Multimodal Analysis of Temporal Affective Variability Within Treatment for Depression

Article *in* Journal of Consulting and Clinical Psychology · July 2024

DOI: 10.1037/ccp0000901

Adar Paz
Geha Mental Health Center

Eshkol Rafaeli
Bar Ilan University

Eran Bar-Kalifa
Ben-Gurion University of the Negev

Eva Gilboa-Schechtman
Bar Ilan University

Multimodal Analysis of Temporal Affective Variability within Treatment for Depression

Accepted: April 15, 2024

Adar Paz[1], Eshkol Rafaeli[1], Eran Bar-Kalifa[2], Eva Gilboa-Schectman[1], Sharon Gannot[3], Shrikanth Narayanan[4], and Dana Atzil-Slonim[1]

[1] Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel

[2] Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel

[3] Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel

[4] Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA

Please address correspondence to: Adar Paz, Bar-Ilan University, Psychology Department, Ramat-Gan, Israel; Phone number: 972-54-6563162; adar.paz@gmail.com

Abstract

Objective: Affective flexibility, the capacity to respond to life's varying environmental changes in a dynamic and adaptive manner, is considered a central aspect of psychological health in many psychotherapeutic approaches. The present study examined whether affective two-dimensional (i.e., arousal and valence) temporal variability extracted from voice and facial expressions would be associated with positive changes over the course of psychotherapy, at the session, client, and treatment levels. ***Method:*** 22,741 mean vocal arousal and facial expression valence observations were extracted from 137 therapy sessions in a sample of 30 clients treated for major depressive disorder by nine therapists. Before and after each session, the clients self-reported their level of well-being on the Outcome Rating Scale. Session-level affective temporal variability was assessed as the mean square of successive differences (MSSD) between consecutive two-dimensional affective measures. ***Results:*** Session outcome was positively associated with temporal variability at the session level (i.e., within clients, between sessions) and at the client level (i.e., between clients). Importantly, these associations held when controlling for average session- and client-level valence scores. In addition, the expansion of temporal variability throughout treatment was associated with steeper positive session outcome trajectories over the course of treatment. ***Conclusions:*** The continuous assessment of both vocal and facial affective expressions and the ability to extract measures of affective temporal variability from within-session data may enable therapists to better respond and modulate clients' affective flexibility; however, further research is necessary to determine whether there is a causal link between affective temporal variability and psychotherapy outcomes.

Keywords: affective dynamics, facial expression, vocal arousal, major depressive disorder, affective flexibility

Public Health Significance Statement

The current findings highlight the potential of computerized facial expression and vocal analyses to capture moment-by-moment processes in psychotherapy sessions. The results suggest that clients' affective flexibility, which in this project was measured by depressed clients' capacity to dynamically shift their emotional states and their arousal levels from one moment to the next, was associated with better session outcomes. The findings support therapeutic interventions aimed at enhancing clients' affective flexibility. Future work is needed to determine the causal links between affective flexibility in depression and outcomes.

Multimodal Analysis of Temporal Affective Variability within Treatment for Depression

Affective flexibility, the capacity to dynamically modulate one's emotional responses from one moment to the next, represents a fundamental component of psychological well-being (Kuppens et al., 2010). Greater flexibility is considered to be indicative of increased sensitivity and adaptiveness to external factors such as environmental changes, and internal factors including regulatory processes. Lower levels of affective flexibility (or rigidness) are considered to underlie mood disorders, and depression in particular (Koval et al., 2012). People who suffer from depression often tend to persist in maintaining negative experiences or stimuli, and avoid new experiences (Kashdan & Rottenberg, 2010; Nolen-Hoeksema et al., 2008).

Many psychotherapy models (e.g., Fonagy et al., 2018; Fosha, 2001; Greenberg, 2012; Hayes et al., 2011; McCullough & Magill, 2009) view increases in clients' emotional flexibility as a key target for intervention. In these approaches, psychotherapy aims at helping clients augment their ability to experience a broader range of positive and negative feelings, rather than numb or mute feelings (e.g., Fosha, 2001; Greenberg, 2012). Empirical support for this theoretical stance has grown considerably in recent decades, based on evidence that greater emotional flexibility is linked to outcomes such as fewer depressive symptoms (Beshai et al., 2018), as well as better quality of experience (Pascual-Leone & Greenberg, 2007; Pascual-Leone, 2009), and functioning (Bar-Kalifa & Atzil-Slonim, 2020).

Since affect fluctuates and varies continuously over time, the study of affect dynamics depends on a fine-grained assessment involving continuous or repeated measures (Kuppens & Verduyn, 2017). To date, however, the empirical literature on affective flexibility in psychotherapy has relied almost exclusively on clients' end-of-session self-reported emotions (e.g., Bar-Kalifa & Atzil-Slonim, 2020). Although self-reports provide important information about participants' subjective experiences, they also have critical shortcomings, especially with regard to evaluating affect dynamics. One-time data of this type fail to capture the high-resolution dynamic nature of affect, and therefore cannot represent the extent to which individuals' affect shifts flexibly (or rigidly) from one moment to the next within therapeutic sessions (Kuppens, 2015). By only utilizing subjective reports, they may miss highly informative (and potentially less biased) sources of information obtained from observed measures of affect. Specifically, people tend to express their affective states through different, easily observable communicative channels. Whereas the voice conveys the arousal of affective states, the face may more effectively convey the valence of emotions (e.g., Bustamante et al., 2015). Thus, even in the absence of continuous self-reported affect, multimodal session recordings provide a rich stream of information on affective valence and arousal.

This ties in with another limitation of many studies on emotional dynamics in psychotherapy that are restricted to a single dimension of affective space at any one time. Many studies have tended to focus either on the valence (e.g., Bar-Kalifa & Atzil-Slonim, 2020) or the arousal (e.g., Mundt et al., 2012) of emotional experiences, despite the broad consensus that affect is best conceptualized as two-dimensional or more (e.g., Jacobson et al., 2021; Russell, 2003; Watson & Tellegen, 1985). For example, Russell's (1980) widely used circumplex model posits that affective states can be situated in a roughly circular organization in a two-dimensional space defined by a positive-negative *valence* dimension and a high-low *arousal* dimension (see also Russell, 2003). Recent discussions on emotional flexibility and affect dynamics have highlighted the importance of examining these qualities within their multidimensional space by for instance considering valence and arousal jointly (e.g., Kuppens et al., 2013).

To address these limitations, we (a) examined affective states within psychotherapy dynamically (i.e., moment to moment within sessions) rather than statically (b) using observed indices rather than subjective self-reports, and (c) exploring these dynamics while attending to more than a single dimension (i.e., valence *or* arousal). We used continuous recordings of psychotherapy sessions (vocal qualities and facial expressions) to examine the dynamics in client affect within the two-dimensional space of valence *and* arousal.

From Facial Expressions and Vocal Analysis to Multimodal Affective Dynamics

Determining Valence from Facial Expressions

Facial expressions are a primary non-verbal emotional communication channel (Cohn et al., 2007; Girard et al., 2013). Recent advances in technology now enable the automatic evaluation of affect, which naturally varies frequently during therapy sessions, from high-quality recordings of visual data. The most common computerized method for extracting emotional information from facial expressions is the *facial action coding system* (FACS; Ekman & Friesen, 1976). FACS describes facial activity in terms of anatomically based action units. Complete facial expressions associated with specific emotions or valence are captured by combinations of FACS codes (Freitas-Magalhães, 2019).

More negative valence in facial emotion expressions is a characteristic of depressed individuals (for a review see Nasser et al., 2020). Although several studies (e.g., Gavrilescu & Vizireanu, 2019; Wang et al., 2018) have differentiated between depressed clients and healthy controls using facial expression valence (FEV), they have all relied on one-time, cross-sectional data. Studies that have made multiple assessments of facial expressions have tended to focus on single expressions (e.g., smiles or contempt; Girard et al., 2013) and/or analyzed video data recorded outside psychotherapy sessions (e.g., before and after deep brain stimulation; Harati et al., 2016). Facial expressions constitute a continuous, unobtrusive, and implicit measure that can capture high-resolution changes in emotional expression, and hence are highly suitable for assessing affective valence dynamics within psychotherapy sessions. Previous studies have demonstrated the usefulness of these measures in assessing psychological states automatically (e.g., Girard & Cohn, 2015; Hu et al., 2021). However, to the best of our knowledge, no study has examined whether the facial expression channel is sensitive to or predictive of within-individual changes in psychotherapy outcomes.

Determining Arousal from Vocal Analysis

The voice serves as a primary nonverbal channel for emotional communication. Two prosodic speech features, specifically vocal pitch (i.e., the auditory perception of tone, indexed using fundamental frequency [f_0]) and intensity (i.e., loudness or volume), have been extensively studied in relation to affective arousal (Cummins et al., 2015; for an evolutionary background to the emergence of vocal features in affective arousal, see also Filippi et al., 2017). In particular, baseline f_0 and deviations from this baseline are strongly correlated with self-reported and physiological indicators of affective arousal (heart rate, blood pressure, and cortisol secretion; Juslin & Scherer, 2005).

Starting with Rice's (1967) pioneering ideas, researchers have used speech- and voice-related measures to study psychotherapy processes. Technological innovations have led to a sharp increase in these studies in recent years (e.g., Baucom et al., 2015; Fischer et al., 2022; Tomicic et al., 2015; Weber et al., 2022). Several vocal features have been explored in psychopathology studies, in particular ones that focus on depression, and include vocal arousal (VA) prosody, level, and variability (Alpert et al., 2001; Yang, Fairbairn & Cohn, 2012), VA range (Breznitz, 1992), and speech-rate and pause variability (Mundt et al., 2012).

Recent work on the analysis of VA suggests that a combination of several features, rather than the f_0 (i.e., a measure of vocal pitch) alone, may reflect human affective arousal more accurately (Bone et al., 2014a; Chaspari et al., 2017). An index combining intensity and pitch was found to be more reliable than separate indices of intensity and pitch (Bone et al., 2014a, 2014b; Chaspari et al., 2017). However, to date, most studies of VA in depression have utilized only one vocal feature (f_0) and have relied on vocal data from single sessions. In addition, to the best of our knowledge, although vocal features have been studied as predictors of affective arousal, no works have investigated VA dynamics within clients and across treatments or explore these dynamics' associations with treatment outcomes.

Combining Facial Expressions and Vocal Analysis

Affective valence as portrayed in facial expressions and affective arousal detected in vocal analyses can be studied in isolation; however, much can be gained by combining the two modalities. Considerable work has recently been devoted to this effort (for a review of multimodal affective computing, see Poria et al., 2017). For example, in the assessment of depression, several studies have shown that two or more modalities outperform any single modality in identifying depression severity levels (e.g., Cohn et al., 2009; Cummins et al., 2013; Dibeklioğlu et al., 2018; Joshi et al., 2013; Scherer et al., 2014).

Multidimensional Affective Dynamics within Psychotherapy Sessions

The continuous recordings of vocal qualities (as an index of affective arousal) and facial expressions (as an index of affective valence) makes it possible to conduct dynamic (rather than static) and multidimensional examinations of affective states within psychotherapy. Key features that reflect affect dynamics over time are variability and inertia (Kuppens et al., 2010). *Variability* measures a client's movement across the affect spectrum; for example, the shift from negative valence to less negative valence within the same session. This is typically operationalized using the standard deviation (SD) metric, which describes the magnitude of deviations from the mean level. *Inertia* represents a client's tendency to retain a certain level of affect from moment to moment; for example, persisting in negative valenced affect from one moment to the next during significant parts of the session. This is typically operationalized using the autoregression (AR) metric, which refers to the extent to which affect is predictable from moment to moment. Both metrics have proven useful in assessing the affect dynamics of one-dimensional data over time. Jahng et al. (2008) introduced the mean square of successive difference (MSSD) as a composite measure of affect dynamics. This metric accounts for both variability and inertia (see also Bos et al., 2019). It addresses several limitations of previous measures (i.e., self-reports) by capturing affective states dynamically (within sessions), is derived from observed measures, and also allows the integration of more than one observed measure (i.e., more than a single dimension). This measure has been shown to capture moment-to-moment fluctuations (e.g., Lazarus et al., 2019; Snir et al., 2017; Trull et al., 2008). Importantly, few studies have used MSSD with *multivariate* affect data (cf. Krone et al., 2018; for an example of multivariate emotion-network density analysis, see Pe et al., 2015), and none, to our knowledge, have used non-self-reported indices recorded continuously.

MSSD measures moment-to-moment variability or fluctuations, and thus, can be considered an index of temporal variability. This index may reflect a positive psychological phenomenon (i.e., flexibility) or a negative one (i.e., instability). On the one hand, Bos et al. (2019; see also Koval et al., 2016) suggested that at shorter time scales (milliseconds to minutes), temporal variability may be indicative of the ability to adapt one's affective system to environmental changes (Carver, 2015; Frijda & Mesquita, 1994; Kashdan & Rottenberg, 2010; Koval et al., 2016; Panksepp, 2012). On the other, on longer time scales (hours to weeks), temporal variability may indicate affect or mood swings associated with negative psychological states (Houben et al., 2015; Koval et al., 2013; Trull et al., 2015). Since our data involved affect that was measured at short (millisecond to second) intervals within therapy sessions, we predicted that our index of temporal variability would reflect clients' responses to interactions with the therapist, and thus would correlate with improved session outcomes.

The Current Study

Our objective was to harness technological advancements, by utilizing vocal and facial expression analysis tools to investigate high-resolution two-dimensional affective temporal variability (operationalized using the 2D-MSSD index) displayed by clients during their treatment for depression. Specifically, we examined how temporal variability related to session outcomes and the overall outcome trajectories over the course of treatment for these clients.

We hypothesized that greater client affective temporal variability would be associated with greater pre-to-post session change in clients' well-being (**Hypothesis 1a: within client**). In addition, we hypothesized that clients with higher affective temporal variability on average would show a greater average improvement in well-being across treatments (**Hypothesis 1b: between clients**). We also expected these associations to hold beyond affective valence. Furthermore, in terms of the association of clients' affective temporal variability on outcome trajectory (the slope of the outcome variable) over the course of treatment, we hypothesized that overall, sessions characterized by greater client affective temporal variability would be marked by greater well-being (**Hypothesis 2: outcome trajectory**).

Method

Participants and Treatment

Clients

In total, 178 candidate participants were screened using the Beck Depression Inventory II (BDI-II; Beck et al., 1996). Of this cohort, 64 individuals with BDI-II scores ≥17 were asked to participate in an intake interview, during which the Mini-International Neuropsychiatric Interview version 5.0 (MINI; Sheehan, 1998) was administered. The inclusion criteria were: (a) a primary diagnosis of major depressive disorder (MDD) as indicated by the MINI and (b) aged 18–67 years. The exclusion criteria were: (a) active suicidality, (b) substance abuse or dependence, (c) current or past bipolar disorder, (d) presence of psychotic features, (e) past severe head injury, (f) pending legal proceedings, and (g) current pregnancy or medical condition warranting hormonal treatment. Out of the initial 35 clients who began treatment, one client withdrew, and another client discontinued therapy before the second session. Three other clients opted to take psychiatric medication and were excluded from the analysis. The final cohort was thus composed of 30 clients (19 men) diagnosed with MDD, with a mean age of 34.63 years (standard deviation [SD]=9.27; range: 21–59 years). Fourteen participants were single, 16 were married or in a long-term relationship, 23 had at least a bachelor’s degree, and all but two were fully or partially employed. Twenty-three clients were native Hebrew speakers, and all the clients indicated that they were Jewish. The clients’ mean BDI-II score at intake was 22.5 (SD=7.75), indicating moderate levels of depression (Beck et al., 1996).

Therapists

Nine therapists (five women) participated in the study (age: mean=33.1; range: 30–41); four therapists treated three to four clients each, whereas the five remaining therapists treated one to two clients each. The therapists were advanced trainees in a university clinic with three to seven years of clinical experience.

Treatments

The clients underwent brief (16 sessions) supportive-expressive psychodynamic therapy (SET; Luborsky & Mark, 1991) adapted for the treatment of depression (Luborsky et al., 1995) primarily implementing supportive techniques such as affirmation and empathic validation, and expressive techniques such as interpretation and confrontation. SET has been reported to be effective in treating depression (Beck et al., 1996; Sheehan, 1998). The therapists were trained and supervised by senior clinicians with extensive expertise in SET and underwent weekly individual and group supervision. All the therapists were native Hebrew speakers, and all were Jewish.

Procedure

The study was conducted in Bar-Ilan University’s out-patient clinic and approved by the associated institutional review board. Data were obtained as part of the routine monitoring used in the clinic. Clients consented to participate voluntarily. They were told they could terminate their participation at any time with no effect on their treatment. The clients completed outcome rating scale (ORS) questionnaires electronically (using computers located in the clinic rooms) before and after each therapy session.

The therapy sessions were video and audio recorded using two cameras and two microphones, with a camera and microphone directed at each speaker. The original audio of the sessions' *working phases* (i.e., the 15 minutes ending five minutes before the end of the session; Watson et al., 2011)[1] were segmented into inter-pausal units (IPU; Levitan & Hirschberg, 2011; for details, see the Data Analysis section) using an automatic speech diarization algorithm (to determine who is speaking at any given moment, allowing the isolation of each person's vocal dynamics) explicitly developed for the imbalanced nature of psychotherapy conversations where clients often speak for extended periods, and therapists frequently respond with shorter utterances. To address imbalance, we used an algorithm based on previous work on speech diarization and separation (Laufer-Goldshtein et al., 2018a, 2018b).[2] The mean VA and FEV were calculated for each client's IPUs. Across the 30 therapy dyads and 137 available sessions, 22,741 mean VA and FEV observations were extracted with an average of M =185 samples per session (SD = 51; Range = [50:294]).[3]

Measures

Outcome Rating Scale (ORS)

The ORS (Miller et al., 2003) is a four-item SBS measure developed as a short-form alternative to the Outcome Questionnaire-45 (Lambert et al., 1996). The ORS is designed to

---

[1] *Working phase* corresponds to the part of the session during which clients are likely to be the most engaged in therapeutic work (Auszra et al., 2013).
[2] Additional information and details on the speech diarization method can be found in the online supplementary material (OSM). https://osf.io/h76r8/?view_only=4f0df1004dac4ff19b85b7364f972d30
[3] The diarization algorithm and vocal arousal extraction were implemented using MATLAB (Version 2019a). The vocal features were extracted using Praat software (Boersma & Weenink, 2017). https://osf.io/h76r8/?view_only=4f0df1004dac4ff19b85b7364f972d30

track client progress session by session. The scale assesses change in three domains of client functioning that are widely considered to be valid indicators of progress in treatment and successful outcomes: individual functioning, interpersonal relationships, and social role performance. The clients completed the ORS by rating items on a visual analog scale anchored at one end by the words “low” and “high.” The items ranged from 0 to 10, with higher scores indicating better functioning. The ORS was completed twice in each session: immediately before and immediately after each session. The pre-to-post ORS change (ORS diff) was calculated as the pre-session ORS subtracted from the post-session ORS. As in previous studies (e.g., Chen, et al., 2018; Fisher et al., 2016; Paz et al., 2021), the ORS was considered an indicator of the client session outcomes. The correlation between the general ORS score and a single item asking clients to rate overall well-being on a continuous scale ranging 0 to10 was $r = 0.94$ ($p < 0.001$).

Vocal arousal (VA)

Following Bone et al. (2014a), VA was computed as a weighted average index of three speech features: (1) intensity, (2) pitch, and (3) HF500 (ratio of energy above 500 Hz divided by the energy between 80 and 500 Hz). A higher VA was defined as vocalizations that were louder, higher in pitch, or harsher and crisper than typical speech. These features were normalized for each speaker for each session, allowing each feature's average level to act as the speaker’s baseline. The final VA score was calculated from the weighted average of the three feature scores. This measure achieved state-of-the-art performance in a cross-corpus automatic arousal recognition competition (Valstar et al., 2016). A total of 22,741 VA observations were obtained (M=0.01; SD=0.43).

Facial expression valence (FEV)

FaceReader-9 was used to extract emotional responses from the video recordings. Based on the circumplex model of affect (Russell, 1980), *valence* was measured on a continuous scale from −1 (most negative emotion) to +1 (most positive emotion), captured at a rate of 25 frames per second (40 ms per sample). The 40 ms samples were aggregated into an average for each speech turn (e.g., Kuppens et al., 2010; Ogbaselase et al., 2020). The validity of FaceReader's affective measures was found to be comparable to external professional annotators (see Skiendziel et al. [2019] who tested for the convergence validity of FaceReader-7). In total, 22,741 FEV observations were obtained (M=−0.05; SD=0.27).

Data Analysis

The basic temporal unit of analysis was the inter-pausal-unit (IPU; Levitan & Hirschberg, 2011); in other words, parts of speech-turns demarcated by pauses lasting at least 50 ms, which themselves are pause-free (i.e., interrupted, at most, by pauses lasting less than 50 ms). In total, 22,741 vocal arousal (VA) and facial expression valence (FEV) data points were extracted from each IPU. These data were collected from 137 sessions, with an average of M =185 samples per session (SD = 51; Range = [50:294]). These served as the two dimensions in a Cartesian arousal-by-valence space; affective temporal variability was measured as the session level dispersion of IPUs in this space and was calculated as a two-dimensional MSSD (2D-MSSD ; Trull et al., 2008; see equation 1). This approach was utilized to quantify the variability of the time-dependent affective two-dimensional movement (vocal arousal, and facial expression valence) from one IPU measurement to the next (Hamaker et al., 2015).

$$2dMSSD\_affect_{sc} = \sqrt{\frac{1}{n} \cdot \sum_{i=2}^{i=MAX(IPU\#)} ((VA_i - VA_{i-1})^2 + (FEV_i - FEV_{i-1})^2)} \quad (1)$$

Statistical model

Given the nested nature of the data (with session [level 1] ratings nested within clients [level 2]), we utilized multilevel modeling (MLM; Raudenbush & Bryk, 2002). For hypothesis 1a and hypothesis 1b, MLM model 1 estimated the association between affective temporal variability (2D-MSSD) and average session outcomes, which were operationalized as changes in the clients' ORS scores pre- and post-session. The model was adjusted for mean FEV:

Level 1

$$ORSdiff_{sc} = \beta_{0sc} + \beta_{1sc} * Session\_FEV_{sc} + \beta_{2sc} * Session\_2dMSSD\_affect_{sc} + e_{sc}$$

$$(\mathrm{e}_{sc}) \sim \mathrm{N}[0, \sigma^2]$$

Level 2

$$\beta_{0sc} = \gamma_{00} + \gamma_{01} * Client\_FEV_{0c} + \gamma_{02} * Client\_2dMSSD\_affect_{0c} + u_{0c}$$

$$\beta_{1sc} = \gamma_{10} \ ; \beta_{2sc} = \gamma_{20}$$

$$(\mathrm{u}_{0c}) \sim \mathrm{N}[0, \tau_{00}{}^2]$$

Session outcomes were estimated using an intercept (γ_{00}) term and two slope terms for session-level mean valence (γ_{10}) and session-level affective temporal variability (γ_{20}). The latter served as a test for the within-client (level 1) hypotheses. In addition, client-level (i.e., level 2) mean FEV (γ_{01}) and mean affective temporal variability (γ_{02}) were included; the latter served as a test of the between-client (level 2) hypotheses. Level 1 residuals (e_{sc}) and level 2 random effects for the intercept (u_{0c}) were also estimated.[4]

[4] Only the intercept term was estimated as a random effect, since estimating the slope terms as random effects did not improve the model fit (*χ^2* [5] = 7.89, *p* = 0.165).

For hypothesis 2, the MLM model 2 estimated the moderation effect of clients' affective temporal variability session outcome trajectory over the course of treatment, which was operationalized as the slope of the clients' post-ORS scores across treatment. The growth model was adjusted for the clients' mean affective temporal variability:

Level 1

$$ORS\_post\_session_{sc} = \beta_{0sc} + \beta_{1sc} * Session\_num_{sc} + \beta_{2sc} * Session\_2dMSSD\_affect_{sc} + \beta_{3sc} * Session\_num_{sc} * Session\_2dMSSD\_affect_{sc} + e_{sc}$$

$$(e_{sc}) \sim N[0, \sigma^2]$$

Level 2

$$\beta_{0sc} = \gamma_{00} + \gamma_{01} * Client\_2dMSSD\_affect_{0c} + u_{0c}$$

$$\beta_{1sc} = \gamma_{10} + \gamma_{11} * Client\_2dMSSD\_affect_{0c} + u_{1c}$$

$$\beta_{2sc} = \gamma_{20};\ \beta_{3sc} = \gamma_{30}$$

$$(u_{0c}, u_{1c}) \sim N\left(0, \begin{bmatrix} {\tau_{00}}^2 & \tau_{00} * \tau_{10} \\ \tau_{10} * \tau_{00} & {\tau_{10}}^2 \end{bmatrix}\right)$$

Post session outcomes (ORS measures) were estimated using an intercept (γ_{00}) term and three slope terms: one for outcome trajectory (γ_{10}), second for the session-level affective temporal variability moderation of the outcome trajectory (γ_{30}), and a third term controlling for the main effect of session-level affective temporal variability (γ_{20}). In addition, client-level (i.e., level 2) mean affective temporal variability (γ_{02}) was included. Level 1 residuals (e_{sc}) and level 2 random effects for the intercept (u_{0c}) and outcome trajectory (u_{1c}) were also estimated.[5]

[5] Only the intercept term and the slope term for session number (Session_num) were estimated as a random effects, since estimating the remainder of the slope terms as random effects did not improve the model fit (*χ^2* [7] = 3.69, *p* = 0.815).

Results

The hypotheses were a-priori but not preregistered. The data and the study analysis code are available upon request from the first author. The descriptive statistics for the study variables are presented in Table 1.

[Insert Table 1 about here]

Table 2 presents the model 1 results predicting session outcomes (i.e., pre-to-post session ORS differences) from clients' mean affective temporal variability while adjusting for FEV. In line with the hypotheses, session outcome was associated with temporal variability at a session level (i.e., Hypothesis 1a: within clients, between sessions) and at the client level (i.e., Hypothesis 1b: between clients). Importantly, these associations remained robust even after accounting for both session- and client-level mean valence scores in the model. Calculation of the effect sizes yielded a standardized estimation of 0.27 for within- clients' effects and a standardized estimation of 0.26 for between- clients' effects. In other words, clients who experienced more affective temporal variability tended to show greater improvement in their pre- to post-session ORS scores. In addition, within clients, sessions characterized by greater affective temporal variability were also characterized by greater improvement in ORS scores.[6]

[Insert Table 2 about here]

The following examples of greater and lesser temporal variability patterns illustrate these data. The two clients were chosen to represent good and poor responses. Client I's well-

[6] In response to the editor's request, we conducted an additional analysis comparing the unified two-dimensional MSSD presented in this paper to separate univariate MSSDs for vocal arousal and facial expression valence, as well as a model including both univariate MSSDs. The results demonstrated that the unified model had a superior fit in terms of fixed effects than each of the univariate MSSD models. Interestingly, the univariate model assessing MSSD from the arousal signal alone explained nearly as much variance as the unified model. The model incorporating both separate univariate MSSDs for arousal and valence explained slightly more variance than the unified two-dimensional MSSD model. A detailed description and discussion of these analyses and findings can be found in the online supplementary material (OSM): https://osf.io/h76r8/?view_only=4f0df1004dac4ff19b85b7364f972d30.

being tended to improve within each session [ORS difference from pre- to post-session: Mean=0.50; SD=0.41], whereas client II's well-being tended to decline [ORS difference: Mean=-0.48; SD=1.06]). Figure 1 shows the affect temporal variability patterns observed in the working phase of 5 of each client's analyzed sessions. The x-axis depicts the valence extracted from the client's facial expressions, and the y-axis represents the arousal extracted from the client's voice recordings. The darker purple arrows correspond to measures from the beginning of the working phases, and the lighter yellow arrows show the latter parts of the working phases.

Client I's average temporal variability across the 5 sessions (Mean=0.62; SD=0.016) was greater than client II's temporal variability (Mean=0.46; SD=0.44). In terms of within-client (i.e., session-level) variability, client II's session d was characterized by greater temporal variability (and improvement in well-being; affective temporal variability = 0.48; ORS difference = 0.4), whereas session e was characterized by less temporal variability (and a decline in well-being; affective temporal variability = 0.41; ORS difference = −2.1).

[Insert Figure 1 here]

Table 3 presents the model 2 results predicting the session outcome trajectory (i.e., post-session ORS) moderation by affective temporal variability while adjusting for clients' overall mean affective temporal variability. In line with the hypotheses, the session outcome trajectory was steeper when affective temporal variability increased throughout therapy. Importantly, this moderation remained robust even when accounting for clients' mean temporal variability across treatment. In other words, clients who exhibited more of a linear growth trend in their temporal variability also exhibited a steeper positive trajectory in their outcome measures on average throughout their treatment. This effect was robust even when accounting the effect of clients' mean temporal variability.

[Insert Table 3 about here]

Discussion

The goal of the present study was to determine whether multidimensional affective temporal variability (i.e., 2D-MSSD of affective arousal and valence) when assessed multimodally (i.e., by examining voice and facial expression data) was associated with positive changes in psychotherapy at both the session and client levels. Data came from sessions drawn from 30 therapeutic dyads in which clients suffering from depression underwent brief SET (Luborsky & Mark, 1991).

In line with hypothesis 1a, the results indicated that sessions characterized by greater affective temporal variability (i.e., greater 2-dimensional successive variability in vocal arousal and valence measured by facial expressions) were associated with greater improvement in clients' self-reported pre-session to post-session well-being. In addition, and in line with hypothesis 1b, *clients* who experienced more affective temporal variability tended to show greater overall session level improvement in their well-being across treatment. These associations remained significant even when accounting for mean affective valence scores at both the session and client levels.

In addition, in line with hypothesis 2, the session outcome trajectory throughout treatment was moderated by clients' affective temporal variability. Clients who exhibited more affective temporal variability throughout treatment (i.e., sessions in the latter parts of treatment were also characterized by greater affective temporal variability) were also characterized by steeper linear improvement in their outcome trajectory across treatment. Overall, these results indicated session and treatment level associations between affective temporal variability and outcome. Nevertheless, it's important to note that the findings for between-client (hypothesis 1b) and treatment-level trajectory (hypothesis 2) should be approached with caution due to the relatively small sample size of 30 clients.

These results strengthen previous findings that have linked affective temporal variability with greater well-being, in general (Coifman & Summers, 2019; Hollenstein, 2015; Kashdan & Rottenberg, 2010) and specifically in individuals suffering from depression (e.g., Koval et al., 2012). They are in line with various emotion theories which argue that temporal variability may reflect healthy flexibility, as it enables adaptive responses to environmental demands (Carver, 2015; Frijda & Mesquita, 1994; Kashdan & Rottenberg, 2010; Panksepp, 2012). This may be particularly true for temporal variability in the micro timescale (i.e., milliseconds to seconds; see Bos et al. [2019] & Koval et al., [2016]).

The results are also consistent with psychotherapy studies that have noted the importance of temporal variability as a predictor of treatment outcomes (e.g., Bar-Kalifa & Atzil-Slonim, 2020; Fisher & Newman, 2016; Pascual-Leone, 2009). However, the diverse range of resolution from between-session dynamics to various levels of within-session dynamics, as well as the array of methodologies employed to assess temporal variability (self-reports at the therapy or session level, complex systems theory methodologies, and external raters), may present challenges in directly extending the results of our project to build upon these prior findings.

The results also support psychotherapy theories that emphasize the central role of inflexible psychological patterns, and particularly affective/emotional rigidity, in psychopathology, such as EFT (Greenberg, 2012), ACT (Hayes et al., 2011), and psychodynamic theories (Fonagy et al., 2018; Fosha, 2001; McCullough & Magill, 2009).

The present study goes beyond these studies by examining temporal variability as multimodal rather than unimodal—as strongly advocated by affect scholars (e.g., Kuppens et al., 2013). A multimodal view is also consonant with the fact that humans communicate through more than one expressive channel. The examination of multimodal expressive data—in our case, vocal and facial expressions—draws on improved techniques for the monitoring of bio-behavioral signals, and has become quite common in affect research. To date, such methods have been used most extensively in the prediction of psychopathology (for a review, see Poria et al., 2017; for applications specific to depression, see Alghowinem et al., 2016; Dibeklioğlu et al., 2015). The results of this project highlight the advantages of using this multimodal approach to capture affect dynamics in psychotherapy research.

Specifically, the collection of multimodal data made it possible to assess affective valence and affective arousal, and examine temporal variability within a two-dimensional affective space. In the context of psychotherapy, clients' affective flexibility can manifest in several ways. Greater temporal variability in the valence dimension may signify clients' flexible capacity to experience a broader spectrum of emotional valence. This can include the ability to traverse a wide range of emotions, such as transitioning from negative to positive emotions, or shifting between secondary and primary negative emotions, as proposed by Greenberg (2012). At the same time, greater temporal variability in the arousal dimension may reflect greater ability to sustain and regulate emotional experiences. For example, clients who experiences sadness with a high level of arousal during a session may be able to downregulate their emotional state to a lower level of arousal while still experiencing sadness. This ability to experience a wider range of emotions combined with the ability to adaptively regulate emotional experiences may allow clients to better communicate and handle their emotions which is one of the main goals of therapy.

To date, most studies on temporal variability have explored it with respect to one dimension of affect, namely, valence or arousal (e.g., Carryer & Greenberg, 2010; Nolen-Hoeksema et al., 2008). Using both affect dimensions allows for the exploration of individuals' ability to both generate or amplify a *range of emotions* as well as to inhibit or dampen them, a capacity that may promote flexible adaptation to challenges as well as routine or daily functioning.

Limitations and Future Directions

Our data suggest that multi-dimensional temporal variability may play a salubrious role in terms of client well-being. This role may be causal in nature, and may suggest that temporal variability fosters well-being by allowing clients to respond flexibly to life circumstances (i.e., to show regulatory flexibility; e.g., Zhu & Bonanno, 2017). However, such causal claims may be premature. One way to explore this issue further would be to use multi-session psychotherapy data and additional measures (unfortunately not available in the present study) for both affective and regulatory flexibility.

If multidimensional temporal variability in affect is indeed causally tied (directly or indirectly) to treatment outcomes, future research will need to explore what facilitates it. In particular, alongside personality (e.g., Rademacher et al., 2022) and contextual factors (Godara et al., 2020), which have been shown to predict temporal variability (albeit unidimensionally), various clinical models offer both theoretical and (some) empirical ideas about altering it. In particular, affective flexibility has been discussed as an outcome of particular interventions that increase cognitive defusion (e.g., ACT: Hayes et al., 2011), work with within-person dialogues (e.g., EFT: Greenberg, 2012; self-compassion therapy: Gilbert, 2009; and schema therapy: Rafaeli et al., 2010), and rely on explicit or implicit models that highlight within-person multiplicity (Lazarus & Rafaeli, 2023).

Future work could fruitfully use the methods presented in the present work with within-session data drawn from these or other modalities to explore which factors contribute to the emergence of adaptive affective dynamics. For example, it would be interesting to use natural language processing (NLP) techniques to automatically extract therapists' interventions speech turn-by-speech turn and examine which intervention sequences enhance clients' affective flexibility (e.g., Warikoo et al., 2022).

While temporal variability was operationalized using successive changes in facial and vocal affect, it is worth considering that fluctuations in affective channels may reflect other processes. Complex systems theory, for example, suggests that affect often exhibits stable states that are occasionally interrupted by less common departures from these states. These departures are typically preceded by *destabilization* of the system, as indicated by increased variability (Hollenstein, 2015; Lewis, 2000; Olthof et al., 2023). Thus, temporal variability may reflect destabilization rather than flexibility, as suggested in studies such as Lichtwarck-Aschoff et al. (2012).

Our results, particularly those related to Hypothesis 2, provide some insights into this question. Specifically, while the results related to Hypotheses 1a and 1b demonstrated an association between temporal variability and session outcomes, the results pertaining to Hypothesis 2 help further orient the interpretation of our temporal variability index towards the flexibility rather than the destabilization interpretation. The interaction effect in Hypothesis 2 indicates that while destabilization could co-occur, affective temporal variability indeed increased throughout treatment and was linked to improved outcomes (which is also in line with the findings in Fisher & Newman [2016]). Importantly, these findings are preliminary and require replication; It is important to note that we did not identify specific states, their distribution, or state shifts in a manner that could answer this question. This opens up a potential avenue for exploration in future studies (see for example Cui & Lichtwarck-Aschoff, 2023).

Various clinical models posit that *interpersonal* (i.e., therapist-client) dynamics may enhance *intrapersonal* flexibility. For example, Fonagy et al. (2018) suggested that clients internalize affective regulation abilities in ways that are similar to those in which children internalize them from their caregivers (see also Feldman, 2015; Fosha, 2001; Wright et al., 2023). Recent theories (Koole & Tschacher, 2016) and meta-analytic reviews (Atzil-Slonim et al., in press) have begun to delineate the conditions in which dyadic co-regulation exerts these effects on individual affect dynamics. For example, in a recent study, Paz et al. (2021) related interpersonal co-regulation to session outcomes and clients' ability to internalize intrapersonal arousal regulation capability throughout treatment.

This points to another limitation of this study which focused on examining clients' affect dynamics but did not explore the potential role of therapists' affective flexibility in facilitating their clients' development of affective flexibility. Previous findings have highlighted the therapeutic gain from therapists' dynamic emotional responsiveness to their clients' changing needs (e.g., Lazarus et al., 2019; Stiles, 2009). In a recent paper, Abargil and Tishby (2021) showed that greater variation in therapist-experienced emotions was linked to therapy outcomes. These findings may encourage researchers to examine whether therapists' affective flexibility influences clients' affective flexibility and vice-versa.

A further limitation of this study has to do with its focus on the vocal and facial (communicative) aspects of affect. Alongside the advantages of bio-behavioral data such as the availability of high-resolution recordable signals that do not require participant self-reporting, they have several drawbacks. One is associated with the non-continuous nature of speech, which creates "blind spots" when assessing clients' affective states during moments without speech. This also poses challenges for conducting consecutive moment-to-moment time-dependent analyses, including the MSSD employed in this study. Additionally, temporal variability in affect may operate differently at levels that are more or less explicit (e.g., Cisler et al., 2010). Future exploration of affective flexibility within psychotherapy should integrate these levels and explore temporal variability across many modalities in tandem.

The generalizability of our findings is constrained by three main issues. First, this study used psychotherapy data in individuals experiencing major depression, who received a specific form of psychotherapy treatment (SET; Luborsky & Mark, 1991). Although previous research has highlighted the relevance of temporal variability in affect to a broad spectrum of psychopathological conditions and well-being parameters (e.g., Kashdan & Rottenberg, 2010), applying our findings to other populations and alternative therapeutic approaches necessitates further replication. Second, the client sample size (N = 30), was relatively small. This may have led to underpowered analyses, making it difficult to detect between-dyad effects accurately. Therefore, conducting replications with a more substantial client sample is imperative to enhance the robustness and generalizability of our results. Third, our psychotherapy data was in Hebrew and the sample was relatively homogeneous, since it mostly consisted of native Hebrew speakers. This may limit the generalizability of the results to other languages and cultures. Fourth, the ORS measure (Miller et al., 2003), while previously validated for assessing clients' session-level outcomes, is applied less often to calculate changes in well-being during a session (the ORS difference) by measuring pre-post session shifts (see Paz et al., 2021 for an example). Despite its strong correlation with clients' well-being this metric raises questions about the extent of change detectable in clients' well-being or functioning within a psychotherapy session. Its narrow scope may limit the generalizability of the findings. Thus, future research would benefit from the inclusion of additional variables to more thoroughly investigate how affective dynamics relate to session outcomes.

Finally, our analyses focused on temporal variability assessed at the session level. Recent technological advances, such as natural language processing, now permit more fine-grained analyses of processes and outcomes in psychotherapy sessions that can be applied to within-session segments (e.g., Atzil-Slonim et al., 2021; Flemotomos et al., 2022). This allows researchers to capture the micro-levels events that precede or follow changes in temporal variability and show how such sequences are associated to treatment outcomes.

Clinical implications

The continuous assessment of both vocal and facial affective expressions, along with the ability to derive indices of temporal variability in affect from within-session data holds promise for both clients and clinicians in psychotherapy. It encourages therapists to be more attentive to their clients' emotional range, including what can be discerned from nonverbal aspects of speech and facial expressions. This awareness can guide therapists toward interventions aimed at enhancing their clients' emotional and psychological flexibility.

The integration of multimodal assessments of affective temporal variability into existing feedback systems can furnish therapists with valuable insights into within session affect dynamics. This includes identifying moments when temporal variability in affect diminishes or increases. This information can significantly enhance the precision and tailoring of therapeutic approaches and interventions.

References

Abargil, M., & Tishby, O. (2021). Fluctuations in therapist emotions and their relation to treatment processes and outcomes. *Journal of Psychotherapy Integration*, *31*(1), 1–18. doi:10.1037/int0000205.

Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Hyett, M., Parker, G., & Breakspear, M. (2016). Multimodal depression detection: Fusion analysis of paralinguistic, head pose and eye gaze behaviors. *IEEE Transactions on Affective Computing*, *9*(4), 478–490. doi:10.1109/TAFFC.2016.2634527.

Alpert, M., Pouget, E. R., & Silva, R. R. (2001). Reflections of depression in acoustic measures of the patient’s speech. *Journal of Affective Disorders*, *66*(1), 59–69. doi:10.1016/s0165-0327(00)00335-9.

Atzil-Slonim, D., Juravski, D., Bar-Kalifa, E., Gilboa-Schechtman, E., Tuval-Mashiach, R., Shapira, N., & Goldberg, Y. (2021). Using topic models to identify clients’ functioning levels and alliance ruptures in psychotherapy. *Psychotherapy, 58(*2), *324–339*. doi:10.1037/pst0000362.

Atzil-Slonim, D., Soma, C. S., Zhang, X., Paz, A., & Imel, Z. E. (in press). Facilitating Dyadic Synchrony. In Hill, C. E., & Norcross, J.C (Eds.). *Psychotherapy Skills and Methods that Work*. The Society for the Advancement of Psychotherapy (D29), APA Press.

Auszra, L., Greenberg, L. S., & Herrmann, I. (2013). Client emotional productivity-optimal client in-session emotional processing in experiential therapy. *Psychotherapy Research*, *23*(6), 732–746. doi:10.1080/10503307.2013.816882.

Bar-Kalifa, E., & Atzil-Slonim, D. (2020). Intrapersonal and interpersonal emotional networks and their associations with treatment outcome. *Journal of Counseling Psychology*, *67*(5), 580–594. doi:10.1037/cou0000415.

Baucom, K. J., Baucom, B. R., & Christensen, A. (2015). Changes in dyadic communication during and after integrative and traditional behavioral couple therapy. *Behaviour research and therapy*, *65*, 18-28.

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory–II. *Psychological assessment*.

Beshai, S., Prentice, J. L., & Huang, V. (2018). Building blocks of emotional flexibility: Trait mindfulness and self-compassion are associated with positive and negative mood shifts. *Mindfulness*, *9*(3), 939–948. doi:10.1007/s12671-017-0833-8.

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer (Version 6.0.24) [Computer program]. Retrieved from http://www.praat.org/

Bone, D., Lee, C. C., & Narayanan, S. (2014a). Robust unsupervised arousal rating: A rule-based framework with knowledge-inspired vocal features. *IEEE Transactions on Affective Computing*, 5(2), 201–213. doi:10.1109/TAFFC.2014.2326393

Bone, D., Lee, C. C., Potamianos, A., & Narayanan, S. S. (2014b). An investigation of vocal arousal dynamics in child-psychologist interactions using synchrony measures and a conversation-based model. *Fifteenth Annual Conference of the International Speech Communication Association*, Singapore, September 14–18, 2014. doi:10.21437/Interspeech.2014-56

Bos, E. H., de Jonge, P., & Cox, R. F. (2019). Affective variability in depression: Revisiting the inertia–instability paradox. *British Journal of Psychology*, *110*(4), 814-827.

Breznitz, Z. (1992). Verbal indicators of depression. *Journal of General Psychology*, *119*(4), 351–363. doi:10.1080/00221309.1992.9921178.

Bustamante, P. A., Lopez Celani, N. M., Perez, M. E., & Quintero Montoya, O. L. (2015, August). Recognition and regionalization of emotions in the arousal-valence plane. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE Publications, *2015*, (6042–6045). doi:10.1109/EMBC.2015.7319769.

Carryer, J. R., & Greenberg, L. S. (2010). Optimal levels of emotional arousal in experiential therapy of depression. *Journal of Consulting and Clinical Psychology*, *78*(2), 190–199. doi:10.1037/a0018401.

Carver, C. S. (2015). Control processes, priority management, and affective dynamics. *Emotion Review*, *7*(4), 301-307.

Chaspari, T., Han, S. C., Bone, D., Timmons, A. C., Perrone, L., Margolin, G., & Narayanan, S. S. (2017, March). Quantifying regulation mechanisms in dating couples through a dynamical systems model of acoustic and physiological arousal. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3001-3005). IEEE. Doi: 10.1109/ICASSP.2017.7952707.

Chen, R., Atzil-Slonim, D., Bar-Kalifa, E., Hasson-Ohayon, I., & Refaeli, E. (2018). Therapists' recognition of alliance ruptures as a moderator of change in alliance and symptoms. *Psychotherapy Research*, 28, 560– 570. http://dx.doi.org/10.1080/10503307.2016.1227104

Cisler, J. M., Olatunji, B. O., Feldner, M. T., & Forsyth, J. P. (2010). Emotion regulation and the anxiety disorders: An integrative review. *Journal of Psychopathology and Behavioral Assessment*, *32*(1), 68–82. Doi:10.1007/s10862-009-9161-1.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. In *Handbook of emotion elicitation and assessment* (pp. 203–221). Oxford: Oxford University Press.

Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., . . . and De la Torre, F. (2009, September). Detecting depression from facial actions and vocal prosody. In *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, *2009* (pp. 1–7). IEEE Publications. Doi:10.1109/ACII.2009.5349358.

Coifman, K. G., & Summers, C. B. (2019). Understanding emotion inflexibility in risk for affective disease: Integrating current research and finding a path forward. *Frontiers in psychology*, *10*, 392. Doi: 10.3389/fpsyg.2019.00392

Cui, J., Hasselman, F., & Lichtwarck-Aschoff, A. (2023). Unlocking nonlinear dynamics and multistability from intensive longitudinal data: A novel method. https://doi.org/10.31234/osf.io/wjzg2

Cummins, N., Joshi, J., Dhall, A., Sethu, V., Goecke, R., & Epps, J. (2013). Diagnosis of depression by behavioural signals: A multimodal approach. In *Proceedings of the 3rd ACM international workshop on audio/visual emotion challenge* (pp. 11–20). Doi:10.1145/2512530.2512535.

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10–49. Doi:10.1016/j.specom.2015.03.004.

Dibeklioğlu, H., Hammal, Z., & Cohn, J. F. (2018). Dynamic multimodal measurement of depression severity using deep autoencoding. *IEEE Journal of Biomedical and Health Informatics*, *22*(2), 525–536. Doi:10.1109/JBHI.2017.2676878.

Dibeklioğlu, H., Hammal, Z., Yang, Y., & Cohn, J. F. (2015, November). Multimodal detection of depression in clinical interviews. In. *Proceedings of the ... ACM International Conference on Multimodal Interaction. ICMI, Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, (307–310). doi:10.1145/2818346.2820776.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, *1*, 56-75. doi: 10.1007/bf01115465

Feldman, R. (2015). Mutual influences between child emotion regulation and parent–child reciprocity support development across the first 10 years of life: Implications for developmental psychopathology. *Development and Psychopathology*, *27*(4 pt. 1), 1007–1023. doi:10.1017/S0954579415000656.

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S., Pašukonis, A., Hoeschele, M., Ocklenburg, S., Boer, B. D., Sturdy, C. B., Newen, A., & Güntürkün, O. (2021, October 29). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings of the Royal Society B: Biological Sciences*, 284(1859), 20170990. http://dx.doi.org/10.1098/rspb.2017.0990.

Fischer, M. S., Baucom, D. H., Abramowitz, J. S., & Baucom, B. R. (2022). Interpersonal emotion dynamics in obsessive–compulsive disorder: Associations with symptom severity, accommodation, and treatment outcome. *Couple and Family Psychology: Research and Practice.* Advance online publication. https://doi.org/10.1037/cfp0000218

Fisher, A. J., & Newman, M. G. (2016). Reductions in the diurnal rigidity of anxiety predict treatment outcome in cognitive behavioral therapy for generalized anxiety disorder. *Behaviour research and therapy*, *79*, 46-55. https://doi.org/10.1016/j.brat.2016.02.006.

Fisher, H., Atzil-Slonim, D., Bar-Kalifa, E., Rafaeli, E., & Peri, T. (2016). Emotional experience and alliance contribute to therapeutic change in psychodynamic therapy. *Psychotherapy*, *53*(1), 105.

Flemotomos, N., Martinez, V. R., Chen, Z., Singla, K., Ardulov, V., Peri, R., . . . and Narayanan, S. (2022). Automated evaluation of psychotherapy skills using speech and language technologies. *Behavior Research Methods*, *54*(2), 690–711. doi:10.3758/s13428-021-01623-4.

Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2018), *Affect regulation, mentalization, and the development of the self*. Routledge.

Fosha, D. (2001). The dyadic regulation of affect. *Journal of Clinical Psychology*, *57*(2), 227–242. doi:10.1002/1097-4679(200102)57:2<227::aid-jclp8>3.0.co;2-1.

Freitas-Magalhães, A. (2019). *Micro Expressions-Brain, Face and the Emotion*. Leya.

Frijda, N. H., & Mesquita, B. (1994). The social roles and functions of emotions. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 51–87). American Psychological Association. https://doi.org/10.1037/10152-002

Gavrilescu, M., & Vizireanu, N. (2019). Predicting depression, anxiety, and stress levels from videos using the facial action coding system. *Sensors*, *19*(17), 3693. doi:10.3390/s19173693.

Gilbert, P. (2009). Introducing compassion-focused therapy. *Advances in Psychiatric Treatment*, *15*(3), 199–208. doi:10.1192/apt.bp.107.005264.

Girard, J. M., Cohn, J. F., Mahoor, M. H., Mavadati, S., & Rosenwald, D. P. (2013). Social risk and depression: Evidence from manual and automatic facial expression analysis. *Proceedings of the ... International Conference on Automatic Face and Gesture Recognition. IEEE International Conference on Automatic Face and Gesture Recognition* 10th *IEEE International Conference and* Workshops *on* Automatic Face *and* Gesture Recognition *(FG)*, 1–8, 1–8. doi:10.1109/FG.2013.6553748.

Godara, M., Sanchez-Lopez, A., Liefooghe, B., & De Raedt, R. (2020). Contextual changes influence attention flexibility towards new goals. *Cognitive Therapy and Research*, *44*(2), 327–344. doi:10.1007/s10608-019-10072-5.

Greenberg, L. S. (2012). Emotions, the great captains of our lives: Their role in the process of change in psychotherapy. *American Psychologist*, *67*(8), 697–707. doi:10.1037/a0029858.

Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review*, 7(4), 316-322.

Harati, S., Crowell, A., Mayberg, H., Kong, J., & Nemati, S. (2016, August). Discriminating clinical phases of recovery from major depressive disorder using the dynamics of facial expression. In. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE Publications, *2016*, (2254–2257). doi:10.1109/EMBC.2016.7591178.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2011), *Acceptance and commitment therapy: The process and practice of mindful change*. New York: Guilford Press.

Hollenstein, T. (2015). This Time, It's Real: Affective Flexibility, Time Scales, Feedback Loops, and the Regulation of Emotion. *Emotion Review*, *7*(4), 308–315. doi:10.1177/1754073915590621.

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological bulletin*, *141*(4), 901.

Jacobson, N. C., Evey, K. J., Wright, A. G. C., & Newman, M. G. (2021). Integration of discrete and global structures of affect across three large samples: Specific emotions within-persons and global affect between-persons. *Emotion*. doi:10.1037/emo0001022.

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*(4), 354.

Joshi, J., Goecke, R., Alghowinem, S., Dhall, A., Wagner, M., Epps, J., . . . and Breakspear, M. (2013). Multimodal assistive technologies for depression diagnosis and monitoring. *Journal on Multimodal User Interfaces*, *7*(3), 217–228. doi:10.1007/s12193-013-0123-2.

Juslin, P. N., & Scherer, K. R. (2005). *Vocal expression of affect*. Oxford University Press. doi:10.1093/acprof:oso/9780198529620.003.0003

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*(7), 865–878. doi:10.1016/j.cpr.2010.03.001.

Koole, S. L., & Tschacher, W. (2016). Synchrony in psychotherapy: A review and an integrative framework for the therapeutic alliance. *Frontiers in psychology*, *7*, 862. doi: http:10.3389/fpsyg.2016.00862

Koval, P., Kuppens, P., Allen, N. B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition and Emotion*, *26*(8), 1412–1427. doi: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*(6), 1132.

Koval, P., Sütterlin, S., & Kuppens, P. (2016). Emotional inertia is associated with lower well-being when controlling for differences in emotional context. *Frontiers in psychology*, *6*, 160705.

Krone, T., Albers, C. J., Kuppens, P., & Timmerman, M. E. (2018). A multivariate statistical model for emotion dynamics. *Emotion*, *18*(5), 739–754. doi:10.1037/emo0000384.

Kuppens, P. (2015). It's About Time: A Special Section on Affect Dynamics. *Emotion Review*, 7(4), 297–300. https://doi.org/10.1177/1754073915590947

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*(7), 984–991. doi:10.1177/0956797610372634.

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*(4), 917–940. doi:10.1037/a0030811.

Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology*, *17*, 22–26. doi:10.1016/j.copsyc.2017.06.004.

Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Ver- meersch, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The reliability and validity of the Outcome

Questionnaire. *Clinical Psychology & Psychotherapy: An International Journal of Theory and Practice*, 3, 249 –258.

Laufer-Goldshtein, B., Talmon, R., & Gannot, S. (2018a). Source counting and separation based on simplex analysis. *IEEE Transactions on Signal Processing*, *66*(24), 6458–6473. doi: 10.1109/TSP.2018.2876349

Laufer-Goldshtein, B., Talmon, R., & Gannot, S. (2018b). Diarization and Separation Based on a Data-Driven Simplex. In *2018 26th European Signal Processing Conference* (EUSIPCO) (pp. 842-846). IEEE. doi:10.23919/EUSIPCO.2018.8552933

Lazarus, G., Atzil-Slonim, D., Bar-Kalifa, E., Hasson-Ohayon, I., & Rafaeli, E. (2019). Clients' emotional instability and therapists' inferential flexibility predict therapists' session-by-session empathic accuracy. *Journal of Counseling Psychology*, *66*(1), 56–69. doi:10.1037/cou0000310.

Lazarus, G., & Rafaeli, E. (2023) Modes: Cohesive personality states and their interrelationships as organizing concepts in psychopathology. *Journal of Psychopathology and Clinical Science*. doi:10.31234/osf.io/9yqr6.

Levitan, R., & Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of INTERSPEECH,* Florence, Italy, August 27–31, 2011 (pp. 3081–3084). doi:10.21437/Interspeech.2011-771.

Lewis, M. (2000). Emotional Self-Organization at Three Time Scales. In M. Lewis & I. Granic (Eds.), *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development* (Cambridge Studies in Social and Emotional Development, pp. 37-69). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511527883.004

Lichtwarck-Aschoff, A., Hasselman, F., Cox, R., Pepler, D., & Granic, I. (2012). A characteristic destabilization profile in parent-child interactions associated with treatment efficacy for aggressive children. *Nonlinear Dynamics-Psychology and Life Sciences*, 16(3), 353.

Luborsky, L., & Mark, D. (1991). Short-term supportive-expressive psychoanalytic psychotherapy. *Handbook of short-term dynamic psychotherapy*, 110–136.

Luborsky, L., Mark, D., Hole, A. V., Popp, C., Goldsmith, B., & Cacciola, J. (1995). Supportive-expressive dynamic psychotherapy of depression: A time-limited version. In J. P. Barber & P. Crits-Christoph (Eds.), *Dynamic therapies for psychiatric disorders (Axis I)* (pp. 13–42). Basic Books.

McCullough, L., & Magill, M. (2009). Affect-focused short-term dynamic therapy: Empirically supported strategies for resolving affect phobias (pp. 249-277). In *Handbook of evidence-based psychodynamic psychotherapy* (pp. 249–277). New York: Humana Press.

Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A., & Claud, D. A. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy*, *2*(2), 91–100.

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological Psychiatry*, *72*(7), 580–587. doi:10.1016/j.biopsych.2012.03.015.

Nasser, S. A., Hashim, I. A., & Ali, W. H. (2020). A review on depression detection and diagnoses based on visual facial cues *3rd International Conference on Engineering Technology and Its Applications (IICETA), 2020* (pp. 35–40). doi:10.1109/IICETA50496.2020.9318860.

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, *3*(5), 400–424. doi:10.1111/j.1745-6924.2008.00088.x.

Ogbaselase, F. A., Mancini, K. J., & Luebbe, A. M. (2022). Indirect effect of family climate on adolescent depression through emotion regulatory processes. *Emotion*, *22*(5), 1017. doi: 10.1037/emo0000899

Olthof, M., Hasselman, F., Oude Maatman, F., Bosman, A. M., & Lichtwarck-Aschoff, A. (2023). Complexity theory of psychopathology. *Journal of Psychopathology and Clinical Science*, 132(3), 314.

Panksepp, J. (2012). What is an emotional feeling? Lessons about affective origins from cross-species neuroscience. *Motivation and Emotion*, *36*, 4-15.

Pascual-Leone, A. (2009). Dynamic emotional processing in experiential therapy: Two steps forward, one step back. *Journal of Consulting and Clinical Psychology*, *77*(1), 113–126. doi:10.1037/a0014488.

Pascual-Leone, A., & Greenberg, L. S. (2007). Emotional processing in experiential therapy: Why “the only way out is through.” *Journal of Consulting and Clinical Psychology*, *75*(6), 875–887. doi:10.1037/0022-006X.75.6.875.

Paz, A. (2023). Affective flexibility in treatment for depression using multimodal analysis OSM. Retrieved from https://osf.io/h76r8/?view_only=4f0df1004dac4ff19b85b7364f972d30

Paz, A., Rafaeli, E., Bar-Kalifa, E., Gilboa-Schectman, E., Gannot, S., Laufer-Goldshtein, B., Narayanan, S., Keshet, J., & Atzil-Slonim, D. (2021). Intrapersonal and interpersonal vocal affect dynamics during psychotherapy. *Journal of Consulting and Clinical Psychology*, 89(3), 227–239. https://doi.org/10.1037/ccp0000623.

Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., Kuppens, P., & Gotlib, I. H. (2015). Emotion-Network Density in Major Depressive Disorder. *Clinical Psychological Science*, 3(2), 292-300. https://doi.org/10.1177/2167702614540645

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98–125. doi:10.1016/j.inffus.2017.02.003.

Rademacher, L., Kraft, D., Eckart, C., & Fiebach, C. J. (2022). Individual differences in resilience to stress are associated with affective flexibility. *Psychological Research*, 1–18. doi:10.1007/s00426-022-01779-4.

Rafaeli, E., Bernstein, D., & Young, J. (2010), *Distinctive features of schema therapy*. London: Routledge.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). sage.

Rice, L., & Wagstaff, A. (1967). Client voice quality and expressive style as indexes of productive psychotherapy. Journal of Consulting Psychology, 31, 557–563. doi:10.1037/h0025164

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. doi:10.1037/h0077714.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. doi:10.1037/0033-295X.110.1.145.

Scherer, S., Stratou, G., Lucas, G., Mahmoud, M., Boberg, J., Gratch, J. (2014). Automatic audiovisual behavior descriptors for psychological disorder analysis. *Image and Vision Computing*, *32*(10), 648–658. doi:10.1016/j.imavis.2014.06.001.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59*(Suppl 20), 22–33.

Skiendziel, T., Rösch, A. G., & Schultheiss, O. C. (2019). Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and Facial Action Coding System Scoring. *PLOS ONE*, *14*(10), e0223905. doi:10.1371/journal.pone.0223905.

Snir, A., Bar-Kalifa, E., Berenson, K. R., Downey, G., & Rafaeli, E. (2017). Affective instability as a clinical feature of avoidant personality disorder. *Personality Disorders*, *8*(4), 389–395. doi:10.1037/per0000202.

Stiles, W. B. (2009). Responsiveness as an obstacle for psychotherapy outcome research: It's worse than you think. *Clinical Psychology: Science and Practice*, 16(1), 86–91. doi:10.1111/j.1468-2850.2009.01148.x.

Tomicic, A., Martinez, C., & Krause, M. (2015). The sound of change: A study of the psychotherapeutic process embodied in vocal expression. Laura Rice's ideas revisited. *Psychotherapy research*, *25*(2), 263-276. doi:10.1080/10503307.2014.892647

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*(3), 647–661. doi:10.1037/a0012532.

Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective dynamics in psychopathology. *Emotion Review*, *7*(4), 355-361.

Valstar, M., Gratch, J., Schuller, B., Ringeval, F., Lalanne, D., Torres, M., Scherer, S., Stratou G., Cowie R., & Pantic, M. (2016). Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In *Proceedings of the 6th international workshop on audio/visual emotion challenge* United States (pp. 3–10). ACM. doi:10.1145/2988257.2988258

Wang, Q., Yang, H., & Yu, Y. (2018). Facial expression video analysis for depression detection in Chinese patients. *Journal of Visual Communication and Image Representation*, *57*, 228–233. doi:10.1016/j.jvcir.2018.11.003.

Warikoo, N., Atzil-Slonim, D., Mayer, T., Eliassaf, A., Gurevych, I., (2022). NLP meets psychotherapy: Using predicted client emotions and self-reported client emotions to measure emotional coherence. *NeurIPS 2022 Workshop: Empowering Communities: A Participatory Approach to AI for Mental Health*.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219–235. doi:10.1037//0033-2909.98.2.219.

Watson, J. C., McMullen, E. J., Prosser, M. C., & Bedard, D. L. (2011). An examination of the relationships among clients' affect regulation, in-session emotional processing, the working alliance, and outcome. *Psychotherapy Research*, *21*(1), 86–96. doi:10.1080/10503307.2010.518637.

Weber, D. M., Baucom, D. H., Wojda-Burlij, A. K., Carrino, E. A., Du Bois, S., & Sher, T. G. (2022). Relationship "jet lag" in long-distance and geographically close relationships: The impact of relationship transitions on emotional functioning. *Couple and Family Psychology: Research and Practice.*

Wright, A. G., Pincus, A. L., & Hopwood, C. J. (2023). Contemporary integrative interpersonal theory: Integrating structure, dynamics, temporal scale, and levels of analysis. Journal of Psychopathology and Clinical Science, 132(3), 263.

Yang, Y., Fairbairn, C., & Cohn, J. F. (2013). Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*, *4*(2), 142–150. doi:10.1109/T-AFFC.2012.38.

Zhu, Z., & Bonanno, G. A. (2017). Affective flexibility: Relations to expressive flexibility, feedback, and depression. *Clinical Psychological Science*, *5*(6), 930–942. doi:10.1177/2167702617717337.

Tables and Figures

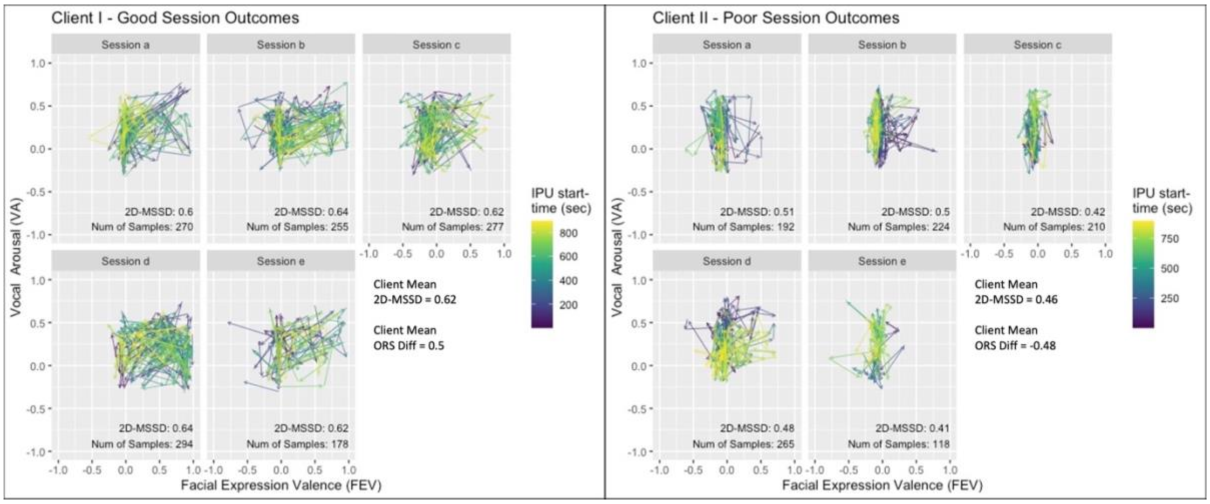

Figure 1. The two-dimensional affect representation, with the x-axis indicating facial expression valence (higher values denoting more positive valenced affect) and the y-axis representing vocal arousal (higher values indicating higher affective arousal). Each panel in the figure depicts the affect dynamics of an individual client across five sessions' working phases. The arrows indicate the movement of affect between consecutive IPUs. The color of the arrows corresponds to the IPU start-time within the working phase in seconds, ranging from dark purple at the beginning to light yellow toward the end of each session's working phase.

Table 1

Means, Standard Deviations, and Inter-Correlations for Study Variables

			Zero-order correlations	
Study variables	*M*	*SD*	*2*	*3*
Affective temporal variability	0.54	*0.057*	0.20 (*p*=0.05)	0.40 (*p*<0.001) ***
FEV	-0.06	*0.130*		0.12 (*p*=0.18)
ORS difference	0.59	*1.233*		

Note. Affective temporal variability: two-dimensional (vocal arousal, facial expression valence) session level MSSD; FEV: facial expression valence; ORS difference: outcome rating scale difference between pre-to-post sessions; zero-order correlations applied the variable means computed across all treatment sessions; $*p < .05$, $**p < .01$, $***p < .001$.

Table 2

Fixed Effect of Temporal variability and Valence of Session and Client Mean as Predictors for Clients' ORS Difference (Between Post- and Pre-Session Reports)

	Est. (SE)	*CI (95%)*	*P*	*Std. Est.*
Intercept (γ_{00})	0.59 (0.15)	[0.30,0.88]	<0.001 ***	-
Session mean-valence (γ_{10})	2.17 (1.25)	[-0.26,4.60]	0.086 ·	0.12
Client mean-valence (γ_{01})	-0.06 (1.33)	[-2.74,2.62]	0.965	-0.01
Session mean-temporal variability (γ_{20})	7.17 (1.79)	[3.68,10.66]	<0.001 ***	0.27
Client mean-temporal variability (γ_{02})	9.73 (4.55)	[0.59,18.87]	0.042 *	0.26

Note. $\cdot p<0.1$. $*p < .05$. $**p < .01$. $***p < .001$.

Table 3

Fixed Effect of Session Outcome (Post Session ORS) Trajectory Moderation by Clients' Affective Temporal variability

	Est. (SE)	*CI (95%)*	*P*	*Std. Est.*
Intercept (γ_{00})	4.65 (0.56)	[3.56,5.73]	<0.000***	-
Session number (γ_{10})	0.17 (0.04)	[0.09,0.25]	<0.001***	0.24
Session mean-temporal variability (γ_{20})	-7.54 (8.58)	[-24.16,9.07]	0.383	-0.16

Client mean-temporal variability (γ_{01})	21.01 (16.31)	[-11.34,53.36]	0.208	0.32
Session number X Session mean-temporal variability (γ_{30})	2.32 (1.09)	[0.20,4.43]	0.038*	0.41
Session number X Client mean-temporal variability (γ_{11})	-0.52 (1.11)	[-2.67,1.63]	0.641	-0.08

Note. · *p*<0.1. **p* < .05. ***p* < .01. ****p* < .001.

Data Transparency

The data reported in this manuscript were previously published in Atzil-Slonim et al. (2022), a study that focused on Oxytocin reactivity during treatment for depression. However, none of the measures reported in the previous study were used in this study.