Indicators of Affect Dynamics: Structure, Reliability, and Personality Correlates

LEON P. WENDT[1]*, AIDAN G.C. WRIGHT[2], PAUL A. PILKONIS[3], WILLIAM C. WOODS[2],
JAAP J.A. DENISSEN[4,5], ANJA KÜHNEL[6] and JOHANNES ZIMMERMANN[1]

[1]*Department of Psychology, University of Kassel, Kassel, Germany*
[2]*Department of Psychology, University of Pittsburgh, Pittsburgh, PA USA*
[3]*Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA USA*
[4]*Department of Developmental Psychology, Tilburg University, Tilburg, The Netherlands*
[5]*Department of Developmental Psychology, Utrecht University, Utrecht, The Netherlands*
[6]*Department of Psychology, MSB Medical School Berlin, Berlin, Germany*

Abstract: Researchers are increasingly interested in the affect dynamics of individuals for describing and explaining personality and psychopathology. Recently, the incremental validity of more complex indicators of affect dynamics (IADs; e.g. autoregression) has been called into question (Dejonckheere et al., 2019), with evidence accumulating that these might convey little unique information beyond mean level and general variability of emotions. Our study extends the evidence for the construct validity of IADs by investigating their redundancy and uniqueness, split-half reliability based on indices from odd-numbered and even-numbered days, and association with big five personality traits. We used three diverse samples that assessed daily and momentary emotions, including community participants, individuals with personality pathology, and their significant others (total N = 1192, total number of occasions = 51 278). Mean and variability of affects had high reliability and distinct nomological patterns to big five personality traits. In contrast, more complex IADs exhibited substantial redundancies with mean level and general variability of emotions. When partialing out these redundancies by using residual variables, some of the more complex IADs had acceptable reliability, but only a few of these showed incremental associations with big five personality traits, indicating that IADs have limited validity using the current assessment practices. © 2020 The Authors. European Journal of Personality published by John Wiley & Sons Ltd on behalf of European Association of Personality Psychology

Key words: affect dynamics; reliability; residual variables; structure; personality traits

INDICATORS OF AFFECT DYNAMICS: STRUCTURE, RELIABILITY, AND PERSONALITY CORRELATES

Major personality models include emotions as part of personality along with motivations, cognitions, and behavioral dispositions (Carver, Sutton, & Scheier, 2000). Researchers have growing interest in how emotions unfold and interact with each other dynamically across time (i.e. affect dynamics) and how these processes relate to diverse psychological phenomena (e.g. Trull, Lane, Koval, & Ebner-Priemer, 2015). For example, affect dynamics have been targeted for describing and explaining well-being (Dejonckheere *et al.*, 2019;

Houben, Van Den Noortgate, & Kuppens, 2015), mood disorders (e.g. Bos, de Jonge, & Cox, 2019), borderline personality pathology (e.g. Mneimne, Fleeson, Arnold, & Furr, 2018), transdiagnostic dimensions of psychopathology (Scott *et al.*, 2020), and normal-range personality differences (e.g. Kalokerinos *et al.*, 2020; Kuppens, Van Mechelen, Nezlek, Dossche, & Timmermans, 2007). Affect dynamics can be measured by various person-specific summary statistics of an emotional time series [e.g. mean of states, standard deviation (*SD*), autoregression, and mean square successive differences (*MSSD*)] using intensive longitudinal research designs (Csikszentmihalyi & Larson, 1987; Hamaker & Wichers, 2017). Hereinafter, we refer to those statistics as indicators of affect dynamics (IADs).

Conceptually, it is presumed that the various IADs each capture distinct and meaningful features of the individuals' emotional experience (in other words, that IADs possess construct validity). The validity of IADs is commonly evaluated by their associations to other relevant constructs (i.e. criterion validity). Critically, past research has demonstrated that manifold redundancies exist between IADs, including mathematical interdependencies (e.g. Jahng, Wood, & Trull, 2008; Mestdagh *et al.*, 2018) and possibly conceptual overlap. These redundancies have to be considered when criterion validity is evaluated, because associations found between IADs and other relevant

*Correspondence to: Leon Wendt, Department of Psychology, University of Kassel, Holländische Str. 36-38, 34127, Kassel, Germany.
E-mail: l.wendt@uni-kassel.de

constructs might be non-specific, thereby undercutting the unique interpretations that presumably justify the use of these metrics. For example, non-specific associations were documented for the criteria of depression (e.g. Koval, Pe, Meers, & Kuppens, 2013), well-being (e.g. Houben *et al*., 2015), and neuroticism (Kalokerinos *et al*., 2020). Indeed, the accumulating evidence casts doubt on the incremental validity of more complex IADs (e.g. autoregression) beyond mean level and general variability of emotions (Bos *et al*., 2019; Dejonckheere *et al*., 2019). To date, there is still insufficient knowledge about the reliability and validity of more complex IADs using the current assessment practices.

In the following, we describe the IADs considered in the current study along with their common interpretation. First, given a sufficient number of repeated measurements, the individual mean of states (*M*) is a stable feature of individuals (Watson & Clark, 1999) and can be regarded as a good approximation of *trait affect* (e.g. Watson & Tellegen, 1985). Second, the individual *SD* is generally referred to as *emotional variability*, with past research indicating that it might be a stable and substantive trait even when controlling for its overlap with the mean (Eid & Diener, 1999). Third, the individual *MSSD* captures the strength of sudden fluctuations in the process. High values of *MSSD* have been interpreted as *emotional instability* (Jahng *et al*., 2008). Fourth, the individual strength of autoregression, which is the likelihood of remaining in a particular affective state from observation to observation, has been interpreted as *emotional inertia* (Kuppens *et al*., 2012; Kuppens, Allen, & Sheeber, 2010). Fifth, several statistics capture types of *emotion differentiation*, defined as the degree to which individuals report distinct emotional states. This concept may be applied to differentiating between affects (e.g. individual contemporaneous correlation between positive affect (PA) and negative affect (NA): affective bipolarity; Dejonckheere *et al*., 2018) or differentiating between more fine-grained emotional states within affects (e.g. the individual average item intercorrelation of positively valenced emotions). Sixth, *cross-lagged effects* can be used to describe how distinct affects predict each other across time as operationalized by temporal networks from dynamic network models (Epskamp, Waldorp, Mõttus, & Borsboom, 2018). Trait affect and also (to a somewhat lesser extent) emotional variability have received most support for their validity. In addition, *M* and *SD* yield the most parsimonious description of an emotional time series when compared against more complex IADs, as the calculation of the former disregards the inherent temporal sequence of repeated measurement. In contrast, more complex IADs do consider the temporal sequence.

The current ambiguity about the validity of IADs impedes research progress on affect dynamics. Three samples were used for secondary analysis in order to shed more light on this issue in several ways: (i) elucidating potential patterns of redundancy by investigating the structure of IADs, (ii) estimating their reliability as a prerequisite for validity, and (iii) extending their nomological network to big five personality traits. Big five personality traits are especially important to consider for tests of criterion validity, as those provide an established framework for capturing major psychological differences between individuals. This study included heterogenous samples in order to achieve generalizability across populations (i.e. community participants, individuals with personality pathology and their significant others) and sampling frames (i.e. daily and momentary data on emotions). In order to establish the incremental information of IADs, we controlled statistically for redundancies with mean level and general variability of an emotional process by using residual variables. More specifically, the *SD* statistics had scale means (i.e. residual variable Type I) and more complex IADs had scale means and *SD*s partialed out (i.e. residual variable Type II). As a result, our residual variables captured the individuals' relative score in relation to what would be expected, given the individuals' mean (and variability) on affect scales.

METHOD

Participants

Participants who completed at least 20 consecutive measurement occasions[1] were selected for the current analyses, resulting in *N* total participants = 1192 and *t* total occasions = 51 278. Big five personality traits at baseline were assessed in all samples. A detailed overview of sample characteristics is given in Table 1.

The first sample was based on the Berlin Diary Study (Denissen & Kühnel, 2008). Participants received daily questionnaires containing retrospective measurement of affect over 30 days that were filled out before going to bed. The second sample (Wright *et al*., 2019) included individuals with a personality disorder diagnosis who completed daily retrospective assessments of affect over 100 days. The third sample (Wright *et al*., 2017) consisted of dyads who completed a 21-day period of ecological momentary assessment (3.7 assessments per day on average) including individuals who were engaged in outpatient psychiatric treatment and their significant others. In Sample 3, momentary affect was assessed multiple times a day following social interactions (i.e. event-contingent assessment).

Measures

Daily and momentary emotions
Emotion adjectives were used to assess daily and momentary affect. In Sample 1, participants were asked to indicate to what degree emotions were descriptive of how they *generally felt today* on a 5-point scale, ranging from 0 (*not at all*) to 4 (*extreme*). In Sample 2, participants were asked about the extent to which they had *felt this way over the past 24 hours* on a 5-point scale, ranging from 0 (*very slightly*) to 4 (*extremely*). In Sample 3, participants were asked multiple times a day to rate their momentary emotions on a 5-point scale ranging from 1 (*very slightly or not at all*) to 5 (*extremely*). IADs were calculated for scales of PA, NA, and hostile affect (HA), as those were consistently identified

[1] The inclusion criterion of 20 measurement occasions is a common, yet arbitrary, threshold.

Table 1. Sample characteristics

Sample #	1	2	3
Further described in	Denissen and Kühnel, 2008	Wright et al., 2019; Wright, Beltz, Gates, Molenaar, & Simms, 2015	Wright et al., 2017
Emotion assessment type	Daily Dairy	Daily Dairy	Momentary assessment
Language	German	English (USA)	English (USA)
Population characteristics	~50% students	Individuals diagnosed with any personality disorder	Outpatients screened for personality pathology and their romantic partners
% of participants that satisfied the inclusion criterion (>20 occasions)	41.7%	89.3%	88.1%
N subjects included	870	100	222
M age	30.6	45.0	29.7
% female	87%	65%	77%
Minimum–maximum days	21–29	43–101	7–33
Total t measurement occasions	21709 days	9017 days	20552 beeps
Average t days	25.0	90.2	22.7
Average t beeps per day	/	/	4.0
M measurement occasions per split half for calculating split-half reliability	/	45.1 days	46.2 beeps
N subjects included for calculating split-half reliability	/	100	143
Number of emotion adjectives	26	10	25
Positive affect items	Active Attentive Determined Inspired Enthusiastic Excited Interested Proud Strong Content Pleased Happy Aroused Hyperactivated	Active Attentive Determined Inspired Alert	Active Attentive Determined Inspired Enthusiastic Excited Interested Proud Strong Alert
Negative affect items	Afraid Ashamed Distressed Guilty Scared Alone Miserable Troubled Unhappy	Afraid Ashamed Nervous	Afraid Ashamed Distressed Guilty Scared Alone Nervous Jittery
Hostile affect items	Hostile Upset Irritable	Hostile Upset	Hostile Upset Irritable Loathing Disgusted Angry Scornful
Assessment of personality traits	Big Five Inventory (Lang, Lüdtke & Asendorpf, 2001)	NEO Five-Factor Inventory (Costa & McCrae, 1992)	Revised NEO Personality Inventory (Costa & McCrae, 1992)
N subjects with assessment of personality traits available	870	99	193

across samples (see Results section). The included emotion adjectives are enlisted in Table 1.

Personality traits

In all samples, we assessed the big five personality traits (i.e. openness to new experiences, conscientiousness, extraversion, agreeableness, and neuroticism). In Sample 1, the German version of the Big Five Inventory (Lang et al., 2001) was used. Participants rate 42 statements on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). In Sample 2, the 60-item NEO Five-Factor Inventory (Costa & McCrae, 1992) was used. In Sample 3, the Revised NEO Personality Inventory (Costa & McCrae, 1992), consisting of 240 items, was used. For the

latter two NEO instruments, items were rated on a 5-point scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*).

Statistical analysis

Selection and computation of IADs

Affect scales were included for calculating IADs when they aligned with within-person factors identified by multilevel exploratory factor analysis (ML-EFA; Reise, Ventura, Nuechterlein, & Kim, 2005). The ideal number of within-person factors was selected with respect to interpretability and a combination of model fit indices, including the comparative fit index (CFI), root mean square error of approximation (RMSEA), and the models' improvements in level-specific fit using standardized root mean square residual (SRMR-within and SRMR-between; Kim, Dedrick, Cao, & Ferron, 2016).

The included IADs comprised univariate statistics (Jahng *et al.*, 2008) and model-based statistics (Epskamp *et al.*, 2018). The calculation of univariate IADs was based on rescaled affect scores with values ranging from 0 to 10 for facilitating cross-sample comparisons. Univariate IADs included individual scale mean (M), SD, $MSSD$, and the average item intercorrelation ($\bar{r}$). We further included corrected indices of emotional variability (i.e. SD^c) and emotional instability (i.e. $MSSD^c$), that have their theoretical maxima adjusted for the observed individual mean (Mestdagh *et al.*, 2018). Model-based IADs were derived from multilevel lag-1 vector autoregressive dynamic networks (Bringmann *et al.*, 2013, 2016; Epskamp *et al.*, 2018). Individual network parameters (also, random effects) including autoregressive effects (i.e. ϕPP, ϕNN, ϕHH), cross-lagged effects (e.g. ϕNP, ϕPN), and contemporaneous partial correlations (i.e. rPN, rPH, rNH) were estimated using person-mean centering of z-standardized daily/momentary affect scores.[2] Non-subsequent measurement occasions were removed from network estimation including overnight lags in case of momentary assessment in Sample 3. Network summary statistics (i.e. node centralities, network density) were not considered, because those were unlikely to be useful for small networks. We evaluated the level and variability of (standardized) individual random effects based on the guidelines of Gignac and Szodorai (2016). The median (Mdn) of individual random effects was regarded as small (~.10), moderate (~.20), or large (~.30), indicating the size of model-based IADs for the average individual. Similarly, the interquartile range (IQR) of individual random effects was regarded as small (~.10), moderate (~.20), or large (~.30), indicating the amount of individual differences in model-based IADs.

Structure of indicators of affect dynamics

Several analytic steps were taken to delineate the structure of the 30 calculated IADs and elucidate their potential redundancies. First, we used parallel analysis and visual inspection of the scree plot (i.e. elbow criterion).[3] Second, the Spearman correlation matrix of IADs was used to extract varimax rotated principal components.[4] Third, we investigated IADs' particular overlap with affect Ms and SDs using the multiple correlation coefficient (R). For emotional variability statistics (i.e. SD, SD^c), we computed their multiple correlation with means, and for more complex statistics (i.e. $MSSD$, $MSSD^c$, $\bar{r}$, individual network parameters), we computed their multiple correlation with scale means and SDs.

Split-half reliability

Split-half reliability was used to evaluate whether IADs consistently measure the same constructs at the between-person level (e.g. Mejía, Hooker, Ram, Pham, & Metoyer, 2014). To this end, residualized IADs were calculated separately based on time series including only odd or even days, and correlations between split-halves were computed. The correlations between split-halves were then corrected using the Spearman–Brown prophecy formula in order to derive reliability estimates of the IADs based on the complete time series (r_{sb}). In Sample 2, split-half intervals consisted of 45.1 days on average (odd days = 44.5, even days = 45.7). In Sample 3, split-half intervals consisted of 11.4 days and 46.2 momentary occasions on average (odd days = 44.1, even days = 48.4). Split-half reliability was not calculated in Sample 1, as fewer measurement occasions per person were available. Reliability was regarded as low (~.50), moderate (~.70), or high (~.90).

Associations with personality traits

The incremental validity of IADs was evaluated by their bivariate correlations with self-report measures of personality traits using residual variables. Pearson correlations were calculated and Fisher z-transformed in each sample separately, before integrating them in a random effects meta-analysis. In the current study, significant meta-analytic correlations indicate that IADs are incrementally associated with big five personality traits (i.e. across populations and sampling frames) beyond mean level and general variability of affect.

Software packages

Openly accessible R scripts and data are provided that allow reproduction of the reported statistical analyses at https://osf.io/6ghcx/. All analyses were executed with the statistical environment R version 3.6.1 (R Core Team, 2019). ML-EFA was estimated using the WLSMV estimator and oblique geomin rotation in Mplus 8.0 (Muthén & Muthén, 2018). SD^c and $MSSD^c$ were calculated using the *relativeVariability* package version 1.0 (Mestdagh *et al.*, 2018). Principal components, scree plots, and parallel analysis were computed by the *psych* package version 1.8.18 (Revelle, 2018). Multilevel dynamic networks were estimated by the *mlVAR* package version 0.4.3 (Epskamp, Deserno, & Bringmann, 2019).

---

[2]The full results on network model parameters (i.e. fixed effects, random effect variances) will be made available by the corresponding author upon request.

[3]Exploratory factor analysis was considered; however, solutions had bad fit or did not converge.
[4]Quartimin rotation was considered and produced similar results.

Table 2. Median, interquartile range, and skew of raw IADs

	Sample 1			Sample 2			Sample 3		
Statistic	Mdn	IQR	Skew	Mdn	IQR	Skew	Mdn	IQR	Skew
PA M	4.49	1.59	−0.23	4.05	2.55	0.26	2.39	2.05	0.76
NA M	1.68	1.98	0.97	1.50	2.39	1.38	0.26	0.54	2.52
HA M	1.77	1.73	0.86	1.65	1.66	1.68	0.35	0.61	2.78
PA SD	1.27	0.57	0.47	1.44	0.70	0.44	1.30	0.64	0.87
NA SD	1.33	0.82	0.20	1.28	0.82	0.17	0.49	0.59	1.20
HA SD	1.77	0.88	0.06	1.64	0.99	0.22	0.74	0.71	0.91
PA SD^c	0.27	0.12	0.42	0.33	0.15	0.71	0.33	0.12	0.57
NA SD^c	0.37	0.15	0.39	0.40	0.13	0.49	0.32	0.14	1.10
HA SD^c	0.49	0.17	0.26	0.51	0.18	0.23	0.42	0.16	0.66
PA $MSSD$	2.36	2.26	1.71	2.99	2.94	1.93	2.32	2.34	1.93
NA $MSSD$	2.45	3.15	1.60	2.28	3.02	1.29	0.38	0.68	2.39
HA $MSSD$	4.85	5.17	1.24	4.06	4.59	1.39	0.86	1.49	2.60
PA $MSSD^c$	0.03	0.03	1.81	0.05	0.05	1.84	0.06	0.05	2.26
NA $MSSD^c$	0.08	0.07	2.42	0.10	0.08	2.51	0.07	0.06	2.98
HA $MSSD^c$	0.15	0.13	2.41	0.17	0.11	0.76	0.12	0.10	1.92
PA $\bar{r}$	.41	.14	0.57	.68	.36	0.15	.50	.39	0.70
NA $\bar{r}$	.47	.16	0.59	.66	.46	0.06	.41	.31	0.82
HA $\bar{r}$	.65	.22	−0.01	1.00	.30	−0.72	.64	.48	0.17
ϕPP	.17	.08	0.19	.26	.17	0.96	.31	.24	0.57
ϕNN	.22	.04	0.59	.26	.10	0.70	.32	.27	1.17
ϕHH	.12	.08	0.53	.13	.09	0.73	.18	.13	1.03
ϕPN	.00	.04	−0.01	−.01	.04	−0.24	−.01	.03	−0.86
ϕPH	.01	.05	0.19	−.01	.05	0.63	−.01	.02	−0.72
ϕNP	−.01	.08	0.59	−.01	.03	0.47	−.02	.07	−0.79
ϕNH	.02	.01	−0.22	.12	.05	−0.58	.14	.08	1.27
ϕHP	.01	.06	0.19	.01	.03	−0.11	.00	.03	0.13
ϕHN	.01	.06	0.10	.02	.04	−0.47	.01	.04	0.76
rPN	−.43	.18	0.14	.04	.14	0.15	−.02	.12	0.47
rPH	−.04	.06	−0.11	.00	.13	0.63	−.05	.08	−0.35
rNH	.41	.20	−0.36	.47	.24	−0.13	.62	.27	−0.15

Note: IADs are based on rescaled values of affect scales in the range of 0 and 10. Mdn, median; IQR, interquartile range; M, mean; SD, standard deviation; $MSSD$, mean square successive differences; SD^c, corrected standard deviation; $MSSD^c$, corrected mean square successive differences; $\bar{r}$, average item intercorrelation of affect scales; ϕ, autoregressions and cross-lagged effects; r, contemporaneous partial correlations.

Random effects meta-analysis was estimated by the *metafor* package version 2.1−0 (Viechtbauer, 2010).

RESULTS

Selection of affect dimensions

Considering both fit and interpretability, ML-EFA solutions were retained that indicated five (Sample 1) or three within-person factors (Samples 2 and 3) achieving acceptable fit, RMSEA $\leq$ .044, CFI $\geq$ .929, SRMR-within $\leq$ .039. PA, NA, and HA were consistently identified in all samples and were therefore used for calculating IADs. Additional factors were identified (i.e. factors of tiredness and calmness) but were not considered for calculating IADs, as those were only present in Sample 1. Further information on ML-EFA models are displayed in the supporting information (see Table S1 for fit statistics and Tables S2–S4 for the estimated within-person factor loadings).

Descriptive statistics

Median, IQR, and skew of raw IADs are displayed in Table 2. Individual mean of affect was high for PA as compared to

NA and HA, indicating that individuals tended to report positive emotions more often than negative emotions. Greater positive skew was observed in NA and HA distributions. *MSSD* statistics had positive skew across the included affect scales (>1.24). Median of the average item intercorrelation of affect scales (i.e. $\bar{r}$) ranged from .41 to .68 (except for HA $\bar{r}$ that had median of 1.00 in Sample 2), indicating that the respective indicators of affect scales were substantially intercorrelated for the average individual (and perfectly intercorrelated for the average individual in terms of HA in Sample 2).[5]

Median of individual autoregressive parameters was ranging from .12 (small) to .32 (large), indicating that affects carried over to the next day/moment for the average individual. IQR of individual autoregressive parameters was small to moderate in daily data (IQR = .04−.17) and small to large in momentary data (IQR = .13−.27). Median of individual cross-lagged parameters was small (Mdn = −.02−.02), except for ϕNH, for which small-to-moderate median was observed in Sample 2 (Mdn = .12) and Sample 3 (Mdn = .14). Individual cross-lagged effects had small IQR, IQR = .01−.08. Median of individual contemporaneous

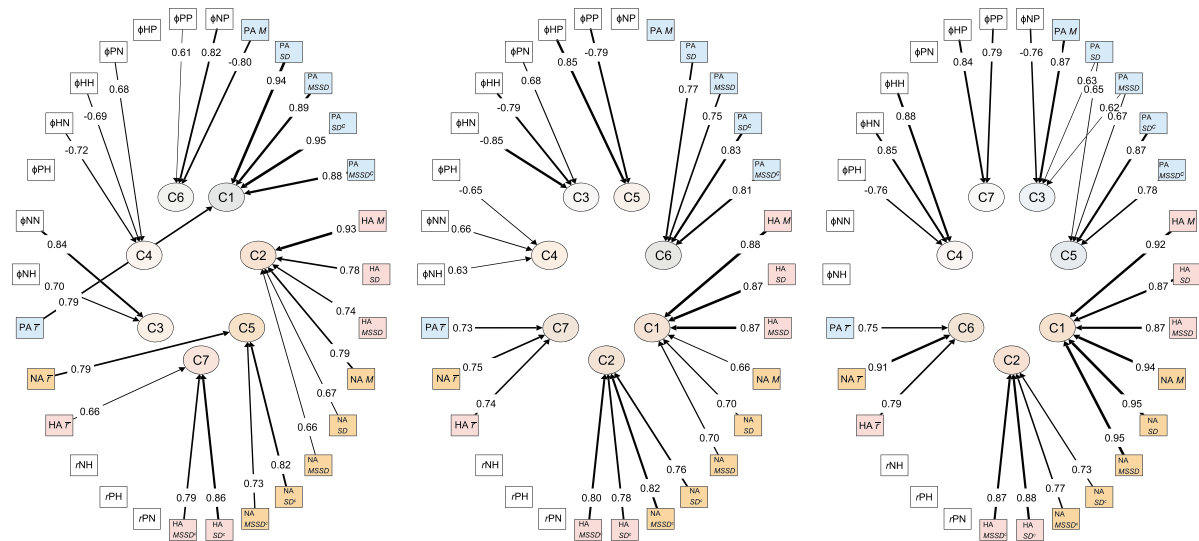[5]Note that the HA scale comprised only two emotion adjectives (i.e. hostile, upset) in Sample 2.

FIGURE 1.   Varimax rotated components of IADs across samples. Component loadings < .60 are suppressed. PA, positive affect; NA, negative affect; HA, hostile affect; *M*, mean; *SD*, standard deviation; *MSSD*, mean square successive differences; *SD*^{c}, corrected standard deviation; *MSSD*^{c}, corrected mean square successive differences; $\bar{r}$, average item intercorrelation of affect scales; φ, autoregressions and cross-lagged effects; *r*, contemporaneous partial correlations. [Colour figure can be viewed at wileyonlinelibrary.com]

partial correlation between NA and HA was large (Mdn = .41–.62), and IQR was moderate to large, IQR = .20–.27. Median of individual contemporaneous partial correlation between PA and NA was negative in Sample 1 (Mdn = −.43, IQR = .18) and close to zero in Sample 2 (Mdn = .04, IQR = .14) and Sample 3 (Mdn = −.02, IQR = .12). Median of individual contemporaneous partial correlation between PA and HA was close to zero (Mdn = −.05–.00) and IQR was small, IQR = .06–.13. Similar distributions of individual random effects were obtained for odd and even days when compared against the networks that were calculated using the complete data (e.g. *r*NH in Sample 2, complete data: .47, odd days: .47, even days: .45; φHH, complete data: .13, odd days: .13, even days: .17).

Structure

Parallel analysis indicated seven components in Sample 2 and nine components in Samples 1 and 3. Visual inspection of scree plots was inconclusive because there were no clear drops in eigenvalues. We base our interpretation of the structure of IADs on the seven-component resolution as it represented the greatest common denominator across samples. Figure 1 provides a sparse graphical display of the varimax rotated principal components. Commonalities were found between univariate IADs of PA (Sample 1, C1; Sample 2, C6; Sample 3, C3, C5), indicating that those tended to be interrelated. Univariate IADs of NA and HA were reflected in several principal components (Sample 1, C2, C5, C7; Sample 2, C1, C2; Sample 3, C1, C2). Temporal effects including autoregressive and cross-lagged parameters formed additional principal components (Sample 1, C3, C4, C6; Sample 2, C3, C4, C5; Sample 3, C4, C7). The average item intercorrelation of affect scales formed principal components in Sample 2 (C7) and Sample 3 (C6).

Figure 2a–c shows variance decompositions of IADs by sample, indicating the particular degree to which variation in IADs was accounted for by scale means and *SD*s.[6] Generally, scale means and *SD*s shared plenty of common variance with the more complex IADs, including autoregressions, cross-lagged effects, and contemporaneous partial correlations, thus, highlighting the need to consider these redundancies for extracting their more unique information.

Split-half reliability

Figure 2b–c displays the estimated variance of IADs that was unique and reliable in Samples 2 and 3 as fractions of the total variance observed. These estimates align with the split-half reliability of residualized IADs (see Figures S1 and S2). Split-half reliability of raw IADs is reported in Table S5. Individual *M* of affect scales was highly reliable, r_{sb} = .94–.99. Residualized *SD* and *SD*^{c} had moderate-to-high reliability (*SD*, r_{sb} = .77–.96; *SD*^{c}, r_{sb} = .67–.95), indicating that those tend to reliably capture incremental features of emotional time series beyond mean of affect. Partialing out scale means and *SD*s tended to reduce the split-half reliability of more complex IADs, indicating that their reliability estimates were inflated because of redundancies with *M* and *SD*. For example, the split-half reliability of PA *MSSD* decreased from .92 to .69 in Sample 2 and the split-half reliability of φNP decreased from .75 to .06 in Sample 3. Nevertheless, some residualized IADs achieved moderate-

---

[6]The total variance was decomposed into the part of variance explained by scale means (i.e. the squared multiple correlation with scale means), the part of variance explained by scale standard deviations beyond what had already been explained by scale means (i.e. the squared multiple correlation with scale means and scale standard deviations minus the squared multiple correlation with scale means), the unique variance that was reliable (i.e. split-half reliability of the residualized variable), and the unique variance that was not reliable (one minus the sum of the aforementioned variance parts).
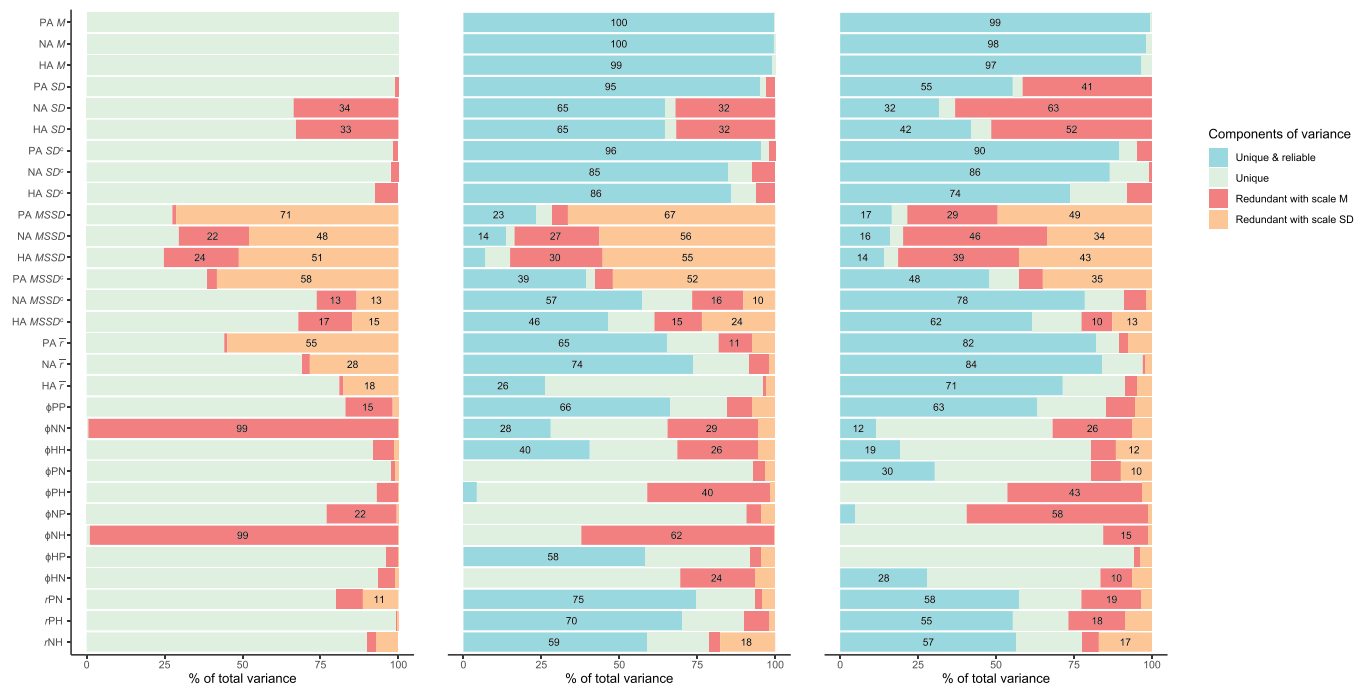
Figure 2. (a–c) Variance decomposition of IADs. From left to right Samples 1–3 are depicted. PA, positive affect; NA, negative affect; HA, hostile affect; *M*, mean; *SD*, standard deviation; *MSSD*, mean square successive differences; *SD*^{c}, corrected standard deviation; *MSSD*^{c}, corrected mean square successive differences; $\bar{r}$, average item intercorrelation of affect scales; ϕ, autoregressions and cross-lagged effects; *r*, contemporaneous partial correlations.

to-high reliability (e.g. *MSSD* and *MSSD*^{c}, $r_{sb} = .61–.88$). The reliability of residualized HA *MSSD* was low in Sample 2, $r_{sb} = .31$. The residualized average item intercorrelations of affect scales had moderate-to-high reliability, $r_{sb} = .64–.85$, except for HA $\bar{r}$ in Sample 2, $r_{sb} = .16$. The reliability of residualized autoregression was moderate in case of ϕPP ($r_{sb} = .59–.65$) and low in case of ϕHH (Sample 2, $r_{sb} = .42$; Sample 3, $r_{sb} = .14$). The residualized autoregression of ϕNN had reliability that was close to zero, $r_{sb} = .09–.27$. Residualized cross-lagged effects also tended towards reliability estimates that were close to zero, $r_{sb} = .00–.13$, except for ϕHP in Sample 2, $r_{sb} = .47$. The reliability of residualized contemporaneous partial correlations was moderate, $r_{sb} = .57–.66$.

Associations with personality traits

Figure 3 displays meta-analytic estimates for the bivariate associations between residualized IADs and personality traits. Tables S6–S7 display sample-specific results for raw and residualized IADs. More desirable mean level of affect (i.e. high PA, low NA, and HA) were associated with extraversion (PA *M*, $r = .32$; NA *M*, $r = −.24$; HA *M*, $r = −.14$) conscientiousness (PA *M*, $r = .23$; NA *M*, $r = −.20$; HA *M*, $r = −.20$), agreeableness (PA *M*, $r = .11$; NA *M*, $r = −.16$; HA *M*, $r = −.26$), and openness (PA *M*, $r = .16$). Less desirable mean level of affect were associated with neuroticism (PA *M*, $r = −.26$; NA *M*, $r = .39$; HA *M*, $r = .30$). When controlling for mean level of affects, greater PA variability and

NA variability were incrementally associated with openness (PA *SD*, $r = .15$; PA *SD*^{c}, $r = .13$; NA *SD*, $r = .06$; NA *SD*^{c}, $r = .06$) and extraversion (PA *SD*, $r = .15$; PA *SD*^{c}, $r = .14$; NA *SD*, $r = .08$). Lower PA variability was incrementally associated with agreeableness (PA *SD*^{c}, $r = −.07$). The residualized statistics of NA and HA variability (*SD*), but not *SD*^{c}, were incrementally associated with neuroticism (NA *SD*, $r = .17$; HA *SD*, $r = .15$).

The bivariate correlations between more complex IADs and personality traits were of smaller size when residual variables were used, indicating that non-specific associations were induced by redundancies with *M* and *SD*. For example, the correlation between NA *SD* and neuroticism decreased from .22 to −.04 (Sample 1), and the correlation between ϕNN and neuroticism decreased from .16 to −.02 (Sample 1). Notwithstanding, some incremental associations between more complex IADs and personality traits were found that reached statistical significance ($p < .05$). After controlling for scale means and *SD*s, lower NA instability and HA instability were incrementally associated with neuroticism (NA *MSSD*^{c}, $r = −.10$; HA *MSSD*^{c}, $r = −.09$). A less differentiated reporting of hostile states was incrementally associated with agreeableness (HA $\bar{r}$, $r = .09$) and a more differentiated reporting of negative emotional states was incrementally associated with neuroticism (NA $\bar{r}$, $r = −.07$). The contemporaneous partial correlation between NA and HA was incrementally associated with agreeableness (rNH, $r = .08$), indicating that highly agreeable individuals exhibit a greater than
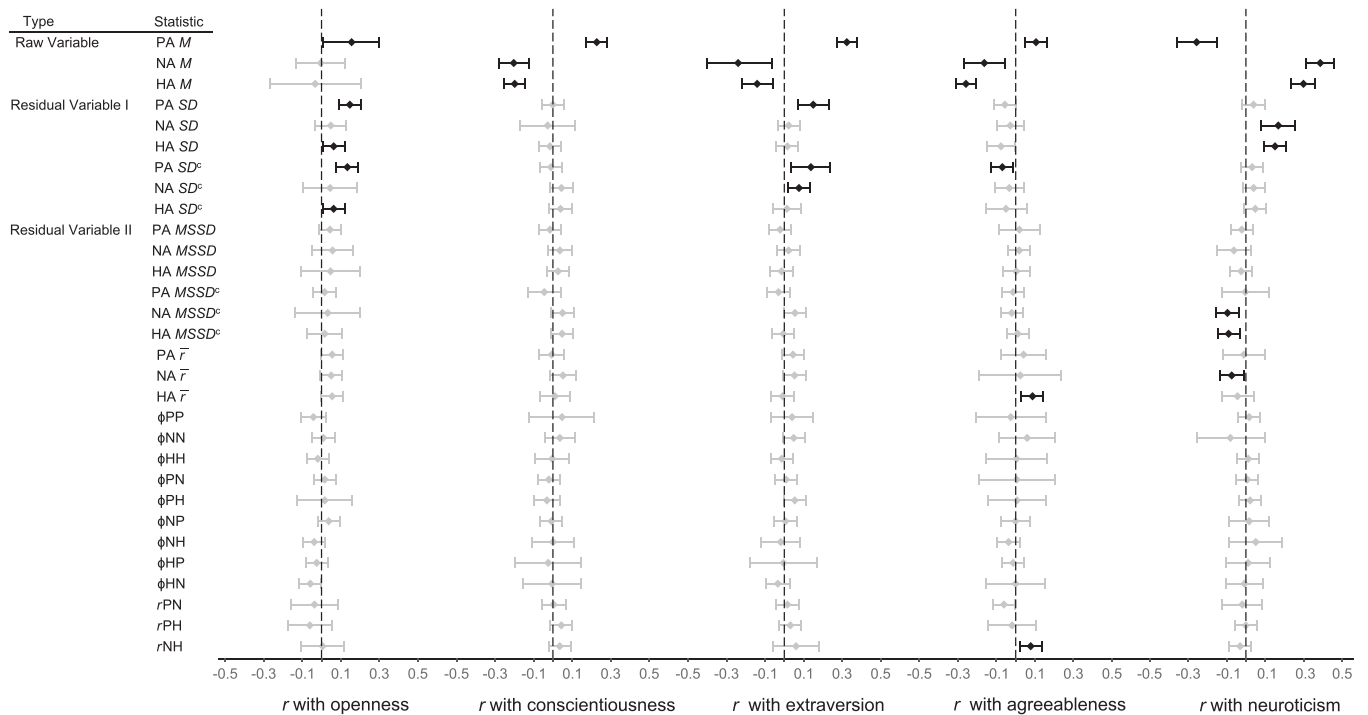
FIGURE 3.   Meta-analytic estimates of the bivariate correlations between residualized IADs and big five traits. Residual variable type I were residualized for scale means. Residual variables type II were residualized for scale means and standard deviations. r, meta-analytic estimate of the bivariate correlation; PA, positive affect; NA, negative affect; HA, hostile affect; M, mean; SD, standard deviation; $MSSD$, mean square successive differences; SD^c, corrected standard deviation; $MSSD^c$, corrected mean square successive differences; $\bar{r}$, average item intercorrelation of affect scales; ϕ, autoregressions and cross-lagged effects; r, contemporaneous partial correlations.

average tendency to experience negative and hostile states in concordance.[7] Temporal parameters (i.e. autoregressive and cross-lagged effects) had no significant incremental associations with personality traits.

DISCUSSION

Failed tests of incremental validity of more complex IADs raised doubt about their validity and usefulness for studying affect dynamics (Dejonckheere *et al.*, 2019). The current study extends the body of evidence by providing a comprehensive analysis of their structure, split-half reliability, and association with personality traits. In line with earlier results, more complex IADs exhibited substantial redundancies with mean level and general variability of emotions. When these redundancies were controlled statistically, the reliability and validity coefficients of more complex IADs shrunk, and in some cases, they became small or non-significant, indicating that many of the more complex IADs convey limited incremental information on affect dynamics using the current assessment practices.

Structure of indicators of affect dynamics

The covariance structures of IADs elucidated similar patterns of redundancy in the included samples, showing that many of the more complex IADs exhibit marked redundancies with mean level and general variability of emotional time series.

[7]Remember that the contemporaneous partial correlation rNH of the average individual was large, .41–.61.

Our analyses indicate that additional redundancies may exist beyond of that, as was exemplified by principal components that summarized additional common variance between IADs related to emotion differentiation (i.e. $\bar{r}$) and temporal effects (e.g. ϕPP).

We discuss possible reasons for the observed redundancies. The statistical overlap between individual M and SD of affect scales was of higher magnitude when the underlying emotion distributions exhibited greater skew, which was the case for negative emotions, hostile emotions, and when affects were measured using momentary assessment. This points to the fact that M and SD are mathematically interdependent in skewed distributions. The redundancies found between $MSSD$ and SD indicate that their raw forms capture similar constructs at the between-person level, that is, general variability. This is not surprising considering that both SD and $MSSD$ may serve as global indices of dispersion (Jahng *et al.*, 2008). However, the residualized form of $MSSD$ and $MSSD^c$ that were used in the current study should have a different interpretation, because their statistical redundancies with M and SD were partialed out (i.e. 10–85% of variance). Thus, the residualized $MSSD$ should mainly reflect the temporal dependency of consecutive measurement occasions, similar to the autoregressive parameter. Thereby, the residualized $MSSD^c$ may delineate a continuum that ranges from emotional instability (i.e. high values) to emotional inertia (i.e. low values).

We have no explanation for the divergent pattern of redundancy with respect to the average item intercorrelation of affect scales (i.e. PA $\bar{r}$, NA $\bar{r}$, HA $\bar{r}$). On the one hand, in Sample 1, greater average item intercorrelations were associated with

greater emotional variability on the respective scales. This might suggest that individuals that respond homogenously to items of one affect scale (i.e. internal consistency) could have an increased probability of producing true score variability. On the other hand, in Samples 2 and 3, the average item inter-correlation of affect scales formed a component of their own, indicating a general form of emotion differentiation.

Network parameters (i.e. autoregressions, cross-lagged effects, and contemporaneous partial correlations) exhibited redundancies primarily with M, some of those were extremely large (i.e. ϕNN and ϕNH in Sample 1). This may be unintuitive, as the networks were estimated using within-person centered variables. However, within-person centering does not change the variance and shape of individual state affect distributions, and, as noted above, many of the affect scales showed little variance (i.e. small IQR) and were skewed, especially the negative items. Thus, the well-known associations between the mean and variability and the role of variance restriction serve as the likely explanation for the observed redundancies between network random effects and individual M (including redundancies of ϕNN and ϕNH in Sample 1). Indeed, this has been an issue that has plagued the psychological network literature, which has often found that many of the most conceptually interesting statistics are highly dependent on observed variances in real world data (Rodebaugh *et al.*, 2018).

Split-half reliability of indicators of affect dynamics

In line with prior research, mean and variability of affects had high reliability indicating that those consistently measure the same constructs at the between-person level. Some of the more complex IADs were moderately reliable (e.g. $MSSD^c$, $\bar{r}$, ϕPP, contemporaneous partial correlations) after controlling for their overlap with means and SDs. Those IADs may reflect true and unique individual variation; however, they also include substantial measurement error. The extent to which IADs are unreliable puts a ceiling on the strength of their associations with other constructs that can possibly be observed. In consequence, such associations will be attenuated and will require larger sample sizes for detecting signals.

Temporal effects (i.e. autoregressions and cross-lagged effects) tended towards having very poor or no reliability (except for ϕPP). This suggests that those are not valid IADs for between-person research using the current assessment practices. Considering that most temporal effects yielded little random effect variances, it seems quite plausible that, in some cases, sampling variation and statistical redundancies may occasionally account for the total of their variance. One reason for the unreliability of some IADs could be that the indicators used here (i.e. the selected pool of emotions items) are not optimal or that measurement occasions were too few. Another reason could be that some IADs reflect more volatile psychological conditions that fluctuate rapidly (e.g. because of contextual factors; e.g. Koval & Kuppens, 2012). This would suggest that their assessment should be based on different assessment frames (e.g. more frequent assessments within a smaller time frame) or that such

IADs should be better studied under more controlled conditions (e.g. experimental designs; Dejonckheere, Mestdagh, Kuppens, & Tuerlinckx, 2020).

Associations between indicators of affect dynamics and big five personality traits

Our findings indicate that big five personality traits are characterized by distinct pattern of affect dynamics that primarily encompass individual differences in Ms and SDs of PA, NA, and HA. Mean affect had correlations with personality traits that were in line with prior research (e.g. Ching *et al.*, 2014; Howell, Ksendzova, Nestingen, Yerahian, & Iyer, 2017; Watson & Clark, 1999), indicating that more adaptive configurations of personality traits (i.e. emotional stability, openness, conscientiousness, agreeableness, and extraversion) are robustly related to more desirable emotional experiences (i.e. high PA, low NA, low HA).

Positive affect variability was positively associated with extraversion and openness. Given that greater variability of positive emotions was a substantive characteristic of extraverted and open individuals, one could speculate that it reflected the exploratory nature of these traits, in other words, a greater tendency to seek potentially rewarding situations (i.e. greater sensitivity to rewards; DeYoung, 2015). This 'high risk high reward' strategy could result in greater variability in the achievement of rewards, and thus, in greater variability of experiencing positive emotions. In contrast, individuals high in agreeableness had less variability but higher mean level of PA. This is notable, because it indicates that personality traits are differentially associated with mean affect and variability, and thus, it provides evidence for their divergent nomological patterns. A competing account of variability measures argues that those might reflect extreme responding (Baird, Lucas, & Donnellan, 2017), which is the tendency to choose more extreme response categories in self-report questionnaires. However, extreme responding has been mainly associated with high extraversion and high conscientiousness (Austin, Deary, & Egan, 2006), indicating that the current results might not be sufficiently explained by this type of method bias. In the current study, neuroticism did not show consistent incremental associations with greater negative emotional variability, because although SD reached statistical significance, SD^c did not. This corroborates findings from a recent meta-analysis (Kalokerinos *et al.*, 2020). On the one hand, the relative indices used here (i.e. SD^c, $MSSD^c$) may be superior for deriving the more unique information about emotional variability and emotional instability in the presence of non-linear dependencies (Mestdagh *et al.*, 2018). On the other hand, it may be, though, that they overcorrect in skewed distributions thereby reducing their validity, and other methods for accounting for the association between mean and variability should be considered in future work.

Some incremental associations were found between more complex IADs and personality traits. Lower instability of negative and hostile emotional states (i.e. NA $MSSD^c$, HA $MSSD^c$) was observed in individuals high in neuroticism. With respect to our interpretation of the residualized $MSSD^c$

that was discussed earlier, this finding suggests that individuals high in neuroticism might be more resistant to change in negative emotional states; thus, adding to the body of conflicting results on how fluctuations in negative emotions are linked to neuroticism-related constructs (also known as the 'instability-inertia paradox', e.g. Bos *et al*., 2019; Bosley, Soyster, & Fisher, 2019; Koval, Kuppens, Allen, & Sheeber, 2012; Koval *et al*., 2013). Furthermore, a more differentiated reporting of negative emotions (i.e. NA $\bar{r}$) was observed in neurotic individuals, and a more differentiated reporting of hostile states (i.e. HA $\bar{r}$) was observed in disagreeable individuals. One hypothesis could be that emotion differentiation is related to attentional processes, such that, individuals high in neuroticism pay greater attention to their negative emotions and individuals high in disagreeableness pay greater attention to their hostile emotions, and thus, they might experience and report specific affects more nuanced. Finally, a greater association between negative and hostile states (i.e. rNH) was observed in agreeable individuals. One explanation could be that agreeable individuals have a larger aversion of interpersonal conflicts, such that hostility caused more negativity (Suls, Martin, & David, 1998). However, we want to emphasize that the associations between more complex IADs and personality traits were small in magnitude, and as we did not adjust for multiple testing, any of the suggested interpretations need to be taken cautiously.

Limitations

The current study has some important limitations with respect to the samples, the measurement of emotions, the intensive longitudinal designs, the methods for assessing personality traits, and the statistical analyses. First, our study might have compared 'apples and oranges' by drawing inferences across samples that were diverse in terms of sampling frames (i.e. daily and momentary emotion data), instruments used, and populations investigated. We regard this limitation a strength, as the heterogeneity across samples contributes to greater generalizability (Yarkoni, 2019). Second, the pool of emotions used might not completely span the affective space. Third, with regard to the involved intensive longitudinal designs, Sample 1 might contain too few measurement occasions for deriving a reliable assessment of network parameters (i.e. 20–30 consecutive measurements per person in Sample 1). Moreover, measurement burst designs (Stawski, MacDonald, & Sliwinski, 2015) can inform researchers about the trait status of IADs, by investigating their stability over more widely spaced temporal intervals, because stability is usually demonstrated over longer periods of time (i.e. 1–2 years) than were used in the current study. Also, IADs might exhibit stronger incremental validity when studied under more controlled contextualized conditions (e.g. experimental designs or event-contingent assessment) because of potentially preferable signal-to-noise ratio (Dejonckheere *et al*., 2020; Lapate & Heller, 2020). Fourth, we only had access to concurrent self-reported personality data. Some of the dynamic indices could be stronger related to personality facets or nuances, future personality or

personality change, informant reports (Finnigan & Vazire, 2018), or digital footprints of personality (Hinds & Joinson, 2019).

Fifth, our study did not model measurement error. Generally, unmodelled measurement error might have attenuated the reliability and validity estimates of IADs (Rouder & Haaf, 2019; Schuurman & Hamaker, 2019). Error may have been introduced when calculating IADs based on daily/momentary affect scores that may not be perfectly reliable measures of state affect. Measurement error might have led to a downward bias of network parameters in particular (Schuurman, Houtveen, & Hamaker, 2015). Error variance may further accumulate when calculating residualized IADs based on affect means and standard variability that may not be perfectly reliable measures of trait affect and emotional variability. Thereby, validity estimates may be biased when residualized IADs are insufficiently cleared of their redundancies with mean level and general variability (Westfall & Yarkoni, 2016), but this might also occur in the presence of non-linear dependencies (Mestdagh *et al*., 2018). Further limitations of our statistical approach were that we did not consider alternative approaches that differ in their assumptions or estimation techniques (e.g. dynamic structural equation modelling, Asparouhov, Hamaker, & Muthén, 2018; Geukes *et al*., 2017; Kuppens, Oravecz, & Tuerlinckx, 2010; Loossens *et al*., 2019) and that we did not model the dyadic data structure in Sample 3 for calculating IADs.

CONCLUSIONS

The current study highlights that observed links between IADs and other constructs might be non-specific, as arising from redundancies between them. More specifically, in research settings in which affect dynamics are linked to between-person constructs, researchers should rule out more parsimonious explanations (e.g. trait affect, affect variability) before attributing incremental value to more complex IADs. Occasionally, researchers have strived for trait interpretations of IADs—explicitly or implicitly (i.e. by investigating or theorizing on their associations with relatively stable traits as was done in the current study). Our results demonstrate that there is scarce evidence for trait interpretations of many of the more complex IADs with respect to their low reliability und unknown stability.

More generally, researchers should consider the limited validity of IADs for research questions at the between-person level, as their reliability might be low, and any true effects might be obscured or attenuated. Notwithstanding, some of those more complex IADs had unique and somewhat reliable variance, including IADs with respect to emotion differentiation (e.g. $\bar{r}$, contemporaneous partial correlations) and emotional instability (e.g. $MSSD^c$). For those, small but incremental associations with personality traits were found. These results point out to the possibility that such IADs might contain substantive between-person variance that may be of interest to researchers for studying individual differences, albeit they may not always meet standard psychometric criteria (Wright & Zimmermann, 2019). The current assessment practices might need refinement in order to further improve the

validity of more complex IADs, for example, by increasing the frequency or duration of emotion assessment. However, it is questionable whether a more intensive or longer enduring assessment would be practically feasible, as longer time frames may increase burden of participation and lead to higher non-compliance rates (Eisele *et al.*, 2020). More research is needed for identifying the conditions under which affect dynamics can be assessed most validly.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Fit Statistics of Multi-Level Exploratory Factor Models (Geomin Rotation) by Sample

Table S2. Geomin Rotated Factor Loadings by ML-EFA in Sample 1

Table S3. Geomin Rotated Factor Loadings by ML-EFA in Sample 2

Table S4. Geomin Rotated Factor Loadings by ML-EFA in Sample 3

Table S5. Split-half Reliability of Raw IADs in Sample 2 and Sample 3

Table S6. Bivariate Correlations between Raw IADs and Big Five Traits by Sample

Table S7. Bivariate Correlations between Residualized IADs and Big Five Traits by Sample

Figure S1. Split-half Reliability of IADs in Daily Data (Sample 2).

Figure S2. Split-half Reliability of IADs in Momentary Data (Sample 3)

REFERENCES

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 359–388. https://doi.org/10.1080/10705511.2017.1406803

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*, 40, 1235–1245. https://doi.org/10.1016/j.paid.2005.10.018

Baird, B. M., Lucas, R. E., & Donnellan, M. B. (2017). The role of response styles in the assessment of intraindividual personality variability. *Journal of Research in Personality*, 69, 170–179. https://doi.org/10.1016/j.jrp.2016.06.015

Bos, E. H., de Jonge, P., & Cox, R. F. (2019). Affective variability in depression: Revisiting the inertia–instability paradox. *British Journal of Psychology*, 110, 814–827. https://doi.org/10.1111/bjop.12372

Bosley, H. G., Soyster, P. D., & Fisher, A. J. (2019). Affect dynamics as predictors of symptom severity and treatment response in mood and anxiety disorders: Evidence for specificity. *Journal for Person-Oriented Research*, 5, 101–113. https://doi.org/10.17505/jpor.2019.09

Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., … Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment*, 23, 425–435. https://doi.org/10.1177/1073191116645909

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, 8, e60188. https://doi.org/10.1371/journal.pone.0060188

Carver, C. S., Sutton, S. K., & Scheier, M. F. (2000). Action, emotion, and personality: Emerging conceptual integration. *Personality and Social Psychology Bulletin*, 26, 741–751. https://doi.org/10.1177/0146167200268008

Ching, C. M., Church, A. T., Katigbak, M. S., Reyes, J. A. S., Tanaka-Matsumi, J., Takaoka, S., … Ortiz, F. A. (2014). The manifestation of traits in everyday behavior and affect: A five-culture study. *Journal of Research in Personality*, 48, 1–16. https://doi.org/10.1016/j.jrp.2013.10.002

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PIR) and NEO Five Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

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/10.1097/00005053-198709000-00004

Dejonckheere, E., Mestdagh, M., Houben, M., Erbas, Y., Pe, M., Koval, P., … Kuppens, P. (2018). The bipolarity of affect and depressive symptoms. *Journal of Personality and Social Psychology*, 114, 323–341. https://doi.org/10.1037/pspp0000186

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

Dejonckheere, E., Mestdagh, M., Kuppens, P., & Tuerlinckx, F. (2020). Reply to: Context matters for affective chronometry. *Nature Human Behaviour*. Advance Online Publication, 1–4. https://doi.org/10.1038/s41562-020-0861-6

Denissen, J. J. A., & Kühnel, A. (2008). *Handbook for the use of data from the diary study at Humboldt Universität zu Berlin*. Retrieved from https://www.psychologie.hu-berlin.de/de/prof/perdev/downloadentwper/diarystudy/Handbook_Diary.pdf

DeYoung, C. G. (2015). Cybernetic big five theory. *Journal of Research in Personality*, 56, 33–58. https://doi.org/10.1016/j.jrp.2014.07.004

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, 76, 662–676. https://doi.org/10.1037/0022-3514.76.4.662

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2020, February 20). *The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population*. https://doi.org/10.31234/osf.io/zf4nm

Epskamp, S., Deserno, M. K., & Bringmann, L. F. (2019). *mlVAR: Multi-level vector autoregression*. R package version 0.4.3. https://CRAN.R-project.org/package=mlVAR

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, 53, 453–480. https://doi.org/10.1080/00273171.2018.1454823

Finnigan, K. M., & Vazire, S. (2018). The incremental validity of average state self-reports over global self-reports of personality. *Journal of Personality and Social Psychology*, 115, 321–337. https://doi.org/10.1037/pspp0000136

Geukes, K., Nestler, S., Hutteman, R., Dufner, M., Küfner, A. C., Egloff, B., Denissen, J. J. A., et al. (2017). Puffed-up but shaky selves: State self-esteem level and variability in narcissists. *Journal of Personality and Social Psychology*, 112, 769–786. https://doi.org/10.1037/pspp0000093

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*, 10–15. https://doi.org/10.1177/0963721416666518

Hinds, J., & Joinson, A. (2019). Human and computer personality prediction from digital footprints. *Current Directions in Psychological Science*, *28*, 204–211. https://doi.org/10.1177/0963721419827849

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

Howell, R. T., Ksendzova, M., Nestingen, E., Yerahian, C., & Iyer, R. (2017). Your personality on a good day: How trait and state personality predict daily well-being. *Journal of Research in Personality*, *69*, 250–263. https://doi.org/10.1016/j.jrp.2016.08.001

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

Kalokerinos, E., Murphy, S. C., Koval, P., Bailen, N. H., Crombez, G., Hollenstein, T., Gleeson, J., … Bastian, B. (2020). Neuroticism may not reflect emotional variability. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 9270–9276. https://doi.org/10.1073/pnas.1919934117

Kim, E. S., Dedrick, R. F., Cao, C., & Ferron, J. M. (2016). Multilevel factor analysis: Reporting guidelines and a review of reporting practices. *Multivariate Behavioral Research*, *51*, 881–898. https://doi.org/10.1080/00273171.2016.1228042

Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, *12*, 256–267. https://doi.org/10.1037/a0024756

Koval, P., Kuppens, P., Allen, N. B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition & Emotion*, *26*, 1412–1427. https://doi.org/10.1080/02699931.2012.667392

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

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*, 984–991. https://doi.org/10.1177/0956797610372634

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., Sheeber, L. B., Yap, M. B., Whittle, S., Simmons, J. G., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, *12*, 283–289. https://doi.org/10.1037/a0025046

Kuppens, P., Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. *Emotion*, *7*, 262–274. https://doi.org/10.1037/1528-3542.7.2.262

Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica*, *47*, 111–121.

Lapate, R. C., & Heller, A. S. (2020). Context matters for affective chronometry. *Nature Human Behaviour. Online advance publication.*, 1–2. https://doi.org/10.1038/s41562-020-0860-7

Loossens, T., Mestdagh, M., Dejonckheere, E., Kuppens, P., Tuerlinckx, F., & Verdonck, S. (2019, September 5). *The affective ising model: A computational account of human affect dynamics*. https://doi.org/10.31234/osf.io/ky23d

Mejía, S., Hooker, K., Ram, N., Pham, T., & Metoyer, R. (2014). Capturing intraindividual variation and covariation constructs: Using multiple time-scales to assess construct reliability and construct stability. *Research in Human Development*, *11*, 91–107. https://doi.org/10.1080/15427609.2014.906728

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, *23*, 690–707. https://doi.org/10.1037/met0000153

Mneimne, M., Fleeson, W., Arnold, E. M., & Furr, R. M. (2018). Differentiating the everyday emotion dynamics of borderline personality disorder from major depressive disorder and bipolar disorder. *Personality Disorders, Theory, Research, and Treatment*, *9*, 192–196. https://doi.org/10.1037/per0000255

Muthén, L. K., & Muthén, B. O. (2018). *Mplus User's Guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for statistical computing, Vienna, Austria. URL https://www.R-project.org/

Reise, S. P., Ventura, J., Nuechterlein, K. H., & Kim, K. H. (2005). An illustration of multilevel factor analysis. *Journal of Personality Assessment*, *84*, 126–136. https://doi.org/10.1207/s15327752jpa8402_02

Revelle, W. (2018) *psych: Procedures for personality and psychological research*, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psychVersion=1.8.12

Rodebaugh, T. L., Tonge, N. A., Piccirillo, M. L., Fried, E. I., Horenstein, A., Morrison, A. S., Goldin, P., … Heimberg, R. G. (2018). Does centrality in a cross-sectional network suggest intervention targets for social anxiety disorder? *Journal of Consulting and Clinical Psychology*, *86*, 831–844. https://doi.org/10.1037/ccp0000336

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*, 452–467. https://doi.org/10.3758/s13423-018-1558-y

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, *24*, 70–91. https://doi.org/10.1037/met0000188

Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in n= 1 psychological autoregressive modeling. *Frontiers in Psychology*, *6*, 1038. https://doi.org/10.3389/fpsyg.2015.01038

Scott, L. N., Victor, S. E., Kaufman, E. A., Beeney, J. E., Byrd, A. L., Vine, V., … Stepp, S. D. (2020). Affective dynamics across internalizing and externalizing dimensions of psychopathology. *Clinical Psychological Science*. Advance online publication, *8*, 412–427. https://doi.org/10.1177/2167702619898802

Stawski, R. S., MacDonald, S. W., & Sliwinski, M. J. (2015). Measurement burst design. *The Encyclopedia of Adulthood and Aging*, 1–5.

Suls, J., Martin, R., & David, J. P. (1998). Person-environment fit and its limits: Agreeableness, neuroticism, and emotional reactivity to interpersonal conflict. *Personality and Social Psychology Bulletin*, *24*, 88–98. https://doi.org/10.1177/0146167298241007

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

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48. URL. http://www.jstatsoft.org/v36/i03/

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*, 219–235. https://doi.org/10.1037//0033-2909.98.2.219

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS ONE*, *11*, e0152719. https://doi.org/10.1371/journal.pone.0152719

Wright, A. G. C., Beltz, A. M., Gates, K. M., Molenaar, P. C. M., & Simms, L. J. (2015). Examining the dynamic structure of daily internalizing and externalizing behavior at multiple levels of analysis. *Frontiers in Psychology*, *6*, 1914. https://doi.org/10.3389/fpsyg.2015.01914

Wright, A. G. C., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. *Psychological Assessment*, *32*, 502–515. https://osf.io/nf5me/

Wright, A. G. C., Stepp, S. D., Scott, L. N., Hallquist, M. N., Beeney, J. E., Lazarus, S. A., & Pilkonis, P. A. (2017). The effect of pathological narcissism on interpersonal and affective processes in social interactions. *Journal of Abnormal Psychology*, *126*, 898–910. https://doi.org/10.1037/abn0000286

Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment*, *31*, 1467–1480. https://doi.org/10.1037/pas0000685

Yarkoni, T. (2019, November 22). *The Generalizability Crisis*. https://doi.org/10.31234/osf.io/jqw35