Journal of Psychopathology and Behavioral Assessment
https://doi.org/10.1007/s10862-019-09766-7

Comparing Signal-Contingent and Event-Contingent Experience Sampling Ratings of Affect in a Sample of Psychotherapy Outpatients

Sindes Dawood[1] · Michael N. Hallquist[1] · Aaron L. Pincus[1] · Nilam Ram[1] · Michelle G. Newman[1] · Stephen J. Wilson[1] · Kenneth N. Levy[1]

Abstract

Experience sampling methods are widely used in clinical psychology to study affective dynamics in psychopathology. The present study examined whether affect ratings (valence and arousal) differed as a function of assessment schedule (signal- versus event-contingent) in a clinical sample and considered various approaches to modeling these ratings. A total of 40 community mental health center outpatients completed ratings of their affective experiences over a 21-day period using both signal-contingent schedules (random prompts) and event-contingent schedules (ratings following social interactions). We tested whether assessment schedules impacted 1) the central tendency (mean) and variability (standard deviation) of valence or arousal considered individually, 2) the joint variability in valence and arousal via the entropy metric, and 3) the between-person differences in configuration of valence-arousal landscapes via the Earth Mover's Distance (EMD) metric. We found that event-contingent schedules, relative to signal-contingent schedules, captured higher average levels of pleasant valence and emotional arousal ratings. Moreover, signal-contingent schedules captured greater variability within and between individuals on arousal-valence landscapes compared to event-contingent schedules. Altogether, findings suggest that the two assessment schedules should not be treated interchangeably in the assessment of affect over time. Researchers must be cautious in generalizing results across studies utilizing different experience sampling assessment schedules.

Keywords Signal-contingent schedules · Event-contingent schedules · Affective variability · Entropy · Earth Mover's Distance

This study examined the effects of time- and event-based assessment schedules on ratings of affect in a sample of psychotherapy outpatients. Over the last three decades, investigators have increasingly used experience sampling methodology (ESM; Csikszentmihalyi and Larson 1987; Mehl and Conner 2012) to investigate the momentary feelings, thoughts, and behaviors of people in their natural environments (Stone and Shiffman 1994; Shiffman et al. 2008). ESM is an umbrella term that encompasses a range of procedures that intensively and repeatedly assess individuals' experiences over specific periods of time (see Conner and Bliss-Moreau 2006; Mohr et al. 2017; Moskowitz et al. 2009; Shiffman et al. 2008; Trull and Ebner-Priemer 2013 for reviews). This includes, but is not limited to, paper diaries, repeated telephone interviews, electronic recording technologies (e.g., smartphones, internet), and wearable sensors (e.g., to measure peripheral physiology). Relative to retrospective self-report measures, ESM minimizes recall bias, improves ecological validity, and allows for the study of temporal processes over hours or days (Moskowitz et al. 2009; Shiffman et al. 2008).

ESM studies typically yield intensive repeated measures data (i.e., many observations per person over hours, days, or weeks), but often differ in the protocol for sampling participants' experiences. Self-report ESM data are commonly collected using one of three assessment schedules: interval-contingent, signal-contingent, or event-contingent (Reis and Gable 2000). In interval-contingent schedules, participants report on experiences or events at specific intervals (e.g., every 20 min over 24 h, at the end of each day for two weeks). In signal-contingent schedules, participants respond to several randomly scheduled prompts each day over a period of observation (e.g., one week, one month). Finally, in event-contingent schedules, participants report on their experiences following a predetermined type of event (e.g., a social interaction, a binge eating episode).

✉ Sindes Dawood
szd157@psu.edu

[1] Department of Psychology, Pennsylvania State University, 361 Bruce V. Moore Building, University Park, PA 16802, USA

Published online: 27 November 2019

ESM Studies of Affective Dynamics

A growing body of research in clinical psychology uses ESM to study the relationship between affective dynamics (e.g., extent of moment-to-moment variability) and psychopathology (Myin-Germeys et al. 2009; Trull et al. 2012). Although there are numerous models of core affect, we focus on the two-dimensional representations reflecting either *valence* (unpleasant/pleasant) and *arousal* (inactive/active) (Posner et al. 2005) or *negative affect* (NA) and *positive affect* (PA) (Watson and Tellegen 1985), which are both commonly employed in ESM studies. We briefly review several studies to demonstrate the substantial variation in assessment schedules used in studies of affective dynamics in clinical populations.

Interval-contingent schedules have been especially popular in studies of affective dynamics within varying forms of psychopathology. For instance, Ebner-Priemer and colleagues (Ebner-Priemer et al. 2006, 2007, 2008; Reisch et al. 2008) prompted patients with borderline personality disorder (BPD) ($N = 50$) and healthy controls ($N = 50$) four times per hour over a 24-h period to assess NA and PA. To assess affective valence, other studies have prompted participants with BPD or with other clinical diagnoses (post-traumatic stress disorder, bulimia nervosa, panic disorder, or major depression) every 15 min over 24 h, or every hour over a 48-h period (Santangelo et al. 2014, 2016). Means and variability of affect have been examined in patients with remitted bipolar disorder or remitted unipolar depressive disorder by way of asking participants to complete a diary measure of NA and PA twice a day, at a designated time each morning and evening, for one week (Knowles et al. 2007). Lastly, Farmer and Kashdan (2014) studied affective dynamics of social anxiety disorder (SAD) by having adults with SAD ($N = 40$) and matched healthy controls ($N = 39$) complete end-of-day reports on their NA and PA for 14 days.

Signal-contingent (e.g. randomly timed prompts) assessment schedules have also been used to study affective dynamics in psychopathology, especially BPD and mood disorders. For example, Stein (1996) randomly prompted participants with and without BPD five times per day within approximately three-hour intervals over 10 days to assess affective states of hedonic valence and activation. In contrast, Trull et al. (2008) scheduled random prompts six times per day over a 28-day period to assess PA and NA in BPD patients and clinically depressed patients. Most recently, Crowe et al. (2018) used signal-contingent schedules in which 31 patients with clinical major depressive disorder (MDD) and 33 healthy controls were randomly prompted 10 times a day within 90-min intervals over six days to complete items measuring (amongst others) PA/NA.

Finally, studies have used event-contingent assessment schedules to study affective dynamics in psychopathology, most notably in the examination of daily social interactions in BPD (e.g., Kopala-Sibley et al. 2012; Russell et al. 2007; Sadikaj et al. 2010, 2013). Moreover, event-contingent schedules have often been used with eating disordered patients (with or without BPD) to record their reported affective experiences surrounding a pre-determined event such as (but not limited to) engaging in dysregulated eating behaviors (e.g., binge eating/purging, bulimic behaviors) and/or self-destructive behaviors (e.g., alcohol intoxication, self-injury, drug abuse) over the course of two weeks (e.g., Berner et al. 2017; Mason et al. 2017; Pisetsky et al. 2016; Selby et al. 2012). ESM studies with eating-disordered individuals can also employ signal-contingent and/or interval-contingent schedules into their investigations.

Taken together, this research demonstrates that ESM designs often employ different assessment schedules at varying time scales to study affective dynamics in daily life. However, making sense of findings across ESM studies of affect is difficult because of potential confounding with the assessment schedule employed. To be more confident that results reflect substantive effects that are robust to assessment schedule, it is important to empirically compare ESM methods within the same study sample rather than descriptively comparing across different samples and ESM methods. Consequently, given that prior studies examining affective dynamics in psychopathology typically employed only one type of assessment schedule, little is known about whether and how the assessment schedule itself influences substantive conclusions. For instance, a clinical participant (e.g., BPD, an anxiety disorder) who tends to interpret positive social interactions negatively may experience more unpleasant affect and arousal in response to social interactions than a healthy participant (Clifton et al. 2007). Thus, when a study assesses patients' affect in response to a social interaction (i.e., an event-contingent schedule), investigators may capture higher average levels of unpleasant affect and arousal than when using signal-contingent schedules.

When the stimuli that precede assessments of affect are unknown (e.g., a person may or may not have engaged in a social interaction or encountered a stressor proximate to a random prompt), signal-contingent schedules may capture greater variability in valence and arousal than event-contingent schedules. This could reflect the heterogeneity of contexts, experiences, and events that modulate affect in randomly timed assessments in daily life. Alternatively, event-contingent schedules involving social interactions (as the event) may capture greater variability in valence and arousal than signal-contingent schedules because these interactions are among the strongest situations encountered in daily life (Pincus et al. in press).

Recently, Himmelstein et al. (2019) tested whether signal-contingent and event-contingent schedules of interpersonal behavior and affect in social situations produce the same quality and quantity of data using a between-person design. Unselected undergraduate students were randomized into either a signal-contingent or event-contingent group. Participants in both conditions were instructed to report on any social interaction lasting at least five minutes between themselves and at least one other person throughout the one-week study period. In the signal-contingent condition, participants were prompted six times per day within at least 90-min intervals to complete momentary interpersonal items and items on PA and NA. In the event-contingent group, participants were instructed to report on their interpersonal behavior and PA/NA immediately following any social interaction lasting at least five minutes and were asked to initiate at least four prompts per day.

These authors found no significant differences in means or variances in momentary interpersonal behaviors or PA/NA by ESM schedule. Moreover, ESM schedules did not impact within-person associations between momentary interpersonal behavior and affect variables. They did find, however, that participants reported, on average, a greater frequency of social interactions when using event-contingent schedules than signal-contingent schedules. Overall, these researchers interpreted their results to suggest signal-contingent and event-contingent schedules can be used interchangeably to make inferences about means, variances, and correlations between momentary constructs when studying social situations. However, more stringent comparisons of signal-contingent and event-contingent schedules of affect require participants to complete assessments using both formats concurrently during the same assessment period. It is also unclear whether findings from an unselected undergraduate sample would generalize to a community clinical sample.

The Current Study

The present study is the first to directly test whether affective ratings in ESM differed as a function of assessment schedule (i.e., signal- versus event-contingent) in a clinical outpatient sample. We also used three different but complementary analytic methods to evaluate our research question. More specifically, we first examined whether the assessment schedule impacted mean levels or variability in each dimension of the affective circumplex (valence and arousal) over the ESM study interval. Given our clinical sample, we hypothesized that there would be higher mean levels of unpleasant affect and arousal in event-contingent ratings following social interactions compared to random signal-contingent ratings (H1). We also hypothesized that there would be greater variability in valence and arousal in random signal-contingent ratings relative to event-contingent ratings (H2).

The ESM studies reviewed above have typically relied on unidimensional analyses of affect, examining within-person means for valence and arousal independently. Given that the affective dimensions of valence and arousal may be related within a person (e.g., a tendency to experience highly arousing negative emotions), it is essential to extend analyses of affective dynamics in psychopathology using modeling approaches that better capture the full valence-arousal affective space. Here, we examined whether the within-person joint variability in valence and arousal, which describes individual differences in blends of valence and arousal over time (e.g., a tendency to concurrently report low valence and high arousal; see Fig. 1), differs across assessment schedules. We summarized each individual's report of affect over time as a two-dimensional valence-arousal landscape (Ram et al. 2013) using entropy, a

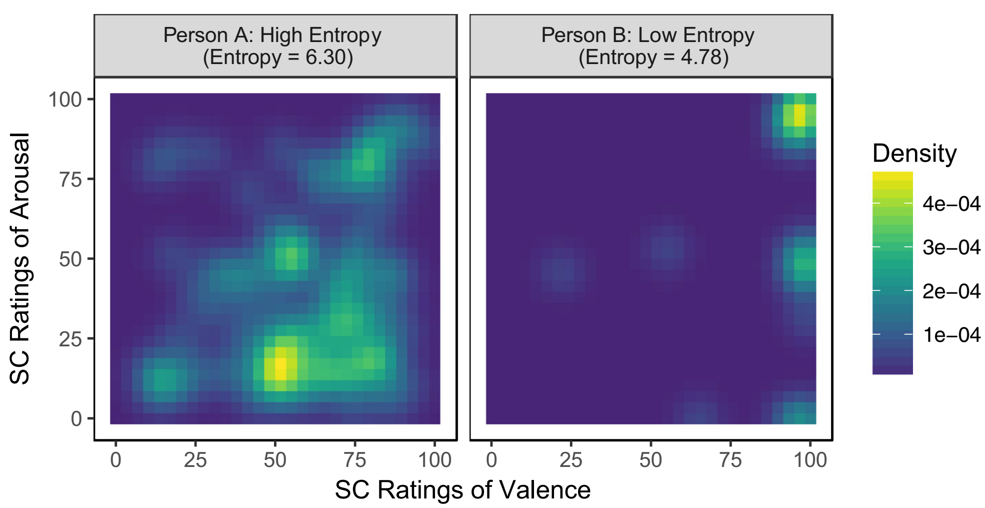

Fig. 1 Example of kernel density plots of entropy of signal-contingent (SC) affective valence-arousal ratings of two patients with borderline personality disorder. Entropy reflects how diffuse or narrow the "island" of a two-dimensional space (here, the affect of valence and arousal) individuals' tend to live in over a chosen assessment period (here, 21-days). Higher entropy values indicate greater spread across the two-dimensional space, as shown in the left panel where, over time, Person A describes their affects using more of the valence-arousal space relative to Person B in the right panel

complexity metric from information theory (Hollenstein 2007; Ram et al. 2012). The entropy metric reflects the total variability in the affective circumplex space without characterization of a particular affective style. Conceptually, entropy describes how compact or diffuse is the area of a two-dimensional valence-arousal space, or the "island" that individuals' tend to endorse over the assessment period (Ram et al. 2017). We hypothesized that the average level of entropy would be higher for randomly prompted signal-contingent affect ratings than event-contingent affect ratings (H3).

Lastly, we used Earth Mover's Distance (EMD; Rubner et al. 2000) as metric to quantify the omnibus similarity of two affective landscapes (Ram et al. 2013, 2017).

That is, EMD allowed for comparisons of the similarity of valence-arousal landscapes between and within assessment schedules. EMD provides a measure of the work needed to equate two landscapes numerically; thus, EMD is zero if two landscapes are identical, but increases in proportion to dissimilarities in two bivariate (here, valence-arousal) landscapes (see Fig. 2). More specifically, we examined the effect of assessment schedule on between-person differences in EMD. We expected between-person differences in affective landscapes would be greater for signal-contingent than event-contingent schedules (H4), thereby extending H3.

Method

Participants

Participants were recruited from a large University-affiliated community mental health center. All participants completed a standard intake evaluation, including two semi-structured clinical interviews, namely the Anxiety Disorders Interview Schedule-IV (ADIS-IV; Brown et al. 1994) to assess for *DSM-IV* Axis I disorders and the International Personality Disorders Examination for *DSM-IV* (IPDE; Loranger 1999) to assess for personality disorders (Axis II disorders). As part of a study designed to address questions outside the scope of the current paper, participants who completed the intake assessment were screened to meet diagnostic criteria for either BPD or any anxiety disorder without a diagnosis of BPD (and did not have a cognitive disorder). After being screened for eligibility, 54 of 173 participants (90% female) met study criteria and were recruited into a 21-day signal-contingent + event-contingent ESM study.

Doctoral graduate students in clinical psychology were trained to reliability in the use of both the ADIS-IV and IPDE, with the supervision of a licensed clinical psychologist. Diagnostic interviews were videotaped and interviews of a subsample of 32 participants were coded by a second rater (see Ellison et al. 2019, for demographic and diagnostic

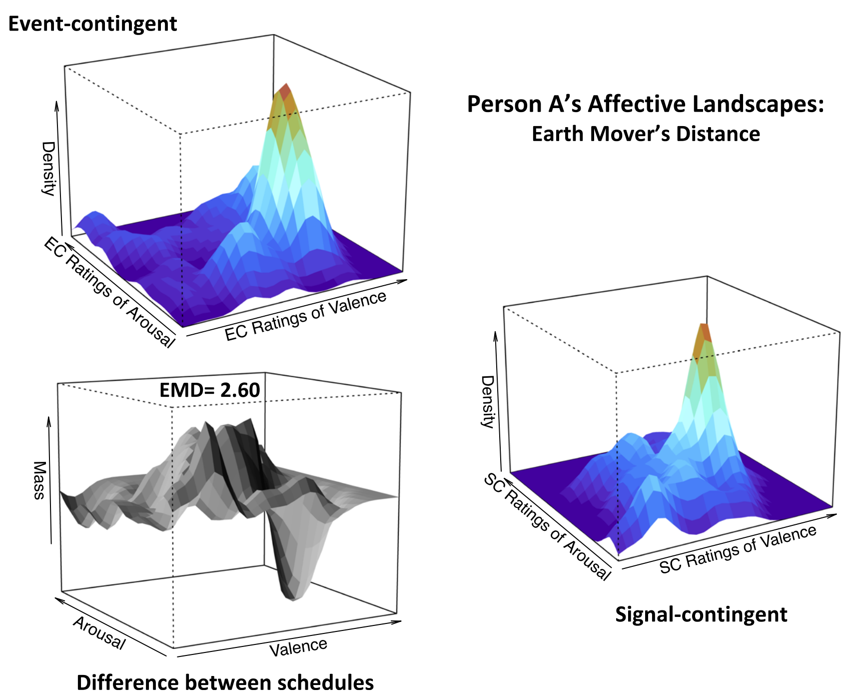

Fig. 2 Earth Mover's Distance (EMD) is the amount of work needed to equate two landscapes numerically. Two affective landscapes for an individual with borderline personality disorder (Person A) are displayed on the diagonal. The upper left landscape depicts the joint occurrence of event-contingent (EC) valence and arousal ratings over the study, whereas the lower right landscape depicts signal-contingent (SC) ratings. The difference between EC and SC landscapes is depicted in the lower left. The overall difference is quantified by EMD, the minimum work needed to make the ground flat. Graphically, this would involve bulldozing the peaks of the landscape into the depressions, thereby flattening the landscape. Thus, EMD is higher when the differences between SC and EC landscapes are greater. The EMD between signal-contingent and event-contingent affect ratings for Person A is 2.60

characteristics for the sample). Overall interrater reliability was substantial (on average κ = 0.62, with, κ = 0.66 for anxiety disorders and κ = 0.68 for BPD).

The present analyses made use of data from participants who completed at least 40 total observations for both signal-contingent and event-contingent ratings. We chose the cut point of 40 observations (i.e., the number of surveys completed) based on visual inspection of the lower tail end of the distribution of two histograms, one for event-contingent surveys and one for signal-contingent surveys.[1] Of the 54 patients, 14 had less than 40 total observations on either one or both assessment schedules and were therefore excluded from the analyses. The final sample consisted of 40 outpatients (36 women and 4 men), with diagnoses of anxiety disorders (75%), borderline personality disorder (60%), and major depressive disorder (40%) occurring most frequently (see Table 1 for diagnostic characteristics).These participants were, on average, 32.70 years of age (*SD* = 11.53), with 90% identifying themselves as Caucasian, 2.5% as African American, 2.5% as Asian, and 5% as "Other." Participants identified themselves as heterosexual (67.5%), bisexual (17.5%), or homosexual (15%); and as single (55%), dating (22.5%), married (17.5%), divorced (2.5%), or separated (2.5%).

Table 1 Diagnostic profile of sample

	Total (*N* = 40)	
	n	%
Current Axis I Diagnoses		
Mood Disorders	21	53
MDD	16	40
Other Mood Disorders	5	13
Anxiety Disorder	30	75
GAD	8	20
PTSD	7	18
Social Phobia	6	15
Other Anxiety Disorders	9	25
Alcohol & Substance Use	4	10
Eating Disorders	4	10
Somatoform Disorders	5	13
Current Axis II Diagnoses		
Any PD	38	96
Paranoid	0	0
Schizoid	0	0
Schizotypal	0	0
Antisocial	2	5
Borderline	24	60
Histrionic	2	5
Narcissistic	0	0
Avoidant	4	10
Obsessive-Compulsive	1	3
Dependent	0	0
PD NOS	5	13
	M	*SD*
No. of Axis I diagnoses	1.60	1.01
No. of Axis II diagnoses	0.95	0.75
No. of IPDE Criteria Met (*n* = 39)	10.95	7.21
IPDE Dimensional Score (*n* = 39)	33.69	18.60
GAF (*n* = 37)	55.59	10.06

MDD = major depressive disorder; GAD = generalized anxiety disorder; PTSD = post-traumatic stress disorder; PD NOS = personality disorder not otherwise specified; IPDE = International Personality Disorder Examination; GAF = Global Assessment of Functioning

Procedure

At the start of the study, participants visited the laboratory where they provided informed consent, completed a number of self-report questionnaires, and were trained by research assistants to complete surveys on a study-provided smartphone (Motorola Razr) that they would carry with them for about 21-days. During the next three weeks, participants were prompted, via an audible signal, to complete *signal-contingent* surveys at six random times during their waking hours. Specifically, a participant-defined 12-h "waking time" period was stratified into six equal intervals and prompts were delivered at a random time within each interval. In parallel, participants also completed *event-contingent* surveys about their affective experience following face-to-face social interactions of more than three minutes. They were required to log, on average, an entry for six interactions per day to achieve an "active participant status". Participants were compensated upon completion of the study. Missing data for the event-contingent reports were generated by coding missing values for the number of interactions patients reported per day in the study with the value *NA* in R. Missing data for signal-contingent reports were generated the same way, except now for the number of random prompts patients reported per day in the study. The final sample of 40 patients did not complete 27.68% of the signal-contingent reports and 29.85% of possible event-contingent reports, but overall, participants provided ratings for, on average, 5.18 social interactions per day. In total, participants included in this analysis provided 8877 reports (4251 signal-contingent; 4626 event-contingent) over 17 to 25 days (median = 21 days). Further details on study procedure have been described previously (see Scala et al. 2018).

[1] We used this approach, rather than Bolger et al. (2003) approach of dropping individuals with entries that are more than two standard deviations below the mean for the sample, because our specific analyses depend on a kernel density estimate of the two-dimensional space (e.g., valence and arousal). That is, a low number of completed surveys does not provide enough data to accurately estimate the person's surface. Thus, we do not think the conventional wisdom applies here.

Affect Measure

Participants' momentary affect experience, operationalized as *valence* and *arousal* (the two core dimensions of the affective circumplex; Russell 2003), was assessed in each randomly timed signal-contingent survey and in each event-contingent (i.e., after relevant social interactions) survey. For the prompted surveys, participants responded to the item, "Which of these best describes how you feel right now?" using two visual analog scales (0–100) rating valence ("Unpleasant" to "Pleasant") and arousal ("Sleepy" to "Activated/Aroused"). In the event-contingent surveys, participants responded to the item, "After this interaction my mood is:" using the same visual analog scales. Participants only saw an anchored slider for each item, not the underlying numeric scale.

Data Analysis

Individuals' repeated ratings of valence and arousal were summarized in a variety of ways: independently (mean or standard deviation of affect), together (entropy), and as two-dimensional affective landscapes (Earth Mover's Distance). Differences across assessment schedule (signal-contingent versus event-contingent) were then examined for each metric and measure.

Average Affect and Variability in Affect For each participant, we quantified their average level of valence and arousal over the 21-day study as two 'iMean' scores that indicated their emotional tendencies for valence and arousal. As well, we quantified the extent of each individuals' variability in valence and arousal ratings over time using the within-person standard deviation, 'iSD'. iMean and iSD measures were computed separately for signal-contingent and event-contingent ratings. Differences in emotional valence and arousal across assessment schedule (signal-contingent versus event-contingent) were examined using within-subjects (repeated measures) ANOVAs. To separate the emotional tendencies from variability over time, we conducted separate ANOVAs for iMeans and iSDs, respectively. The *ez* (Lawrence 2016) package in *R* was used to perform the ANOVA and compute effect size measures.

Entropy Although our mean and standard deviation analyses provide information about individual tendency (i.e., iMean) and variability (i.e., iSD) in valence and arousal, they do not explicitly accommodate the two-dimensional nature of core affect. Variability in the explicitly two-dimensional space was thus measured using Shannon's entropy metric. Essentially, entropy describes the total variability in the bivariate space spanning valence and arousal (see Fig. 1) such that individuals who experience a broader range of emotional experiences over the study would have higher entropy.

To calculate the entropy of each individual's repeated reports of valence and arousal, we first computed the density (relative frequency) of observations at each location in the two-dimensional space over the 21-day study using a nonparametric kernel density estimator (see Fig. 1 for an example). Following Ram et al. (2017), we divided the 0–100 range of valence and arousal into a 30 × 30 grid (i.e., covering approximately 3 valence and arousal points per cell) and used a Gaussian kernel with bandwidth of $h = 5$ to produce a smoothed, weighted estimate of the number of observations in each cell. We applied this procedure to each participant's ESM data to obtain an individual two-dimensional matrix (valence x arousal). Nonparametric density estimation was completed using the *bkde2function* in the *KernSmooth* package in *R* (Wand 2015). We then conducted a within-subjects ANOVA, where the dependent measure was the entropy values of affect ratings and the within-subject variable was type of assessment schedule.

Earth Mover's Distance As articulated above (and see Fig. 2), one can calculate the dissimilarity of two landscapes (i.e., two-dimensional density distributions) by estimating the work needed to equate them using EMD (see Ram et al. 2017 for computational details). To test for within-subject differences in the *configuration* of the affective (valence-arousal) landscape as a function of the type of assessment schedule, we computed the Earth Mover's Distance (EMD) between an individual's signal-contingent and event-contingent landscapes using the *emdist* package in *R* (Urbanek and Rubner 2012). For example, if Person A had largely identical landscapes for signal-contingent and event-contingent ratings of arousal and valence, the EMD would approach zero, indicating that the ratings were not sensitive to the assessment schedule.

To test whether there is a non-negligible effect of assessment schedule on affective landscapes, we used a one-sample t-test on within-person EMD values. Although the one-sample *t*-test provides information about whether the average EMD is larger than zero between signal- and event-contingent schedules, it does not facilitate an interpretation of the magnitude of such differences (e.g., effect size). To inform an understanding of variability of the EMD between signal- and event-contingent schedules, we conducted a landscape permutation procedure where a person's landscape was compared against a permuted copy of itself, varying the proportion of cells shuffled between 10% and 50% in 5% increments. This provides a useful benchmark for the EMD between schedules that quantifies how much the EMD differs as the level of similarity is parametrically decreased by permutation. Of note, the permutation procedure was designed to isolate differences in EMD, and thus did not change the summary statistics of the landscape, including entropy, because the permutation was performed by exchanging density estimates on the two-dimensional grid without altering their numerical values. For each level of shuffling (10–50%), we repeated the procedure 100 times, generating an empirical cumulative density function at each level. To interpret the observed difference in event- versus signal-contingent EMD, we computed the percentile of the observed mean against the EMD mean statistics from the permuted data.

Results

Average Affect We found that the iMeans for pleasantly valenced affect were significantly higher in event-contingent ratings relative to signal-contingent ratings, $F(1,39) = 37.12$, $p < 0.001$, generalized $\eta^2 = 0.09$, which contradicts our hypothesis. However, consistent with H1, the iMeans for arousal were significantly higher in event-contingent relative to signal-contingent ratings, $F(1,39) = 59.76$, $p < 0.001$, generalized $\eta^2 = 0.09$ (see Table 2).

Variability in Affect Inconsistent with H2, there were no significant effects of schedule on variability in valence, ps > 0.10. In contrast, and consistent with our hypothesis, we found that variability in arousal was significantly higher in signal-contingent ratings of arousal relative to event-contingent ratings, $F(1, 39) = 15.79$, $p < 0.001$, generalized $\eta^2 = 0.06$ (see Table 3).

Entropy Consistent with H3, we found that the average level of entropy was significantly higher for signal-contingent relative to event-contingent affect ratings, $F(1,38) = 8.88$, $p = 0.005$, generalized $\eta^2 = 0.04$ (see Table 4).

Earth Mover's Distance We found that the average EMD between signal-contingent and event-contingent affect ratings was 3.65 and was non-negligible (95% CI = 2.87, 4.42; $t(39) = 9.51$, $p < 0.0001$). To interpret the magnitude of this dissimilarity, we compared the mean EMD statistic against distributions of within-schedule permutations of landscapes in which 10% to 50% of cells were randomly shuffled. This analysis indicated that the level of dissimilarity *between* schedules was comparable to 45% of the cells in the landscape having been randomly shuffled (50th percentile of observed EMD mean versus the 45% shuffled distribution). Extending H3 and consistent with H4, we found a significant main effect of assessment schedule on between-person EMD, $X^2(1) = 20.27$, $p < 0.0001$, such that event-contingent landscapes (estimated marginal $M = 5.27$, $SE = 0.35$) were more similar between subjects than signal-contingent landscapes (estimated marginal $M = 5.91$, $SE = 0.35$).

Table 2 Within-person mean (iMean) ratings for valence and arousal as a function of type of assessment schedule

	iMean	
Variable	*M*	*SD*
Valence		
Event-Contingent	72.73[a]	12.40
Signal-Contingent	64.04[b]	16.06
Arousal		
Event-Contingent	63.98[a]	16.55
Signal-Contingent	54.01[b]	15.95

The iMean represents the average of assessment schedule ratings over the 21-day interval. Different superscripts indicate significant differences at $p \leq 0.05$

Table 3 Within-person standard deviation (iSD) ratings for valence and arousal as a function of type of assessment schedule

	iSD	
Variable	*M*	*SD*
Valence		
Event-Contingent	19.91	7.22
Signal-Contingent	20.26	2.11
Arousal		
Event-Contingent	20.95[a]	8.08
Signal-Contingent	25.09[b]	7.82

iSD represents the standard deviation of assessment schedule ratings over the 21-day interval. *M* iSD is the average within-person variability in the sample, whereas *SD* iSD reflects the between-person variability in iSD. Different superscripts indicate significant differences at $p \leq 0.05$

Discussion

Experience sampling with signal- and event-based assessment schedules is frequently used to study affective dynamics in psychopathology. To our knowledge, this is the first study to directly test whether affective ratings differ as a function of assessment schedule (signal- versus event-contingent) when concurrently assessed in the same sample. Briefly, signal-contingent schedules refer to ratings in response to randomly

Table 4 Entropy ratings for valence-arousal as a function of the type assessment schedule

	Entropy	
Variable	*M*	*SD*
Valence-Arousal Affect		
Event-Contingent	5.53[a]	0.60
Signal-Contingent	5.73[b]	0.39

Different superscripts indicate significant differences at $p \leq 0.05$

scheduled prompts, whereas event-contingent schedules refer to ratings following a predetermined event (here, a social interaction). In a general sample of psychotherapy outpatients, we examined the effects of assessment schedule using three complementary approaches to modeling affect ratings. First, we tested whether assessment schedule influence the central tendency (intraindividual mean) and variability (intraindividual standard deviation) of patients' repeated reports of their momentary valence and arousal. Second, we explored whether assessment schedule impacts the joint variability in arousal and valence, as measured by entropy. Lastly, we examined the effect of assessment schedule on between-person differences in the configuration of valence-arousal landscapes, as measured by Earth Mover's Distance (EMD).

As anticipated, we found that mean levels of arousal were higher for event-contingent ratings relative to signal-contingent ratings. However, contrary to expectations, we found mean levels of valence were higher for event-contingent ratings relative to signal-contingent ratings. Prior studies have found that individuals tend to report more friendly social interactions than unfriendly ones (Wang et al. 2014), and more hostile interactions may be briefer than 3 min. This may in part help us to understand why our data indicates individuals report on average more positive emotional valence and greater emotional activation following social interactions over time. Moreover, one possibility is that context (here, social interaction) moderates affective ratings. Hence, deviations from mean pleasantness may be of particular interest to take note of as they are uncommon, even within-person (see Liu et al. 2019 for review and meta-analysis). Although we chose to use social interactions as a prompt for affect ratings, our findings may have implications for other psychopathology research that used different eliciting events for such ratings. For example, affect ratings linked with binge-eating episodes (Smyth et al. 2007) or non-suicidal self-injury (Muehlenkamp et al. 2009) may elicit greater negative affect on average.

Additionally, we found greater variability in arousal, but not valence, in signal-contingent ratings relative to event-contingent ratings. Other studies using signal-contingent schedules to examine variability in valence in psychopathology have yielded mixed findings. For instance, some ESM studies using signal-contingent schedules to measure valence as two independent dimensions (a separate scale for PA and a separate scale for NA) found BPD patients exhibited greater variability in PA and NA than clinically depressed patients (Trull et al. 2008) or that MDD patients displayed greater variability in PA and NA than healthy controls (Crowe et al. 2018). These studies did not include a measure of arousal. In contrast, other ESM studies using signal-contingent schedules to measure affect as represented by the two dimensions valence (unpleasant to pleasant) and arousal (inactive to activated) of an affective space show a pattern similar to our results (e.g., Stein 1996; Vansteelandt et al. 2013). For example, Stein (1996) found pleasant and unpleasant affect did not differ between BPD patients and eating disordered patients; however, BPD patients did report greater variability for high activation emotions (i.e., aroused/surprised) compared to eating disordered patients. Thus, the type of measure used to assess affect (e.g., valence/arousal vs. negative affectivity/positive affectivity) may also impact affective variability ratings. This remains a question for future research. Overall, our results suggest affective arousal is sensitive to the type of ESM assessment schedule employed. This is important, as affective arousal is understudied relative to affective valence.

As hypothesized, we also found that the average level of entropy was higher for signal-contingent than event-contingent affect ratings. This indicates that patients reported more diverse affective experiences over time when assessed by signal-contingent schedules. Furthermore, using EMD we found that event-contingent valence-arousal landscapes were more similar across individuals than signal-contingent landscapes. Taken together, these findings demonstrate that signal-contingent schedules generally capture more variability *within* individuals (i.e., individuals rate affective experiences spanning more of the valence-arousal affective space) and more variability *between* individuals when modeled as a two-dimensional affective space (i.e., the landscape). When interpreting results, investigators should keep in mind that event-contingent schedules increase the homogeneity of events sampled in daily life relative to signal-contingent designs. An advantage of considering the joint variability in valence and arousal (entropy), rather than treating each affect dimension individually, is that researchers can more precisely model the co-occurrence of these core dimensions of affect.

In addition, this study's findings contrast to those of Himmelstein et al. (2019) whom suggested signal-contingent and event-contingent schedules can be used interchangeably to make conclusions about means, variances, and correlations between momentary constructs (e.g., affect and interpersonal behavior) when studying social situations. The current results suggest the opposite conclusion. However, there are several important differences between the two studies. For example, Himmelstein and colleagues employed a between-subjects design whereas the present study employed a within-subjects design. Other important differences include sample type, measure of affect, and the frequency and duration of sampling, all of which may contribute to the different pattern of results. Thus, future ESM studies in this area would be helpful to confirm our findings.

Assessment Implications and Future Directions

The present study shows that when using ESM, newer modeling methods (e.g., entropy, EMD) can be used to study two-dimensional constructs such as affect landscapes in clinical populations, which has real-world implications. For instance, examining differences in two-dimensional affective landscapes may be particularly useful in that it allows clinicians to see whether the patient's configuration of valence-arousal landscape changes across individualized or standardized treatment. To test whether such changes are clinically meaningful (e.g., indicate the extent to which affect surrounding a person's social life is different than the rest of his/her life), the establishment of benchmarks for within-person entropy and EMD in a "healthy" population may be helpful. Specifically, EMD may provide clinicians with a quantification of the degree to which an individual with a psychological disorder moves toward a healthy configuration of affective experience with appropriate treatment. The information gleaned from an individual's affective landscapes as assessed by ESM could be incorporated into patient diagnosis, treatment planning, and progress monitoring as appropriate (Roche and Pincus 2016; van Os et al. 2013; Wright et al. 2019). Additional research is needed to test such possibilities.

This study also adds to the importance of assessing affective dynamics in psychological disorders more generally, as there is a growing literature suggesting that affective variability may be a transdiagnostic marker of psychopathology (e.g., Santangelo et al. 2014). Although affective variability was initially a focus in studies of BPD (Trull et al. 2008), accumulating research indicates that affective variability does not differ across a number of diagnostic groups. Beyond diagnoses, affective variability is also associated with other mental health problems including depressive symptoms (Koval et al. 2013), alcohol consumption (Mohr et al. 2015), and suicidal ideation (Palmier-Claus et al. 2012). Accordingly, event-contingent and signal-contingent schedules are useful tools for studying affective processes and their role in individuals' psychological well-being. This is also consistent with advances in transdiagnostic treatment approaches that broadly focus on affective dysregulation (e.g., Barlow and Farchione 2018).

Another area for future research is to understand how eliciting events (context) moderate signal-contingent affect ratings. Our findings suggest event-contingent and signal-contingent schedules should not be treated as interchangeable. However, signal-contingent data that are proximate to a social interaction may resemble event-contingent data. Should this be the case, it suggests that, with appropriate assessment of context, the moderation of ratings by context (here, the presence or absence of a social interaction) can be analyzed within signal-contingent assessment schedules to yield similar information derived from event-contingent data. This possibility presumes that the assessment occurs at a high enough temporal sampling rate, and for a long enough period of time, to randomly encounter and code social interactions and other contexts of interest. However, in many circumstances the contingent event of interest (e.g., a binge eating episode) will be too infrequent, thus there remains a place for both types of assessment schedules.

Limitations and Conclusion

This study is not without its own problems and limitations, which should be considered when interpreting study findings. First, given the combined signal-contingent and event-contingent design creates high participant burden, the sample size was relatively small. Consequently, our study may have been underpowered to detect some effects. Null findings should be interpreted cautiously and replication in larger clinical samples is warranted. Second, we used a convenience sample of outpatients in this study. The sample was predominately female and Caucasian and consisted only of patients from a community mental health center who were oversampled for either BPD or any anxiety disorder. Future studies are needed to replicate our findings with a larger more diverse sample and to determine whether our findings would generalize to individuals with other demographic characteristics, as well as to nonclinical and other clinical populations from different settings (e.g., inpatient hospital, outpatient psychiatric clinic, long-term care). Third, we used only one type of eliciting event (face to face social interactions) in our event-contingent assessment, thus the current results may not generalize to other types of eliciting events commonly studied in clinical research such as eating binges and episodes of non-suicidal self-injury. Fourth, our results may not generalize to studies that use different assessment schedules (e.g., interval-contingent), time sampling strategies, and measures of affect other than those used in the present study. For instance, we only considered valence and arousal dimensions of affect in the current investigation, but other two-dimensional models of affect (e.g., Watson and Tellegen 1985; Thayer 1989) and even three-dimensional models (e.g., Carver and Scheier 1998; Schimmack and Grob 2000) could be employed. These could be tested in future ESM studies examining affective dynamics in psychopathology. Furthermore, the inability of our data to separate out PA and NA may also be considered a limitation given that PA and NA are usually distinct (if anticorrelated), but can become bipolar under stress (Zautra et al. 2002). Consequently, the methodological choice to use one scale to capture valence than separating out PA and NA may have influenced findings (e.g., Zautra et al. 2002) and therefore should be considered in future study designs. Fifth, the signal-contingent and event-contingent item stems for affect were slightly different, however, the visual analogue scale anchors and response options were identical. Finally, our results focusing on affective variability (iSD, entropy) may not generalize to findings examining affective instability, which accounts both within-person variability and temporal dependency (e.g., Dawood and Pincus 2018; Ebner-Priemer et al. 2007; Santangelo et al. 2014).

Despite these limitations, the present study makes a novel contribution to the ESM literature on affective dynamics and psychopathology. Notably, we found that assessment schedules impact ESM affect ratings in our clinical sample. Specifically, event-contingent schedules homogenize experience sampling relative to more random signal-contingent schedules. In the present study, the event-contingent context of social interactions was associated with more pleasant valence and emotional arousal in affect ratings. We also found that signal-contingent schedules captured greater variability within individuals and between individuals on arousal-valence landscapes relative to event-contingent schedules. Taken together, context appears to moderate affective ratings. Importantly, our findings suggest that event-contingent schedules and signal-contingent schedules should not be treated interchangeably in the assessment of affect over time; and concurrent modeling of two-dimensional landscapes (Ram et al. 2013, 2017) offers new ways to examine ESM ratings of affect for research and clinical purposes. Ultimately, we hope our study encourages ESM researchers interested in describing dynamic affective processes give thoughtful consideration to what assessment schedules and modeling approaches would be optimal for their particular scientific questions.

Acknowledgements We thank J. Wesley Scala, M.S., and Emily A. Dowgwillo, M.S., for assistance in developing training material and coordinating the collection of data, and Kathleen Bohomey, Colin Carey, Caroline Curran, Wendi Falk, Sarah Forsythe, Laura Frey, Caroline Gooch, Jessica Grom, Brittani Hollern, Lauren Lipner, Kristin McLaughlin, Megan Moyer, Joanna Pantelides, Megan Parker, Jacqueline Proczynski, Carolina Ribo, Silvia Rizkallah, Aimee Sohnleitner, and Alyssa Spaw for their assistance in data collection. We also thank Jennifer Fox and Allison Clark for their assistance in the recruitment of participants. Finally, we thank Drs. William D. Ellison, Peter Molenaar and Joshua Smyth for consultation.

Funding Information This research was supported by grants from The Pennsylvania State University Social Science Research Institute (principal investigator: Levy), and secondary grants to William D. Ellison from a dissertation grant from The Pennsylvania State University Research and Graduate Studies Office, and grants from American Psychoanalytic Association and International Psychoanalytic Association.

Compliance with Ethical Standards

Conflict of Interest Sindes Dawood, Michael N. Hallquist, Aaron L. Pincus, Nilam Ram, Michelle G. Newman, Stephen J. Wilson, and Kenneth N. Levy declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

Barlow, D. H., & Farchione, T. J. (2018). *Applications of the unified protocol for transdiagnostic treatment of emotional disorder*. New York: Oxford University Press.

Berner, L. A., Crosby, R. D., Cao, L., Engel, S. G., Lavender, J. M., Mitchell, J. E., & Wonderlich, S. A. (2017). Temporal associations between affective instability and dysregulated eating behavior in bulimia nervosa. *Journal of Psychiatric Research, 92*, 183–190. https://doi.org/10.1016/j.jpsychires.2017.04.009.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030.

Brown, T. A., Di Nardo, P. A., & Barlow, D. H. (1994). *Anxiety disorders interview schedule for DSM-IV (adult version)*. Albany: Graywind.

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. New York: Cambridge University Press.

Clifton, A., Pilkonis, P. A., & McCarty, C. (2007). Social networks in borderline personality disorder. *Journal of Personality Disorders, 21*, 434–441. https://doi.org/10.1521/pedi.2007.21.4.434.

Conner, T., & Bliss-Moreau, E. (2006). Sampling human experience in naturalistic settings. In S. Hesse-Biber & P. Leavy (Eds.), *Emergent methods in social research* (pp. 109–129). Thousand Oaks: Sage.

Crowe, E., Daly, M., Delaney, L., Carroll, S., & Malone, K.M. (2018). *Psychiatry Research.* Advanced online publication. https://doi.org/10.1016/j.psychres.2018.02.032.

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/10.1007/978-94-017-9088-8_3.

Dawood, S., & Pincus, A. L. (2018). Pathological narcissism and the severity, variability, and instability of depressive symptoms. *Personality Disorders: Theory, Research, and Treatment, 9*, 144–154. https://doi.org/10.1037/per0000239.

Ebner-Priemer, U. W., Kuo, J., Welch, S. S., Thielgen, T., Witte, S., Bohus, M., & Linehan, M. M. (2006). A valence-dependent group specific recall bias of retrospective self-reports: A study of borderline personality disorder in everyday life. *Journal of Nervous and Mental Disease, 194*, 774–779. https://doi.org/10.1097/01.nmd.0000239900.46595.72.

Ebner-Priemer, U. W., Kuo, J., Kleindienst, N., Welch, S. S., Reisch, T., Reinhard, I., Lieb., Linehan, & Bohus (2007). State affective instability in borderline personality disorder assessed by ambulatory monitoring. *Psychological Medicine, 37*, 961–970. https://doi.org/10.1017/S0033291706009706 .

Ebner-Priemer, U. W., Kuo, J., Scholtz, W., Kleindienst, N., Rosenthal, Z., Detterer, L., Linehan, M., & Bohus, M. (2008). Distress and affective dysregulation in patients with borderline personality disorder: A psychophysiological ambulatory monitoring study. *Journal of Nervous and Mental Disorders, 196*, 314–320. https://doi.org/10.1097/NMD.0b013e31816a493f.

Ellison, W.D., Levy, K.N., Newman, M.G., Pincus, A.L., Wilson, S.J., & Molenaar, P.C.M. (2019). Dynamics among borderline personality and anxiety features in psychotherapy outpatients: An exploration of nomothetic and idiographic patterns. *Personality Disorders: Theory, Research, and Treatment.* Advanced online publication. doi: https://doi.org/10.1037/per0000363.

Farmer, A. S., & Kashdan, T. B. (2014). Affective and self-esteem instability in the daily lives of people with generalized social anxiety disorder. *Clinical Psychological Science, 2*, 187–201. https://doi.org/10.1177/2167702613495200.

Himmelstein, P. H., Woods, W. C., & Wright, A. G. C. (2019). A comparison of signal- and event-contingent ambulatory assessment of interpersonal behavior and affect in social situations. *Psychological Assessment, 31*, 952–960. https://doi.org/10.1037/pas0000718.

Hollenstein, T. (2007). State space grids: Analyzing dynamics across development. *International Journal of Behavioral Development, 31*, 384–396. https://doi.org/10.1177/0165025407077765.

Knowles, R., Tai, S., Jones, S. H., Highfield, J., Morriss, R., & Bentall, R. P. (2007). Stability of self-esteem in bipolar disorder: Comparisons among remitted bipolar patients, remitted unipolar patients and healthy controls. *Bipolar Disorder, 9*, 490–495. https://doi.org/10.1111/j.1399-5618.2007.00457.x.

Kopala-Sibley, D. C., Zuroff, D. C., Russell, J. J., Moskowitz, D. S., & Paris, J. (2012). Understanding heterogeneity in borderline personality disorder: Differences in affective reactivity explained by the traits of dependency and self-criticism. *Journal of Abnormal Psychology, 121*, 680–691. https://doi.org/10.1037/a0028513.

Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affective dynamics in relation to depressive symptoms: Variable, unstable, or inert? *Emotion, 13*, 1132–1141. https://doi.org/10.1037/a0033579.

Lawrence, M. A. (2016). ez: Easy analysis and visualization of factorial experiments. URL. https://cran.r-project.org/package=ez.

Liu, H., Xie, Q. W., & Lou, V. W. Q. (2019). Everyday social interactions and intra-individual variability in affect: A systematic review and meta-analysis of ecological momentary assessment studies. *Motivation and Emotion, 43*, 339–353. https://doi.org/10.1007/s11031-018-9735-x.

Loranger, A. W. (1999). *International Personality Disorder Examination (IPDE)*. Odessa: Psychological Assessment Resources.

Mason, T., Lavender, J. M., Wonderlich, S. A., Steiger, H., Cao, L., Engel, S. G., Mitchell, J. E., & Crosby, R. D. (2017). Comfortably numb: The role of momentary dissociation in the experience of negative affect around binge eating. *The Journal of Nervous and Mental Disease, 205*, 335–339. https://doi.org/10.1097/NMD.0000000000000658.

Mehl, M. R., & Conner, T. S. (2012). *Handbook of research methods for studying daily life*. New York: The Guildford Press.

Mohr, C. D., Aprin, S., & McCabe, C. T. (2015). Daily affect variability and context-specific alcohol consumption. *Drug and Alcohol Review, 34*, 581–587. https://doi.org/10.1111/dar.12253.

Mohr, C. D., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology, 13*, 23–47. https://doi.org/10.1146/annurev-clinpsy-032816-044949.

Moskowitz, D. S., Russell, J. J., Sadikaj, G., & Sutton, R. (2009). Measuring people intensively. *Canadian Psychology, 50*, 131–140. https://doi.org/10.1037/a0016625.

Muehlenkamp, J. J., Engel, S. G., Crosby, R. D., Wonderlich, S. A., Simonich, H., & Mitchell, J. E. (2009). Emotional states preceding and following acts of non-suicidal self-injury in bulimia nervosa patients. *Behavior Research and Therapy, 47*, 83–87. https://doi.org/10.1016/j.brat.2008.10.011.

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespau, P., & van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological Medicine, 39*, 1533–1547. https://doi.org/10.1017/S0033291708004947.

Palmier-Claus, J. E., Taylor, P. J., Gooding, P., Dunn, G., & Lewis, S. W. (2012). Affective variability predicts suicidal ideation in individuals at ultra-high risk of developing psychosis: An experience sampling study. *Journal of Clinical Psychology, 51*, 72–83. https://doi.org/10.1111/j.2044-8260.2011.02013.x.

Pincus, A. L., Hopwood, C. J., & Wright, A. G. C. (in press). The interpersonal situation: An integrative framework for the study of personality, psychopathology, and psychotherapy. In D. Funder, J. F. Rauthmann, & R. Sherman (Eds.), *Oxford handbook of psychological situations*. New York: Oxford University Press.

Pisetsky, E. M., Crosby, R., Cao, L., Fitzsimmons-Craft, E. E., Mitchell, J. M., Engel, S. G., & Peterson, C. B. (2016). An examination of affect prior to and following episodes of getting drunk in women with bulimia nervosa. *Psychiatry Research, 240*, 202–208. https://doi.org/10.1016/j.psychres.2016.04.044.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17*, 715–734. https://doi.org/10.1017/S0954579405050340.

Ram, N., Conroy, D. E., Pincus, A. L., Hyde, A. L., & Molloy, L. (2012). Tethering theory to method: Using measures of intraindividual variability to operationalize individuals' dynamic characteristics. In G. Hancock & J. Harring (Eds.), *Advances in longitudinal modeling in the social and behavioral sciences* (pp. 81–110). Charlotte: Information Age Publishing.

Ram, N., Coccia, M., Conroy, D. E., Lorek Dattiko, A., Orland, B., Pincus, A. L., Sliwinski, M., & Gerstorf, D. (2013). Behavioral landscapes and change in behavioral landscapes: A multiple timescale density distribution approach. *Research in Human Development, 10*, 88–110. https://doi.org/10.1016/j.jrp.2016.06.010.

Ram, N., Benson, K., Brick, T. R., Conroy, D. E., & Pincus, A. L. (2017). Behavioral landscapes and earth mover's distance. A new approach for studying individual differences in density distributions. *Journal of Research in Personality, 69*, 191–205. https://doi.org/10.1016/j.jrp.2016.06.010.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). New York Cambridge Univ. Press.

Reisch, T., Ebner-Priemer, U. W., Tschacher, W., Bohus, M., & Linehan, M. M. (2008). Sequences of emotions in patients with borderline personality disorder. *Acta Psychiatrica Scandinavica, 118*, 42–48. https://doi.org/10.1111/j.1600-0447.2008.01222.x.

Roche, M. J., & Pincus, A. L. (2016). Precision assessment. An individualized and temporally dynamic approach to understanding patients in their daily lives. In U. Kumar (Ed.), *The Wiley handbook of personality assessment* (pp. 192–204). Chichester: Wiley, Ltd.

Rubner, Y., Tomasi, C., & Guibas, L. (2000). The earth mover's distance as a metric for image retrieval. *International Journal on Computer Vision, 40*, 99–121.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145–172. https://doi.org/10.1037/0033-295X.110.1.145.

Russell, J. J., Moskowitz, D. S., Zuroff, D. C., Sookman, D., & Paris, J. (2007). Stability and variability of affective experiences and interpersonal behavior in borderline personality disorder. *Journal of Abnormal Psychology, 116*, 578–588. https://doi.org/10.1037/0021-843X.116.3.578.

Sadikaj, G., Russell, J. J., Moskowitz, D. S., & Paris, J. (2010). Affect dysregulation in individuals with borderline personality disorder: Persistence and interpersonal triggers. *Journal of Personality Assessment, 92*, 490–500. https://doi.org/10.1080/00223891.2010.513287.

Sadikaj, G., Moskowitz, D. S., Russell, J. J., Zuroff, D. C., & Paris, J. (2013). Quarrelsome behavior in borderline personality disorder:

Influence of behavioral and affective reactivity to perceptions of others. *Journal of Abnormal Psychology, 122*, 195–207. https://doi.org/10.1037/a0030871.

Santangelo, P., Reinhard, I., Mussgay, L., Steil, R., Sawitzki, G., Klein, C., et al. (2014). Specificity of affective instability in patients with borderline personality disorder compared to posttraumatic stress disorder, bulimia nervosa, and healthy controls. *Journal of Abnormal Psychology, 123*, 258–272. https://doi.org/10.1037/a0035619.

Santangelo, P. S., Limberger, M. F., Stiglmayr, C., Houben, M., Coosemans, J., Verleysen, G., Kuppens, P., Tuerlinckx, F., Vanpaemel, W., & Ebner-Priemer, U. W. (2016). Analyzing subcomponents of affective dysregulation in borderline personality disorder in comparison to other clinical groups using multiple e-diary datasets. *Borderline Personality Disorder and Emotion Dysregulation, 3*, 1–13. https://doi.org/10.1186/s40479-016-0039-z.

Scala, J. W., Levy, K. N., Johnson, B. N., Kivity, Y., Ellison, W. D., Pincus, A. L., et al. (2018). The role of negative affect and self concept clarity in predicting self-injurious urges in borderline personality disorders using ecological momentary assessment. *Journal of Personality Disorders, 32*, 36–57. https://doi.org/10.1521/pedi.2018.32.supp.36.

Schimmack, U., & Grob, A. (2000). Dimensional models of core affect: A quantitative by means of structural equation modeling. *European Journal of Personality, 14*, 325–345. https://doi.org/10.1002/1099-0984(200007/08)14:4<325::AID-PER380>3.0.CO;2-I.

Selby, E. A., Doyle, P., Crosby, R. D., Wonderlich, S. A., Engel, S. G., Mitchell, J. D., & Le Grange, D. (2012). Momentary emotion surrounding bulimic behaviors in women with bulimia nervosa and borderline personality disorder. *Journal of Psychiatric Research, 46*, 1492–1500. https://doi.org/10.1016/j.jpsychires.2012.08.014.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415.

Smyth, J. M., Wonderlich, S. A., Heron, K. E., Sliwinski, M. J., Crosby, R. D., Mitchell, J. E., & Engel, S. G. (2007). Daily and momentary mood and stress are associated with binge eating and vomiting in bulimia nervosa patients in the natural environment. *Journal of Consulting and Clinical Psychology, 75*, 629–638. https://doi.org/10.1037/0022-006X.75.4.629.

Stein, K. F. (1996). Affect instability in adults with a borderline personality disorder. *Archives of Psychiatric Nursing, 10*, 32–40. https://doi.org/10.1016/S0883-9417(96)80084-7.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine, 16*, 199–202. https://doi.org/10.1093/abm/16.3.199.

Thayer, R. E. (1989). *The biopsychology of mood*. New York: Oxford University Press.

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology, 9*, 151–176. https://doi.org/10.1146/annurev-clinpsy-050212-185510.

Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, 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.

Trull, T. J., Ebner-Priemer, U. W., Brown, W. C., Tomko, R. L., & Scheiderer, E. M. (2012). Clinical psychology. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 620–635). New York: Guilford Press.

Urbanek, S., and Y. Rubner, 2012. *emdist: Earth Mover's Distance.* R package version 0.2-1. URL http://CRAN.R-project.org/package=emdist.

van Os, J., Delespaul, P., Wigman, J., Myin-Germeys, I., & Wichers, M. (2013). Beyond DSM and ICD: Introducing "precision diagnosis" for psychiatry using momentary assessment technology. *World Psychiatry, 12*, 113–117. https://doi.org/10.1002/wps.20046.

Vansteelandt, K., Claes, L., Muehlenkamp, J., Claes, L., Muehlenkamp, K., De Cuyper, K., Lemmens, J., Probst, M., Vanderlinden, J., & Pieters, G. (2013). Variability in affective activation predicts non-suicidal self-injury in eating disorders. *European Eating Disorders Review, 21*, 143–147. https://doi.org/10.1002/erv.2220.

Wand, M. (2015). KernSmooth: Functions for kernel smoothing supporting Wand and Jones (1995). R package version 2.23-15, URL http://CRAN.R-project.org/package=KernSmooth.

Wang, S., Roche, M. J., Pincus, A. L., Conroy, D. E., Rebar, A., & Ram, N. (2014). Interpersonal dependency and emotion in every day life. *Journal of Research in Personality, 53*, 5–12. https://doi.org/10.1016/j.jrp.2014.07.007.

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.

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, 31*, 502–515. https://doi.org/10.1037/pas0000617.

Zautra, A. J., Berkhof, J., & Nicolson, N. A. (2002). Changes in affect interrelations as a function of stressful events. *Cognition and Emotion, 16*, 309–318. https://doi.org/10.1080/02699930143000257.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.