# Neuroticism may not reflect emotional variability

Elise K. Kalokerinos[a,1,2], Sean C. Murphy[a,1], Peter Koval[a,b], Natasha H. Bailen[c], Geert Crombez[d], Tom Hollenstein[e], John Gleeson[f], Renee J. Thompson[c], Dimitri M. L. Van Ryckeghem[d,g], Peter Kuppens[b], and Brock Bastian[a]

[a]Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, VIC 3010, Australia; [b]Faculty of Psychology and Educational Sciences, KU Leuven, 3000 Leuven, Belgium; [c]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO 63130-4899; [d]Faculty of Psychology and Educational Sciences, Ghent University, 9000 Ghent, Belgium; [e]Department of Psychology, Queen's University, Kingston, ON K7L 3N6, Canada; [f]School of Behavioural and Health Sciences, Australian Catholic University, Fitzroy, VIC 3065, Australia; and [g]Faculty of Psychology and Neuroscience, Maastricht University, 6229 Maastricht, The Netherlands

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 12, 2020 (received for review November 12, 2019)

**Neuroticism is one of the major traits describing human personality, and a predictor of mental and physical disorders with profound public health significance. Individual differences in emotional variability are thought to reflect the core of neuroticism. However, the empirical relation between emotional variability and neuroticism may be partially the result of a measurement artifact reflecting neuroticism's relation with higher mean levels—rather than greater variability—of negative emotion. When emotional intensity is measured using bounded scales, there is a dependency between variability and mean levels: at low (or high) intensity, it is impossible to demonstrate high variability. As neuroticism is positively associated with mean levels of negative emotion, this may account for the relation between neuroticism and emotional variability. In a metaanalysis of 11 studies ($N$ = 1,205 participants; 83,411 observations), we tested whether the association between neuroticism and negative emotional variability was clouded by a dependency between variability and the mean. We found a medium-sized positive association between neuroticism and negative emotional variability, but, when using a relative variability index to correct for mean negative emotion, this association disappeared. This indicated that neuroticism was associated with experiencing more intense, but not more variable, negative emotions. Our findings call into question theory, measurement scales, and data suggesting that emotional variability is central to neuroticism. In doing so, they provide a revisionary perspective for understanding how this individual difference may predispose to mental and physical disorders.**

neuroticism | negative emotion | emotional variability | personality | experience sampling

Variation in human personality is commonly described in terms of a handful of organizing dimensions (1). A dimension that features in most, if not all, taxonomies of personality is neuroticism (1). Neuroticism is typified by negative emotionality (2), and, as such, is central for understanding differential risk in mental and physical health (3) and has profound public health significance (4). Neuroticism is such a powerful predictor of future emotional disorder that some scholars have proposed that clinical efforts shift toward directly targeting neuroticism (3). However, this requires a full understanding of how neuroticism manifests in everyday emotional experience, which we argue may be lacking.

The negative emotionality central to neuroticism is thought to manifest not only in higher mean levels of negative emotion, but also in greater emotional variability (5). Emotional variability is a core part of Eysenck's foundational conceptualization of neuroticism (6), which described neuroticism as hyperreactivity manifesting in emotional volatility. This centrality of variability to neuroticism has inspired more recent research (e.g., ref. 7), and there is even a body of research testing whether emotional variability and neuroticism are separable concepts (5, 8–10). This link is also strongly reflected in the measurement of neuroticism: most scholars use "emotional stability" as the inverse of neuroticism (e.g., refs. 11 and 12), and scales often include variability as a facet of hierarchical models of neuroticism (e.g., ref. 13). In addition, the major assessment scales have items tapping emotional variability (13–17), meaning that almost all neuroticism research incorporates variability.

Emotional variability is commonly operationalized as the within-person SD of repeated emotion assessments. A meta-analysis of 61 effects found a small-to-medium positive association between neuroticism and the negative emotion within-person SD (18), providing evidence for this link. However, we argue that findings linking neuroticism to emotional variability may be partly the result of a methodological artifact. Variability in a construct can be dependent on mean levels of the same construct, especially when measurements are bounded within scales (19). For example, consider a study in which emotions are repeatedly assessed on a scale from 0 (no emotion) to 100 (strong emotion). Here, a person's mean will always fall between 0 and 100. If their mean is low (e.g., 10) or high (e.g., 90), their variability is limited by the scale endpoints. That is, a person with a mean of 10 (or a mean of 90) cannot demonstrate as much variability as somebody with a mean of 50, since the scores of the latter individual are less constrained by the scale boundaries. In line with this, low mean levels of emotion are associated with lower emotional variability (20, 21).

## Significance

**Neuroticism is the personality trait most closely linked with mental health challenges. Thus, it is important to understand how neuroticism manifests in everyday experience. Neuroticism has been characterized by greater variability between high and low levels of negative emotion. However, the way negative emotion is often measured means that there is a dependency between variability and the mean, which is problematic because neuroticism is also associated with high mean levels of negative emotion. In a metaanalysis of 11 studies that investigated emotion in everyday life, we found that, after accounting for the mean, neuroticism was not associated with emotional variability. This calls into question the definition of neuroticism and therefore how its association with mental illness is best understood.**

Author contributions: E.K.K., S.C.M., P. Koval, P. Kuppens, and B.B. designed research; E.K.K., S.C.M., P. Koval, N.H.B., G.C., T.H., J.G., R.J.T., D.M.L.V.R., P. Kuppens, and B.B. performed research; E.K.K. and S.C.M. analyzed data; and E.K.K. and S.C.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: Our data are available on the Open Science Framework at https://osf.io/gvfdx/.

[1]E.K.K. and S.C.M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: elise.kalokerinos@unimelb.edu.au.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1919934117/-/DCSupplemental.

First published April 15, 2020.

This mean–variability dependency is likely to interfere with the association between neuroticism and negative emotional variability for three reasons. First, measurements of emotional experience are bounded by scale endpoints. Second, emotional variability is usually calculated using reports of emotion in everyday life (18), where most people encounter few intense negative events (22), meaning negative emotional experience is often near the scale floor. This results in low mean scores, precluding participants from demonstrating high variability. Third, low neuroticism is associated with low mean levels of negative emotion (e.g., ref. 23). As such, less neurotic individuals may be constrained to have lower emotional variability simply because they experience low average levels of negative emotion. Thus, findings showing emotional variability is linked with neuroticism may be a byproduct of the link with mean levels of negative emotion, with mean levels providing a more parsimonious account of the relationship.

To address this issue, Eid and Diener (5) controlled for mean levels of negative emotion and still found an association between negative emotional variability and neuroticism. However, controlling for the mean is problematic, as the high correlation between the mean and SD can lead to multicollinearity, skewing conclusions (24). Moreover, this method does not consider nonlinear dependencies between the mean and SD. To address these issues, Mestdagh et al. (24) proposed a relative variability index, which measures variability as a proportion of the maximum possible variability given a participant's mean and the scale endpoints. This index is based on the assumption that the mean constrains the SD (rather than vice-versa), a decision made because the mean is a more parsimonious statistic than the SD. Using this index, we determine whether associations with variability could also be explained by mean levels. If variability does not add anything to our understanding over and above mean levels, we believe the more parsimonious mean should take precedence.

To systematically examine whether a dependency between mean levels and variability has led to an overestimation of the link between neuroticism and negative emotional variability, we used the relative variability index. We tested this association metaanalytically in 11 studies using diary and experience sampling methods (ESMs). In using these methods, we echo most studies of emotional variability (18). These methods are ideal for capturing variability because they allow for the collection of many data points (increasing the reliability of variability indices) and index emotions across varied contexts. Using these methods, participants are not asked to self-report emotional variability; instead, variability is computed by calculating variance across momentary reports.

We focused on negative emotion, as neuroticism is theoretically centered on negative, rather than positive, emotionality (15). Our key measure of variability was the within-person SD of negative emotion (18). We replicated our analyses using the mean squared successive difference (MSSD), which captures the temporal aspect of instability between measurements (25) and suffers from the same issues as the within-person SD (24).

Our analyses consisted of five steps. In the first two steps, we investigated whether the mean–variability dependency was likely to be problematic. We tested the associations between neuroticism and mean levels of negative emotion and between mean levels and within-person SDs of negative emotion. We hypothesized that both associations would be positive, underscoring a need to account for mean levels when examining how neuroticism is related to variability.

In the third step, we replicated previous findings, investigating the association between neuroticism and the within-person SD of negative emotion, hypothesizing a positive association. In the final two steps, we investigated whether this association held when accounting for the aforementioned measurement problems. To investigate the role of the lower scale boundary in constraining negative emotion, we tested the link between neuroticism and variability in daily maximum negative emotion. Focusing on daily maxima removed many occasions on which participants scored near the scale floor. If the neuroticism–variability association is strengthened by many observations near the lower boundary, we should find an attenuated association when looking at variability in daily maxima.

Finally, we tested the link between neuroticism and negative emotional variability computed using the relative variability index, which mathematically corrects for the dependency between variability and mean levels (24). We hypothesized an attenuation of the neuroticism–variability link. This would suggest that the link between neuroticism and negative emotional variability may



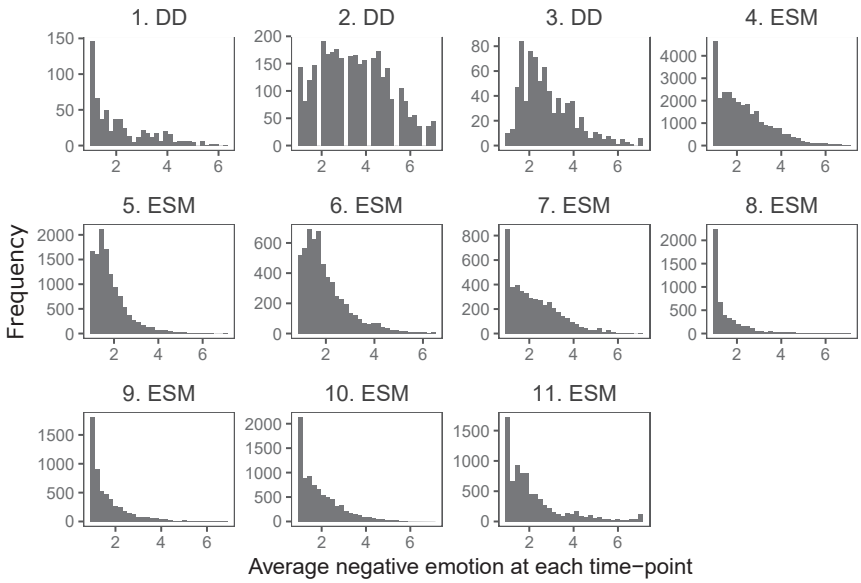


**Fig. 1.** Histograms depicting the frequency of momentary negative emotion for each dataset. The *y* axis represents the frequency of observations and is different for each dataset because they have a different total number of observations. DD, daily diary; ESM, experience sampling method.

be result of the dependency between variability and mean levels of negative emotion.

## Results

We performed a series of five random-effects metaanalyses across 11 datasets using the metafor package in R (26).

**Evidence that Mean Levels Are Implicated in the Relationship between Neuroticism and Negative Emotional Variability.** Fig. 1 contains histograms of negative emotion scores across all participants for each dataset (descriptive statistics are provided in *SI Appendix*, Table S3). Distributions of negative emotion scores were right-skewed: participants very frequently reported negative emotion levels near the scale floor. The exceptions were dataset 2, which focused on participants high in depressive symptoms, and dataset 3, which asked participants to report on their most negative daily event. These two datasets were focused on situations where negative emotion was likely to be more frequent, moving observations away from the lower boundary.

In these analyses, we investigated whether the mean–variability dependency was likely to be problematic. First, we metaanalyzed the association between neuroticism and mean levels of negative emotion (Fig. 2*A*). As hypothesized, there was a significant medium-sized correlation ($r = 0.36$) between neuroticism and mean negative emotion, suggesting neuroticism is characterized by the experience of more intense negative emotions in daily life.

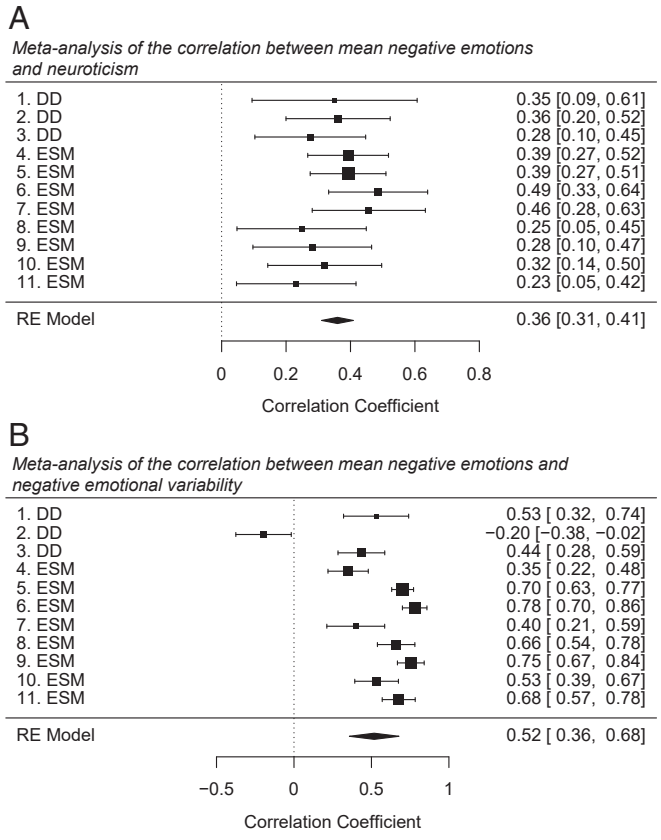


**Fig. 2.** Forest plots of the relationships between mean negative emotions and neuroticism (*A*) and negative emotional variability (*B*), demonstrating the potential for mean levels to cloud the relationship between neuroticism and negative emotional variability. For each dataset, we provide a correlation bounded by a 95% CI. The area of each square is proportional to the weight of the study in the metaanalysis. The results of the random-effects metaanalysis are depicted at the bottom of the figure (RE model), with the width of the rhombus representing the 95% CI. The dotted line represents no effect. DD, daily diary; ESM, experience sampling method.

Second, we metaanalyzed the association between the mean and within-person SD of negative emotions (Fig. 2*B*). As hypothesized, there was a significant large correlation ($r = 0.52$) between mean negative emotion and negative emotional variability. The only exception was dataset 2: this daily diary study focused on participants with high levels of depressive symptoms, and many participants reported negative emotions near the scale ceiling. Here, we saw the opposite problem from the scores at the scale floor that were our focus in the Introduction: it is likely that these participants could not demonstrate variability because range was restricted by the scale ceiling, resulting in a negative correlation between their mean and SD.

**Neuroticism and Negative Emotional Variability.** In these analyses, we investigated the links between neuroticism and negative emotional variability. Third, we metaanalyzed the association between neuroticism and the within-person SD of negative emotions, with no correction for mean negative emotion (Fig. 3*A*). Replicating previous work, neuroticism had a significant medium-sized positive correlation ($r = 0.28$) with negative emotional variability.

Fourth, for the eight ESM datasets with multiple measures per day, we metaanalyzed the association between neuroticism and variability in the daily maximum of negative emotions (Fig. 3*B*). To check the validity of this measure, we calculated the metaanalytic association between variability based on the daily maxima and variability based on all time points. We found a significant and large metaanalytic correlation ($r = 0.68$; *SI Appendix*, Fig. S10), suggesting the two indices are tapping the same construct, but are not redundant with each other. As predicted, the association between neuroticism and the within-person SD of the daily maxima was smaller than the association between neuroticism and the within-person SD across all data points: there was a small positive correlation ($r = 0.10$). In addition, this correlation did not fall within the confidence intervals of the association between neuroticism and the within-person SD (0.21 to 0.35), providing evidence that the correlations were statistically different in size. This analysis provided initial evidence that the observed neuroticism–variability relationship may be partially driven by the many scores at the lower scale boundary.

Fifth, for all studies, we metaanalyzed the association between neuroticism and the relative SD (24) of negative emotions (Fig. 3*C*). This measure statistically corrects for the dependency between variability and mean levels, conceptualizing variability as a proportion of the maximum variability possible given the mean. As predicted, when using the relative SD, the association between neuroticism and negative emotional variability was weak ($r = 0.05$) and had a 95% CI including zero. This suggests the association between neuroticism and negative emotional variability can be explained more parsimoniously by mean levels.

**Supplemental Analyses.** To check the robustness of these results, we ran three additional sets of analyses. First, the issues with the within-person SD also apply to the mean squared successive difference (MSSD), a measure of moment-to-moment instability also linked to neuroticism (18). We also ran our main analyses using the MSSD rather than the within-person SD (*SI Appendix*, Figs. S1–S3). When using the relative MSSD to correct for mean levels, we found that the direction of the association reversed: there was a small negative association between neuroticism and negative emotional variability, providing further evidence that mean levels can obscure the effect of instability.

Second, we ran analyses separately for the two neuroticism measures (subscales from the Big Five Inventory, $n = 5$; and the Ten Item Personality Inventory [TIPI], $n = 6$). We did this because, in some datasets, the TIPI had low reliability, a known issue with this brief scale (27). Results were similar for the two sets of analyses (*SI Appendix*, Figs. S4–S9): the confidence

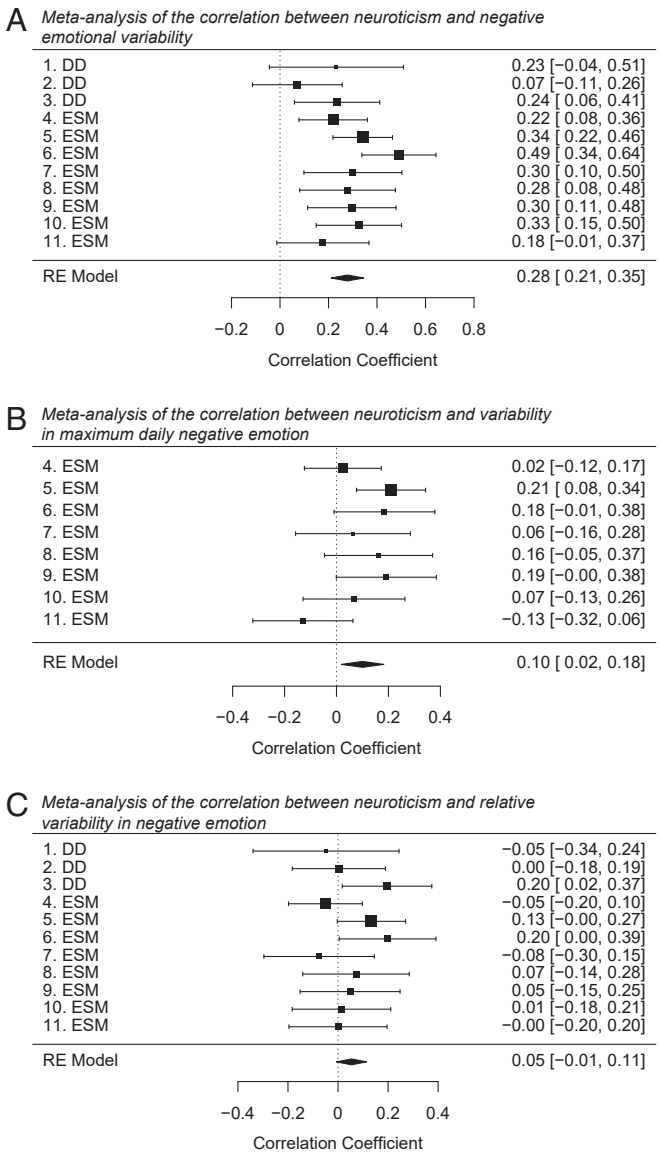


**Fig. 3.** Forest plots of the relationship between neuroticism and negative emotional variability (the within-person SD; *A*), variability in maximum daily negative emotion (for ESM datasets with multiple daily measures only; *B*), and relative variability in negative emotion (*C*). For each dataset, we provide a correlation bounded by a 95% CI. The area of each square is proportional to the weight of the study in the metaanalysis. The results of the random-effects metaanalysis are depicted at the bottom of the figure (RE model), with the width of the rhombus representing the 95% CI. The dotted line represents no effect. DD, daily diary; ESM, experience sampling method.

intervals of the metaanalytic effects overlapped, and reflected the overall findings reported here.

Third, to provide additional evidence for the primacy of mean levels over variability, we metaanalyzed the adjusted correlation between neuroticism and the within-person SD after partialing out mean levels (*SI Appendix*, Fig. S11). In these analyses, the association between neuroticism and the within-person SD was significant, but smaller in size than in the analyses not accounting for mean levels (dropping from $r = 0.28$ to $r = 0.12$). We also metaanalyzed the adjusted correlation between neuroticism and mean levels after partialing out the within-person SD (*SI Appendix*, Fig. S12). This association was significant (dropping from $r = 0.36$ to $r = 0.23$), and the adjusted association was stronger than the adjusted association with the within-person SD ($r =$ 0.12), providing additional evidence for the primacy of the mean. These analyses are less conservative than the analysis using the relative variability index, since they do not account for curvilinear relationships or multicollinearity, but present additional evidence for the predominance of the mean.

## Discussion

Neuroticism is thought to be typified not only by more intense negative emotions, but also by greater negative emotional variability (28). However, we hypothesized that the association between neuroticism and negative emotional variability was driven by a dependency between variability and mean levels of negative emotion. In a metaanalytic investigation of 11 daily life studies, we found evidence supporting this hypothesis.

In our first two analyses, we found evidence for a dependency between variability and mean levels in negative emotion that could cloud the relationship between neuroticism and variability. First, neuroticism was associated with higher mean levels of negative emotion, suggesting that the lower scale boundary may preclude those low in neuroticism from demonstrating variability. Second, mean levels of negative emotion were positively associated with negative emotional variability. Thus, individuals who experienced low mean levels of negative emotion (i.e., most participants in the studies of daily emotional experience) are restricted in how emotionally variable they can be.

Third, replicating previous work (18), we found a medium-sized positive association between neuroticism and negative emotional variability. The final two analyses set out to determine whether this association could be more parsimoniously explained using mean levels of negative emotion. First, we investigated the association between neuroticism and variability in daily maximum negative emotion, which removed many of the occasions on which participants could not demonstrate variability because they had scores near the scale floor. As predicted, we found an attenuated small positive association between neuroticism and variability in daily maximum negative emotion.

Finally, we investigated this association using the relative variability index, which removes the dependency between mean levels and variability. It does so by calculating variability as a proportion of the maximum variability theoretically possible given the observed mean and scale end points (24). When using this index, we found no reliable evidence of an association between neuroticism and negative emotional variability.

Negative emotional variability is implicated in the definition, labeling, and assessment of neuroticism (e.g., refs. 2, 13, and 29). However, once we corrected for the mathematical dependence between the mean and variability, neuroticism was no longer associated with greater variability in negative emotions. This was likely because neuroticism had a strong positive association with mean levels of negative emotion, meaning that highly neurotic individuals were less likely than less neurotic individuals to use the lower scale boundary when rating their negative emotional experience. This suggests associations with variability could be more simply explained using the mean, and that we should move away from highlighting emotional (in)stability as a core feature of neuroticism.

We are not able to infer the exact nature of the relation between the mean and the SD using these data. That is, it is not possible to know whether people with mean levels near the scale boundaries truly have low variability, whether the scale boundaries cause low variability, or whether there is some mixture of both. However, given that both constructs reflect the same information, we believe that mean levels represent the most parsimonious statistic and theoretical account of these data without the need to calculate the more complex within-person SD. These data provide support for the mean as the primary statistic. First, mean levels were more strongly associated with neuroticism than variability. Second, analyses using the daily maxima, using the

relative variability index, and partialing out mean levels from variability all supported the idea that the mean best accounts for these findings. With this in mind, we believe it unwise to rely on the counter-assumption that variability is independent of the mean, and this is what researchers do when they undertake the common practice of reporting variability statistics without accounting for mean levels. We suggest that, where variability is theoretically important, it should be considered, but interpreted cautiously and with reference to mean levels.

These findings highlight the need to wrestle with negative emotion measurement. In daily life, with healthy participants, we expect low levels of negative emotion to be the norm. Thus, the bigger question is whether there could be true variability present even when participants report repeated scores at the lower scale boundary. Are participants who are using the lower boundary truly experiencing no negative emotion, and is that lower boundary indexing the same level of negative emotion every time it is used? Indeed, another interpretation of our results is that there is an association between neuroticism and negative emotional variability, but measurement issues prevent it from being demonstrated independently of mean levels. With these points in mind, we believe decoupling measurement issues from variability is critical in determining if there is utility in studying variability as a feature of neurotic emotional experience. We see three avenues to address these measurement issues, which we outline alongside some potential challenges.

First, to address the many zero scores observed in community samples, studies could focus on clinical samples characterized by heightened negative emotion (e.g., ref. 30). However, some of our data suggest that this approach may also result in dependency between variability and the mean. In dataset 2, comprising participants high in depressive symptoms, we initially found that neuroticism was associated with reduced variability. This association disappeared when using the relative variability index.

Second, studies could sample people in emotionally intense situations, where there are fewer scores at the scale floor. This taps into a bigger issue: when we follow normal people in their daily lives, we may often be assessing generalized affect or mood, rather than specific emotional reactions to stimuli (31). Thus, the most common method of testing the neuroticism–variability relationship may not be a good test of theory, which is based in the idea that high neuroticism is associated with stronger emotional reactivity (6). We attempted an investigation of more emotionally intense situations by testing the association between neuroticism and variability in daily maximum negative emotion. We found a positive association ($r = 0.10$), which was considerably smaller than the association calculated using all data points ($r = 0.28$). This suggests that, if we can find the true association by refocusing on stronger emotion, the relationship between emotional variability and neuroticism may be weaker than our initial estimates.

Third, studies could measure negative emotion in a way that avoids potential floor effects. We see three options. First, researchers could use a bipolar scale, ranging from negative to positive. However, this assessment method disregards that positive and negative emotions have different antecedents, consequences, and functions. Second, researchers could collect relative emotional intensity ratings by asking participants whether they feel more or less negative emotion than usual. This option may be a useful way to capture emotional variability, but integrating multiple sources of information might make the method more burdensome for participants, as well as requiring an analytic rethink on the part of researchers. Third, researchers could devote more effort to disentangling emotional intensity and frequency in their measures (32), better clarifying to participants that a score of 0 means no emotion rather than low emotion. Such a measure may still result in dependence between the mean and variability, but we would have a better idea of the role of measurement in that dependence.

Most of the studies included were conducted over weeks (the longest was 30 d), and some might wonder whether the association between neuroticism and negative emotional variability is only visible over longer timescales. While this is possible, it would not fit with the original theorizing around this link, which posited that the variability displayed by those with high neuroticism resulted from episodic reactivity (e.g., ref. 21).

More broadly, these results suggest we need to reexamine what everyday neurotic emotional experience looks like. To do so, future research will need to assess other emotional processes, which may also help us disentangle the dependency between the mean and variability. The hypothesis that neuroticism should manifest in emotional variability was based in the idea that neuroticism is characterized by emotional reactivity (6). If reactivity alone was implicated, we would expect to see associations with both variability and the mean, as reactivity should lead to spikes in responding. The centrality of mean levels could be because poor emotional recovery is the other piece of the puzzle: neuroticism is associated with poorer regulation (33) and more emotional inertia (34). Greater reactivity combined with a poorer ability to recover may manifest in the high mean levels of emotion reflected in the data, but not necessarily in more variable negative emotion.

These results may also call into question the study of variability in other fields. A potential dependency between the mean and the variability is a problem for all work using bounded scales. However, the issue is compounded when assessing negative emotion in daily life for two reasons: there are many scores at the lower scale boundary, and there is a focus on substantively interpreting variability. Indeed, emotion is often seen as inherently variable, and has been used as a benchmark to determine whether it is worth conceptualizing other psychological concepts as variable states (35). Because emotions are dynamic, researchers have mined and interpreted variability in emotion more so than in other domains. Indeed, in the case of neuroticism, emotional variability is part of the core of the construct. Thus, while these findings highlight potentially broader problems and have general implications for work on variability, they also form part of the call for caution when using complex measures in emotion research (36).

In sum, across 11 daily life studies, we found that the association between neuroticism and negative emotional variability disappeared when accounting for the dependency between variability and mean levels of negative emotion. These findings provide evidence that the association between neuroticism and negative emotional variability cannot be demonstrated using existing assessment methods, and, more broadly, call into question whether there is any meaningful relationship between these two constructs.

## Materials and Methods

We metaanalyzed associations across 11 daily life studies. These studies were originally conducted to test other research questions, but all included measurements of both negative emotion in daily life and trait neuroticism. Three studies (datasets 1 to 3) used a daily diary design, and eight studies (datasets 4 to 11) used a momentary assessment design with multiple time points per day. All studies had ethical clearance at the university where they were conducted.

**Participants and Procedure.** Table 1 outlines the participants and procedure for each of the studies (additional methodological details are provided in *SI Appendix*, Table S1). We excluded participants who completed less than 50% of the momentary measurements because within-person variability measurements are based on how much a person fluctuates across time, and are likely to be less reliable with a large amount of missing data. Using this rule excluded very few participants ($n = 34$ of 1,239; *SI Appendix*, Table S1 provides more details), and the results are unchanged if these participants

**Table 1. Methodological information about each dataset**

| | | | Study details | | | | Participant details | | |
|---|---|---|---|---|---|---|---|---|---|
| Dataset | *N* | Obs. | Type | Days | Obs./d | Context of emotion assessment | Mean age, y (SD) | Gender, F/M | Reference with more information |
| 1 | 46 | 604 | Daily diary | 14 | 1 | That day | 45.07 (12.05) | 37/9 | Van Ryckeghem et al. (37) |
| 2 | 112 | 3,091 | Daily diary | 30 | 1 | That day | 34.27 (9.83) | 54/58 | Dejonckheere et al. (38) |
| 3 | 112 | 765 | Daily diary | 7 | 1 | Most negative event that day | 35.23 (11.87) | 57/57 | Kalokerinos et al. (39) |
| 4 | 176 | 29,100 | ESM | 21 | 9–10 | Momentary | 27.15 (9.03) | 117/58 | Grommisch et al. (40) |
| 5 | 200 | 12,085 | ESM | 7 | 10 | Momentary | 18.32 (0.96) | 110/90 | Erbas et al. (41) |
| 6 | 96 | 5,819 | ESM | 7 | 10 | Momentary | 19.05 (1.28) | 60/36 | Brans et al. (42) |
| 7 | 79 | 4,645 | ESM | 10 | 7 | Momentary | 22.18 (5.29) | 79 F | Holland et al. (43) |
| 8 | 85 | 4,946 | ESM | 14 | 5 | Momentary | 23.67 (4.25) | 85 F | Koval et al. (44) |
| 9 | 97 | 5,531 | ESM | 14 | 5 | Momentary | 26.45 (6.06) | 97 F | Koval et al. (44) |
| 10 | 100 | 8,557 | ESM | 14 | 7 | Momentary | 24.12 (6.88) | 77/22* | Dejonckheere et al. (45) |
| 11 | 101 | 8,268 | ESM | 9 | 10 | Momentary | 18.64 (1.45) | 87/14 | Kalokerinos et al. (46) |

Obs., momentary observations.
*One participant did not specify gender.

are included. After applying this exclusion, we were left with 1,205 participants and 83,411 observations, a sample size which allowed us to detect small correlations. The studies ranged from 7 to 30 d in length, and included community members, students, online samples, and participants selected based on clinical symptoms. They were collected in Belgium, Australia, and the United States.

**Materials and Measures.** Negative emotion was measured using different items in all studies (exact items are detailed in *SI Appendix*, Table S2). For each study, we created a mean of negative emotion at each measurement occasion using all available items within that study. We made this decision because all negative emotion items tapped the same broader construct of interest. All studies included at least one low arousal (e.g., sad) and high arousal (e.g., angry) item (31). The exception was dataset 11, which assessed generalized negative emotion using a single item. This dataset also included a multi-item negative emotion measure, but this measure was directed at a specific context, and so the single generalized item was chosen because it mapped more closely onto the other datasets. Negative emotion showed acceptable reliability both between person ($R_{KF}$ 0.96 to 0.99) and within person ($R_C$ 0.62 to 0.86; *SI Appendix*, Table S2). To compare findings, we rescaled these measures so that the range was the same across all datasets. Negative emotion was rescaled to a 1-to-7 scale (original scales ranged from 1 to 5 to 0 to 100). Neuroticism was rescaled to a 1-to-5 scale (original scales were either 1 to 5 or 1 to 7). This rescaling does not change the results; it was purely to allow for greater interpretability.

**Negative Emotional Variability.**

***Within-person SD.*** To measure variability, we used the within-person SD of negative emotion across time. This was the most common measure in published work (18). We calculated the SD in negative emotion for each participant across all measurement occasions.

***Relative variability index.*** To correct for the dependency between variability and mean levels, we calculated the relative variability index for negative emotion (24). Given the mean and end points of a set of measurements, this index calculates the maximum possible variability. It then divides the actual variability by this maximum to get a score between zero and one, where one indicates that actual variability matches the maximum theoretically possible.

The mean squared successive difference (MSSD) is also used as a measure of emotional variability and includes a temporal component (25). Thus, in *SI Appendix, Supplementary Materials*, we also present our primary analyses using the MSSD and the relative variability index for the MSSD. In calculating the MSSD, we excluded overnight lags (i.e., where the last survey of the previous day was used to predict the first survey of the next day).

**Neuroticism.**

***BFI.*** Five studies used the Neuroticism subscale of the Big Five Inventory (BFI) (2), which consists of eight items (e.g., "I see myself as someone who can be moody"), rated either on a 5- or 7-point scale ranging from "disagree strongly" to "agree strongly." This scale showed good reliability in all studies ($\alpha$ = 0.83 to 0.95; *SI Appendix*, Table S2).

***TIPI.*** Six studies used the Emotional Stability subscale of the Ten-Item Personality Inventory (TIPI) (27). This scale consists of two items ("I see myself as anxious, easily upset" and "I see myself as calm, emotionally stable") rated on a 7-point scale (1 = disagree strongly, 7 = agree strongly). We reverse-scored the "calm, emotionally stable" item to form an index of neuroticism. There was variability in the reliability of this scale, with some datasets having low reliability ($\alpha$ = 0.42 to 0.83; *SI Appendix*, Table S2), but our *SI Appendix, Supplemental Analyses* demonstrated that the results are not driven by this issue.

**Data and Code Availability.** Data and code for all analyses are publicly available on the Open Science Framework at https://osf.io/gvfdx/.

**ACKNOWLEDGMENTS.** This research was supported in part by an Australian Research Council (ARC) grant awarded to B.B. and P. Kuppens (DP140103757), the KU Leuven Research Fund (C14/19/054), a Horizon 2020 Marie Skłodowska-Curie Fellowship (704298) awarded to E.K.K., and a Washington University in St. Louis Psychological and Brain Science's Department Graduate Student Research Grant awarded to N.H.B., E.K.K., and P. Koval are supported by ARC Discovery Early Career Researcher Awards (DE180100352 and DE190100203).

1. J. M. Digman, Personality structure: Emergence of the five-factor model. *Annu. Rev. Psychol.* **41**, 417–440 (1990).
2. O. P. John, L. P. Naumann, C. J. Soto, "Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues" in *Handbook of Personality: Theory and Research*, O. P. John, R. W. Robins, L. A. Pervin, Eds. (Guilford Press, 2008), pp. 114–158.
3. D. H. Barlow, S. Sauer-Zavala, J. R. Carl, J. R. Bullis, K. K. Ellard, The nature, diagnosis, and treatment of neuroticism: Back to the future. *Clin. Psychol. Sci.* **2**, 344–365 (2014).
4. B. B. Lahey, Public health significance of neuroticism. *Am. Psychol.* **64**, 241–256 (2009).
5. M. Eid, E. Diener, Intraindividual variability in affect: Reliability, validity, and personality correlates. *J. Pers. Soc. Psychol.* **76**, 662–676 (1999).
6. H. J. Eysenck, M. W. Eysenck, *Personality and Individual Differences: A Natural Science Approach* (Springer, US, 1985).
7. M. D. Robinson, M. Tamir, Neuroticism as mental noise: A relation between neuroticism and reaction time standard deviations. *J. Pers. Soc. Psychol.* **89**, 107–114 (2005).
8. C. Kamen, L. R. Pryor, E. T. Gaughan, J. D. Miller, Affective lability: Separable from neuroticism and the other big four? *Psychiatry Res.* **176**, 202–207 (2010).
9. S. Marwaha *et al.*, How is affective instability defined and measured? A systematic review. *Psychol. Med.* **44**, 1793–1808 (2014).
10. J. D. Miller, P. A. Pilkonis, Neuroticism and affective instability: The same or different? *Am. J. Psychiatry* **163**, 839–845 (2006).
11. L. R. Goldberg, An alternative "description of personality": The big-five factor structure. *J. Pers. Soc. Psychol.* **59**, 1216–1229 (1990).
12. R. R. McCrae, P. T. Costa Jr, Validation of the five-factor model of personality across instruments and observers. *J. Pers. Soc. Psychol.* **52**, 81–90 (1987).
13. C. G. DeYoung, L. C. Quilty, J. B. Peterson, Between facets and domains: 10 aspects of the Big Five. *J. Pers. Soc. Psychol.* **93**, 880–896 (2007).
14. H. J. Eysenck, S. B. G. Eysenck, *Manual of the Eysenck Personality Questionnaire (Junior and Adult)* (Hodder and Stoughton, 1975).
15. C. J. Soto, O. P. John, The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *J. Pers. Soc. Psychol.* **113**, 117–143 (2017).

16. O. P. John, S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives" in *Handbook of Personality: Theory and Research*, L. A. Pervin, O. P. John, Eds. (The Guilford Press, 1999), pp. 102–138.
17. P. T. Costa Jr, R. R. McCrae; The Revised NEO Personality Inventory (NEO-PI-R), *The SAGE Handbook of Personality Theory and Assessment, Vol 2: Personality Measurement and Testing* (SAGE Publications, Thousand Oaks, CA, 2008), pp. 179–198.
18. M. Houben, W. Van Den Noortgate, P. Kuppens, The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychol. Bull.* **141**, 901–930 (2015).
19. B. M. Baird, K. Le, R. E. Lucas, On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *J. Pers. Soc. Psychol.* **90**, 512–527 (2006).
20. C. McConville, C. Cooper, Mood variability and personality. *Pers. Individ. Dif.* **13**, 1213–1221 (1992).
21. D. G. Williams, Effects of psychoticism, extraversion, and neuroticism in current mood: A statistical review of six studies. *Pers. Individ. Dif.* **11**, 615–630 (1990).
22. J. P. David, P. J. Green, R. Martin, J. Suls, Differential roles of neuroticism, extraversion, and event desirability for mood in daily life: An integrative model of top-down and bottom-up influences. *J. Pers. Soc. Psychol.* **73**, 149–159 (1997).
23. R. J. Larsen, T. Ketelaar, Personality and susceptibility to positive and negative emotional states. *J. Pers. Soc. Psychol.* **61**, 132–140 (1991).
24. M. Mestdagh *et al*., Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychol. Methods* **23**, 690–707 (2018).
25. S. Jahng, P. K. Wood, T. J. Trull, Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychol. Methods* **13**, 354–375 (2008).
26. W. Viechtbauer, Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
27. S. D. Gosling, P. J. Rentfrow, W. B. Swann, A very brief measure of the Big-Five personality domains. *J. Res. Pers.* **37**, 504–528 (2003).
28. H. J. Eysenck, *A Model for Personality* (Springer, 1981).
29. L. R. Goldberg, A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Pers. Psychol. Eur.* **7**, 7–28 (1999).
30. L. A. Clark, D. Watson, Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *J. Abnorm. Psychol.* **100**, 316–336 (1991).
31. J. A. Russell, Core affect and the psychological construction of emotion. *Psychol. Rev.* **110**, 145–172 (2003).
32. U. Schimmack, E. Diener, Affect intensity: Separating intensity and frequency in repeatedly measured affect. *J. Pers. Soc. Psychol.* **73**, 1313–1329 (1997).
33. J. J. Gross, O. P. John, Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *J. Pers. Soc. Psychol.* **85**, 348–362 (2003).
34. J. Suls, P. Green, S. Hillis, Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Pers. Soc. Psychol. Bull.* **24**, 127–136 (1998).
35. W. Fleeson, Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *J. Pers. Soc. Psychol.* **80**, 1011–1027 (2001).
36. E. Dejonckheere *et al*., Complex affect dynamics add limited information to the prediction of psychological well-being. *Nat. Hum. Behav.* **3**, 478–491 (2019).
37. D. Van Ryckeghem, S. Rost, N. Decoene, G. Crombez, *Ghent Attention and Self-Regulation in Fibromyalgia-I Study (ASEF-1)* (Protocol, 2014).
38. E. Dejonckheere, B. Bastian, E. I. Fried, S. C. Murphy, P. Kuppens, Perceiving social pressure not to feel negative predicts depressive symptoms in daily life. *Depress. Anxiety* **34**, 836–844 (2017).
39. E. K. Kalokerinos, M. Tamir, P. Kuppens, Instrumental motives in negative emotion regulation in daily life: Frequency, consistency, and predictors. *Emotion* **17**, 648–657 (2017).
40. G. Grommisch *et al*., *Modeling Individual Differences in Emotion Regulation Repertoire in Daily Life with Multilevel Latent Profile Analysis. Emotion* (Advance Online Publication, 2019).
41. Y. Erbas *et al*., Why I don't always know what I'm feeling: The role of stress in within-person fluctuations in emotion differentiation. *J. Pers. Soc. Psychol.* **115**, 179–191 (2018).
42. K. Brans, P. Koval, P. Verduyn, Y. L. Lim, P. Kuppens, The regulation of negative and positive affect in daily life. *Emotion* **13**, 926–939 (2013).
43. E. Holland, P. Koval, M. Stratemeyer, F. Thomson, N. Haslam, Sexual objectification in women's daily lives: A smartphone ecological momentary assessment study. *Br. J. Soc. Psychol.* **56**, 314–333 (2017).
44. P. Koval, *et al*., How does it feel to be treated like an object? Direct and indirect effects of exposure to sexual objectification on women's emotions in daily life. *J. Pers. Soc. Psychol.* **116**, 885–898 (2019).
45. E. Dejonckheere, E. K. Kalokerinos, B. Bastian, P. Kuppens, Poor emotion regulation ability mediates the link between depressive symptoms and affective bipolarity. *Cogn. Emotion* **33**, 1076–1083 (2019).
46. E. K. Kalokerinos, Y. Erbas, E. Ceulemans, P. Kuppens, Differentiate to regulate: Low negative emotion differentiation is associated with ineffective use but not selection of emotion-regulation strategies. *Psychol. Sci.* **30**, 863–879 (2019).