Mood that lingers and mood that bounces back were not shown to be separate traits

A preregistered test of emotional inertia against recovery in four open experience-sampling datasets, with a time-shuffle control and a one-parameter AR(1) null

Zain Dana Harper

Version 1.0, 9 November 2026. Finding F27 in the findings ledger. Honest null (tier D).

Abstract

Emotional inertia is the carryover of mood from one moment to the next. Recovery is how fast mood returns to a person's usual level after a sharp high or low. We asked whether these are two separate traits or one trait seen twice. In four open phone-survey datasets from the openESM database (71, 49, 99 and 259 eligible people), the two measures were computed on disjoint sets of prompt pairs and correlated across people. Across people, inertia and recovery were negatively correlated in all four datasets, r -0.44, -0.41, -0.28 and -0.29. The coupling was weaker than a one-parameter AR(1) null in two datasets, 0008 clearly and 0018 by 0.001, and sat inside that null in the other two. A factor model with two factors never fit better than one. Split-half reliability ran from 0.03 to 0.47, so the factor test had little power. The preregistered prediction that the two are separate was not supported in any dataset on both of its tests. We report it as an honest null with low power, and with one methods warning: with the usual all-pairs inertia, which reuses the spike pairs, the correlation reached 0.67 to 0.81 in size.

1 What this means for a reader

A person whose mood tends to linger from hour to hour also tends, in these data, to come back more slowly after a big swing. We could not show that these are two different things. That is weak evidence that they are one, because the person-level measures were noisy. If you read a study that reports a strong link between inertia and recovery, check whether both were computed from the same moments; in our data that overlap alone pushes the correlation from about 0.3 to about 0.7.

2 Introduction

Emotional inertia was introduced as a person-level property by Kuppens, Allen and Sheeber [2], who found higher inertia of both positive and negative emotions in people with depression and low self-esteem, across two naturalistic studies. Inertia is usually read off as an autocorrelation, so it measures how self-predictable a person's mood is from one prompt to the next.

Koval and colleagues [3] asked how inertia relates to the events people meet. In 200 first-year students, measured in daily life and in a film-clip task, the most consistent result was that higher inertia of negative affect went with slower recovery after negative stimuli. Their abstract reads this as inertia being driven mostly by weaker repair of mood after negative events. Schreuder and colleagues [5] measured recovery more directly: in a preregistered study of 68 adults with burst sampling, they timed returns to baseline after positive and negative episodes with survival analysis, and trait resilience did not predict return time.

A broader caution comes from Dejonckheere and colleagues [4]. Across many datasets, measures of mood dynamics such as autocorrelation added little to the prediction of well-being once mean affect and its variability were known. A dynamic measure has to earn its place as a separate quantity.

Our question sits between these. If inertia and recovery are two traits, they should vary across people more loosely than a model with one parameter would allow. If they are one trait, the coupling should look like what a single parameter produces on the same prompt schedules. The search we ran found adjacent work but no test of this form; that is what the search returned, not proof that none exists.

3 Preregistration and design

The preregistration was written on 29 October 2026, before any inertia, recovery, correlation or factor model was computed on the real data. Before writing it we had seen column names, row counts, the most common raw values of each mood column, and the number of prompts per person. All design choices were tested on synthetic populations first.

The idea and the prediction came from the author, who framed mood as an audio chain: a reverb, how long a feeling rings on, and a compressor release, how fast a large swing returns. The prediction had two parts, tested in each dataset separately:

- P1, separate variation: the across-person correlation of inertia and recovery has a 95 percent interval inside -0.7 to 0.7, and its size is below the 5th percentile of the same quantity from a one-parameter AR(1) null.
- P2, two factors beat one: a two-factor model of odd-day and even-day inertia and recovery beats a one-factor model, with a chi-square difference test p below 0.05 and a lower BIC.

The verdict rules were fixed in advance. A dataset is "supported" when the shuffle check P0, P1 and P2 all pass; "partly" when P0 passes and exactly one of P1 and P2; "null" when P0 passes and both fail; "instrument failure" when P0 fails; "untestable" with fewer than 30 eligible people. There was no pooling across datasets, because their scales and samples differ. The preregistration's SHA-256 is given in the reproduction section.

4 Data and provenance

The four datasets come from openESM [1], a database of open experience sampling data, all under the Creative Commons Attribution 4.0 licence (CC BY 4.0). Only group statistics are reported here; nothing describes an individual participant.

- 0018, Bailon 2020 (CoVidAffect, Spain, spring 2020): a valence slider from 0 to 100. There are no prompt numbers, so prompts are ordered by answer time and a pair needs the same day and a gap of at most 6 hours.
- 0062, Neubauer 2024 (STECCO, German first-year students, 2021): valence is the mean of four positive items minus the mean of four negative items, each 1 to 7, from the five momentary prompts a day. Two 7-day bursts six weeks apart.
- 0008, Westhoff 2024 (young adults, Germany): a valence item from 0 to 100, five prompts a day for 21 days.
- 0076, Bayer (mCog, US adults with Android phones, 2019): positive affect from 1 to 5, six prompts a day for 14 days.

The four input files were hashed by the run, and the hashes are in its result file. The samples are convenience samples, each on its own scale, which is why each is analysed alone.

5 Methods

Pairs, spikes and recovery

A lag pair is two adjacent prompts on the same study day. For a person with mean m and standard deviation s over all their valid prompts, the deviation at prompt t is

$$d_t = x_t - m \tag{1}$$

A spike is the lead prompt of a lag pair with a deviation of at least $1.5s$ in size. For each spike, the share of the deviation that has gone at the next prompt is

$$u_t = 1 - \frac{d_{t+1}}{d_t} \tag{2}$$

clipped to the range -1 to 2. A person's recovery is the median of u_t over their spikes, so 1 means back at the mean one prompt later and 0 means no change.

Inertia on disjoint pairs

Primary inertia is the Pearson r between x_t and x_{t+1} over the lag pairs whose lead prompt is not a spike. The reason was settled on synthetic data before the run. Inertia over all pairs is dominated by large deviations, so it reads the same spike pairs that define recovery. In a synthetic population with separate rates for small and large deviations, all-pairs inertia still correlated with recovery at -0.77 to -0.83; the disjoint version gave -0.23 to -0.29, against -0.54 to -0.61 in a one-parameter population. All-pairs inertia was kept as a secondary measure.

A person was eligible with at least 30 valid prompts, 20 lag pairs, 15 non-spike pairs and 5 spikes. A dataset was tested with at least 30 eligible people.

The across-person correlation

For each dataset, r is the Pearson correlation of inertia and recovery across eligible people. Its 95 percent interval is a person bootstrap with 2,000 resamples.

The one-parameter null

The decisive comparison is with an AR(1) process, in which each person has a single parameter φ:

$$d_{t+1} = \varphi d_t + \varepsilon_{t+1} \tag{3}$$

where ε is fresh Gaussian noise. Each eligible person is simulated on their exact prompt schedule, with the same days, prompts and missed prompts, a fresh stationary start each day, their own m and s, and values rounded and clipped to the scale. The person's φ is their all-pairs inertia corrected for short-series Kendall bias,

$$\varphi_K = \varphi + \frac{1 + 3\varphi}{n} \tag{4}$$

with n the number of pairs, then shrunk toward the dataset mean so that its spread excludes sampling noise, and clipped to -0.5 to 0.95. The whole pipeline is rerun on the simulated people, and the size of r is recorded. Two hundred draws give the null distribution.

The factor test

Each person gives four indicators: inertia and recovery computed on odd study days and on even study days, each half needing at least 10 non-spike pairs and 3 spikes. A one-factor model (four loadings, four unique variances, 2 degrees of freedom) is compared with a two-factor model with inertia halves on one factor, recovery halves on the other and a free factor correlation (1 degree of freedom). The two-factor model wins only if the chi-square difference has p below 0.05 and its BIC is lower. The one-factor model is the two-factor model with the factor correlation at plus or minus 1, a boundary value, so the 1-degree-of-freedom p is conservative. The same halves give each measure's split-half reliability.

p values from draws

Each control's p value is the share of draws at or above the real value, with one added to the count and the number of draws:

$$p = \frac{k+1}{n+1} \tag{5}$$

At 200 draws the smallest value is 0.005.

6 Controls and planted checks

Three controls ran per dataset, 200 draws each, from one random generator with seed 20261029.

- C1, time shuffle: each person's mood values are permuted over their own prompt slots, and pairs and spikes are recomputed. The real mean inertia must exceed the shuffle's 95th percentile (rule P0). This checks that the instrument reads real carryover.
- C2, the one-parameter AR(1) null of eq. 3.
- C3, each person cut to a random block of 40 prompts, to see how the coupling behaves in short series.

The AR(1) null was itself a planted check before the run. In a synthetic one-parameter world the observed correlation fell inside this null in 6 of 6 runs, and in a two-parameter world below its 5th percentile in 6 of 6 runs, on both continuous and 5-point scales. So the null can tell the two worlds apart when the data are as clean as the simulation.

The shuffle check passed in all four datasets. Mean primary inertia was 0.236, 0.404, 0.270 and 0.162, against shuffle 95th percentiles of 0.013, 0.011, 0.013 and 0.000, each with p 0.005, the smallest value 200 draws allow.

7 Results

Table 1 gives the main numbers. Across people, inertia and recovery were negatively correlated in all four datasets, r -0.44, -0.41, -0.28 and -0.29. Each 95% interval excludes 0 and lies inside -0.7 to 0.7. People whose mood carries over more tend to recover less of a spike by the next prompt.

Table 1. Across-person correlation of inertia and recovery, with the one-parameter AR(1) null of its size. Intervals are person-bootstrap 95 percent intervals; the null runs 200 draws (156 usable in 0062, where some draws had fewer than 30 eligible people).

Dataset	Eligible	r (95% interval)	Null 5th pct, mean	P1	Factor r
0018	71 of 999	-0.44 (-0.63 to -0.19)	0.440, 0.573	pass	0.95
0062	49 of 321	-0.41 (-0.68 to -0.09)	0.215, 0.512	fail	untestable
0008	99 of 114	-0.28 (-0.45 to -0.09)	0.367, 0.475	pass	0.98
0076	259 of 519	-0.29 (-0.41 to -0.15)	0.168, 0.283	fail	-1.00

The coupling was weaker than a one-parameter AR(1) null in two datasets, 0008 clearly and 0018 by 0.001, and sat inside that null in the other two. In 0008 the size of r was 0.279 against a null 5th percentile of 0.367. In 0018 it was 0.4391 against 0.4399, a pass that a different seed could reverse. In 0062 it was 0.414 inside a null running from 0.215 to 0.682, and in 0076 it was 0.289, next to the null mean of 0.283. Figure 1 draws these against the null and beside the all-pairs value.

A factor model with two factors never fit better than one. In 0018 (53 people with both halves) the chi-square difference was 0.055 with p 0.81 and the BIC favoured one factor by 3.9; the factor correlation was 0.95. In 0008 (60 people) the difference was 0.001 with p 0.97, BIC +4.1, factor correlation 0.98. In 0076 (152 people) the difference was 0.00 with p 1.0, BIC +5.0, and the factor correlation sat at the boundary, -1.00. In 0062 only 5 people had enough spikes on both odd and even days, because the study ran in two 7-day bursts, so the model could not be fitted.

Split-half reliability ran from 0.03 to 0.47, so the factor test had little power. For inertia it was 0.38, 0.47 and 0.03 in 0018, 0008 and 0076; for recovery 0.33, 0.11 and 0.17. Two factors cannot separate when each indicator carries this much noise.

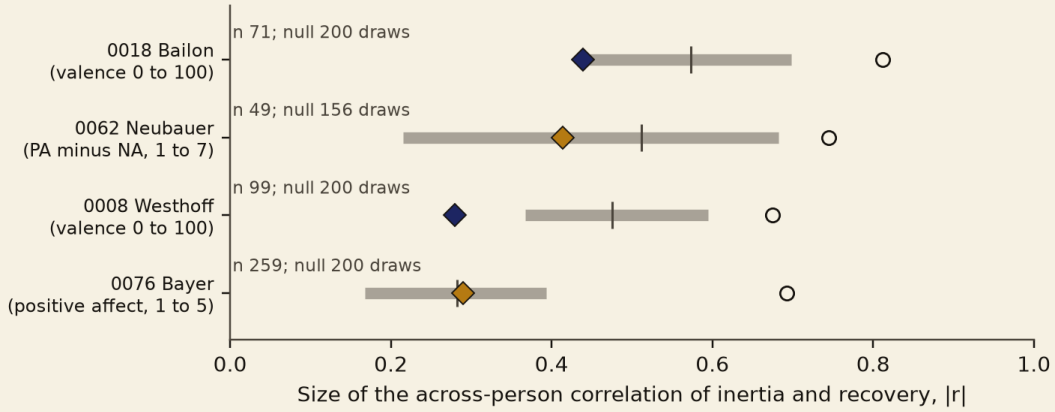


Figure 1. Size of the across-person correlation of inertia and recovery (diamonds; blue where below the null's 5th percentile) against the one-parameter AR(1) null (grey bar from 5th to 95th percentile, tick at the mean). Open circles: the same correlation with all-pairs inertia, which shares the spike pairs with recovery. Drawn from the run's result file.

The definition of inertia matters more than any other choice. With the usual all-pairs inertia, which reuses the spike pairs, the correlation reached 0.67 to 0.81 in size. The values were -0.81, -0.75, -0.67 and -0.69. Under that definition P1 would have failed in every dataset on its first condition. Other preregistered sensitivity checks changed no conclusion: Kendall-corrected inertia gave r -0.43, -0.42, -0.28 and -0.29, and a spike threshold of 1.0 standard deviations gave -0.37, -0.32, -0.27 and -0.16 with more people eligible (79, 138, 112 and 272).

Coupling grew with series length in three datasets (Figure 2). Split by prompts per person into thirds, *r* ran -0.14, -0.45 and -0.78 in 0018, -0.19, -0.27 and -0.40 in 0008, and -0.25, -0.24 and -0.37 in 0076. In 0062 it ran -0.31, -0.78 and +0.25 with 13 to 19 people per third. This pattern fits attenuation: in short series both measures are noisy and the correlation shrinks toward zero. The thirds are small and have no intervals. The 40-prompt control C3 was usable only in 0076, where the mean size of *r* was 0.235 (5th to 95th percentile 0.104 to 0.370); in the other three, too few people stayed eligible in 40 prompts.

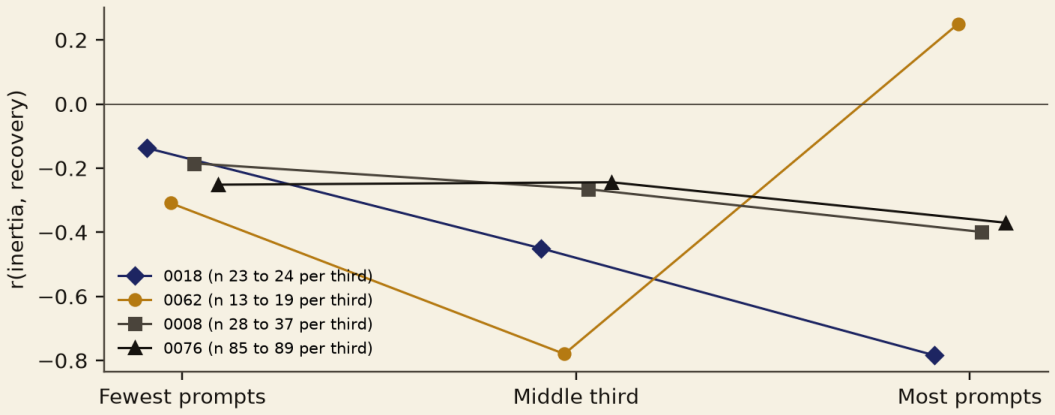


Figure 2. Across-person r of inertia and recovery within thirds of prompts per person. Cells are small and carry no intervals. Drawn from the run's result file.

Table 2 sets the forecasts written in the preregistration against the outcomes.

8 Discussion

Read by the preregistered rules, 0018 and 0008 are "partly" (P1 passed, P2 failed), 0076 is "null", and 0062 failed P1 with P2 untestable. No dataset supported both parts of the prediction. The headline is therefore that

Table 2. Forecasts fixed in the preregistration against what the run gave.

Quantity	Forecast	Outcome
Shuffle check P0 passes in every tested dataset	probability 0.9	4 of 4
Mean primary inertia between 0.10 and 0.40	each dataset	3 of 4 (0062: 0.404)
r negative, about -0.2 to -0.5	each dataset	4 of 4
95% interval inside -0.7 to 0.7	probability 0.8 each	4 of 4
Size of r below the AR(1) null 5th percentile	probability 0.35 each	2 of 4
Two factors beat one	probability 0.3 each	0 of 3 tested
Split-half reliability of recovery 0.1 to 0.3	expected	2 of 3 (0.33, 0.11, 0.17)
0018 untestable or thin	probability 0.4	tested, 71 people

inertia and recovery were not shown to be separate traits.

That does not show they are one trait, and two features of the design lean toward a pass rather than a fail. The AR(1) null has no measurement-noise term. Noise in the real data weakens the observed coupling, which makes a real value below the null's 5th percentile easier to reach. The two P1 passes are weak evidence for that reason, and the 0018 pass sits at the edge. The factor test, on the other side, had too little reliability to detect two factors even if they exist.

The direction agrees with Koval and colleagues [3], whose abstract reports higher negative-affect inertia going with slower recovery. It also fits the caution of Dejonckheere and colleagues [4]: a dynamic measure computed from few prompts carries little information of its own, and here two such measures could not be told apart. Recovery measured in clock time with dense sampling, as in Schreuder and colleagues [5], is a better instrument for the question than one-prompt-later recovery from five or six prompts a day.

The methods result stands apart from the null. Inertia computed over all pairs shares data with recovery, and in these four datasets that overlap alone moves the size of the correlation from 0.28 to 0.44 up to 0.67 to 0.81. Any study that relates autocorrelation to recovery after peaks should compute them from disjoint pairs, or report how much of the link the overlap creates.

9 Limitations

- Split-half reliability was 0.03 to 0.47, so the factor test had little power, and one dataset could not be fitted at all.
- The AR(1) null has no measurement-noise term, which favours a pass on P1.
- Recovery rests on few spikes per person, and one prompt later is a coarse clock.
- The four datasets are convenience samples on different scales and items, and three of the four eligible samples are small.
- The 40-prompt control was usable in one dataset only.
- A second round with a noise term was preregistered on the same day. No run of it is on record, so its outcome is unknown.
- Scoring was done by our own code and was not blind.

10 Verified

Verified: on 9 November 2026 the person counts, the four correlations with their intervals, the shuffle and AR(1) null quantiles and draw counts, the factor-model statistics and correlations, the split-half reliabilities, the all-pairs, Kendall-corrected and 1.0 standard deviation correlations, and the tertile values were re-read from the run's result file. They match the findings ledger entry and the study's own ledger. The preregistration's SHA-256 was recomputed and matches the hash recorded in the result file. The synthetic checks and the forecasts were read from the preregistration. Both figures were drawn by a script that reads only the result file. The study was

not re-run for this paper. The literature in the introduction was found by searches run on 9 November 2026 and read at the level of abstracts and publisher or repository pages; the full texts were not read. For Kendall [6] only the bibliographic record was seen, and the bias formula used here is the one written in our preregistration.

11 Corrections

- 31 October 2026: a review of our ledger noted that the F27 entry named no result file. The result file exists and is the source of every number in this paper.
- 9 November 2026: first published version. The study's notes give 24 people per third in 0018; the result file gives 24, 23 and 24, which this paper uses.

12 Reproduction

- Data: download openESM datasets 0018, 0062, 0008 and 0076 (CC BY 4.0).
- For each person, form lag pairs of adjacent prompts on the same day, flag spikes at 1.5 standard deviations from the person's mean, and compute recovery with eq. 2 and inertia over non-spike pairs.
- Correlate inertia and recovery across eligible people with a 2,000-resample person bootstrap.
- Run the time shuffle, the AR(1) null of eqs. 3 and 4, and the 40-prompt blocks, 200 draws each, seed 20261029.
- Fit one-factor and two-factor models to odd-day and even-day halves.
- Preregistration SHA-256: c26b85b57ec1c6495ee297009a4f7fa1e47184cfa1ee10b4aba77cef4280ce81. Plan SHA-256: 86d93b056d781b181f3873196c9a1ac049d019b151c230a16de2693e0ff51201. Run 20261029T130553Z.
- Cost: the timings in the result file sum to about 63 seconds on one desktop CPU.

13 Appendix A. The preregistered verdicts

Each verdict is a set of rules written before the run, read off by code. Values are from the run's result file. P0 compares mean inertia with the shuffle's 95th percentile; P1 compares the size of r with the AR(1) null's 5th percentile (its interval conditions held everywhere); P2 needs a chi-square difference p below 0.05 and a lower two-factor BIC; S needs the 40-prompt control's mean size of r below 0.7.

Table 3. Preregistered verdicts per dataset, with the deciding values.

Rule	0018	0062	0008	0076
P0	pass: 0.236 over 0.013	pass: 0.404 over 0.011	pass: 0.270 over 0.013	pass: 0.162 over 0.000
P1	pass: 0.439 under 0.440	fail: 0.414 over 0.215	pass: 0.279 under 0.367	fail: 0.289 over 0.168
P2	fail: p 0.81, BIC +3.9	untestable: 5 people	fail: p 0.97, BIC +4.1	fail: p 1.0, BIC +5.0
S	untestable	untestable	untestable	pass: 0.235
Reading	partly	P1 null, P2 untestable	partly	null

14 Glossary

Each term is defined where it is first used in the text. The same definitions, with the field each comes from and what each excludes, are collected here.

- **honest null**. A working test that found no effect beyond a stated bound, a forecast that failed, or an error of our own found and fixed, reported with the same weight as a positive result. Field: research method (our usage). Excludes a void run, and a result left unpublished because it was negative.
- **emotional inertia**. How far a person's mood at one prompt predicts their mood at the next, measured as the lag-one correlation of their own mood series. Field: affective science. Excludes average mood, how much

mood varies, and any claim about what causes the carryover.

- **preregistration**. A dated written plan for a study, fixed and hash-locked before any result is seen: the question, the data, the statistic, the controls, the forecasts and the kill conditions. Field: research method (meta-science). Excludes a plan revised after looking at the outcome, and a protocol written only for the record after the run.
- **chi-square difference test**. A test of whether a model with more free parameters fits nested data better than a simpler model, using the drop in the chi-square fit statistic against its degrees of freedom. Field: statistics (model comparison). Excludes comparisons of models that are not nested.
- **verdict**. The pass or fail of one preregistered rule set against the run's numbers, read off by code. Field: research method (our usage). Excludes an interpretation written after reading the numbers.
- **experience sampling**. A study design in which people answer short questions on a phone several times a day, so that mood is recorded close to the moment it is felt. Field: psychology (methods). Excludes one-off questionnaires and recall of past weeks.
- **lag pair**. Two adjacent prompts on the same study day, the earlier one called the lead; a missed prompt or a night breaks the pair. Field: time-series analysis. Excludes pairs that cross a night or skip a prompt.
- **spike**. A lead prompt whose mood sits at least 1.5 of the person's own standard deviations from the person's own mean. Field: our usage. Excludes deviations judged against other people or against a population norm.
- **recovery**. The median, over a person's spikes, of the share of each spike's distance from the person's mean that has gone by the next prompt, with 1 meaning back at the mean and 0 meaning no change. Field: affective science (our operational form). Excludes recovery measured in clock time, and recovery after events that did not produce a spike.
- **Pearson r**. The correlation coefficient between two quantities measured on the same units, from -1 through 0 for no linear relation to 1. Field: statistics. Excludes non-linear relations and any statement about cause.
- **person bootstrap**. An interval found by drawing people with replacement many times, recomputing the statistic each time, and taking the middle of the resulting spread. Field: statistics (resampling). Excludes resampling of single prompts, which would treat one person's prompts as independent.
- **AR(1) process**. A series in which each value is the mean plus a fixed fraction of the previous value's distance from the mean, plus fresh random noise; the fraction is the single parameter. Field: time-series statistics. Excludes models with separate rates for small and large deviations, and models with measurement noise on top.
- **Kendall bias**. The downward bias of a lag-one correlation estimated from a short series, about (1 + 3 phi)/n for a series of n values with true value phi (Kendall 1954). Field: time-series statistics. Excludes bias from measurement noise, which is a separate attenuation.
- **null distribution**. The spread of scores a control gives over many random draws; its 95th percentile is the bar a real score must clear. Field: statistics. Excludes a theoretical distribution assumed rather than drawn from the data.
- **factor model**. A model that explains the correlations among several measured indicators by a smaller number of unseen common factors, fitted by maximum likelihood. Field: psychometrics (confirmatory factor analysis). Excludes exploratory searches for the number of factors, and models of change over time.
- **BIC**. The Bayesian information criterion, a fit score that charges each extra parameter a penalty growing with sample size; the lower value is preferred. Field: statistics (model selection). Excludes a test of whether either model is true.
- **split-half reliability**. The correlation across people between a measure computed on one half of each person's data and the same measure on the other half; here the halves are odd and even study days. Field: psychometrics. Excludes test-retest reliability over months, and the validity of the measure.
- **control**. A version of the data with the structure under test removed (for example shuffled or pooled), scored

the same way as the real data, so the real score has something to beat. Field: statistics and experimental design. Excludes a comparison with a different dataset or with a published number.

- **planted check**. Data built so that the right answer is known, run through the same analysis before the real data, to show that the analysis can see the effect when it is there. Field: research method; called a positive control in experimental science. Excludes any check made on the real data, which cannot show what the analysis misses.
- **95% interval**. The range that holds the middle 95 percent of the bootstrap or control draws for a number; we report it beside the number. Field: statistics. Excludes a statement that the true value lies inside with 95 percent probability.
- **attenuation**. The shrinking of an observed correlation toward zero when the measures that enter it are noisy. Field: psychometrics. Excludes a true weak relation between the underlying quantities.
- **findings ledger**. The append-only, dated list of every result we have recorded, numbered F1, F2 and on; corrections are added as new dated lines and old entries are never rewritten. Field: our records. Excludes a curated list of successes.

15 References

[1] Siepe, B. S., Haslbeck, J. M. B., Kloft, M., Büchner, A., Zhang, Y., Fried, E. I. and Heck, D. W. (2026). Introducing openESM: a database of openly available experience sampling datasets. Behavior Research Methods 58, 240. doi:10.3758/s13428-026-03112-y. Datasets 0018, 0062, 0008 and 0076, CC BY 4.0.

[2] Kuppens, P., Allen, N. B. and Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. Psychological Science 21, 984-991. doi:10.1177/0956797610372634

[3] Koval, P., Brose, A., Pe, M. L., Houben, M., Erbas, Y., Champagne, D. and Kuppens, P. (2015). Emotional inertia and external events: the roles of exposure, reactivity, and recovery. Emotion 15, 625-636. doi:10.1037/emo0000059

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

[5] Schreuder, M. J., Ariens, S., Lafit, G. and Ceulemans, E. (2024). Bouncing back from emotional ups and downs: insights in emotional recovery using survival analyses of burst ESM data. Cognition and Emotion 39, 910-919 (online 2024). doi:10.1080/02699931.2024.2423706

[6] Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. Biometrika 41, 403-404. doi:10.1093/biomet/41.3-4.403