**Chapter 8: Leveraging Theoretical and Technological Innovations to Study the Mechanisms that Underlie Therapeutic Change.**

Dana Atzil-Slonim

Psychology Department, Bar-Ilan University, Israel

To appear in: Castonguay, L. G., Atzil-Slonim, D., Barkham, M., & Lutz, W. (Eds.). *Practice-based evidence in the psychological therapies: Towards clinical, training, and public policy guidelines*. Oxford University Press.

**Leveraging Theoretical and Technological Innovations to Explore the Dynamics of Therapeutic Change**

The enduring question that has long intrigued clinicians and researchers—what works for whom in therapy, and when—remains as vital as ever. What has evolved is our capacity to explore this question more effectively and to begin uncovering more robust answers. The practice-oriented research approach, which emphasizes the importance of conducting research in naturalistic settings that mirror actual clinical practice, facilitates the collection of extensive datasets. These rich datasets, capturing both verbal and non-verbal information from psychotherapy sessions, are essential for examining the complex dynamics that unfold in therapy. The concurrent maturation of innovative technologies for monitoring the therapeutic process, the rise of AI techniques to analyze massive amounts of data, along with advances in clinical science all provide opportunities for dramatic progress in understanding the mechanisms that account for psychotherapy gains. To fully harness these theoretical and technological advancements, and to ensure their integration into clinical practice enhances patient well-being, the practice-oriented research approach underscores the importance of strong, collaborative partnerships between clinicians and researchers.

This chapter aims to illustrate how adopting the practice-oriented research approach has enabled our interdisciplinary team of clinicians and researchers to work in close collaboration, integrating innovations in clinical science with cutting-edge technology to tackle key challenges in the field and deepen our understanding of the mechanisms that underlie therapeutic change.

The chapter begins with a prologue that outlines the motivations behind our decision to establish a research clinic rooted in the practice-oriented research approach. The subsequent section addresses some of the critical challenges in the conceptualization, treatment, and research of mental health. It then describes significant theoretical shifts in clinical science, such as the transition from general treatment models to transtheoretical interventions and processes, and the shift from one-person to two-person psychology, which have opened new avenues for addressing these challenges. The fourth section details how our team has leveraged these theoretical innovations, alongside advancements in multi-modal analysis and AI, to study intrapersonal (within patients) and interpersonal (between patients and therapists) dynamics associated with improved treatment outcomes. Finally, the chapter concludes by discussing the clinical and training implications derived from our studies.

**Prologue**

The first time I considered crossing to the other side of the river, the waters were so turbulent that it seemed impossible to know if, or how, the crossing could be made. The emotional intensity and deep conviction each side held in their own truth fascinated me as I stood amidst a debate between clinicians at the university counseling center and researchers in the psychology department of the same university in Israel. As an experienced clinician, I felt a strong connection to my colleagues in the counseling services. However, I was also embarking on my PhD focused on psychotherapy research, and as I listened, I recognized significant merit in the arguments from both sides.

On one hand, I resonated with my fellow clinicians who argued that therapeutic interactions are incredibly complex, with each patient-therapist encounter being unique and influenced by myriad factors such as momentary mindset and mood, transference, countertransference, culture, personal values, life events, and numerous other variables. This complexity makes studying what works for whom and when a formidable challenge. They asserted that randomized controlled trials (RCTs) demonstrating the efficacy of specific manualized treatments for specific diagnosis do not adequately address the needs of their real-world patients, who often present with comorbidities. They felt that the pressure to adhere to manualized treatments, as suggested by such studies, hindered their ability to flexibly tailor interventions to their patients’ immediate needs

On the other hand, the researchers' arguments also made sense. They questioned how psychotherapy could remain unstudied, emphasizing our ethical obligation to determine what works and what doesn’t, ensuring the treatments we offer are effective. While acknowledging the uniqueness and complexity of therapeutic encounters, they maintained that achieving a certain level of generalization was still possible (for more on similar debates, see Safran, 2012).

The opportunity to bridge these polarized voices arose when, following my PhD advisors Orya Tishby and Gaby Shefler, and later my postdoctoral mentor, Hadas Wiseman, I became acquainted with the Society for Psychotherapy Research (SPR) community. It was through SPR that I discovered the practice-oriented research approach. The idea of conducting research in naturalistic settings, without imposing constraints on clinical routines, while fostering ongoing dialogue between clinicians and researchers (Castonguay, 2011), opened up the possibility of conducting systematic research that respond to clinicians’ need to understand how they can improve their ability to help their patients achieve better therapeutic outcome.

A few years later, when I began my early steps as a faculty member in the Department of Psychology at Bar-Ilan University (BIU), these ideas inspired me and my colleagues -Eshkol Rafaeli, Eva Galboa-Schechtman, Tuvia Peri, Ilanit Hasson-Ohayon, and Rivka Tuval-Mashiach - to lead a transition in our departmental clinic. Our team transformed the clinic, which has provided subsidized services to over 300 clients at any given time for more than 50 years, into a modern research clinic where all clinical activities are available for scientific inquiry. We designed a clinical research protocol enabling session-by-session monitoring, within-session recording, and therapist feedback. This transition was greatly influenced by our close collaboration with Wolfgang Lutz from Trier University, whose generous sharing of research measures and procedures for continuous clinical monitoring was invaluable.

Transitioning a large mental health clinic into one with intensive research capabilities is challenging on many levels, including personal and ideological ones. While the younger generation tended to embrace the project with enthusiasm, more senior members tended to express concerns. Wolfgang's extensive experience in establishing such a center was instrumental in helping us navigate these challenges. Following his advice, we implemented a gradual transition process, offering the option to participate in the research and fostering collaborative discussions where concerns could be raised and solutions brainstormed. This approach gradually reduced resistance. Discussion groups composed of clinicians and researchers were formed to collaboratively shape the research direction in our training clinic, determine the feedback we wanted to provide to students and interns, and identify the research questions we wished to explore. It was particularly surprising and encouraging to see clinicians who had not previously been involved in research actively participating in these discussions.

Since I first crossed the bridge between practice and research, I’ve done so repeatedly, with the practice-oriented research approach providing the conceptual framework that helped me integrate these two facets of my professional identity. The ongoing dialogues within our interdisciplinary team at BIU—comprising clinicians and researchers with expertise in psychotherapy and computer science—enable allow questions emerging from clinical practice and theoretical thinking to shape the research. In turn, the research findings are fed back to the clinicians, refining practice and expanding the theoretical frameworks over time. Our work has been profoundly influenced by significant shifts in clinical science that emphasize the importance of identifying transtheoretical intrapersonal (within patients) and interpersonal (between patients and therapists) dynamics associated with improved treatment outcomes. Traditional research methods were insufficient for studying these complex dynamics, so we leveraged advances in artificial intelligence and multimodal analyses. The large-scale, naturalistic data we gradually collected enabled us to use computational tools to address these issues, feeding back into our discussions and generating new questions. This iterative process has helped us sharpen some of the key challenges in practice and research that our field faces, initiating a collective effort to overcome them.

**Some Major Challenges for Mental Health Practice and Research**

The prevalence of mental health issues is increasing, affecting millions of people worldwide (World Health Organization, 2022). At present, three major issues hinder the effectiveness of mental health services: First, the need for services greatly exceeds the availability of adequate providers (Kazdin, 2021). A major cause of this gap is the over-reliance on the dominant treatment delivery model of 1-on-1 in-person sessions conducted by highly-trained professionals.

Second, when individuals seek help from mental health services they are often sub-optimally diagnosed. Most clinicians and psychotherapy researchers administer semi-structured interviews following the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013) or the International Classification of Diseases (World Health Organization, 2022) for diagnostic purposes. However, these diagnostic systems have been increasingly criticized for focusing on categorical syndromes and failing to capture the high comorbidity, considerable interpersonal heterogeneity, and very pronounced intrapersonal variability of mental distress (Hofmann & Hayes, 2019).

Third, current mental health treatments are not sufficiently effective. Only 50% of the patients receiving psychological, pharmacological, or combined interventions show adequate improvement in symptom severity following the treatment. Moreover, among those who benefit from treatment, many experience relapse within a year (Cuijpers et al., 2021).

The urgent need to identify the mechanisms facilitating therapeutic change and to find ways to optimize healthcare practices has prompted many researchers to study which therapy processes lead to positive treatment outcomes (Crits-Christoph & Gibbons, 2021). However, these efforts are constrained by several key limitations. Most previous studies rely on self-report measures. Although subjective measures are essential in psychotherapy research, they are limited by patients’ self-awareness, motivation, and restricted scope of potential predictors (Kazdin, 2008). Vast amounts of data from intake interviews and psychotherapy sessions, including both verbal and non-verbal behaviors of patients and therapists, remain largely untapped. Utilizing this data could significantly enhance diagnostic and treatment procedures, improve intervention selection, and deepen our understanding of the mechanisms driving therapeutic change.

To tap what occurs within psychotherapy sessions, researchers have developed numerous observer coding systems (e.g., Pascual-Leone & Greenberg, 2005). However, since observational human coding is very labor-intensive and expensive to implement, these studies have typically focused on a small number of therapeutic components in a relatively small sample of patients and at limited time points, which considerably limits progress in studying complex processes in psychotherapy (Crits-Christoph & Gibbons, 2021). Additionally, efforts to determine which patient characteristics are associated with treatment outcomes have often yielded inconclusive results, possibly due to an emphasis on pre-treatment evaluations and insufficient attention to the multifaceted and dynamic nature of mental distress (Constantino et al., 2021). Similarly, attempts to identify what makes therapists' interventions effective have produced inconclusive results, likely due to a focus on single techniques and their association with treatment outcomes, which may oversimplify the therapeutic process (Hill & Norcross, 2023).

To truly understand what works for whom and when in psychotherapy, it is critical to consider the dynamic nature of the therapeutic interaction and the diverse contexts in which interventions are implemented (Hill & Norcross, 2023). In actual clinical practice, clinicians attend to moment-by-moment verbal and non-verbal information to guide their clinical action. Importantly, therapists differ widely in effectiveness (Wampold & Owen, 2021). Expert therapists have the ability to quickly and flexibly tailor their interventions to meet their patients’ immediate needs and facilitate superior treatment outcomes. However, given the heterogeneous nature of mental distress and the vast array of potential interventions available to therapists at any given time, this goal remains a formidable challenge that research methodologies have struggled to tackle until recently.

To tackle these challenges, our team has combined a theory-based approach with a data-driven methodology to deepen our understanding of the intrapersonal and interpersonal dynamics underlying therapeutic change. By leveraging our extensive, multifaceted dataset, along with interdisciplinary collaborations and advancements in clinical science and technology, we have developed new approaches to explore these intricate dynamics.

**Leveraging Advances in Clinical Science to Study Intrapersonal and Interpersonal Dynamics in Psychotherapy**

Recent theoretical shifts in clinical science have significantly influenced our work. The first major shift is the transition from manualized, disorder-focused treatments to a transtheoretical approach that emphasizes building a consensual knowledge base that integrates evidence-based interventions and processes that are central across diverse therapeutic models and tailoring interventions to meet patients’ specific needs at specific times (Gaines & Goldfried, 2021; Hofmann & Hayes, 2019). Rather than perpetuating the longstanding competition between different schools of psychotherapy, contemporary research advocates for a unified language that synthesizes empirical evidence from across the spectrum of therapeutic approaches (Castonguay et al., 2021).

The second major shift in psychotherapy literature is often referred to as "the relational turn" (Atzil-Slonim & Tschacher, 2020). This shift marks the transition from a one-person psychology, focused solely on the patient's mental experience, to a two-person psychology that acknowledges the therapist's significant involvement in the process of change (Muran et al., 2019). Traditionally, process-outcome research viewed the therapeutic process as unidirectional, where the therapist intervenes, and the patient is affected. However, there is growing recognition that the therapist's role is dynamically intertwined with the patient's experience (e.g., Koole & Tschacher, 2016). Relational psychotherapy theorists argue that the complexity of the therapeutic process can only be understood through the dynamic interplay between patient and therapist, who are mutually but asymmetrically shaped and transformed over time (e.g., Aron & Harris, 2014).

Building on these theoretical advancements, I have developed the self-other dynamic approach, which shifts focus from categorical classification of trait-like psychopathology to modifiable intrapersonal dynamics, and from manualized disorder-focused treatments to interactive interpersonal dynamics tailored to patients' needs at specific moments (Atzil-Slonim, 2024). This approach draws on theories that conceptualize human experience as consisting of multiple self-states that fluctuate and evolve over time (Beck et al., 2021; Bromberg, 2004; Stiles, 2001). These self-states, constitute identifiable units characterized by specific combinations of Affect, Behavior, Cognition, and Desire (ABCD; Revelle, 2007) that tend to be coactivated in a meaningful manner for limited periods (Lazarus & Rafaeli, 2023). Self states are a dynamic phenomenon, fluctuating and changing over time. At any given moment, a particular self-state may dominate, with individuals varying in their ability to dynamically shift between self-states reflect upon and foster dialogue among their self-states (Bromberg, 1996). Patients often seek therapy when they become stuck in a rigid self-state, where one dominant self-state overshadows the internal dialogue among diverse self-states or voices. These self-states often elicit complementary responses from others, creating cyclical self-other dynamics. Some of these dynamics are adaptive, fostering positive interactions, while others are maladaptive, detrimental to both the individual and their relationships. Modern psychotherapy models increasingly emphasize the importance of enhancing patients' ability to transition from maladaptive to more adaptive self-other dynamics (e.g., Fosha, 2001; McCullough, 2003). By attuning to patients' shifting self-states, clinicians can help patients regulate and dampen maladaptive ABCD elements while amplifying adaptive ones. Furthermore, clinicians aim to deepen patients' understanding of these maladaptive states, thereby increasing their sense of agency and choice in life. These therapeutic interactions, which occur through both verbal and non-verbal channels, are expected to lead to symptom reduction and enhanced well-being (e.g., Atzil-Slonim et al., 2023).

Despite these theoretical advancements, attempts to model patients’ self-states and identify effective moment-to-moment interventions have been limited. Previous efforts have often relied on predefined theoretical frameworks (e.g., Arntz & Jacob, 2017), rather than on empirical, data-driven analysis of how self-state elements cluster, the contextual factors that correlate with these states, and their temporal dynamics. The complexity of potential ABCD element combinations, the varying contexts in which they emerge, and the diverse interventions therapists might employ pose significant challenges that traditional research methodologies have struggled to address. However, recent advances in AI, Natural Language Processing (NLP), signal processing, and machine learning offer tremendous potential to aid this process.

**Leveraging Advances in Multi-Modal Analysis and AI to Study Intrapersonal and Interpersonal Dynamics in Psychotherapy**

Communication between patients and therapists occurs through multiple channels, both verbal and nonverbal. While the words spoken in therapy sessions offer valuable insights into internal thoughts and emotions, the manner in which these words are delivered—through facial expressions, vocal tones, and body language—is no less important. These nonverbal cues provide critical information about an individual’s mental state and the therapeutic process, yet they remain largely underutilized in improving diagnoses and treatment outcomes. Intake interviews and psychotherapy sessions contain rich, untapped data that could significantly enhance our understanding and treatment of mental health issues.

Recent technological advances in signal processing, NLP, and computer vision, combined with the rise of AI and machine learning, have shown great promise in assessing mental health constructs by employing quantitative methods to capture and model key behavioral and physiological signals (Cohn et al., 2018). These technologies have the potential to increase precision in diagnosis and deepen our understanding of the mechanisms of change in psychotherapy by enabling the coding of moment-by-moment verbal and nonverbal responses from both patients and therapists. This granular focus on smaller units within therapy sessions allows researchers to process larger datasets, leading to more comprehensive insights into psychotherapy processes and outcomes.

AI-based models are particularly adept at handling the complexity of mental health data, allowing models to refine their predictions based on previous data, thereby maximizing accuracy for new patients (Delgadillo & Atzil-Slonim, 2022). Among the various AI techniques available, deep learning models—especially transformer-based language models—have emerged as the leading approach in recent years (Devlin et al., 2018). These models undergo pre-training on massive datasets of unlabeled text, using a process called unsupervised learning, where the model learns to predict randomly masked words. This enables the model to grasp the underlying structure and context of language. After pre-training, the model can be fine-tuned for specific tasks, such as emotion recognition in psychotherapy texts, by adjusting its parameters for optimal performance on the targeted task. The versatility and effectiveness of transformer-based language models have led to their widespread application across various research domains, including mental health (Delgadillo & Atzil-Slonim, 2022).

Recent advancements in computing power and deep learning have also propelled progress in generative AI and large language models (LLMs), such as GPT-3/4/5 (Bommasani et al., 2021). These models excel in understanding and generating human language, detecting intricate patterns, and capturing contextual nuances. By fine-tuning these models for specific tasks, researchers can significantly enhance their performance and accuracy (Stade et al., 2023).

Our research team, which integrates expertise in mental health and computer science, leverages these advanced AI and machine learning techniques to analyze both verbal and nonverbal data from recorded session with the aim of identifying patients’ intrapersonal and patient-therapist interpersonal dynamics associated with improved psychotherapy outcomes.

**Using Multi-Modal Analysis and AI to Study Intrapersonal Dynamics**

Intrapersonal dynamics refer to the ways in which internal elements—such as emotions, behaviors, cognitions, desires, and physiological responses—fluctuate and interact within an individual over time (Bar-Kalifa & Atzil-Slonim, 2020). These dynamics encompass various aspects, including variability (the range or amplitude of an internal element across time), instability (the magnitude of changes in an internal element from one moment to the next), and inertia (how well an internal element can be predicted from its previous state). Such dynamics reflect individuals’ ability to regulate their affect, behaviors, cognitions, and desires in response to environmental changes, playing a crucial role in psychological well-being and psychopathology (Houben et al., 2015).

Our team has used advanced analytics to explore various intrapersonal dynamics and their links to treatment outcomes. A key dynamic consistently associated with positive outcomes across modalities is affective flexibility—the capacity to dynamically modulate emotional responses from one moment to the next—which is fundamental to psychological well-being (Houben et al., 2015). Greater flexibility suggests increased sensitivity and adaptability to both external factors, such as environmental changes, and internal factors, including regulatory processes. Conversely, lower levels of affective flexibility, or rigidity, often underlie various psychopathologies.

In one study (Bar-Kalifa & Atzil-Slonim, 2020), we utilized multilevel vector autoregressive network analysis to examine patients’ intrapersonal emotional dynamics from session to session. This study, involving 95 patients treated by 58 therapists, aimed to identify differences in the network structures—particularly the density of emotional connections—between treatment responders and non-responders. Our findings revealed that patients who did not improve during therapy exhibited a denser temporal intrapersonal emotional network, indicating less flexibility in their emotional responses over time.

This study relied on patients' subjective reports of their emotions at the end of each session. However, emotions fluctuate continuously, and patients may not fully recall the full range of emotions experienced during a session. Relying solely on subjective reports may overlook valuable data that can be obtained from observed measures of affect. Facial expressions, for example, provide a rich stream of information on intrapersonal affective dynamics. Clinicians often study these expressions to gauge differences in expressivity among patients. Computer vision methods that automatically analyze facial expressions can detect mood disorders and capture their dynamic nature (Girard et al., 2015). In one study that focused on facial expressions (Atzil-Slonim et al., 2023), we examined the link between emotional flexibility and depression by analyzing 283 video-recorded sessions from 58 patients undergoing psychodynamic psychotherapy. Emotional flexibility was measured using the FaceReader (Höfling et al., 2020), an automated facial expression recognition system, while depression levels were assessed with the Beck Depression Inventory-II (BDI-II; Beck et al., 1996). Our results indicated that patients with higher depressive symptoms exhibited lower emotional flexibility during treatment. Additionally, sessions with higher depressive symptoms were marked by lower emotional flexibility, and vice versa. Importantly, the association between emotional flexibility and depressive symptoms was significant beyond the mean valence of the experienced emotions, both at the patient and session levels. These findings underscore the role of emotional flexibility as a central factor in depression and extend previous research by capturing moment-to-moment emotional fluctuations with greater time resolution.

While facial expressions convey the valence of emotions, vocal tone conveys the arousal of affective states. In a recent study, we investigated intrapersonal affective flexibility by focusing on both facial expressions and vocal channels (Paz et al., 2024). The sample included 30 patients diagnosed with major depressive disorder who underwent 137 sessions of supportive-expressive psychodynamic therapy (Luborsky et al., 1995). We utilized multimodal measures, including computerized facial expression valence assessed through FaceReader and computerized vocal arousal measure which uses an index that combines vocal pitch, intensity, and energy features (Paz et al., 2021). In the analysis of the voice, we followed Levitan and Hirschberg's (2011) approach, which highlights that vocal features near turn-switches reveal more about affective interaction than overall vocal scores from entire speech turns. They recommend focusing on interpausal units (IPUs), which are segments of speech separated by pauses of at least 50ms and free of any pauses longer than 50ms. Our findings showed that patients who exhibited higher flexibility in these affective measures during sessions experienced greater improvements in well-being, both within individual sessions and across the course of treatment. Figure 1 illustrates the affect measures during five sessions throughout the therapies of two patients drawn from our sample. The x-axis represents the measured valence that was extracted from facial expressions, whereas the y-axis represents the measured arousal extracted from the voice recordings of the patients. The darker purple arrows represent measures from the beginning of the session, while the lighter yellow arrows relate to the latter parts of the session. Patient I’s flexibility was greater than patient II’s flexibility. Accordingly, patient I’s average well-being improved during treatment while patient II suffered a reduction in well-being across treatments.

Importantly, during the treatment of patient II, session d was marked by greater flexibility and greater improvement in well-being, while session e was marked by less flexibility and accordingly, a reduction in well-being.

Figure 1: Intrapersonal multimodal flexibility

Scatter chart

Description automatically generated

Note. Each panel presents the affect dynamics of a single patient over the course of five sessions. Each arrow represents the affective movement between consecutive IPUs. The time within the session is represented by the color of the arrows: from dark purple at the beginning to light yellow toward the end of each session.

Physiological measures provide another promising channel for examining affect dynamics in psychotherapy, as they provide a more implicit measure of emotional arousal and regulation (Barrett Feldman et al., 2016). In one study (Goren et al., 2022), we explored Respiratory Sinus Arrhythmia (RSA), a measure of parasympathetic nervous system activity, as a biomarker in depression treatment. This study involved 28 patients diagnosed with major depressive disorder who underwent short-term psychodynamic therapy for depression. We measured both resting RSA (i.e., RSA levels during a baseline relaxation period before the session start) and RSA reactivity (i.e., the difference between RSA during the session and RSA at baseline) during five pre-selected sessions throughout the treatment. Depression severity was assessed using the BDI-II, and session outcomes were assessed using the Session Evaluation Scale (SES; Hill & Kellems). Our findings revealed that lower depression severity at the start of a session was associated with higher resting RSA, suggesting that resting RSA may serve as a state-like biomarker of depression severity. Additionally, the interaction between resting RSA and RSA reactivity was linked to session outcomes, with greater reactivity during sessions indicating better therapeutic progress when resting RSA was higher. These findings underscore the value of implicit physiological measures as potential biomarkers for treatment processes and outcomes. In another study (Atzil-Slonim et al., 2022) using the same dataset, we explored the potential of oxytocin, a hormone known for its role in social bonding and stress regulation, as a biomarker for symptom change in psychotherapy for depression. Oxytocin reactivity was assessed by measuring salivary oxytocin levels before and after five sessions, with depressive symptoms tracked using the BDI-II. The results indicated that patients with higher oxytocin reactivity during therapy experienced greater improvement in depressive symptoms over the course of treatment, suggesting oxytocin reactivity as a crucial biomarker of therapeutic progress.

Whereas facial expression, vocal tone, and physiological measures provide insights into non-verbal intrapersonal dynamics, text analysis offers a window into the verbal aspects of therapy. The language patients use in psychotherapy sessions reflects their internal thoughts and emotions, revealing important information about their self-states. Many of our studies have utilized state-of-the-art NLP techniques to analyze text from transcribed psychotherapy sessions. For instance, in one study (Atzil-Slonim et al., 2024), we employed transformer-based emotion recognition models to automatically label patients’ emotions at the utterance level in a dataset comprising 139,061 utterances. We then used these labeled data to examine the coherence between verbally expressed emotions and self-reported emotions, employing multilevel modeling to assess the associations between emotional coherence and patients’ improvement in functioning throughout treatment. The emotion recognition model demonstrated moderate performance, and the findings indicated a significant association between verbally expressed emotions and self-reported emotions. Coherence in clients’ negative emotions was associated with improvement in functioning, underscoring the importance of this coherence to overall well-being.

In another study (Atzil-Slonim et al., 2021), we used topic modeling—a data-driven machine learning technique—to automatically extract latent topics from textual data and determine which topics best identified patients' functioning levels and alliance ruptures in psychotherapy sessions, as well as to assess whether changes in these topics were linked to changes in treatment outcomes. This study, which involved 872 sessions from 68 patients and 59 therapists, revealed that topic modeling produced semantically meaningful topics. A sparse multinomial logistic regression model identified alliance ruptures and patients' functioning levels with 65% and 75% test accuracy, respectively. Additionally, multilevel growth models indicated that changes in topic trajectories were associated with changes in outcome trajectories. These studies demonstrate the usefulness of computerized methods and AI techniques to utilize the rich verbal and non-verbal data from psychotherapy sessions and capture significant intrapersonal dynamics and processes that are associated with beneficial treatment outcomes.

AI techniques are also highly effective for monitoring, early detection and prevention of mental health problems. In a study with our AI experts collaborators (Chim et al., 2024), LLMs were used to automatically identify suicidal risk in 125 social media users, with risk levels (low, moderate, or high) labeled by expert clinicians. The models highlighted text supporting these classifications and generated brief summaries explaining the risk. The study, involving multiple research groups, found accuracy rates of 92% for high risk, 90% for moderate, and 88% for low risk. Summary quality was assessed using a natural language inference (NLI) model (Laurer et al., 2024), which is designed to determine whether one statement logically follows from another. The best model (Mistral) achieved 94% consistency with expert-created summaries (Jiang et al., 2023). These findings underscore the potential of advanced LLMs to identify at-risk individuals and provide explainable evidence for clinicians.

In another study (Song et al., 2024), we used a novel method for creating clinically meaningful formulation of individuals’ mental states from their consented social media timelines. The findings highlight the potential of this approach to assist mental health professionals in monitoring and understanding individuals progress over time, offering a valuable tool for clinical decision-making.

In a study by Bilu et al. (2023), we used machine learning to predict depression in middle-aged adults without prior psychiatric history, using data from the UK Biobank (N = 245,036). After a year, 2.18% of the population developed depression. A model based on one mental health questionnaire had an AUC of 0.66, while incorporating data from 100 questionnaires improved it to 0.79. The results were consistent across demographics, suggesting that machine learning can moderately predict depression risk a year in advance by combining multiple features.

The above studies focused on intrapersonal dynamics and processes, however many intrapersonal dynamics occur in an interpersonal context, and psychotherapy is quintessentially such a context.

**Leveraging Multi modal Analysis and AI to Study Interpersonal Dynamics**

Interpersonal dynamics refer to the processes by which the affect, cognition, behavior and desire of interacting individuals are interconnected and mutually influenced (Butler, 2015). When people interact, they naturally coordinate their emotions, perceptions, behavior and physiology, leading to an interweaving of their states that enhance connection and communication (Feldman, 2012). For example, when people are engaged in a naturally flowing conversation, their vocal tones tend to become aligned, and their bodies tend to start moving in the same cadence.

These dynamics manifest in various forms, such as synchrony (the alignment in time of the affect, physiology or behavior of interacting individuals), coregulation (time-lagged associations in which the emotional arousal of one member of the dyad at one time point influences the emotional arousal of the other member of the dyad at the next time point, in a way that leads to a stable state), convergence (the gradual minimization of differences between conversational partners over time) and congruence (the correspondence between dyad members in a more static state). Interpersonal dynamics occur across numerous relational contexts, including between children and their caregivers (e.g., Feldman, 2012), romantic partners (e.g., Butler, 2015), and even among strangers (Cwir et al., 2011). Moreover, these bio-behavioral linkages are associated with a wide range of meaningful outcomes, such as healthy development (e.g., Feldman, 2012) and overall relationship quality (e.g., Butler, 2015).

Many psychotherapy theories emphasize the importance of interpersonal dynamics between patients and therapists (e.g., Aron & Harris, 2014; Fosha, 2001). These theories suggest that therapy provides patients, whose development may have lacked attuned interactions, with a corrective emotional experience that mirrors more optimal developmental conditions. Through attuned interactions, the bond between patient and therapist deepens, enabling patients to explore and process their emotions and perceptions more effectively. This shared experience allows patients to develop better emotional and perceptual regulation by leveraging the combined resources of the therapeutic relationship, which are eventually internalized. When patients and therapists dynamically attune to each other’s affect, behavior, cognitions, motivations and physiology, it fosters a stronger therapeutic relationship, enhancing the patient’s regulation abilities and leading to better therapeutic outcomes.

The importance of studying interpersonal dynamics, given their pervasive, documented association with well-being across the life span, has led our team to take a deeper interest in these dynamics as a way to learn about the mechanisms of change in psychotherapy (Atzil-Slonim et al., 2023).

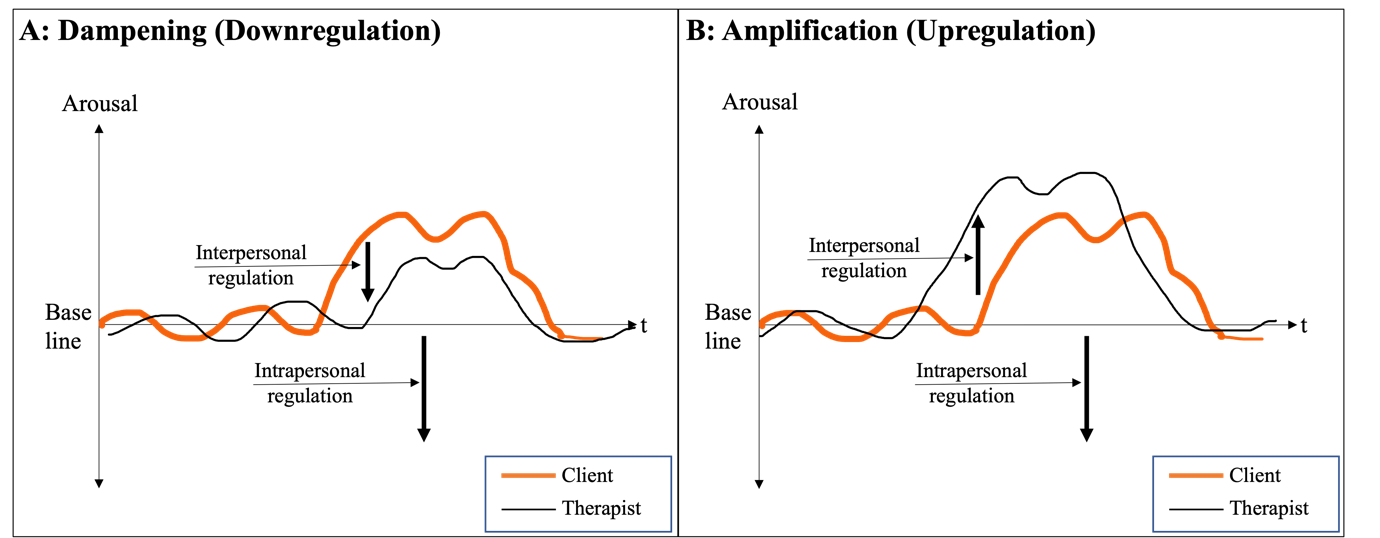
One such study (Atzil-Slonim et al., 2018) examined the emotional congruence between patient and therapist emotions as they co-fluctuated session-by-session, and the association between emotional congruence and treatment outcome. The results indicated that patients and therapists were congruent in the emotions they experienced, and that therapists experienced less intense positive (but not negative) emotions compared to their patients. Importantly, incongruence between patients and therapists in both positive and negative emotions predicted worsened symptoms in the next session. In another study (Atzil-Slonim et al., 2019), we examined therapists’ empathic accuracy by assessing therapists’ actual ability to accurately infer their patients’ states of mind session by session over the course of treatment. We also examined the associations between empathic accuracy and treatment outcomes. The results showed that therapists accurately tracked their patients’ negative (and, to a lesser extent, positive) emotions, and tended to overestimate their patients’ negative emotions and underestimate their patients’ positive emotions. In addition, therapists' own emotions partially mediated the association between patients' emotions and therapists' assessments; and therapists' inaccuracy in assessing their patients' positive emotions was associated with higher reported symptoms in the next session. These results demonstrate the importance of studying the contribution of both patients and therapists to the therapeutic relationship and the ways in which these contributions combine to impact treatment outcome. The results also point to the risk on the part of therapists to neglect patients’ positive emotions and stress the importance of attending to these emotions.

This line of research has gained momentum through advances in technology that made it possible to collect dense within-session repeated measurement data, such as audio, video and physiology, and new sophisticated data analytic methods that can handle the complexity of psychotherapy interactions (for review see, Atzil-Slonim et al., 2023).

However, as psychotherapy studies examining interpersonal dynamics have started to amass, it has become clear that these dynamics are not always associated with favorable therapy outcomes (e.g., [Reich et al., 2014; Schoenherr et al., 2021)](https://paperpile.com/c/W18ZHk/x4vd+o2bw). These mixed results suggested that more studies are needed to provide a better understanding of the modalities and the circumstances in which interpersonal dynamics leads to better (or worse) therapeutic outcomes. Our recent systematic review and meta-analysis (Atzil-Slonim et al., 2023) concluded that several directions could contribute to a better understanding of the role of interpersonal dynamics in psychotherapy. First, the context in which interpersonal dynamics occur is critical to understand when a specific dynamic is beneficial and when it is not. For example, synchrony between dyads in situations of distress can lead to escalation and higher distress, while synchrony in situations of emotional pain or in the case of positive emotions may lead to a feeling of shared experience and closeness. In one study that implemented this direction, we studied physiological synchrony between patients and therapists, while considering the emotional context in which synchrony occurred. Synchrony was assessed by measuring RSA, an autonomic index that is known to be associated with emotion regulation, from 28 patient-therapist dyads in five pre-selected sessions during short term dynamic therapy for depression. The results showed that compared to moments of unproductive emotional experience, greater synchrony was observed during moments of positive emotional experiences. These patterns of synchrony were associated with clients’ favorable evaluations of the session. Another potential context could be the type of intervention used by the clinician. Bar-Kalifa et al. (2019) continuously monitored the electrodermal activity of both patients and therapists during sessions to explore the role of physiological synchrony during emotion-focused techniques compared to cognitive-behavioral techniques. The findings revealed that greater synchrony in the emotion-focused segments (but not in the cognitive-behavioral segments) was linked to a stronger therapeutic alliance. These findings provide a fine-grained picture of physiological synchrony and its potential effects on therapy.

Another direction pointed out by our systematic review is that effective dyadic processes may depend on the therapists’ ability to induce coregulation patterns, perhaps through their capacity to self-regulate their own emotions in the session. In one study that implemented this direction, (Paz et al., 2021), we used an automatic method of assessing arousal in vocal data as well as dynamic system models to explore self-regulation and coregulation affect dynamics within psychotherapy sessions and to determine whether these dynamics are associated with treatment outcomes. Figure 2 illustrates the simultaneous effects of intra- and interpersonal pull forces of two kinds. In the left panel, affect dampening (i.e., downregulation of arousal) is depicted as resulting from the synergistic action of (a) the patient's intrapersonal self-regulatory force toward homeostasis, and (b) an interpersonal coregulatory force in which the therapist's affect "pulls" patients’ affect toward their baseline. In the right panel, affect amplification (i.e., upregulation of arousal) is depicted as resulting from the contrasting forces of (a) the patient's intrapersonal self-regulatory force pulling toward homeostasis, vs. (b) an interpersonal coregulatory force in which the therapist's affect "pulls" the patient's affect away from the baseline.

Figure 2. Illustration of Dampening and Amplification Models of Arousal Regulation

**** *Note*. Panel A represents the *Dampening* model in which both the intrapersonal and interpersonal regulatory “pull forces” conjoin to dampen arousal towards the affective baseline. Panel B represents the *Amplification* model in which the intrapersonal regulatory force pulls towards the baseline while the interpersonal regulatory force pulls away from the affective baseline.

The data from 21,133 mean vocal arousal observations were extracted from 279 therapy sessions in a sample of 30 clients treated by 24 therapists. Before and after each session, patients self-reported their wellbeing level, using the Outcome Rating Scale (ORS; Miller et al., 2003). The results indicated that both the patients’ and the therapists’ levels of arousal were “pulled” towards the other party’s arousal level, and patients were “pulled” by their therapists’ vocal arousal towards their own baseline. Higher levels of interpersonal downregulation were associated with better session outcomes. The findings advance the idea that therapists who are synchronized with their patients, but at the same time downregulate their own and their patients’ affect, may be more successful in helping their patients develop better affective regulation capabilities.

In addition, our systematic review highlighted that while most previous studies assessed dyadic processes in one modality (e.g., movement, physiology, voice), it is crucial to include a multi-modal assessment of dyadic processes that takes different behavioral and biological processes into account and to examine how these processes interact with other processes throughout therapy. Figure 3 illustrates the importance of considering synchrony in multiple modalities simultaneously.

Graphical user interface, chart

Description automatically generated*Figure 3. Illustration of Dyadic Synchrony in Different Modalities.*

*Note*.Illustration of dyadic signals extracted from 15 minutes of a successful psychotherapy session. Four types of signals are presented: the Electrodermal Activity (EDA; panel A), the Respiratory Sinus Arrhythmia (RSA; panel B), the affective valence extracted from facial expression (panel C), and the affective arousal extracted from vocal features (panel D). The physiological channels (Panels A and B) indicate that the dyad had moments of relative synchrony; however, these moments also present a time lag between signals, indicating the therapist’s effort to emotionally attune to the client, and the dyad’s dynamic movement in-and-out of synchrony. The facial expression channel (Panel C) shows that the session began with client’s negative emotional expressions that were met with relatively neutral affect by the therapist. Subsequently, the client expressed more positive emotions that were met with stronger positive expressions by the therapist. The vocal arousal channel (Panel D) shows that when the client experienced high arousal the therapist had lower arousal, a dynamic which could be evidence of a pattern of coregulation.

Furthermore, the systematic review suggested that differences in patterns of dyadic processes should be considered. Synchrony is commonly examined as a positive association between two persons’ stream of bio-behavioral signals. One disadvantage of this approach is that it overlooks the fact that successful interpersonal coordination requires partners to dynamically move toward and away from each other (Mayo & Gordon, 2020). Recent models of synchrony suggest that people tend to move in and out of synchrony while they interact, and that these interpersonal dynamics are complementary and adaptive (Feldman, 2021; Mayo & Gordon, 2020).

In one recent study that implemented these suggested directions (Sayda et al., 2024), we examined in a sample of 58 patient-therapist dyads whether flexibility in synchrony in five different modalities (body movement, facial expressions, heart rate, respiratory sinus arrhythmia, and electrodermal activity) is associated with session-level outcome. The results indicated that higher body movement flexibility in synchrony significantly predicted better session outcomes. Facial expressions flexibility in synchrony moderated the synchrony-outcome association, with higher synchrony predicted better outcomes when flexibility in synchrony was high, and poorer outcomes when flexibility in synchrony was low. No significant effects were found for physiological modalities. These findings suggest that multimodality and flexibility in synchrony may elucidate mixed findings on when synchrony is beneficial and when it is not.

Many of our studies focused on the dialogue between patients and therapists and used advanced NLP technique to first scale up psychotherapy data by automatically annotating key aspect of the therapeutic interaction, and then drilling down to explore processes associated with improved therapeutic outcome. In one such study, (Tsakalidis et al., 2020), we trained a logistic regression model to automatically detect the occurrence of a rupture from the text of the session, as rated by patients and therapists at the end of the session. The results indicated that the models achieved 83% accuracy in automatically identifying sessions in which both the patient and the therapist reported a rupture. We were particularly interested in sessions where the patient reported a rupture, but the therapist did not recognize its occurrence. Our models identified 40% of these instances. Interestingly, a qualitative analysis revealed that it was easier for therapists in our sample to identify confrontational ruptures (i.e., the patient moves against the therapist by expressing anger or dissatisfaction with the therapist or treatment, or by trying to apply pressure on the therapist) than to identify withdrawal ruptures (i.e., the patient moves away from the therapist and the work of therapy). This finding is consistent with previous research (Eubanks et al., 2018) showing that therapists tend to better recognize confrontational ruptures and highlights the importance of using automated methods to capture ruptures that are challenging for therapists to detect.

In another study (Mayer et al., 2024), we used advance deep learning models, to identify patients’ emotions and therapists’ interventions. 196 sessions were manually labelled, speech turn by speech turn. Then, the models were trained to automatically label the rest of the data. The results indicated that the models achieved 66% accuracy in detecting patients’ emotions and 64% in detecting therapists’ interventions. These scaling up of the data allows us to explore in the next steps, sequences of therapists’ intervention and patients’ emotional responses that are predictive of positive outcomes. These insights into intrapersonal and interpersonal dynamics predictive of positive treatment outcomes could be integrated into AI-based support systems for therapists and self-help tools for patients. Utilizing LLM-based systems for monitoring, assessment, and interventions opens novel pathways to ensure that more individuals receive the help they need. The abilities of LLMs to emulate human interaction could be used to analyze large-scale corpora of psychotherapy data, identify patients’ states and therapists’ discrete interventions, and then recommend those interventions most likely to prove effective in the context of specific states. By giving patients direct access to psychological insights regarding what is most likely to be effective in specific contexts, and by augmenting therapists’ ability to select effective courses of intervention, this approach can help democratize access to evidence-based clinical knowledge.

**Summary Conclusions and Implications for Practice and Training**

Several key insights emerged from our studies. On the intrapersonal level, our findings highlight the central role of flexibility as a robust predictor of treatment outcome. Flexibility in emotional experience, facial expressions, voice, cognitions, and even physiology, was associated with fewer symptoms and better treatment outcomes. These findings suggest that therapist may be encouraged to pay more attention to what occurs in both verbal and non-verbal channels, to help their clients increase their psychological flexibility, thereby helping them achieve better well-being. Our findings also revealed the importance of clients’ experience of positive emotions and therapists’ tendency to neglect these emotions and to fucus on negative emotions. By underestimating the importance of positive emotions, therapists may miss opportunities to leverage these emotions for positive change.

Our studies also demonstrated the potential of advanced NLP methods to identify patterns in verbal communication predictive of positive outcomes. For example, we showed the usefulness of these methods in identifying topics associated with positive outcomes and detecting signs of suicidal risk. Such measures could be integrated into monitoring systems, providing a more precise assessment of individuals' progress over time.

On the interpersonal level, we found that synchrony between patients and therapists across multiple communication channels—such as movement, physiology, voice, and facial expression—was often linked to positive outcomes. However, our studies also suggested that the benefits of synchrony depend on additional factors, including the therapists' ability to induce coregulation patterns and the emotional context in which synchrony occurs. Effective therapeutic outcomes appear to hinge on a flexible movement between synchronized and unsynchronized states. Implementing the practice-oriented research approach, we have translated these insights back to the clinicians in our departmental clinic. For instance, based on our findings on the importance of therapists' empathic accuracy, our feedback system visualizes the extent to which therapists were empathically accurate with their patients, allowing them to enhance their attunement to their patients' fluctuating emotions from session to session.

Moreover, we demonstrated the ability of advanced NLP techniques to automatically detect important information about patient states and therapeutic interactions. The models we developed automatically annotated nuanced elements within psychotherapy sessions, such as patients' emotions and therapists' interventions and their performance was comparable to human inter-rater reliability. This automatic annotation offers a cost-effective means of analyzing large datasets and yielding more reliable insights into the processes driving therapeutic progress. Recent findings from our studies also highlight the usefulness of LLMs in generating clinically meaningful information, such as formulations of individuals’ mental states. These innovations can be integrated in session-by-session support tools, as well as in training and supervision, to enhance therapists skills. For example, with the help of their supervisors, therapists can explore the feedback provided by the AI system and identify their clients’ adaptive and mal-adaptive states as well as their own helpful and less helpful interventions to better select the most appropriate interventions for a specific therapist to employ with a specific client at a specific time. We are now developing an advanced feedback system that uses LLMs to generate clinically meaningful summaries, patients’ formulation and suggested sequences of interventions, based on the session verbal and non verbal data.

Despite the exciting potential of AI and LLMs to assist in optimizing mental health services, several challenges must be addressed to ensure their effective and ethical integration. LLMs are prone to factual errors, often called 'hallucinations' and have known biases (Bang et al. 2023). As these systems continue to advance in their power and capabilities, there are concerns that they may pursue objectives that do not fully align with patient needs. Additionally, the "black box" nature of AI models can hinder their adoption by mental health professionals, making it imperative that AI suggestions are explainable and transparent. These concerns underscore the importance of the practice-oriented research approach in fostering clinician-researcher collaboration for integrating AI-based monitoring and feedback systems into clinical routines. These systems must be developed with input from clinicians, guided by patient preferences, and aligned with clinical knowledge. It is crucial that therapists are adequately trained to use these technologies, ensuring that AI systems augment rather than replace clinical decision-making. There is still much work to be done in deepening our understanding of the mechanisms driving therapeutic change and refining our ability to tailor interventions to the specific needs of patients. By grounding our research questions in real-world practice and fostering interdisciplinary collaboration between psychotherapy researchers and AI experts, we can develop more appropriate methods to explore the dynamics that underlie therapeutic change. The ability to study these questions using large, naturalistic samples rich with multimodal data, capturing both verbal and non-verbal interactions between therapists and patients, brings us closer to uncovering more effective solutions to these pressing challenges.

**References**

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders. Fifth edition, 21(21), 591-643.‏

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2021). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research: Journal of the Society for Psychother­apy Research*, *31*(1), 92–116. [https://doi.org/10.1080/10503307. 2020.1808729](https://doi.org/10.1080/10503307.%202020.1808729)

Arntz, A., & Jacob, G. (2013). Schema therapy in practice: An introductory guide to the schema mode approach. Wiley-Blackwell.

Aron, L., & Harris, A. (2014). *Relational psychoanalysis: Vol. 5*. *Evolution of process*. Routledge.

Atzil-Slonim, D. (2024). Self-Other Dynamics: A Transtheoretical Coding Manual.‏ Preprint PsyArXiv,  <https://doi.org/10.31234/osf.io/tk268>

Atzil-Slonim, D., Eliassaf, A., Warikoo, N., Paz, A., Haimovitz, S., Mayer, T., & Gurevych, I. (2024). Leveraging natural language processing to study emotional coherence in psychotherapy. *Psychotherapy*.‏

Atzil-Slonim, D., Soma, C. S., Zhang, X., Paz, A., & Imel, Z. E. (2023). Facilitating dyadic synchrony in psychotherapy sessions: Systematic review and meta-analysis. *Psychotherapy Research*, *33*(7), 898-917.‏

Atzil-Slonim, D., Stolowicz-Melman, D., Bar-Kalifa, E., Gilboa-Schechtman, E., Paz, A., Wolff, M., ... & Feldman, R. (2022). Oxytocin reactivity to the therapeutic encounter as a biomarker of change in the treatment of depression. *Journal of Counseling Psychology*, *69*(5), 755.‏

Atzil-Slonim, D., & Tschacher, W. (2020). Dynamic dyadic processes in psychotherapy: Introduction to a special section. *Psychotherapy Research*, *30*(5), 555-557.‏

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., ... & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. arXiv preprint arXiv:2302.04023.‏

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.‏

Bar-Kalifa, E., Prinz, J. N., Atzil-Slonim, D., Rubel, J. A., Lutz, W., & Rafaeli, E. (2019). Physiological synchrony and therapeutic alliance in an imagery-based treatment. Journal of Counseling Psychology, 66(4), 508–517.

Barrett Feldman, L., Lewis, M., & Haviland-Jones, M. J. (2016). Handbook of emotions: Vol. 1 (4th ed.). Guilford Press.

Beck, A. T., Steer, R. A.,& Brown, G. K. (1996). Beck Depression Inventory (BDI-II) (Vol. 10, p. s15327752jpa6703\_13). Pearson.

Beck, A. T., Finkel, M. R., & Beck, J. S. (2021). The theory of modes: Applications to schizophrenia and other psychological conditions. Cognitive Therapy and Research, 45, 391-400. <https://doi.org/10.1007/s10608-020-10098-0>

Bilu, Y., Kalkstein, N., Gilboa-Schechtman, E., Akiva, P., Zalsman, G., Itzhaky, L., & Atzil-Slonim, D. (2023). Predicting future onset of depression among middle-aged adults with no psychiatric history. *BJPsych Open*, *9*(3), e85.‏

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

Bromberg, P. M. (1998). Standing in the spaces: Essays on clinical process, trauma, and dissociation. Analytic Press.

Butler, E. A. (2015). Interpersonal affect dynamics: It takes two (and time) to tango. Emotion Review, 7(4), 336–341.

Castonguay, L.G. (2011) Psychotherapy, psychopathology, research and practice: Pathways of connections and integration. *Psychotherapy Research, 21*, 125-140, DOI: [10.1080/10503307.2011.563250](https://doi.org/10.1080/10503307.2011.563250)

Castonguay, L. G., Barkham, M., Youn, S. J., & Page, A. C. (2021). Practice-based evidence- findings from routine clinical settings. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and garfield’s handbook of psychotherapy and behavior change* (pp. 129–186). Wiley.

Chim, J., Tsakalidis, A., Gkoumas, D., Atzil-Slonim, D., Ophir, Y., Zirikly, A., ... & Liakata, M. (2024). Overview of the clpsych 2024 shared task: Leveraging large language models to identify evidence of suicidality risk in online posts. In *Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024)* (pp. 177-190).‏

Cohn, J. F., Cummins, N., Epps, J., Goecke, R., Joshi, J., & Scherer, S. (2018). Multimodal assessment of depression from behavioral signals. In S. Oviatt, B. Schuller, P. R. Cohen, D. Sonntag, G. Potamianos, & A. Krüger (Eds.), The handbook of multimodal multisensory interfaces: Signal processing, architectures, and detection of emotion and cognition (Vol. 2, pp. 375–417). https://doi.org/10.1145/3107990.3108004

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

Constantino, M. J., Boswell, J. F., & Coyne, A. E. (2021). Patient, therapist, and relational factors. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and garfield’s handbook of psychotherapy and behavior change* (pp. 225–262). Wiley.

Crits-Christoph, P., & Gibbons, C. MB (2021). Psychotherapy process–outcome research: Advances in understanding causal connections. Bergin and Garfield’s handbook of psychotherapy and behavior change, 263-295.‏

Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., ... & Furukawa, T. A. (2021). Psychotherapies for depression: a network meta‐analysis covering efficacy, acceptability and long‐term outcomes of all main treatment types. World Psychiatry, 20(2), 283-293. <https://doi.org/10.1002/wps.20860>

Cwir, D., Carr, P. B., Walton, G. M., & Spencer, S. J. (2011). Your heart makes my heart move: Cues of social connectedness cause shared emotions and physiological states among strangers. *Journal of Experimental Social Psychology*, *47*(3), 661-664.‏

Delgadillo, J., & Atzil-Slonim, D. (2022). Artificial Intelligence and Machine Learning in Psychotherapy. Encyclopedia of Mental Health (third edition). Elsevier.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pretraining of deep bidirectional transformers for language understanding. arXiv Preprint arXiv:1810.04805.

Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance rupture repair: A meta-analysis. Psychotherapy, 55(4), 508.‏

Feldman, R. (2012). Bio-behavioral Synchrony: A Model for Integrating Biological and Microsocial Behavioral Processes in the Study of Parenting. In *Parenting* (Vol. 12, Issues 2-3, pp. 154–164). <https://doi.org/10.1080/15295192.2012.683342>

Feldman, R. (2021). Social Behavior as a Transdiagnostic Marker of Resilience. In *Annual Review of Clinical Psychology* (Vol. 17, Issue 1, pp. 153–180). https://doi.org/10.1146/annurev-clinpsy-081219-102046

Fosha, D. (2001). The dyadic regulation of affect. Journal of Clinical Psychology, 57.

Gaines, A. N., & Goldfried, M. R. (2021). Consensus in psychotherapy: Are we there yet?. *Clinical Psychology: Science and Practice*, *28*(3), 267.‏

Girard, J. M., Cohn, J. F., Jeni, L. A., Sayette, M. A., & De la Torre, F. (2015). Spontaneous facial expression in unscripted social interactions can be measured automatically. Behavior Research Methods, 47(4), 1136–1147. <https://doi.org/10.3758/s13428-014-0536-1>

Goren, O., Paz. A., Bar-Kalifa, e., Gilboa-Schechtman, E., & Atzil-Slonim, D., (2022). Respiratory Sinus Arrhythmia (RSA) as a within session biomarker in depression psychotherapy: integrating resting RSA and RSA reactivity. A paper presented at the European chapter conference of the Society for Psychotherapy Research. Rome.

Hill, C. E., & Kellems, I. S. (2002). Development and use of the Session Evaluation Questionnaire. In G. S. Tryon (Ed.), *Counseling based on process research: Applying what we know* (pp. 235-246). Allyn & Bacon.

Hill, C. E., & Norcross, J. C. (Eds.). (2023). *Psychotherapy skills and methods that work*. Oxford University Press.

Höfling, T. T. A., Gerdes, A. B. M., Föhl, U., & Alpers, G. W. (2020). Read my face: Automatic facial coding versus psychophysiologi­cal indicators of emotional valence and arousal. *Frontiers in Psy­chology*, *11*. https://doi.org/10.3389/fpsyg.2020.01388. https:// [www.frontiersin.org/article/](http://www.frontiersin.org/article/)

Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. Clinical Psychological Science, 7(1), 37-50. https://doi.org/10.1177/2167702618772296

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), 901–930. <http://dx.doi.org/10.1037/a0038822>

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. D. L., ... & Sayed, W. E. (2023). Mistral 7B. arXiv preprint arXiv:2310.06825.‏

Kazdin, A. E. (2008). Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. American psychologist, 63(3), 146. <https://doi.org/10.1037/0003-066X.63.3.146>

Kazdin, A. E. (2021). Extending the scalability and reach of psychosocial interventions. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), Bergin and Garfield’s handbook of psychotherapy and behavior change (pp. 763–790). Wiley.

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

Lazarus, G., & Rafaeli, E. (2023). Modes: Cohesive personality states and their interrelationships as organizing concepts in psychopathology. Journal of Psychopathology and Clinical Science, 132(3), 238. <https://doi.org/10.1037/abn0000699>

Laurer, M., Van Atteveldt, W., Casas, A., & Welbers, K. (2024). Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and BERT-NLI. Political Analysis, 32(1), 84-100.‏

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).

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.

Mayer, T., Warikoo, N., Eliassaf, A., Atzil-Slonim, D., & Gurevych, I. (2024). Predicting Client Emotions and Therapist Interventions in Psychotherapy Dialogues. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1463-1477).‏

Mayo, O., & Gordon, I. (2020). In and out of synchrony—Behavioral and physiological dynamics of dyadic interpersonal coordination. In Psychophysiology (Vol. 57, Issue 6). https://doi.org/10.1111/psyp.13574

McCullough, L., Kuhn, N., Andrews, S., Valen, J., Hatch, D., & Osimo, F. (2003). The reliability of the Achievement of Therapeutic Objectives Scale (ATOS): A research and teaching tool for psychotherapy. Journal of Brief Therapy, 2, 75-90

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.

Pascual-Leone, A., & Greenberg, L. S. (2005). Classification of affective-meaning states (CAMS). In A. Pascual-Leone (Ed.), Emotional processing in the therapeutic hour: Why the only way out is through (pp. 289–367). Unpublished doctoral thesis. Toronto, ON: York University.

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.

Paz, A., Rafaeli, E., Bar-Kalifa, E., Gilboa-Schectman, E., Gannot, S., Narayanan, S., & Atzil-Slonim, D. (2024). Multimodal Analysis of Temporal Affective Variability within Treatment for Depression.‏

Reich, C. M., Berman, J. S., Dale, R., & Levitt, H. M. (2014). Vocal Synchrony in Psychotherapy. In *Journal of Social and Clinical Psychology* (Vol. 33, Issue 5, pp. 481–494). https://doi.org/10.1521/jscp.2014.33.5.481

Revelle, W. (2007). Experimental approaches to the study of personality. In B. Robins, R. C. Fraley, & R. F. Krueger (Eds.), Handbook of Research Methods in Personality Psychology (37–61). Guilford Press.

Safran, J. D. (2012). Doublethinking or dialectical thinking: A critical appreciation of Hoffman's “doublethinking” critique. Psychoanalytic Dialogues, 22(6), 710-720.‏

Sayda, D., Bar-Kalifa, E., & Atzil-Slonim, D., (2024). Can multimodality and flexibility explain mixed findings on synchrony-outcome association? A paper presented at the conference of the Israeli Society for psychotherapy Research.

Schoenherr, D., Strauss, B., Stangier, U., & Altmann, U. (2021). The influence of vocal synchrony on outcome and attachment anxiety/avoidance in treatments of social anxiety disorder. *Psychotherapy* , *58*(4), 510–522.

Song, J., Chim, J., Tsakalidis, A., Ive, J., Atzil-Slonim, D., & Liakata, M. (2024, August). Combining Hierachical VAEs with LLMs for clinically meaningful timeline summarisation in social media. In *Findings of the Association for Computational Linguistics ACL 2024* (pp. 14651-14672).‏

Summers, R. F., Barber, J. P., & Zilcha-Mano, S. (2024). *Psychodynamic Therapy*. Guilford Publications.‏

Stade, E., Stirman, S. W., Ungar, L. H., Yaden, D. B., Schwartz, H. A., Sedoc, J., DeRubeis, R., Willer, R., & Eichstaedt, J. C. (2023). *Artificial intelligence will change the future of psychotherapy: A proposal for responsible, psychologist-led development*.

Stiles, W. B. (2001). Assimilation of problematic experiences. Psychotherapy: Theory, Research, Practice, Training, 38(4), 462. <https://doi.org/10.1037/0033-3204.38.4.462>

Tsakalidis, A., Atzil-Slonim, D., Polakovski, A., Shapira, N., Tuval-Mashiach, R., & Liakata, M. (2021). Automatic Identification of Ruptures in Transcribed Psychotherapy Sessions. In Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access (pp. 122-128).‏

Wampold, B. E., & Owen, J. (2021). Therapist effects: History, methods, magnitude. Bergin and Garfield's handbook of psychotherapy and behavior change, 297-326.

World Health Organization. (2022). International Classification of Diseases: World Health Organization. *Geneva, Switzerland: Author*.‏