



Affective variability in depression: Revisiting the inertia–instability paradox

Elisabeth H. Bos* , Peter de Jonge and Ralf F. A. Cox

Department of Developmental Psychology, Behavioural and Social Sciences,
University of Groningen, The Netherlands

How can depression be associated with both instability and inertia of affect? Koval *et al.* (2013, *Emotion*, 13, 1132) showed that this paradox can be solved by accounting for the statistical overlap between measures of affect dynamics. Nevertheless, these measures are still often studied in isolation. The present study is a replication of the Koval *et al.* study. We used experience sampling data (three times a day, 1 month) of 462 participants from the general population and a subsample thereof ($N = 100$) selected to reflect a uniform range of depressive symptoms. Dynamics measures were calculated for momentary negative affect scores. When adjusting for the overlap among affect dynamics measures, depression was associated with ‘dispersion’ (SD) but not ‘instability’ (RMSSD) or ‘inertia’ (AR) of negative affect. The association between dispersion and depression became non-significant when mean levels of negative affect were adjusted for. These findings substantiate the evidence that the presumed association between depression and instability is largely accounted for by the SD , while the association between dispersion and depression may largely reflect mean levels of affect. Depression may thus not be related to higher instability per se, which would be in line with theories on the adaptive function of moment-to-moment fluctuations in affect.

Major depressive disorder is one of the mood (or affective) disorders and is characterized by an alteration in mood or affect. Typically, this alteration involves heightened levels of depressed mood and negative affect (American Psychiatric Association, 2013; Barge-Schaapveld, Nicolson, Berkhof, & de Vries, 1999; Watson, Clark, & Carey, 1988). More recently, depression research has also focused on the *dynamic* aspects of affect, stimulated by technological advancements facilitating the repeated assessment of variables in daily life (Csikszentmihalyi & Larson, 1987; Ebner-Priemer & Trull, 2009). Such experience sampling studies have shown that depressed patients do not feel equally bad all the time; affect levels fluctuate, even in people with severe depression (Ben-Zeev & Young, 2010; Thompson *et al.*, 2012; Van Os *et al.*, 2017).

Feelings probably fluctuate for a reason. Theories on the functions of affect and emotions¹ in non-depressed people suggest that fluctuations are adaptive, serving to help the individual to adequately respond to environmental changes and demands (Carver,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

*Correspondence should be addressed to Elisabeth H. Bos, Department of Developmental Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712TS, Groningen, The Netherlands (email: elske.bos@rug.nl).

¹The terms ‘emotion’ and ‘affect’ are often used interchangeably in the literature (Koval, Sütterlin, & Kuppens, 2016; Kuppens, 2015).

2015; Frijda & Mesquita, 1994; Panksepp, 2012). Affective responses occur if an event is appraised as relevant to the individual's concerns, leading to changes in action readiness and concomitant physiological changes. As a result, the individual may interact with the environment, or not, depending on several regulation processes. In this way, changes in affect and emotion serve to monitor and safeguard the individual's needs, goals, and general well-being (Frijda & Mesquita, 1994; Panksepp, 2012). These theories also suggest that fluctuations in affect and emotions are functional only within certain boundaries; regulatory forces serve to prevent them from exceeding dysfunctional thresholds (Carver, 2015; Kuppens & Verduyn, 2017; Larsen, 2000).

If affect dynamics are adaptive and vital to well-being, structural alterations in these dynamics may be related to psychological dysfunction. Indeed, recent studies in depressed individuals have indicated that depression is not only related to alterations in mean levels of affect, but also to altered dynamics of affect (Houben, van den Noortgate, & Kuppens, 2015; Trull, Lane, Koval, & Ebner-Priemer, 2015; Wichers, Wigman, & Myin-Germeys, 2015). The majority of these studies have focused on three measures of affect dynamics. The first measure is the within-person standard deviation (*SD*) or variance. This is a measure of the general dispersion of the scores (Trull *et al.*, 2015). The *SD* is often referred to with the term 'variability' (Houben *et al.*, 2015; Koval, Pe, Meers, & Kuppens, 2013), but we will use the term 'dispersion', to avoid confusion with the broader use of the term 'variability' (as in 'intra- and interindividual variability', 'heart-rate variability', and 'moment-to-moment variability'). The second measure of affect dynamics is the first-order autocorrelation (*AR*), which is a measure of the temporal dependency of affect, or its resistance to change. This measure is referred to with the term 'inertia of affect' (Koval *et al.*, 2013; Trull *et al.*, 2015). The third measure is the mean square successive difference (*MSSD* or its square root: *RMSSD*), which captures moment-to-moment fluctuations (Jahng, Wood, & Trull, 2008). This measure is often called 'affective instability' (Koval *et al.*, 2013; Trull *et al.*, 2015). The *MSSD* captures both the magnitude and temporal dependency of affective fluctuations, in contrast to the *SD* (which captures the magnitude but not the temporal dependency) and the *AR* (which captures the temporal dependency but not the magnitude of fluctuations; Jahng *et al.*, 2008).

Paradoxically, the studies done thus far have indicated that depression is associated with higher dispersion, higher instability, *and* higher inertia of affect (see Houben *et al.*, 2015; Koval *et al.*, 2013; Trull *et al.*, 2015). This is a paradox, since inertia implies 'resistance to change' and thus would be expected to go along with rigidity rather than instability and variability. Koval *et al.* (2013) provided an elegant solution to this paradox, showing that the three measures of affect dynamics are not independent. They are mathematically related to each other with the formula $MSSD = 2 * SD^2(1 - AR)$ (see also Jahng *et al.*, 2008). Most studies have investigated these measures in isolation, which makes it hard to decide how each measure is independently related to depression. For example, the positive association between the *MSSD* and depression may either be due to higher dispersion (*SD*) or lower inertia of affect (*AR*) in depressives, or both. Therefore, Koval *et al.* used multiple regression analyses to examine how each measure of affect dynamics was uniquely related to depression while the other measures were adjusted for. Using experience sampling measures of negative affect in a sample of 100 students, they found that depression was indeed related to higher dispersion (*SD*), but not to inertia (*AR*) or instability (*RMSSD*) of affect when adjusting for dispersion. If this finding can be replicated, it makes a strong case for resolving the paradox of inertia versus instability. In that case, depression would not be related to higher instability per se, but mainly to higher dispersion. This would be in line with theories on the adaptive function of moment-to-moment fluctuations in affect.

In this study, we aimed to replicate the findings by Koval *et al.* (2013) in a general population sample. Our sample consisted of individuals who participated in the diary study of the HowNutsAreTheDutch project (www.hownutsarethedutch.nl). Following Koval *et al.* (2013), we selected 100 participants evenly distributed over the whole range of baseline depression severity. We also analysed the full sample ($N = 462$), in which depression scores were not evenly distributed but skewed to the right. We hypothesized that higher depression scores are related to higher dispersion, but not to higher instability or higher inertia of negative affect when adjusting for dispersion. We also examined to what extent the observed associations are accounted for by mean levels of negative affect, because affect dynamics measures may be confounded with the mean (Mestdagh *et al.*, 2018; Wagenmakers & Brown, 2007).

Method

Participants

Participants were sampled from an initial pool of 1,273 individuals from the general population who took part in the online diary study of the project 'HowNutsAreTheDutch' (www.hownutsarethedutch.nl; Van der Krieke *et al.*, 2015, 2016). These 1,273 individuals were those who participated between 22 May 2014 (launching of the diary study) and 19 December 2017 (end of 4-year wave of the 'HowNutsAreTheDutch' project). Inclusion criteria were age 18 years or older, having a smartphone, and approval of use of their anonymized data for research purposes (all assessed online). During the inclusion procedure, participants also stated not to be engaged in shift work and not to anticipate a major disruption of daily routines during the study period (e.g., a holiday). We first selected the participants who completed at least 75% of the diary measurements (≥ 68 of the 90 observations), which were 462 participants. From these, we sampled 100 participants with scores evenly distributed over the scale of depressive symptoms. We did so to obtain a sample representing a wide and uniform range of depression severity, replicating the procedure used by Koval *et al.* (2013). To this end, we divided the baseline (pre-diary) depression score (Quick Inventory of Depressive Symptoms, QIDS; Rush, Gullion, Basco, Jarrett, & Trivedi, 1996) into four roughly equal segments, and selected a random sample of 25 participants from each segment. The sample size of 100 was chosen by Koval to ensure sufficient power to detect moderate effect sizes. We repeated the analyses in the full sample ($N = 462$). In this sample, depression scores were skewed to the right. Median QIDS score of the full sample of 462 participants was 6.0 (interquartile range (IQR) = 6.0; range 0–24). Mean QIDS score of the subsample of 100 randomly selected participants was 9.4 ($SD = 5.5$; median = 9.5, IQR = 10.5; range 1–21). The $N = 100$ sample consisted of 83% females. Mean age was 42.7 years ($SD = 13.9$; range 20–75), and educational level was rather high (low 4.0%, middle 9.0%, high 83.0%; missing 4.0%). These characteristics were similar to those of the original HowNutsAreTheDutch sample and the sample of $N = 462$ participants (i.e., those who completed at least 75% of the measurements). The study was conducted in accordance with the Declaration of Helsinki. The protocol was evaluated by the Medical Ethics Committee of our institution and judged as exempt.

Materials and procedure

Dutch inhabitants were informed about the HowNutsAreTheDutch project by means of articles in newspapers and magazines, public lectures, and radio broadcasts. They were

invited to visit the website www.hoegekis.nl and to participate in a cross-sectional study, a diary study, or both. In the diary study, participants monitored their affect, behaviour, and cognitions three times a day for a period of 30 days (up to 90 measurements) by means of an electronic diary. The three time points were separated by equidistant intervals of 6 hr (e.g., 10.30 am, 4.30 pm, and 10.30 pm), the exact schedule depending on the participants' sleep-wake schedule. At each prompt, a text message with a link to the diary questionnaire was sent to the participants. Participants were instructed to fill out the diary questionnaire immediately after the prompt, but at least within 1 hr. Median response time was 10.4 min. For further details on the diary study, see Van der Krieke *et al.* (2016). Data of the HowNutsAreTheDutch study are available upon request.

Depressive symptoms

The Quick Inventory of Depressive Symptoms (QIDS-SR16; Rush *et al.*, 1996, 2006) was used to measure baseline depressive symptoms. This questionnaire was filled out just before the start of the diary study, during the online inclusion procedure. The QIDS-SR16 is a reliable and well-validated self-report questionnaire consisting of 16 items reflecting the nine DSM-IV symptoms of major depressive disorder (sad mood, loss of interest, concentration problems, self-criticism, suicidal ideation, energy/fatigue, sleep disturbance, appetite or weight change, and psychomotor agitation or retardation). Respondents are asked to rate the severity and frequency of these symptoms over the last 7 days. The total score can range from 0 to 27.

Momentary affect

Momentary negative affect was assessed using six items of the circumplex model of affect (Barrett & Russell, 1998; Yik, Russell, & Barrett, 1999). The circumplex model aims to assess both the valence (pleasantness) and activation (arousal) dimension of affect. In our study, low-arousal negative affect was assessed with the items 'gloomy', 'dull', and 'tired'. High-arousal negative affect was assessed with the items 'anxious', 'nervous', and 'irritable'. Participants rated the extent to which they felt so on a continuous slider scale ranging from 0 (not at all) to 100 (very much). For the purpose of the present study, we pooled the negative affect items in a composite 'negative affect' score, in line with Koval *et al.* (2013). The composite negative affect score was calculated by taking the mean of the six items (range 0–100). Cronbach's alpha for this measure was 0.86, which is a mix of between-subject and within-subject reliability in longitudinal designs. Separating this mix following the recommendations by Shrout and Lane (2012) yielded a between-subject reliability of 0.99 and a within-subject reliability of 0.71.

Affect dynamics measures

We calculated the measures of affect dynamics separately for each individual. Following Koval *et al.* (2013), the within-person *SD* was used as a measure of dispersion of affect. As a measure of affective instability, the RMSSD was calculated (the square root of the MSSD; Jahng *et al.*, 2008). A high RMSSD represents high moment-to-moment variability. The first-order AR was calculated as a measure of temporal dependency or inertia of affect (Kuppens, Oravecz, & Tuerlinckx, 2010). Also multilevel methods have been proposed to estimate the measures of affect dynamics (e.g., Bernstein, Curtiss, Wu, Barreira, & McNally, 2018; Trull *et al.*, 2008), but we did not use these for two reasons: First, we

wanted to follow the procedure used in Koval *et al.* (2013) and most other studies done on this topic (cf. Houben *et al.*, 2015). Second, these multilevel methods do not allow for the examination of all three measures of affect dynamics in one and the same model.

Statistical analyses

Mean number of completed observations of the 100 participants was 77.9 (range 68–90, $SD = 5.6$). In the full sample of $N = 462$, this was quite similar (77.6, $SD = 5.6$, range 68–90). The total number of completed observations in the $N = 100$ sample was 7,786; 1,214 of the 9,000 measurements were missing (13.5%). In the full sample, the total number of completed observations was 35,842 (13.7% missing). In our main analyses, we did not impute these missing values. In the calculation of the affect dynamics measures, missing values were excluded by leaving an empty row in the data file, in order to maintain equal intervals for the calculation of RMSSD and AR. First, we present Pearson's correlations between the baseline QIDS score and all affect dynamics measures (SD , RMSSD, AR). Second, we examined the associations between the affect dynamics measures and depression using three multiple regression models, adding two of the three measures simultaneously in a model, following Koval *et al.* (2013). We adjusted these models for mean levels of affect in a subsequent step. In view of the expected correlations between the predictors, we checked the collinearity diagnostics to detect problematic multicollinearity; VIF values were all below 10, and condition indices were all below 30 (highest VIF was 4.48, and highest condition index was 20.3), suggesting no serious problem with multicollinearity (Hair, Anderson, Tatham, & Black, 1998). The analyses were performed in SPSS25, and a p -value of .05 was used as the significance level.

In the analysis of the full sample ($N = 462$), baseline QIDS scores were skewed to the right. Also, the distributions of the predictor variables were somewhat skewed in this larger sample, especially mean levels of negative affect. To prevent violation of model assumptions and undue influence of the heavy tails, robust regression analysis with bootstrapped confidence intervals was used for these analyses. These analyses were performed in R, using the functions 'lmrob' from the *robust* package and 'bootEst' from the *complmrob* package.

We also performed two sensitivity analyses in the $N = 100$ sample. First, we examined the potential influence of missing data by repeating the analyses in an imputed data set. We imputed the missing observations using Random Forest imputation (Stekhoven & Bühlmann, 2012), for each participant separately. Random Forest imputation is a non-parametric machine learning method for imputation of multivariate data. For this imputation, we used all diary questionnaires (except skip questions) and their lagged versions, negative affect and its lagged version, a variable denoting the measurement point and its square, daypart, day of week, baseline QIDS scores, age, and sex. Imputation was done with the R package *missForest* (Stekhoven & Bühlmann, 2012). Second, we performed an analysis in which the interval from the evening to the morning measurement was excluded from the calculation of the RMSSD and AR. We did so because this interval is longer than the other intervals, since participants could not be assessed during the night.

Results

Correlations between depression severity and affect dynamics

Table 1 presents descriptive statistics and Pearson's correlations between baseline depression severity (QIDS score) and the measures of affect dynamics for the $N = 100$

sample. Dispersion (*SD*) of negative affect was significantly and positively associated with depression severity ($r = .30, p = .003$). Thus, more depressive symptoms went along with higher dispersion of negative affect. Instability (RMSSD) was also positively associated with depression severity ($r = .25, p = .011$). The correlation between inertia (AR) of affect and depression was not significant ($r = .09, p = .388$). Mean levels of negative affect were also positively associated with depression severity ($r = .68, p < .001$). As could be expected given the formula $MSSD = 2 * SD^2(1 - AR)$, there were several significant intercorrelations among the affect dynamics measures. In particular, the correlation between the *SD* and the RMSSD stands out; this correlation was large ($r = .88, p < .001$). Noteworthy is also the moderate correlations between the mean and the *SD* ($r = .25, p = .013$) and the mean and the RMSSD ($r = .24, p = .017$), reflecting some floor effects in the negative affect variable (the distribution of this variable was a bit skewed to the right).

Multiple regression analyses

To examine how depression severity was related to the affect dynamics measures while accounting for the overlap among the latter, we performed multiple regression analyses. Table 2 shows the results for the $N = 100$ sample. In each model, we included two of the three affect dynamics measures as predictors (step 1). In step 2, we also included the mean level of negative affect. The betas from these models are the standardized regression coefficients. These can be interpreted as effect sizes for the independent associations between each measure of affect dynamics while adjusting for one of the other measures (step 1) and while additionally adjusting for mean levels of affect (step 2).

In model 1, depression severity was regressed on dispersion (*SD*) and inertia (AR) of affect. Only dispersion of affect was independently associated with depression severity in this model. The effect size was moderate: 0.29. When mean level of negative affect was added to the model in step 2, this effect size was substantially reduced ($\beta = 0.13$) and not significant anymore. Mean level itself was a very strong predictor of depression severity

Table I. Descriptive statistics and Pearson's correlations for depression severity and negative affect dynamics ($N = 100$)

Measure	Mean (<i>SD</i>)	QIDS	Correlations			
			negative affect dynamics			
			<i>M</i>	<i>SD</i>	AR	RMSSD
QIDS	9.4 (5.5)					
negative affect						
Mean level (<i>M</i>)	31.1 (16.0)	.68				
Dispersion (<i>SD</i>)	10.8 (3.1)	.30	.25			
Inertia (AR)	0.29 (0.18)	.09	.05	.19		
Instability (RMSSD)	12.6 (3.7)	.25	.24	.88	-.26	

Notes. QIDS = Quick Inventory of Depressive Symptoms (baseline assessment); negative affect dynamics are person-based measures calculated from momentary negative affect scores. AR = autocorrelation; *M* = mean level; RMSSD = root mean square successive difference; *SD* = standard deviation.

Correlations in bold are significant at $p < .05$ (two-tailed).

Table II. Results of multiple regression models predicting depression severity from dynamics and mean level of negative affect ($N = 100$)

	Step 1			Step 2 (including mean level)		
	B (SE)	β	p	B (SE)	β	p
Model 1						
Dispersion (<i>SD</i>)	0.52 (0.18)	0.29	.004	0.24 (0.14)	0.13	.089
Inertia (AR)	1.00 (2.98)	0.03	.736	0.87 (2.26)	0.03	.701
Mean level	–	–	–	0.22 (0.03)	0.65	<.001
Adjusted R^2	.071			.467		
Model 2						
Dispersion (<i>SD</i>)	0.60 (0.37)	0.33	.106	0.39 (0.28)	0.22	.160
Instability (RMSSD)	–0.06 (0.30)	–0.04	.843	–0.14 (0.23)	–0.09	.544
Mean level	–	–	–	0.22 (0.03)	0.65	<.001
Adjusted R^2	.070			.468		
Model 3						
Inertia (AR)	4.99 (3.04)	0.17	.104	2.57 (2.33)	0.09	.271
Instability (RMSSD)	0.44 (0.15)	0.30	.004	0.18 (0.12)	0.12	.126
Mean level	–	–	–	0.22 (0.03)	0.64	<.001
Adjusted R^2	.071			.464		

Notes. Linear regression analyses. Dependent variable = Quick Inventory of Depressive Symptoms (QIDS) score at the baseline assessment. AR = autocorrelation; RMSSD = root mean square successive difference; *SD* = standard deviation; SE = standard error. $N = 100$.

Effects with $p < .05$ in bold.

($\beta = 0.65$). The percentage of explained variance increased from 7.1 to 46.7% when mean level was included in the model.

In model 2, depression severity was regressed on dispersion (*SD*) and instability (RMSSD) of negative affect. Neither of these predictors were significant. The regression coefficient for instability was almost zero in this model ($\beta = -0.04$), in contrast with the positive significant correlation between instability and depression in Table 1. Instability did not explain any extra variance above what was explained by dispersion in this model (checked by entering the predictors one by one). So, the apparent association between instability and depression seems to be accounted for by the ‘*SD* part’ of the RMSSD.

In model 3, depression severity was regressed on inertia (AR) and instability (RMSSD) of affect. In this model, instability was a significant predictor of depression, inertia was not. When mean level was included in the model, instability was not significantly related to depression anymore. The effect size for instability dropped from 0.30 to 0.12 in this second step.

We repeated these analyses in the full sample ($N = 462$; Table 3). The results were highly similar, except that the effect of dispersion (*SD*) was now also significant in the first step of model 2 and inertia (AR) was now significant in the first step of model 3. In the mean-adjusted models, none of the dynamics measures were significant.

Sensitivity analyses

First, we repeated the analyses for the $N = 100$ sample in a data set in which missing values were imputed. The results were essentially the same; in all models, the same effects were

Table III. Full sample ($N = 462$) results of multiple regression models predicting depression severity from dynamics and mean level of negative affect

	Step 1				Step 2 (including mean level)			
	B	β	Bootstrapped 95% CI		B	β	Bootstrapped 95% CI	
			Lower	Upper			Lower	Upper
Model 1								
Dispersion (<i>SD</i>)	0.40	0.28	0.28	0.51	0.10	0.07	-0.01	0.22
Inertia (AR)	-0.51	-0.02	-2.42	1.44	-0.06	-0.002	-1.96	1.82
Mean level	-	-	-	-	0.20	0.58	0.17	0.23
Adjusted R^2	0.087				0.37			
Model 2								
Dispersion (<i>SD</i>)	0.28	0.20	0.02	0.54	0.03	0.02	-0.23	0.32
Instability (RMSSD)	0.11	0.09	-0.10	0.33	0.06	0.05	-0.16	0.29
Mean level	-	-	-	-	0.20	0.58	0.17	0.23
Adjusted R^2	0.088				0.37			
Model 3								
Inertia (AR)	2.58	0.11	0.80	4.43	0.74	0.03	-1.22	2.72
Instability (RMSSD)	0.34	0.29	0.25	0.44	0.10	0.08	-0.0005	0.19
Mean level	-	-	-	-	0.20	0.58	0.17	0.23
Adjusted R^2	0.089				0.37			

Notes. Robust regression analyses with bootstrapped 95% confidence intervals. Dependent variable = Quick Inventory of Depressive Symptoms (QIDS) score at the baseline assessment. AR = autocorrelation; CI = confidence interval; RMSSD = root mean square successive difference; SD = standard deviation; SE = standard error. $N = 462$.

Effects with $p < .05$ in bold.

significant as in the original models and the coefficients were only slightly different. Second, we repeated the analyses excluding the interval from the evening to the morning measurement in the calculation of the RMSSD and AR. The results of these analyses were also highly similar, except that dispersion (*SD*) now was a significant predictor in model 2, step 1 ($B = .70$, $\beta = 0.39$, $p = .035$). In step 2, when mean levels were adjusted for, this effect was not significant anymore ($B = .47$, $\beta = 0.26$, $p = .063$). In sum, the sensitivity analyses yielded highly similar results as the original analyses: When the affect dynamics measures were adjusted for each other, only dispersion of affect (*SD*) was associated with depression, and the latter effect in turn disappeared when mean levels were adjusted for. Detailed results of all sensitivity analyses are available from the first author upon request.

Discussion

In this study, we aimed to replicate the study by Koval *et al.* (2013), who solved the paradoxical finding that depression is positively associated with both instability and inertia of affect. Koval *et al.* showed that when statistically adjusting for the overlap among affect dynamics measures, depressive symptoms are associated with dispersion (*SD*) but not with instability (RMSSD) or inertia (AR) of negative affect in daily life. Our study showed the same; instability and inertia of negative affect were not associated with

depressive symptoms when dispersion was adjusted for. This replication of the Koval *et al.* study substantiates the evidence that the presumed association between depression and instability is mainly accounted for by the *SD*.

This resolution of the paradox may shed a different light on earlier results. In a recent meta-analysis, Houben *et al.* (2015) reported an overall correlation of -0.21 between instability of affect (MSSD/RMSSD) and psychological well-being (obtained from 146 associations found in 79 articles). However, this correlation was based on zero-order correlations because most studies did not report associations adjusted for the overlap among the affect dynamics measures. The authors mentioned this as an important limitation and caveat in their discussion. It remains to be seen whether the association between the MSSD and psychological well-being will hold true in meta-analyses of studies that do adjust for the overlap among dynamics measures.

If it is true that it is mainly the '*SD* part' of the MSSD that accounts for its association with depression, the term 'instability' to denote the MSSD may not be adequate and too negative; the more neutral term 'moment-to-moment variability' might be more appropriate, as fluctuations in affect do not seem to be unfavourable per se. Important emotion theories suggest that some degree of moment-to-moment variability in affect may even be healthy and adaptive, because it enables us to adequately respond to environmental demands (Carver, 2015; Frijda & Mesquita, 1994; Kashdan & Rottenberg, 2010; Panksepp, 2012). It is therefore not surprising that in some parts of the literature, 'moment-to-moment variability' is referred to with the term 'flexibility' (e.g., Hollenstein, 2015; Kashdan & Rottenberg, 2010), which has a positive connotation. Fluctuations in affect may especially be functional at the shorter timescale, that is, seconds or minutes. At this timescale, moment-to-moment variability may reflect functional responsiveness to situational contingencies (Koval *et al.*, 2016). At longer timescales (e.g., several hours or days), moment-to-moment variability may rather reflect mood swings and may thus be unfavourable (Koval *et al.*, 2013). The fact that several different timescales are used in studies on this topic may therefore be one reason for the existence of such contrasting terms for the (R)MSSD. Remarkably, the meta-analysis by Houben *et al.* (2015) did not show an effect of time interval on the association between the (R)MSSD and psychological well-being (maximum interval of included studies was 1 week). This might be due to the fact that the *SD* was not adjusted for and/or that studies with shorter timescales were underrepresented in this meta-analysis.

Another reason for the use of both positive and negative terms for the (R)MSSD may be that both extremes in moment-to-moment variability might be deleterious. Both too much instability and too much rigidity in affect are thought to be detrimental, and both are observed in depression (Booij, Snippe, Jeronimus, Wichers, & Wigman, 2018; Peeters, Nicolson, Berkhof, Delespaul, & de Vries, 2003; Rottenberg, 2005; Trull *et al.*, 2008; Wichers *et al.*, 2010). When both excessive and blunted emotional reactivity are associated with psychopathology, healthy emotion regulation is probably a matter of achieving an optimum (Hollenstein, 2015; Kashdan & Rottenberg, 2010; Kuppens & Verduyn, 2017). This would imply an inverted *U*-shape as regards the association between measures of variability of affect and depression. Most studies thus far have only examined linear associations (Houben *et al.*, 2015), and we did so as well because we wanted to replicate Koval *et al.* (2013). What level of moment-to-moment variability is optimal will also depend on the context: Whether or not an event has occurred and what was the nature of that event (Houben *et al.*, 2015; Thompson *et al.*, 2012). Like many other studies in this field, we did not consider this context in our analysis, which is a major limitation.

We did find a positive association between dispersion of affect (*SD*) and depression in our study. However, adjusting for mean levels of affect substantially reduced this association, rendering it non-significant. Also in the Koval *et al.* (2013) study, this association became non-significant after mean adjustment. It is well known that the *SD* (and also the MSSD/RMSSD) can be related to the mean, especially if variables have skewed distributions or show a restriction of range (e.g., Mestdagh *et al.*, 2018; Wagenmakers & Brown, 2007). In ESM data, positively skewed distributions are quite common, especially for negative affect or symptom variables assessed in general population samples. Indeed, our negative affect variable showed some positive skewness, and the mean and *SD* of this variable were correlated ($r = .25$). In the Koval *et al.* (2013) study, this correlation was also substantial, even as high as $r = .81$. This potential mean-level dependence of the *SD* may have consequences for the interpretation of earlier studies. The meta-analysis by Houben *et al.* (2015), for example, showed an overall correlation of -0.18 between dispersion of affect (*SD*) and measures of psychological well-being, but this effect size was not adjusted for mean levels because most studies did not report these. It may also explain why associations between dispersion in positive emotions and psychological well-being are often absent or less strong (Houben *et al.*, 2015; Koval *et al.*, 2013); ESM data of positive emotions are often not skewed, and the *SD* of these variables will thus be less strongly related to the mean. Future studies thus should account for mean levels or use measures that are independent of the mean, for example, the recently proposed relative variability index (Mestdagh *et al.*, 2018), or use complex dynamics system models in which several dynamics measures can be studied in concert (e.g., Kuppens *et al.*, 2010).

Nevertheless, in one of the models the association between dispersion of negative affect and depression showed a trend in the expected direction even after adjustment for mean levels of affect. The same trend was observed in the larger sample ($N = 462$). However, the effect size was rather small ($\beta = 0.13$), five times smaller than the effect size of mean levels of negative affect ($\beta = 0.65$), and the same was true in the larger sample ($\beta = 0.07$ and 0.58 for *SD* and mean, respectively). Further, if this effect would be real, it would tell us that depression goes along with a greater range or dispersion in negative affect, which, after all, may not be a very surprising insight. Trull *et al.* (2015) have called the *SD* 'an inappropriate summary index to capture a dynamics process' (p356), because it does not account for the temporal dependencies between successive observations. Shifting the temporal order of a series does not yield a different *SD*. Further, a high *SD* may reflect frequent fluctuations, slow but extreme fluctuations, or an extreme trend in affect over time (Jahng *et al.*, 2008; Larsen, 1987). If affective variability is to be an important construct in our emotion theories, we need to discriminate individuals showing such different patterns of affect dynamics.

In our study, the AR was not related to depression scores, except in the unadjusted model 3 in the larger sample. This effect disappeared when mean levels were adjusted for. Interestingly, the AR has been found to be related to depression in laboratory studies using a film task, even after adjustment for dispersion and mean levels of affect (Koval *et al.*, 2013, 2016). In this film task, the emotional content to which people were exposed was under the experimenter's control. This may be a crucial difference with the present ESM study, in which emotional exposure varied across participants. Also, the fact that in the film task different emotional responses (positive, negative, neutral) were induced may be relevant, since this may increase fluctuations in affective states and reduce floor and ceiling effects, as a result of which differences in inertia of affect might become more pronounced. Also, the very short timescale (seconds, minutes) of the film task may be a

distinguishing feature here (Koval *et al.*, 2013; Kuppens, 2015). Future research is needed to shed more light on this issue.

Limitations of the present study include the following. First, generalizability of the sample may have been limited. Although the source sample came from the general population, the majority of the participants were female and highly educated. Second, our sampling frequency was lower than in the Koval *et al.* study (three vs. ten times a day). Due to our larger interval, we may have missed rapid changes in affect (Trull *et al.*, 2008, 2015) and may have tapped more into changes due to mood swings. Although the meta-analysis by Houben *et al.* (2015) did not show an effect of time interval on the association between the dynamics measures and psychological well-being, the longer time interval in our study may have impacted the size of our estimates. A third limitation is that the interval between the evening and the morning measurement was longer than the other two intervals. Although one could argue that ‘experienced time’ in this interval was actually not longer because participants were asleep, we did perform a sensitivity analysis in which we excluded this interval. This yielded highly similar results, and it did not change our main conclusions. A final issue was the high correlation between the *SD* and the RMSSD. This correlation was also very high in the Koval *et al.* (2013) study. In other studies, this correlation may be lower, depending on the nature of the studied variables. In our study, it might have produced some multicollinearity in the regression models. Although the collinearity diagnostics suggested no serious collinearity problems, there is always *some degree* of multicollinearity if variables are correlated (Baguley, 2012). Although this leaves the regression estimates unbiased, it does reduce the effective sample size (Baguley, 2012). Indeed, in the larger sample the effect of the *SD* in the unadjusted model 2 became significant. This effect was non-significant in the mean-adjusted model and did not change the conclusions of the present paper. The high correlation between the variables also reflects a logical problem: The *SD* is by definition part of the RMSSD. The degree to which, however, is the point at stake. Two series may have the same *SD* but different RMSSDs, if the AR component of these series is different (Jahng *et al.*, 2008). The major issue is that in many of the studies who found evidence for an association between the RMSSD and depression or other outcomes, it is unclear to what extent this effect could just as well have been explained by the *SD*.

To conclude, this study showed that the presumed association between depression and ‘instability’ (MSSD/RMSSD) of affect is merely accounted for by the *SD* and that the association between ‘dispersion’ (*SD*) and depression is mainly accounted for by the mean. Therewith, we replicated Koval *et al.* (2013) in their solution of the instability–inertia paradox and additionally highlighted the potential confounding by mean levels in this area of research.

Acknowledgements

We are grateful to all participants of the HowNutsAreTheDutch study. We thank Koval *et al.* for their thoughtful study that inspired us to do this research, and we acknowledge two anonymous reviewers for their fruitful comments.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

- Baguley, T. (2012). *Serious Stats: A guide to advanced statistics for the behavioral sciences*. UK: Palgrave MacMillan. <https://doi.org/10.1007/978-0-230-36355-7>
- Barge-Schaapveld, D. Q. C. M., Nicolson, N. A., Berkhof, J., & de Vries, M. W. (1999). Quality of life in depression: Daily life determinants and variability. *Psychiatry Research*, *88*, 173–189. [https://doi.org/10.1016/S0165-1781\(99\)00081-5](https://doi.org/10.1016/S0165-1781(99)00081-5)
- Barrett, L. F., & Russell, J. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*, 967–984. <https://doi.org/10.1037/0022-3514.74.4.967>
- Ben-Zeev, D., & Young, M. A. (2010). Accuracy of hospitalized depressed patients' and healthy controls' retrospective symptom reports. An experience sampling study. *Journal of Nervous and Mental Disease*, *198*, 280–285. <https://doi.org/10.1097/NMD.0b013e3181d6141f>
- Bernstein, E. E., Curtiss, J. E., Wu, G. W. Y., Barreira, P. J., & McNally, R. J. (2018). Exercise and emotion dynamics: An experience sampling study. *Emotion*, *19*(4), 637–644. <https://doi.org/10.1037/emo0000462>
- Booij, S. H., Snippe, E., Jeronimus, B. F., Wichers, M., & Wigman, J. T. W. (2018). Affective reactivity to daily life stress: Relationship to positive psychotic and depressive symptoms in a general population sample. *Journal of Affective Disorders*, *225*, 474–481. <https://doi.org/10.1016/j.jad.2017.08.051>
- Carver, C. S. (2015). Control processes, priority management, and affective dynamics. *Emotion Review*, *7*, 301–307. <https://doi.org/10.1177/1754073915590616>
- Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, *175*, 526–536. <https://doi.org/10.1017/CBO9780511663246.006>
- Ebner-Priemer, U. W., & Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment*, *21*, 463–475. <https://doi.org/10.1037/a0017075>
- Frijda, N. H., & Mesquita, B. (1994). The social roles and functions of emotions. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 51–87). Washington, DC: American Psychological Association. <https://doi.org/10.1037/10152-000>
- Hair, Jr., J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Hollenstein, T. (2015). This time it's real: Affective flexibility, time scales, feedback loops, and the regulation of emotion. *Emotion Review*, *7*, 308–315. <https://doi.org/10.1177/1754073915590621>
- Houben, M., van den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*, 901–930. <https://doi.org/10.1037/a0038822>
- Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, *13*, 354–375. <https://doi.org/10.1037/a0014173>
- Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*, 865–878. <https://doi.org/10.1016/j.cpr.2010.03.001>
- Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, *13*, 1132–1141. <https://doi.org/10.1037/a0033579>
- Koval, P., Sütterlin, S., & Kuppens, P. (2016). Emotional inertia is associated with lower well-being when controlling for differences in emotional context. *Frontiers in Psychology*, *6*, 1997. <https://doi.org/10.3389/fpsyg.2015.01997>
- Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, *99*, 1042–1060. <https://doi.org/10.1037/a0020962>
- Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology*, *17*, 22–26. <https://doi.org/10.1016/j.copsyc.2017.06.004>

- Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, 7, 297–300. <https://doi.org/10.1177/1754073915590947>
- Larsen, R. J. (1987). The stability of mood variability: A spectral analytic approach to daily mood assessments. *Journal of Personality and Social Psychology*, 52, 1195–1204. <https://doi.org/10.1037/0022-3514.52.6.1195>
- Larsen, R. J. (2000). Toward a science of mood regulation. *Psychological Inquiry*, 11, 129–141. https://doi.org/10.1207/S15327965PLI1103_01
- Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelineing the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, 23(4), 690–707. <https://doi.org/10.1037/me0000153>
- Panksepp, J. (2012). What is an emotional feeling? Lessons about affective origins from cross-species neuroscience. *Motivation and Emotion*, 36(1), 4–15. <https://doi.org/10.1007/s11031-011-9232-y>
- Peeters, F., Nicolson, N. A., Berkhof, J., Delespaul, P., & de Vries, M. (2003). Effects of daily events on mood states in major depressive disorder. *Journal of Abnormal Psychology*, 112(2), 203–211. <https://doi.org/10.1037/0021-843X.112.2.203>
- Rottenberg, J. (2005). Mood and emotion in major depression. *Current Directions in Psychological Science*, 14, 167–170. <https://doi.org/10.1111/j.0963-7214.2005.00354.x>
- Rush, A., Bernstein, I., Trivedi, M., Carmody, T., Wisniewski, S., Mundt, J., . . . Fava, M. (2006). An evaluation of the quick inventory of depressive symptomatology and the Hamilton rating scale for depression: A sequenced treatment alternatives to relieve depression trial report. *Biological Psychiatry*, 59, 493–501. <https://doi.org/10.1016/j.biopsych.2005.08.022>
- Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The inventory of depressive symptomatology (IDS): Psychometric properties. *Psychological Medicine*, 26, 477–486. <https://doi.org/10.1017/S0033291700035558>
- Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 302–320). New York: Guilford Press.
- Stekhoven, D. J., & Bühlmann, P. (2012). MissForest – non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology*, 121, 819–829. <https://doi.org/10.1037/a0027978>
- Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective dynamics in psychopathology. *Emotion Review*, 7, 355–361. <https://doi.org/10.1177/1754073915590617>
- Trull, T. J., Solhan, M. B., Tragesser, S. L., Jangh, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology*, 117, 647–661. <https://doi.org/10.1037/a0012532>
- Van der Krieke, L., Blaauw, F. J., Emerencia, A. C., Schenk, H. M., Slaets, J. P. J., Bos, E. H., de Jonge, P., & Jeronimus, B. F. (2016). Temporal dynamics of health and well-being: A crowdsourcing approach to momentary assessments and automated generation of personalized feedback. *Psychosomatic Medicine*, 79(2), 213–223. <https://doi.org/10.1097/PSY.0000000000000378>
- Van der Krieke, L., Jeronimus, B. F., Blaauw, F. J., Wanders, R. B. K., Emerencia, A. C., Schenk, H. M., de Vos, S., Snippe, E., Wichers, M., Wigman, J. T. W., Bos, E. H., Wardenaar, K. J., & de Jonge, P. (2015). HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research*, 25, 123–144. <https://doi.org/10.1002/mpr.1495>
- Van Os, J., Verhagen, S., Marsman, A., Peeters, F., Bak, M., Marcelis, M., Drukker, M., Reininghaus, U., Jacobs, N., Lataster, T., Simons, C., Investigators, E. S. M.-M. E. R. G. E., Lousberg, R., Gülöksüz, S., Leue, C., Groot, P. C., Viechtbauer, W., & Delespaul, P. (2017). The experience sampling method

- as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *Depression & Anxiety*, 34, 481–493. <https://doi.org/10.1002/da.22647>
- Wagenmakers, E. J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, 114, 830–841. <https://doi.org/10.1037/0033-295X.114.3.830>
- Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology*, 97, 346–353. <https://doi.org/10.1037/0021-843X.97.3.346>
- Wichers, M., Peeters, F., Geschwind, N., Jacobs, N., Simons, C. J. P., Derom, C., Thiery, E., Delespaul, P. H., & van Os, J. (2010). Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: A momentary assessment study. *Journal of Affective Disorders*, 124(1–2), 191–195. <https://doi.org/10.1016/j.jad.2009.11.010>
- Wichers, M., Wigman, J. T. W., & Myin-Germeys, I. (2015). Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review*, 7, 362–367. <https://doi.org/10.1177/1754073915590623>
- Yik, M. S. M., Russell, J. A., & Barrett, L. F. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, 77, 600–619. <https://doi.org/10.1037/0022-3514.77.3.600>

Received 15 September 2018; revised version received 28 November 2018