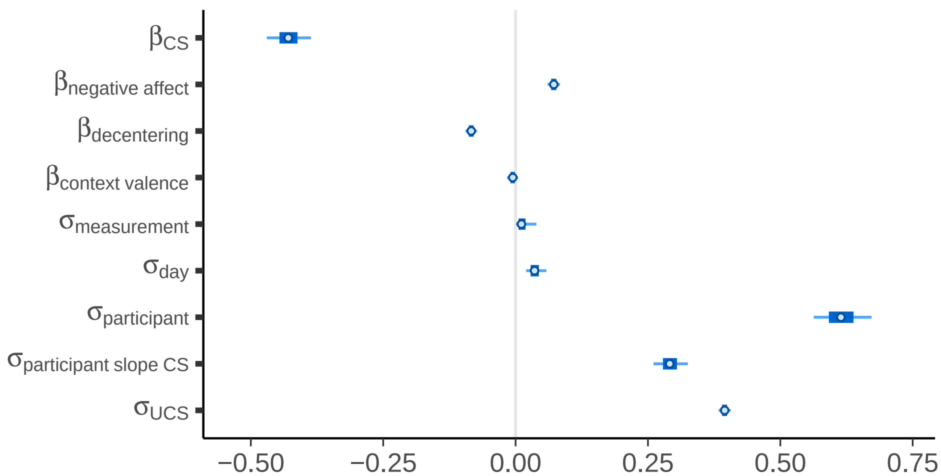


Figure 2: The estimates reflect both inter-individual differences and intra-individual variations across days and moments. The bars represent 89% credibility intervals. $\beta_{CS}$ denotes the coefficient for CS; $\beta_{\text{negative affect}}$, $\beta_{\text{context valence}}$, and $\beta_{\text{decentering}}$ represent coefficients for negative affect, contextual valence, and decentering, respectively; $\sigma_{\text{measurement}}$, $\sigma_{\text{day}}$, and $\sigma_{\text{participant}}$ correspond to the standard deviations of random effects across measurements, days, and participants. $\sigma_{\text{participant slope CS}}$ captures the variability in CS slopes across participants; $\sigma_{UCS}$ reflects residual population-level variability.

The analysis confirmed a strong negative association between CS and UCS ($\beta = -0.43$, 89% CI [-0.47, -0.39]), reinforcing the bipolar nature of the construct. Negative affect had a small positive effect on UCS ($\beta = 0.07$, 89% CI [0.06, 0.08]), while decentering showed an inverse relationship ($\beta = -0.08$, 89% CI [-0.09, -0.07]). Event unpleasantness had little impact on UCS, with its 89% credible interval including zero ($\beta = -0.005$, 89% CI [-0.016, 0.005]).

Random effects revealed substantial variability between participants ($\sigma_{\text{participant}} = 0.61$, 89% CI [0.56, 0.67]), reflecting meaningful individual differences in self-compassion responses. Day-to-day fluctuations in UCS were minimal ($\sigma_{\text{day}} = 0.04$, 89% CI [0.02, 0.06]), and measurement error was negligible ($\sigma_{\text{measurement}} = 0.01$, 89% CI [0.001, 0.039]). Participant-level variability in the effect of CS on UCS was moderate ($\sigma_{\text{participant slope CS}} = 0.29$, 89% CI [0.26, 0.32]). Residual variability in UCS ($\sigma_{\text{UCS}} = 0.39$, 89% CI [0.38, 0.41]) indicated that additional unmeasured factors may contribute to UCS.

In conclusion, the results from Study 2 provide further evidence for the Bipolar Continuum Hypothesis. The negative association between CS and UCS underscores the conceptualization of state self-compassion as a bipolar construct. Contextual influences, such as negative affect and decentering, played a key role in shaping momentary positions along this continuum. While temporal fluctuations in UCS were small, substantial between-participant variability highlights the importance of individual differences in state self-compassion dynamics.

### 2.9 Analysis 3: Levels of Personal Concern

This section describes the statistical modeling approach used to test the hypothesis that the relationship between the CS and UCS components of state self-compassion varies depending on the level of personal concern, operationalized by proximity to an academic exam (**H3**). The analysis included data from three distinct time points:

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.

The model incorporated a multivariate response structure for CS and UCS, with random effects accounting for within- and between-subject variability across multiple time scales. The model specification was as follows:

```
f_joint <- brm(
  data = alldata,
  family = student,
  bf(
    mvbind(SC, USC) ~ 0 + exam_day +
      (1 + exam_day | user_id) +
      (1 | user_id:day) +
      (1 | user_id:day:time_window),
    sigma ~ 0 + exam_day
  ) +
    set_rescor(TRUE),
  iter = 8000,
  warmup = 2000,
  chains = 4,
  cores = 8,
  seed = 210191,
  backend = "cmdstanr"
)
```

This model explicitly included fixed effects for the variable `exam_day`, reflecting the proximity to the exam (e.g., pre-exam, post-exam, or baseline), and allowed random slopes for `exam_day` at the participant level (`user_id`). Additional random intercepts were specified for `user_id:day` and `user_id:day:time_window`, capturing nested within-day and within-time-window variability. The residual correlation structure between CS and UCS was estimated using the `set_rescor(TRUE)` option, and heteroscedasticity was modeled through the inclusion of `sigma ~ 0 + exam_day`.

The model was fit using the `cmdstanr` backend in `brms`, employing a robust Student's t-distribution for residuals to account for potential outliers. The sampling process utilized 4 chains, 8000 iterations (2000 warmup), and 8 cores, ensuring robust convergence diagnostics and precise parameter estimation.

### 2.10 Analysis 4: Decentering and SC and USC Correlation

The Bipolar Continuum Hypothesis posits a stable inverse relationship between the CS (Compassionate Self) and UCS (Uncompassionate Self) components of state self-compassion. In this analysis, decentering—a mindfulness-related process—was examined as a potential moderator of this relationship. The purpose was to investigate whether mindfulness processes introduce flexibility into the CS-UCS coupling, suggesting that the strength of the bipolar structure may vary across individuals or contexts.

Decentering, characterized by the non-judgmental observation of thoughts and emotions, is theorized to promote psychological resilience. It reframes negative thoughts and emotions as transient experiences rather than fixed aspects of identity, fostering acceptance and self-compassion. Specifically, decentering was expected to enhance CS by encouraging kindness and acceptance of negative emotions while reducing UCS by mitigating self-critical and over-identifying tendencies. If decentering moderates the CS-UCS relationship, it would provide evidence that mindfulness processes reinforce the bipolar structure by strengthening the inverse coupling between these two poles.

#### 2.10.1 Hypotheses and Measurement

The analysis hypothesized that higher levels of decentering would be associated with a stronger inverse relationship between CS and UCS (**H2**). This moderated relationship reflects greater psychological flexibility, with decentering enhancing self-compassion and reducing uncompassionate responses.

To assess decentering, four items adapted from prior EMA research (Bennett et al. 2021; Bernstein et al. 2015) were incorporated into Study 2. These items enabled a nuanced evaluation of how mindfulness processes shape self-compassion components.

#### 2.10.2 Bayesian Hierarchical Model

To test this hypothesis, we employed a Bayesian hierarchical regression model predicting UCS as a function of CS, decentering, and their interaction. This model accounted for variability across three levels—person, day, and moment—capturing both individual differences and temporal dynamics. The interaction term (CS $\times$ Decentering) was central to determining whether decentering moderates the CS-UCS relationship.

The model specification is detailed below:

```
model_interaction <- brm(
  bf(UCS ~ dec * CS + (1 + CS | user_id)) +
    bf(CS ~ dec + (1 | user_id)),
  data = d1,
  family = student(),
  chains = 4,
  cores = 8,
  threads = threading(2),
  backend = "cmdstanr",
```

```
  control = list(adapt_delta = 0.95, max_treedepth = 12)
)
```

- **Fixed Effects:** UCS was predicted by CS, decentering, and their interaction to test moderation.
- **Random Effects:** Random intercepts and slopes for CS were included at the person, day, and moment levels to capture variability in the CS-UCS relationship across individuals and time points.
- **Student's t-Distribution:** A robust Student's t-distribution was used for UCS to account for potential outliers.

### 2.10.3 Results and Interpretation

The analysis yielded the following findings:

1. **Person-Level Effects:**
   The interaction between CS and decentering was negative ( = -0.05; 89% CI: [-0.08, -0.02]), indicating that individuals with higher levels of decentering showed a stronger inverse relationship between CS and UCS. This supports the hypothesis that decentering reinforces the bipolar structure of self-compassion at a trait level, suggesting that mindfulness-related processes stabilize the CS-UCS coupling.

2. **Day- and Moment-Level Effects:**
   At the day and moment levels, the interaction effects were close to zero, with 89% credible intervals including zero. These findings suggest that the moderating role of decentering is not evident at shorter timescales and is more relevant to stable, trait-like patterns of self-compassion rather than transient fluctuations.

### 2.10.4 Contribution to the Test of the Bipolar Continuum Hypothesis

This analysis provides important evidence supporting the Bipolar Continuum Hypothesis. By demonstrating that decentering strengthens the inverse relationship between CS and UCS at the person level, the findings validate the hypothesis that self-compassion operates as a dynamic yet stable bipolar construct. The results further highlight the role of mindfulness processes, such as decentering, in reinforcing the bipolar structure and promoting psychological flexibility.

Decentering's impact on the CS-UCS relationship underscores its potential as a mindfulness-related mechanism for fostering self-compassion. While day-to-day and moment-to-moment effects were minimal, the trait-level findings suggest that decentering contributes to resilience by stabilizing self-compassion components over time. These insights align with the broader research aim of testing and validating the Bipolar Continuum Hypothesis while emphasizing the importance of mindfulness processes in self-compassion dynamics.

# 3 Both Studies Combined

## 3.1 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.

### 3.1.1 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;
```

#### 3.1.1.1 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 |

#### 3.1.1.2 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

#### 3.1.2 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;
```

#### 3.1.2.1 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 |

#### 3.1.2.2 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

### 3.1.3 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);
```

#### 3.1.3.1 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)** | | |

| Item | Within-Level Standardized Loadings | Between-Level Standardized Loadings |
|---|---|---|
| **SCN2** | 0.473 | 0.626 |
| **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.

#### 3.1.3.2 Interpretation

**General Factor.** The **general factor** exhibited strong and consistent loadings across items, particularly at the between-subject level (ranging from 0.143 to 0.935), reflecting a robust overarching construct of self-compassion across individuals. The within-subject loadings were also notable, albeit slightly weaker (0.210 to 0.623). This suggests that the general factor captures substantial common variance among items both across and within individuals.

- Items SCP7 ("At this moment I am able to accept my flaws and weaknesses"; between-level loading: 0.935) and SCP1 ("I'm giving myself the caring and tenderness I need"; 0.860) demonstrated the strongest loadings at the between-subject level, indicating their critical role in representing the general self-compassion construct.
- SCN8 ("At this moment I let myself be carried away by my emotions") exhibited weak loadings on the general factor at both levels (within: 0.210; between: 0.143), implying limited shared variance with the general factor and potentially unique characteristics for this item.

**Specific Factors.** The **positive and negative factors** captured variance unique to the positive and negative components of self-compassion, beyond what is explained by the general factor.

1. **Positive Factor**:

    - Within-subject loadings for SCP1 and SCP7 were modest (0.219 and 0.246, respectively), while SCP3 ("I'm remembering that there are lots of others in the world feeling like I am") and SCP6 ("I'm keeping things in perspective") had negligible or negative loadings (-0.313 and -0.009, respectively). This suggests limited within-subject differentiation of the positive factor from the general factor.
    - Between-subject loadings were also modest but slightly stronger for SCP7 (0.322) and SCP1 (0.215), highlighting these items' modest roles in capturing positive-specific variance at the between-person level.

2. **Negative Factor**:

    - Negative items (e.g., SCN2 ("I'm obsessing and fixating on everything that's wrong"), SCN4 ("I feel like I'm struggling more than others right now"), SCN5 ("I feel intolerant and impatient toward myself"), SCN8 ("At this moment I let myself be carried away by my emotions")) had moderate loadings on the negative factor at both levels, with slightly higher between-subject loadings (e.g., SCN2: 0.626, SCN4: 0.378). This indicates that the negative-specific factor captures meaningful unique variance, particularly across individuals.

**Implications.**

- The strong loadings on the general factor, especially at the between-subject level, suggest that state self-compassion is predominantly a unidimensional construct at the higher level.
- The specific factors capture additional nuance, with the negative-specific factor demonstrating relatively more consistent loadings than the positive-specific factor. This may reflect a more distinct role for negative-specific items in characterizing individual differences in state self-compassion.
- Weak or negative loadings on specific factors (e.g., SCP3 on the positive factor) suggest that some items primarily contribute to the general factor, reinforcing the overarching nature of state self-compassion.

In summary, the bifactor model successfully captures both the shared variance across all state self-compassion items and the unique variance attributable to positive and negative components. The results emphasize the predominance of a general state self-compassion factor while acknowledging the contribution of specific dimensions, particularly for the negative-specific factor.

### 3.1.4 Goodness-of-fit indices

The goodness-of-fit indices for each of the three multilevel CFA models described in the previous sections 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.

#### 3.1.5 Practical Importance

While the previous section focused on the psychometric properties of the three multilevel CFA models under consideration, it is essential to emphasize that dimensionality assessments should not rely exclusively on statistical fit indices. Instead, they should also account for the substantive relevance and interpretability of each factor (Rodriguez, Reise, and Haviland 2016; Pere J. Ferrando and Lorenzo-Seva 2018). Although more complex models often achieve superior statistical fit, they risk overfitting the data, particularly when an excessive number of factors are included (Pere Joan Ferrando and Lorenzo-Seva 2019). Striking a balance between model complexity and interpretability is therefore critical to ensure that the conclusions drawn are both meaningful and practically applicable.

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 ($\text{Gen}_w$) 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 ($\text{POS}_w$ and $\text{NEG}_w$) 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., $\text{POS}_w$ on SCP3 = -0.313, $\text{NEG}_b$ 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.

## 3.2 Idionomic Analysis

The analyses presented in the previous section support the "essential unidimensionality" of state self-compassion (Reise, Bonifay, and Haviland 2013). Using a multilevel framework that accounted for repeated measurements over a three-month period, we showed that compassionate and uncompassionate self-responding (CS and UCS) are inversely related at the *nomothetic level*—when examined across the entire sample. These findings align with Neff's hypothesis of a bipolar continuum, suggesting that CS and UCS function as opposing dimensions.

Our psychometric evaluations, relying on *internal criteria* such as fit indices for item scores, showed that more complex models improved statistical fit. However, these improvements did not substantially alter the overarching conclusion. While such models revealed minor deviations, their practical importance was limited, and the core pattern remained consistent with the bipolar continuum hypothesis.

Further validation using *external criteria*—specifically, the effects of covariates such as momentary negative affect and context evaluation, as analyzed in the two studies presented in this manuscript—confirmed that these situational factors did not alter the core inverse relationship between CS and UCS. These results collectively strengthen the evidence for the robustness of the proposed bipolar continuum.

Nevertheless, it should be noted that all these prior analyses were conducted at a *nomothetic level*, focusing on group-level patterns. This aggregated perspective, while informative, risks obscuring the ways self-compassion may vary across individuals and contexts. In fact, the dynamics of self-compassion in everyday life—how CS and UCS interact within the unique contexts of individual experiences—may not be fully captured when averaged across the sample. Understanding this variability is crucial for a more complete picture of how self-compassion operates.

To address this limitation, we conducted an *idionomic analysis* to investigate the relationship between CS and UCS at the individual level (Ciarrochi et al. 2024; Ferrari et al. 2022; Sahdra et al. 2024). This approach shifts the focus from group-level averages to person-specific dynamics, enabling the exploration of heterogeneity in self-compassion processes. By examining the interaction of CS and UCS within individuals over time, the idionomic analysis provides a more nuanced understanding of how these constructs function in daily life and reveals patterns that may be masked in nomothetic analyses.

### 3.2.1 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 separatedly, 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 $\times$ 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.

#### 3.2.1.1 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.

### 3.2.2 Discussion

Our idionomic analysis offers critical insights into the momentary relationship between CS and UCS, providing an individual-level test of Neff's Bipolar Continuum Hypothesis. The group-level findings largely supported the Bipolar Continuum Hypothesis, indicating that CS and UCS function as inversely related dimensions for most participants. However, the model also revealed substantial variability in this relationship across individuals, suggesting that the Bipolar Continuum may not universally apply.

The majority of participants demonstrated a negative association between CS and UCS, consistent with the Bipolar Continuum Hypothesis. However, a notable subset exhibited either no clear association or even positive correlations, where CS and UCS varied together. These findings challenge the universality of the Bipolar Continuum Hypothesis, suggesting that self-compassion does not always function as two strictly opposing dimensions. Such variability underscores the importance of idiographic analyses, which can reveal nuanced patterns obscured by aggregate, group-level approaches (Ferrari et al., 2023; Ullrich et al., 2020).

The model described below provided a test of contextual and temporal influences on the CS-UCS relationship. Negative affect was associated with higher UCS levels, indicating that heightened negative emotional states amplify uncompassionate self-responding. However, there was no credible evidence that negative affect moderated the inverse relationship between CS and UCS, suggesting that momentary emotional states do not alter how CS influences UCS. Similarly, while context evaluations were linked to UCS levels, they did not meaningfully disrupt the CS-UCS dynamic, supporting the stability of the bipolar structure.

The inclusion of lagged CS as a predictor allowed for the examination of temporal dependencies within the same day. The findings suggest that prior CS levels had a modest influence on UCS, indicating some degree of temporal continuity in self-compassion processes. However, the constrained effect of lagged CS also highlights the stability of the core relationship between CS and UCS, which remains largely unaffected by immediate past states.

Overall, these results strengthen the evidence for the Bipolar Continuum Hypothesis while acknowledging its limitations. The variability observed across individuals suggests that the continuum may be a stable structure for some but more flexible or context-dependent for others. By incorporating idiographic methods, this study demonstrates that the relationship between CS and UCS is both robust and context-sensitive, highlighting the value of moving beyond group-level analyses to explore the diverse ways self-compassion operates in daily life.

This analysis reinforces the value of the Bipolar Continuum framework while highlighting the need for a more nuanced understanding of its applicability across individuals and contexts. In the following section, we outline the methodological details of the idionomic analysis.

#### 3.2.2.1 Step 1

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

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.

#### 3.2.2.2 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.

#### 3.2.2.3 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.

#### 3.2.3 Step 2

In the second step of our idionomic 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.

##### 3.2.3.1 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`.

```
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)
)
```

**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.

#### 3.2.3.2 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.

### 3.2.4 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.

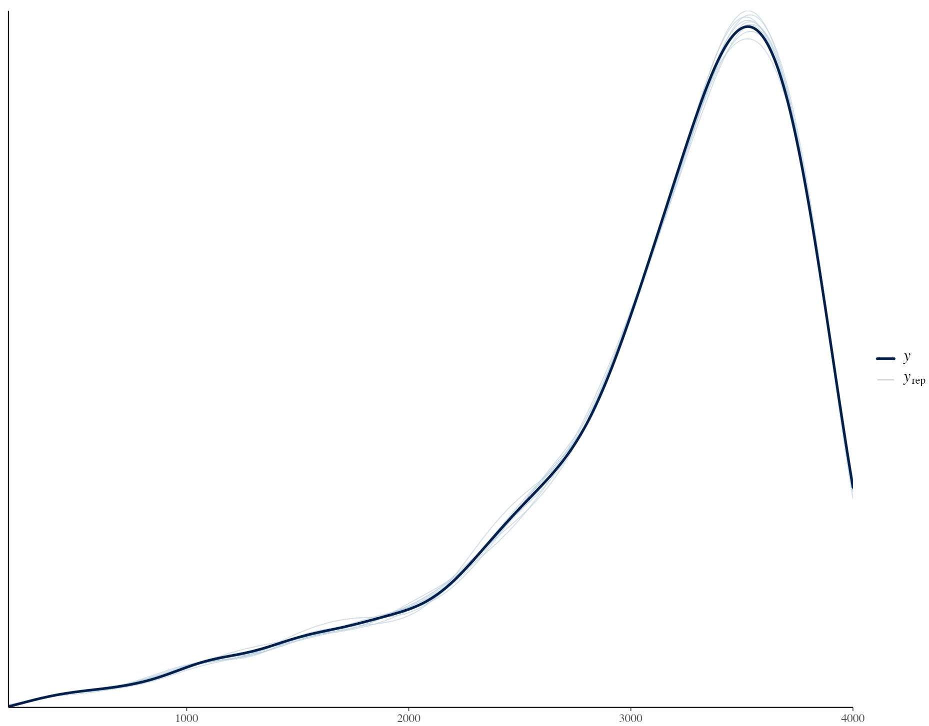


Figure 3: Idiographic analysis. Posterior Predictive Check for the hierarchical model fitted to the proportion of posterior draws of the `gamma_CS`parameter that were negative.

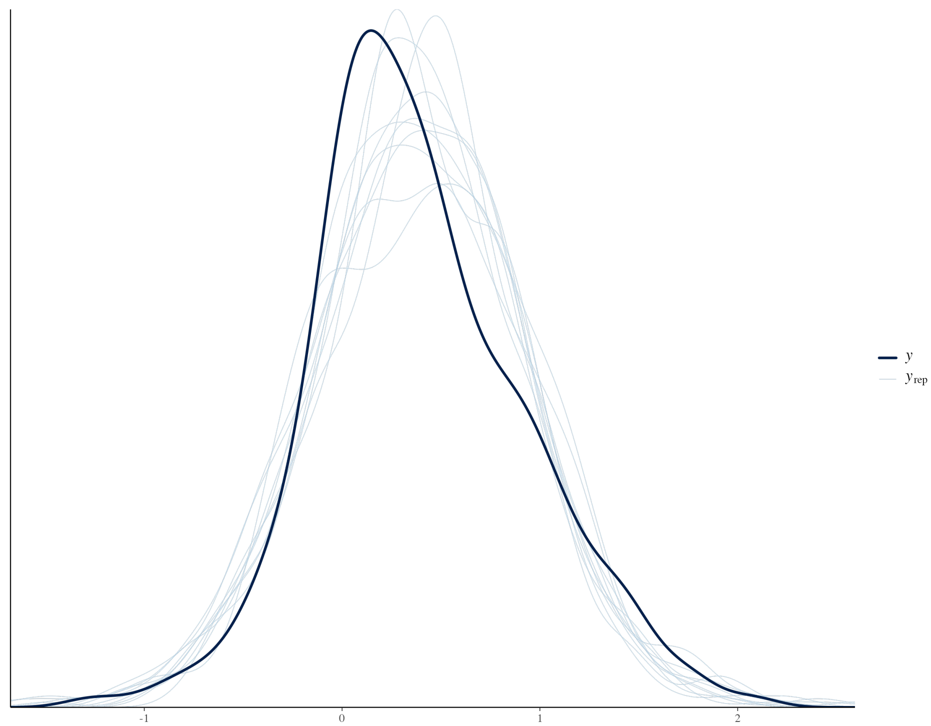


Figure 4: Idiographic analysis. Posterior Predictive Check for the hierarchical model fitted to the mean of the posterior draws of the `gamma_neg_aff` parameter.

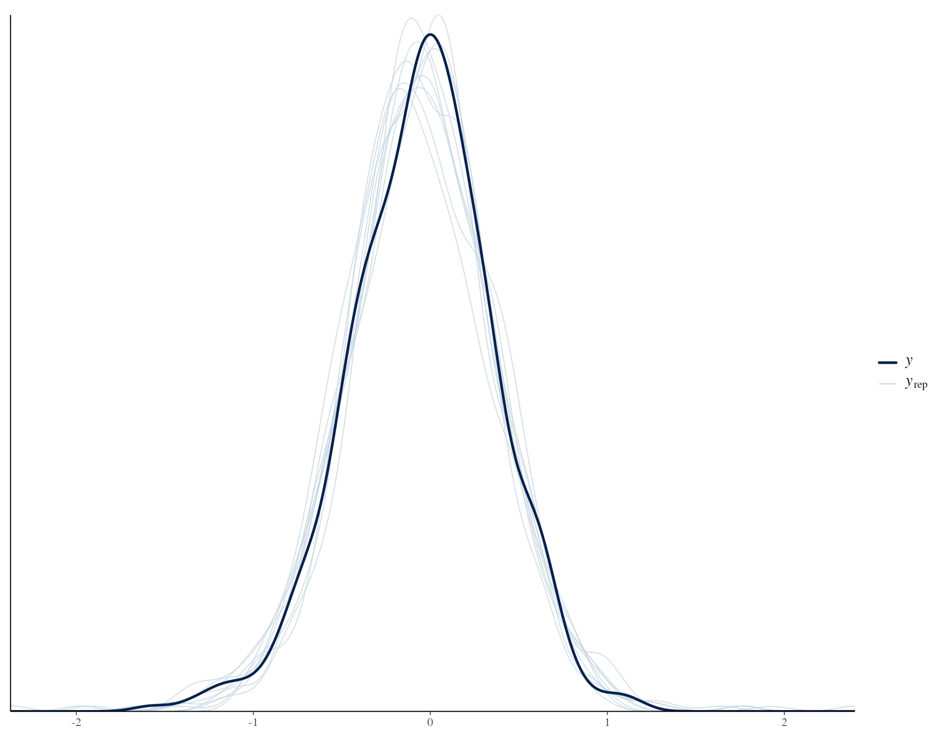


Figure 5: Idiographic analysis. Posterior Predictive Check for the hierarchical model fitted to the mean of the posterior draws of the `gamma_context` parameter.

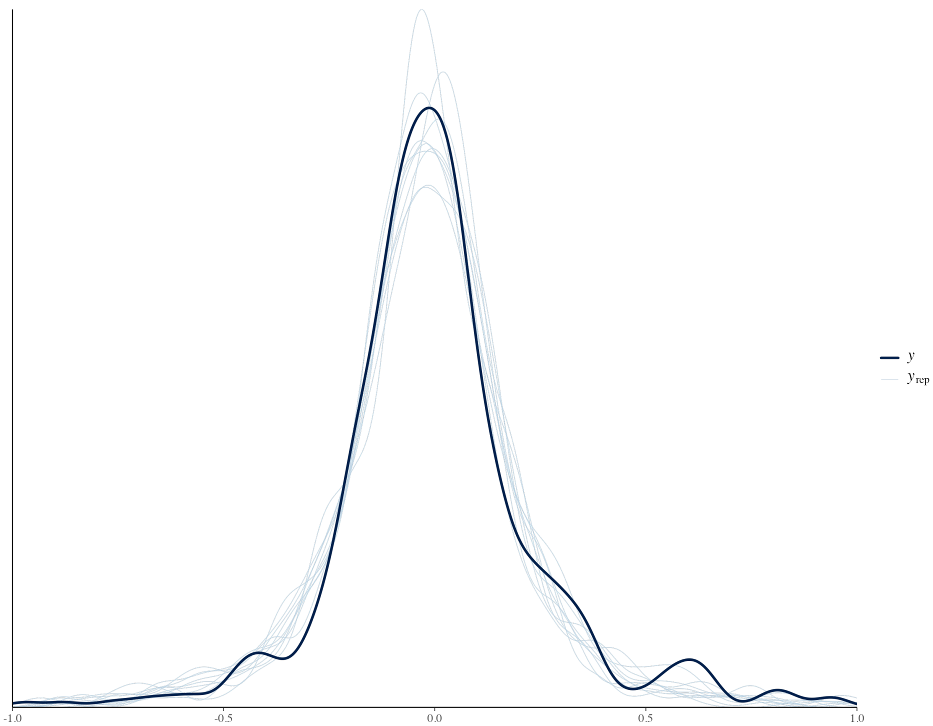


Figure 6: Idiographic analysis. Posterior Predictive Check for the hierarchical model fitted to the mean of the posterior draws of the `gamma_interaction` parameter.

## 3.3 Examining Response Bias

In a final analysis, we explored the potential impact of response bias. Our focus was on the subset of participants who exhibited a positive association between the CS and UCS components of the state self-compassion Scale. This positive relationship—where higher levels of CS coincide with higher levels of UCS—directly contradicts the Bipolar Continuum Hypothesis, which posits an inverse relationship between these components. A Reviewer suggested that this unexpected pattern could stem from response biases, such as “careless responding” or “insufficient effort,” where participants may repeatedly select the same response option for both components, irrespective of their actual subjective state.

To assess this possibility, we analyzed four established indices of careless responding: the Longstring Index, Intra-Individual Response Variability (IRV), Even-Odd Inconsistency Index, and Mahalanobis Distance. These indices were compared between two groups: participants with a positive UCS-CS association and all other participants. To ensure sufficient statistical power, this analysis utilized the combined dataset from both studies ($N = 495$).

### 3.3.1 Analysis Steps

**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.

**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.

**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.

### 3.3.2 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 9: 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 |

### 3.3.3 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.

### 3.3.4 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:

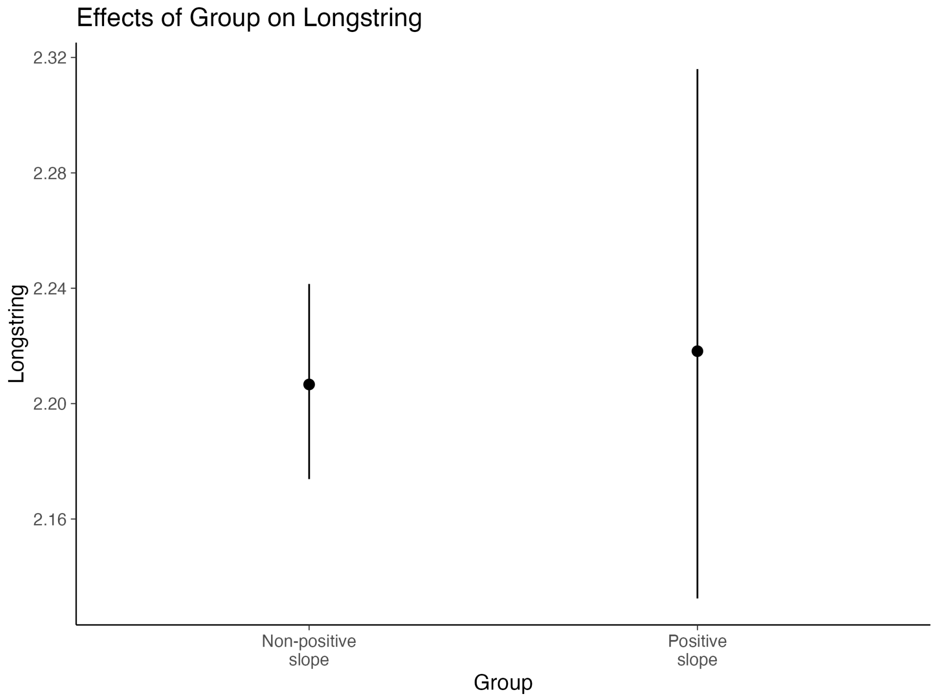


Figure 7: Conditional effect of Group on Longstring Index. Vertical bars represent 89% credibility intervals.

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.

### 3.3.5 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.

### 3.3.6 Time to Completion

TODO

## 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.