The Mechanisms of Change in the Treatment of Borderline Personality Disorder With Transference Focused Psychotherapy

Kenneth N. Levy  
*Pennsylvania State University and Joan and Sanford I. Weill Medical College of Cornell University*

John F. Clarkin and Frank E. Yeomans  
*Joan and Sanford I. Weill Medical College of Cornell University*

Lori N. Scott and Rachel H. Wasserman  
*Pennsylvania State University*

Otto F. Kernberg  
*Joan and Sanford I. Weill Medical College of Cornell University*

We address how Transference Focused Psychotherapy (TFP) conceptualizes mechanisms in the cause and maintenance of borderline personality disorder (BPD) as well as change mechanisms both within the patient and in terms of specific therapists’ interventions that engender patient change. Mechanisms of change at the level of the patient involve the integration of polarized representations of self and others; mechanisms of change at the level of the therapist’s interventions include the structured treatment approach and the use of clarification, confrontation, and “transference” interpretations in the here and now of the therapeutic relationship. In addition, we briefly review evidence from our group regarding the following hypothesized mechanisms of change: contract setting, integration of representations, and changes in reflective functioning (RF) and affect regulation. © 2006 Wiley Periodicals, Inc. J Clin Psychol 62: 481–501, 2006.

Keywords: mechanisms of change; Transference Focused Therapy; borderline personality disorder

We would like to thank Lindsay L. Hill for assistance with the article.  
Correspondence concerning this article should be addressed to: Kenneth N. Levy, Department of Psychology, Pennsylvania State University, 240 Moore Bldg., University Park, PA 16802; e-mail: klevy@psu.edu.
Borderline personality disorder (BPD) is a serious and prevalent psychiatric problem characterized by affective instability, angry outbursts, frequent suicidality and parasuicidality, and marked deficits in the capacity to work and to maintain meaningful relationships. Epidemiological, prevalence, and longitudinal studies suggest that BPD affects approximately 1–4% of the general population, 10–15% of psychiatric outpatients, and up to 20% of psychiatric inpatients (Lenzenweger, Loranger, Korfine, & Neff, 1997; Paris, 1999; Torgersen, Kringlen, & Cramer, 2001; Weissman, 1993; Widiger & Frances, 1989; Widiger & Weissman, 1991; Zimmerman, Rothschild, & Chelminski, 2005). In adult clinical outpatient and inpatient samples, the majority of patients are women; however, both forensic and veteran populations reflect high levels of BPD in men (Southwick, Yehuda, & Giller, 1993, Timmerman & Emmelkamp, 2001), and community samples find a relatively even distribution of men and women (Lenzenweger et al., 1997). One study examining prevalence in a primary care waiting room found 6% of patients met the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) criteria for BPD (Gross et al., 2002).

BPD is highly comorbid with other personality disorders, as well as with a number of Axis I disorders, most notably depression, anxiety, eating disorders, posttraumatic stress disorder, and substance abuse (Zanarini et al., 1999). Zanarini and colleagues (Zanarini et al., 1999) found that BPD could be depicted by a pattern of what she called complex comorbidity, characterized by multiple comorbid diagnoses that included both internalizing and externalizing disorders. Consistently with this finding, Grilo and colleagues (Grilo, Becker, Walker, Edell, & McGlashan, 1997) found that 86% of those meeting criteria for major depression and substance abuse were comorbid for BPD. This is particularly problematic in relation to the finding that treatment outcome studies of Axis I disorders that included comorbid BPD patients have found that BPD has detrimental effects on the treatment of the Axis I disorders—negatively affecting both the psychotherapeutic and psychopharmacological treatment efficacy for these disorders (see Clarkin, 1996). Thus, much of what we know about empirically supported treatments for Axis I disorders can be discarded when the patient has a comorbid diagnosis of borderline personality disorder.

Not surprisingly, patients who have borderline personality disorder utilize higher levels of services in emergency rooms, day hospital and partial hospitalization programs, outpatient clinics and inpatient units (Bender et al., 2001). For example, although borderline patients made up only 1% of the patient population seen in a psychiatric emergency room, they accounted for 12% of all visits (Bongar, Peterson, Golann, & Hardiman, 1990) and 20% of psychiatric hospitalizations (Zanarini & Frankenburg, 2000). In addition, patients who have BPD constitute up to 40% of frequent recidivists in psychiatric hospitals (Geller, 1986; Swigar, Astrachan, Levine, Mayfield, & Radovich, 1991). A survey of Australian psychiatrists found that although patients who have personality disorders represented only 6 percent of the patients in treatment, they accounted for 13 percent of the psychiatrists’ treatment time (Andrews & Hadzi-Pavlovic, 1988). Moran, Jenkins, Tylee, Blizard, and Mann (2000) found that those who have personality disorders (PDs) are more likely than those who do not have them to consult their general medical practitioner on an emergency basis.

Patients who have BPD typically experience profound impairment in general functioning (Bender et al., 2001; Skodal et al., 2002); for example, these patients are often unemployed or underemployed in relation to their capacities, training, and socioeconomic status. Perhaps the most perplexing symptom is the high level of parasuicidality (Clarkin, Widiger, Frances, Hurt, & Gilmore, 1983; Zisook, Goff, Sledge, & Schucter, 1994), which ranges from 69% to 80%. Parasuicidal behavior predicts suicidality and patients
who have BPD have an estimated suicide completion rate of between 8% and 10% (McGlashan, 1986; Stone, 1983).

Further compounding these problems, patients who have borderline personality disorder are notoriously difficult to treat. The disorder is characterized by high rates and chaotic use of medical and psychiatric services, repeated patterns of dropout, erratic psychotherapy attendance, refusal to take medications as prescribed, and pervasive non-compliance (Gunderson et al., 1989; Kelly, Soloff, Cornelius, George, & Lis, 1992; Skodal, Buckley, & Charles, 1983; Waldinger & Frank, 1989; Waldinger & Gunderson, 1984). Thus, BPD is a debilitating and life-threatening disorder that represents a serious clinical and public health concern.

Several psychotherapy studies have reported evidence for the efficacy (Bateman & Fonagy, 1999; Koons et al., 2001; Linehan, Armstrong, Suarez, Allmon, & Heard, 1991; Linehan, Kanter, & Comtois, 1999; Linehan et al., 1999; Linehan et al., 2002; Turner, 2000; Verheul et al., 2003) and effectiveness (Blum, Pröhl, & St. John, 2002; Brown, Newman, Charlesworth, 2004; Clarkin, et al., 2001; Ryle & Golyenkina, 2000; Stevenson & Meares, 1992) of specific treatments for patients who have BPD. Furthermore, studies testing the effectiveness and efficacy of new treatments have recently been completed (Giesen-Bloo & Arntz, 2000; Clarkin, Levy, Lenzenweger, & Kernberg, 2005) or are currently being conducted (Markowitz, Skodol, Bleiberg, & Strasser-Vorus, 2004).

One such treatment that has garnered effectiveness and efficacy data in clinical trials is Transference Focused Psychotherapy (TFP; Clarkin et al., 1999, 2001, 2005), a highly structured, twice-weekly modified psychodynamic treatment based on Kernberg’s (1984) object relations model of BPD. Recent studies have demonstrated TFP’s effectiveness in using patients as their own controls (Clarkin et al., 2001) and in comparison to a treatment-as-usual BPD group (Levy, Clarkin, Foelsch, & Kernberg, 2004). In addition, a randomized control trial (Clarkin, Levy, Lenzenweger, & Kernberg, 2004) comparing TFP, Dialectical Behavioral Therapy (DBT), and supportive psychotherapy (SPT) found reduced suicidality and anger in patients treated with TFP and DBT, but not in those treated with SPT; all three treatments were effective in reducing depression and anxiety and in improving global functioning and social adjustment. Only TFP was consistently related to reductions in aggression (Clarkin et al., 2005) and only TFP showed increases in personality change as indicated by changes in attachment coherence and reflective function1 (Levy, Kelly, Meehan, Reynoso, Clarkin, Lenzenweger, & Kernberg, 2005). Both the findings about aggression and personality organization are important given the focus in TFP on aggression, reflectiveness, and increased integration of representations of self and others.

What is becoming clear is that although BPD is a chronic problem functionally, it is also a highly treatable disorder (Leichsenring & Leibing, 2003; Oldham et al., 2001; Perry, Banon, & Ianni, 1999). What remain uncertain, however, are the mechanisms in the development and maintenance of BPD, the processes of change within patients during treatment, and the specific therapeutic techniques that bring about such changes. Therefore, despite the support for the effectiveness and even efficacy of existing treatments for BPD, researchers are still confronted with a high degree of uncertainty about the underlying processes of change.

In order to conceptualize change mechanisms in psychotherapy, one must address the question of how borderline personality disorder develops. In a series of important articles for conceptualizing child psychotherapy, which are also relevant to the adult literature, Kazdin (1999, 2000, 2001, 2004) proposed that the first stages in treatment

---

1Fonagy (Fonagy et al., 1997) notes that the term reflective function (RF) refers to the psychological processes underlying the capacity to mentalize, a concept that has been described in both the psychoanalytic (Fonagy, 1991; Fonagy & Higgitt, 1989) and cognitive psychology literatures (Morton & Frith, 1995).
development are the following: (1) to elaborate the core affective, cognitive, and behavioral mechanisms involved in the development and maintenance of a specific clinical problem; (2) to study multifinality and equifinality for understanding the heterogeneity of a disorder; (3) to understand the relationship between social environments and biological predispositions; and (4) to understand developmental processes, pathways, and the various courses that the disorder takes. Understanding these questions leads to two additional questions about mechanisms of change: (1) What changes occur within a person? (2) How and why do treatments work for a specific population of clients who have a particular disorder? Or, as Gordon Paul proposed: “What treatment, by whom, is most effective for this individual with that specific problem, and under which set of circumstances?” (Paul, 1967, p. 111). Thus, change mechanisms can be conceptualized at two levels: (1) what is hypothesized to change in the patient (e.g., increased self-esteem or emotion regulation skills, increased emotional stability, or increased mindfulness) and (2) what are the active ingredients in the treatment that elicit the change in the patient (e.g., the teaching of new skills, interpretation of transference, or provision of emotional support). In the present article we address how TFP conceptualizes mechanisms in the cause and maintenance of BPD as well as change mechanisms both within the patient and in terms of what the therapists does to engender change in the patient.

How Does TFP Conceptualize the Cause and Maintenance of BPD?

As do other theories of BPD (e.g., Bateman & Fonagy, 2003; Linehan, 1993), TFP conceptualizes the basic etiological elements of BPD as an interaction between constitutional and environmental factors that results in a personality structure or organization characterized by identity disturbance; use of immature or low-level defense mechanisms such as projective identification, splitting, and omnipotent control; and deficits in social reality testing (although perceptual reality testing is generally maintained). Regarding the interaction between biological constitution and environment, Kernberg (1984) posits that BPD patients have difficulty integrating disparate representations of themselves and others, in part, because negative emotions, particularly aggression, disrupt one’s capacity to integrate these partial representations. Strong unmetabolized or unprocessed emotions have the capacity to overwhelm positive representations. Kernberg further hypothesizes that individuals may therefore be motivated to keep these representations separate or split in an effort to protect the positive representations of themselves and others (or some combination of self and other representations). These high levels of negative emotionality and aggression can be constitutional or engendered through experience, or some combination of the two. Regardless of origin, high levels of aggression interfere with the normative developmental process of integrating disparate representations, and instead the high levels of aggression result in a division between positive and negative representations. Likewise, Siever and colleagues (Gurvits, Koenigsberg, & Siever, 2000) point out that affective instability may interfere with the ability to develop stable perceptions of self and others. They note that both the specific role of aggression and the more general role of affect lability may make the developmental task of integrating stable representations of self and others more difficult to accomplish. However, Kernberg and colleagues (Clarkin, Yeomans, & Kernberg, 2006) also note that emotional instability in borderline personality disorder can be secondary to a lack of differentiation and integration of internal images of self and others, which leads to instability in one’s sense

---

2 Although some authors (e.g., Doss, 2004) have differentiated between processes and mechanisms of change, these distinctions are often arbitrary and are unnecessary for our purposes; therefore, the terms are used interchangeably in this paper.
of self and ultimately to affective instability. Thus, the relationship between lack of integration of representations and affective instability may operate in a vicious circle, as the intensity of early affects results in a split experience of self and others to protect positive representations, which then may lead to further affective instability by failing to provide a foundation from which to understand oneself and others.

The relative influences of constitutional and environmental factors can vary in relation to each other. For example, the higher the constitutional disadvantage, the lower the threshold for environmental perturbations to overwhelm the child’s capacity to assimilate and accommodate to his or her environment. Conversely, a child who has a low constitutional load may be resilient to greater perturbations. In addition, family stressors may affect the developing child directly and through the effects on caregivers. Finally, the development of undifferentiated and unintegrated representation models of self and others may leave one vulnerable to life’s stressors and traumatic experiences. Research, which is consistent with this idea, suggests that patients who have borderline personality disorder can be differentiated from other psychiatric patients not on the basis of trauma, which is relatively high across psychiatric diagnoses, but instead on the basis of their lack of resolution of traumatic experiences and lack of capacity to reflect on such experiences (Fonagy et al., 1996; Patrick, Hobson, Castle, Howard, & Maughan, 1994). Patrick and colleagues (1994) found that BPD patients, compared with depressed patients, were no more likely to have had a history of trauma but were more likely to lack resolution of trauma events (75% vs. 20%). Fonagy and coworkers (1996) found that 97% of patients who have a history of abuse and low reflective functioning met criteria for BPD, whereas only 17% of abused patients who had high reflective function did. Thus, potentially traumatic experiences become traumatizing when the individual cannot adequately reflect on or integrate the experience into a fuller context.

Common factors are usually only conceptualized as elements of psychotherapy that converge across different treatment approaches; however, common factors may also exist in etiological theories. One potential common etiological mechanism in BPD, although articulated in subtly divergent ways, is what Kernberg (1984) calls identity diffusion (an absence of identity consolidation), what Bateman and Fonagy (2003) call a deficit or inhibition of mentalization, and what Linehan (1993) calls a deficit in mindfulness. All of these terms describe a lack of metasocial-cognitive ability to observe, reflect, and describe emotional states; predict and understand behavior; and recognize the difference between inner and outer reality and the capacity to reconcile opposing thoughts or mental states. Likewise, Kernberg, Fonagy, and Linehan all suggest environmental contributions to the development of BPD. Linehan (1993) posits an environment in which the child’s emotional experiences have been invalidated. Fonagy and colleagues (Fonagy, Target, Gergely, Allen, & Bateman, 2003) suggest a similar process by which the parent fails to comprehend the child’s mind. This process results in the child’s having difficulty thinking about his or her own mind and that of others and leaves the child holding the caregiver’s mind. Fonagy contends that the child experiences the caregiver’s projections as “alien” or as an “alien self,” which in Winnicott’s words is experienced as a “false self.” This alien self leaves the child feeling disconnected from his or her true internal world and needing to project the alien self onto others. Likewise, Kernberg describes similar experiences in childhood. Moving beyond these similarities, there are differences in the techniques of how these problems should be addressed and the question of whether these patients can achieve “normal” personality functioning.3 Kernberg’s (1984) concept of identity

---

3By normal personality we mean the achievement of satisfactory love relationships, work investments, and emotion regulation at levels consistent with those of nonpatients.
consolidation, which is characterized by the integration of mental representations of self, others, and affective experience, also appears to be strongly related to patterns of affect and self-regulation. Levy (2000) found that less differentiated and integrated representations of self and other were significantly related to the self-reported use of more maladaptive strategies (e.g., self-injurious behaviors, promiscuous sex, illicit drug use, and violent fantasies and behaviors) to regulate negative affective states.

What Putatively Changes in the Patient Treated With TFP and How Does the Therapist Facilitate Those Changes?

In TFP, hypothesized mechanisms of change derive from Kernberg’s (1984) developmentally based theory of BPD, which conceptualizes the disorder in terms of unintegrated and undifferentiated affects and representations (or concepts) of self and other. Partial representations of self and other are paired and linked by an affect in mental units called object relations dyads. These dyads are elements of psychological structure. In borderline pathology, the lack of integration of the internal object relations dyads corresponds to a “split” psychological structure in which totally negative representations are split off/segregated from idealized positive representations of self and other (Figure 1). The putative global mechanism of change in patients treated with TFP is the integration of these polarized affect states and representations of self and other into a more coherent whole. Through the exploration and integration of these “split off” cognitive-affective units of self and other representations, Kernberg postulates that the patient’s awareness and experience in life become more enriched and modulated, and the patient develops the capacity to think more flexibly, realistically, and benevolently. The integration of the split and polarized concepts of self and others leads to a more complex, differentiated, and realistic sense of self and others that allows for better modulation of affects and in turn clearer thinking (Figure 2). Therefore, as split-off representations become integrated, patients tend to experience increased coherence of identity, relationships that are balanced and not

Figure 1. Split organization: Separation of positive and negative representations and affects.
at risk of being overwhelmed by aggressive affect, greater capacity for intimacy, reduction in self-destructive behaviors, and general improvement in functioning. This initial conceptualization has been elaborated in light of an evolving developmental and neuroscience empirical literature (Clarkin, in press; Posner et al., 2003).

According to Kernberg and his colleagues (Clarkin et al., in press), in TFP the putative mechanisms of change at the level of the therapist’s interventions begin with the structured treatment approach (or what Bateman and Fonagy call a “theoretically coherent” treatment approach, e.g., the use of a treatment manual, treatment contract, hierarchy of problems addressed, and group supervision for therapists) and the use of clarification, confrontation (honest inquiry pointing out disparate information), and “transference” interpretations in the here and now of the therapeutic relationship (very similar to Kohlenberg’s (1994) work on functional analytic psychotherapy) (Figure 3). Using the triad of clarifications, confrontations and interpretations, the TFP therapist is thought to provide the patient with the opportunity to integrate cognitions and affects that were previously split and disorganized. In addition, the highly engaged, interactive, and emotionally intense stance of the therapist is thought to be experienced by patients as emotionally holding (i.e., containing) because the therapist conveys that he or she can tolerate the patient’s negative affective states. Furthermore, the therapist’s expectation of the patient’s ability to have a thoughtful and disciplined approach to emotional states (i.e., that the patient is a fledgling version of a capable, responsible, and reflective adult) is thought to be experienced as cognitively holding. The therapist’s timely, clear, and tactful interpretations of the dominant, affect-laden themes and patient enactments in the here and now of the transference are hypothesized to shed light on the reasons that representations remain split off and thus facilitate integrating polarized representations of self and others.

With regard to the flow of treatment, the structured frame of TFP facilitates the full activation of the patient’s distorted internal representations of self and other in the ongoing
The relationship between patient and therapist, which constitutes the transference. It is expected that the unintegrated representations of self and other will be activated in the treatment setting as they are in every aspect of the patient’s life. These partial representations are constantly active in determining the patient’s experience of real life interactions and in motivating the patient’s behavior. The difference in the therapy is that the therapist both experiences the patient’s representation of the interaction and nonjudgmentally observes and comments on it (within the psychoanalytic literature, which is known as the third position). This process is facilitated by the therapist’s establishing a treatment frame (e.g., contract), which, in addition to providing structure and holding for the patient and a consensual reality from which to examine acting out behavior, minimizes the therapist’s potential for acting in iatrogenic ways. The therapist does not respond to the patient’s fragmented one-dimensional partial representation but helps the patient observe it and the implied other that is paired with it.

As these internal object relations unfold in the relation with the therapist, the TFP therapist seeks cognitive clarification of the patient’s internal experience because the patient may not have a clear representation of his or her own experience. This technique of clarification appeals for explication of internal states and for reflection. However, in most cases this technique alone does not lead to integration because clarification alone does not address the conflicts that keep the partial representations separated. Confrontation—the technique of inquiring about the elements of the patient’s verbal and nonverbal communications that are in contradiction with each other—and interpretation of obstacles to integration are needed to allow the patient to progress beyond the level of split organization. Interpretation includes helping the patient see that he or she identifies at different moments in time with each pole of the predominant object relations dyads within him or her, that is, that self-representation and object representation can switch, often without the patient’s awareness. Increasing the patient’s awareness of his or her range of identifications increases his or her ability to integrate the different parts.
Movement toward integration initially causes anxiety because of the existence of the internal barriers that keep conflicting affects separate. The scenarios reexperienced with the therapist in the transference are not simply a literal reproduction of what the patient experienced in the past, but a mix of what happened, how the patient perceived what happened, and what the patient defensively set up to avoid awareness of aspects of conflicts that are consciously intolerable. Interpretations are hypotheses in which the therapist offers a cognitive formulation of the temporally split-off object relations as they are activated in the transference, and of the reasons that they remain separated. As the therapy advances, interpretations can address deeper levels of conflicts within the patient.

On the practical level, the relationship with the therapist in TFP is structured under controlled conditions in order to allow the patient to experience affects without their overwhelming the situation and destroying communication. The negotiation of a treatment frame provides a safe setting—a containment or holding environment—for the reactivation of the internalized relation paradigms. The safety and stability of the therapeutic environment permit the patient to begin to reflect about what is going on in the present with another person, in light of these internalized paradigms. The process is similar to what attachment theorists would describe as a safe haven, which along with the guidance of an attachment figure allows for the exploration of the content of the mind. With guidance from the therapist, the patient becomes aware of the extent to which his or her perceptions are based more on internalized representations than on what is occurring now. The therapist’s help to structure cognitively what at first seemed chaotic also provides a containing function for the patient’s affects.

TFP fosters change by allowing this reactivation of unintegrated object relations under controlled circumstances that inhibit the vicious circle of setting off reactions in others that often occurs when the patient behaves with emotion dysregulation in the “real” world (often eliciting the very responses that the patient fears from others). The objective and nonjudgmental attitude of the therapist (therapeutic neutrality)\(^4\) assists in the reactivation of the internalized experience patterns, their containment, and their exploration for new understandings. In this way TFP suspends the ordinary reaction of the social environment in reaction to a disturbed patient and lets the patient live out his or her internal representations in the treatment setting. Then, instead of attempting to deter these behaviors by educative means, TFP draws the patient’s attention to the internal mental representations behind them, with the goal of understanding, modifying, and integrating them. It is believed that this focus on the activation of the patient’s internal world in the therapeutic setting generally leads to a decrease in the level of acting out and chaos in the patient’s life outside the therapy. The therapist is careful to monitor conditions in the patient’s life at the same time as he or she focuses on what plays out in the therapy.

The treatment focus is on the current *psychic reality*,\(^5\) based on the split structure described, that is a fundamental motivational factor in the patient’s life. This structure is

\(\text{\cite{neutrality}}\)
the focus of modification in the treatment and is the reason a fundamental mechanism of change is calling the patient’s attention to the reactivation of split internal object relations in the relationship with the therapist. Key to this process is the development of introspection or self-reflection: the patient’s increase in reflection is an essential mechanism of change. The disorganization of the patient involves not only internal representations of self and others, relationships with self and others, and predominance of primitive affects, but also the processes that prevent reflection and full awareness. These primitive defensive processes that characterize a split psychological structure erase and distort awareness and thinking. BPD patients manifest a fragmentation and disconnection of thinking with attacks on the linking of thoughts (Bion, 1967), so the very thought processes are affected. Thought processes can be so powerfully distorted that affects, particularly the most negative ones, are expressed in action without cognitive awareness of their existence. The affect is only in the action, not in cognitive awareness.

We have described how the TFP therapist observes the material in the patient’s presentation and actions, seeking to explore, with the patient, the underlying object relations that motivate acting out and that constitute character structure. Mechanized, automated behavior is understood in terms of the internal relationship image(s) that gave origin to it, what attachment theorists call the internal working models. The concept of an internalized relational scenario that encompasses an image of self in interaction with another and that involves expectations of interpersonal transactions is common to object relations dyads and to the internal working model of attachment. Given the primitive disorganization of affects and of their connection with cognitive processes, the therapist’s helping delineate these primitive scenarios helps to contain the affect and, at the same time, facilitates the patient’s development of the cognitive capacity to represent affect. The therapist assists the patient in connecting cognition with abnormally dissociated and disorganized affect.

Evidence for Mechanisms of Change: Reflective Functioning as a Measure of Change

Both experimental research and psychotherapy research are useful for studying putative mechanisms of change. In this section, we present research findings from our group related to the hypothesized mechanisms of change explicated previously.

Treatment Contracting

One of the important tactics in TFP is the use of treatment contracts, which occurs before the treatment begins. The function of the contract is to define the responsibilities of patient and therapist, protecting the therapist’s ability to think clearly and reflect, provide a safe place for the patient’s dynamics to unfold, set the stage for interpreting the meaning of deviations from the contract as they occur later in therapy, and provide an organizing therapeutic frame that permits therapy to become an anchor in the patient’s life. The contract specifies the patient responsibilities, such as attendance and participation, paying of fees, and reporting of thoughts and feelings without censoring. The contract also specifies the therapist’s responsibilities, including attending to the schedule; making every effort to understand and, when useful, comment; clarifying the limits of his or her involvement; and predicting threats to the treatment. Essentially, the treatment contract makes the expectations of the therapy explicit (Clarkin, 1996). There is some controversy regarding the value of treatment contracting. The American Psychiatric Association guidelines
recommend that the therapist base the contract on issues of safety (Oldham et al., 2001). Others (Sanderson, Swenson, & Bohus, 2002) have suggested that the evidence contraindicates their use and shows them to be ineffective (Kroll, 2000). However, Kroll’s study (2000) was designed to determine the extent of no-suicide contracts (which was found to be 57%) and although 42% of psychiatrists who used no-suicide contracts had patients who either committed suicide or made a serious attempt, the design of the study does not allow for assessment of the efficacy of no-suicide contracts. Other data suggest the utility of contracting around self-destructive behavior and treatment threats (Clarkin et al., 2001; Clarkin et al., 2005; Levy et al., 2005; Smith, Koenigsberg, Yeomans, Clarkin, & Selzer, 1995; Yeomans et al., 1994). For example, Yeomans and colleagues (Yeomans et al., 1994) in a pre–post study of 36 BPD patients found that the quality of the therapist’s presentation and handling of the patient’s response to the treatment contract correlated with treatment alliance and the length of treatment. In addition, in our earlier work (Smith et al., 1995), when we did not stress treatment contracting, our dropout rates were high: 31% and 36% at the 3-month and 6-month marks of treatment. However, on the basis of the findings of Yeomans and associates (1994), we further systematized and stressed the importance of the treatment contract, and in later studies (Clarkin et al., 2001, 2005; Levy et al., 2005) we found lower rates of dropout (19%, 13%, and 25%) even over a year-long period of treatment. We suggest that these findings taken together suggest that the treatment contract may have the desired effect of resulting in less dropout and longer treatments. Future research will need to address the issue of treatment contracts more directly.

Integration and Reflectiveness

As the patient progresses in the course of TFP from split-off contradictory self-states to reflectiveness and integration, from action to reflection, this increase in reflectiveness involves two specific levels. The first level entails an articulation and reflection of what one feels in the moment. The patient increases in his or her ability to experience, articulate, and contain an affect and to contextualize it in the moment. This contextualizing of affect in the moment can involve complex and accurate perceptions of both what one (the patient) is experiencing and an understanding of what the other (e.g., the therapist) is experiencing. This level of reflective functioning (RF) may correspond to the therapeutic work of clarifying cognitively what the patient is experiencing in the moment.

A second, more advanced level of reflection is the ability to place the understanding of momentary affect states of self and others into a general context of a relationship between self and others across time. This level of RF reflects the establishment of an integrated sense of self and others—a sense against which momentary perceptions can be compared and put in perspective. We suggest this level of RF is achieved when the therapist moves on from the stage of clarifying the patient’s momentary perceptions of self and other to confronting the contrast and contradictions between different states within the patient’s psyche and interpreting the reasons that these internal states have remained split off. Borderline patients are quite sensitive, for example, to any behavior of others (i.e., the therapist) that suggests disrespect, a personal slight, or abandonment. Early in TFP, one patient experienced the therapist’s arriving 3 minutes late to a session on a snowy morning as proof that the therapist did not like her and did not want to see her. She reacted with rage and hatred in the moment. Gradually, she understood that the “pure” quality of her rage and hatred coexisted, but did not mix, with moments of experiencing her therapist as the “perfect” provider. She further understood that in her system,
negative affect could not come near the pure positive image of her therapist for fear that it would overwhelm and destroy the latter. As this understanding fostered a growing integration of her image of the therapist as someone who was well intentioned but not perfect, and toward whom she could experience anger, the patient could place frustrating interactions with him in the context of a relationship in which there were also support and consistency from him, so that her reaction at any moment was no longer determined solely by the event of the moment but was modulated by understanding of each specific event in the context of a broader internal image.

In a recently completed randomized control trial (Clarkin et al., 2004) comparing TFP, DBT, and supportive psychotherapy (SPT), only those patients randomized to the TFP condition showed increases in attachment coherence and reflective function after 1 year of treatment (Levy et al., 2005). In order to assess integration and reflectiveness of representation we used the Adult Attachment Interview (AAI), a semistructured clinical interview designed to elicit thoughts, feelings, and memories about early attachment experiences and to assess the individual’s state of mind or internal working model with regard to early attachment relationships. The interview consists of 20 questions asked in a set order with standardized probes. Individuals are asked to describe their childhood relationship with their parents, choosing five adjectives to describe each relationship and supporting these descriptors with specific memories. To elicit attachment-related information they are asked how their parents responded to them when they were in physical or emotional distress (e.g., during times when they were upset, injured, and sick as children). They are also asked about memories of separations, loss, experiences of rejection, and times when they might have felt threatened, including, but not limited to, those involving physical and sexual abuse. The interview requires that patients reflect on their parents’ styles of parenting and that they consider how their childhood experiences with their parents have influenced their life. The technique has been described as having the effect of “surprising the unconscious” (George, Kaplan, & Main, 1985) and allowing numerous opportunities for the interviewee to elaborate upon, contradict, or fail to support previous statements. The AAI is transcribed verbatim, and trained coders first score the transcripts by using subscale ratings, which are then used to assign individuals to one of five primary attachment classifications (secure/autonomous, dismissive, preoccupied, unresolved, and cannot classify). The unresolved classification can be the primary classification or the secondary classification in addition to the assignment of an organized style. The AAI is also scored with the Reflective Function Scale (Fonagy, Steele, Steele, & Target, 1997), a 9-point scale, which ranges from −1 (negative RF, in which interviews are overly concrete, totally barren of mentalization, or grossly distorting of the mental states of others) to 9 (exceptional RF, in which interviews show unusually complex, elaborate, or original reasoning about mental states). We believe that the coherence of narrative score and the reflective functioning (RF) score obtained from the Adult Attachment Interview are appropriate operationalized measures of the multilayered integrative and reflective process that Kernberg describes. We believe that the reason for change in coherence and reflective function in the TFP treatment, but not in the other two treatments, is that TFP specifically focuses on the integration of disparate mental states and representations. We hypothesize that the TFP therapists would make more bids or appeals for RF. Of course, a question arises about the value of RF as a construct. Fonagy and colleagues (Fonagy, Steele, Moran, Steele, & Higgitt, 1991) found that RF mediated the relationship between parental attachment security and infant attachment security in Ainsworth’s Strange Situation laboratory procedure (Ainsworth, Blehar, Waters, & Wall, 1978). Slade and colleagues (Grienenberger, Kelly, & Slade, 2005; Slade, Grienenberger, Bernbach, Levy, & Locker, 2005) found that a mother’s RF mediates the relationship
between atypical maternal behaviors and attachment security in her infants. Fonagy and colleagues (Fonagy et al., 1996) found that among psychiatric patients reporting abuse, those who scored low on RF were more likely to be diagnosed with BPD compared to those who were abused but scored high on RF. However, the importance of these findings is clouded by the relative paucity of research on RF. Certainly RF has been established as an important construct for thinking about parenting behavior; however, less is known about its broader significance. Establishing the validity of the concept is an important step to providing empirical support for the psychoanalytic notion of structural change as a central aspect of change in psychotherapy. Later we present data on the validity of the RF construct by relating it to external measures of neurocognitive functioning, specifically impulsivity and concept formation as measured by the Continuous Performance Task and the Wisconsin Card Sorting Test, respectively.

**Reflectiveness, Coherence, and Measures of Affect Regulation.** We have recently examined the relationship between reflective function scores and measures of affect regulation, including neurocognitive measures known to tap affect regulation (Levy et al., 2005). Three samples of participants were used in the study: patients diagnosed with BPD ($N = 24$), a nonclinical comparison group matched to the patient group in terms of impulsivity and negative affect, and a “comparison” group who had normal levels on these indices. Comparison participants were excluded from the study if they met criteria for any Axis II personality disorder. These samples were evaluated in terms of RF, scored from the Adult Attachment Interview, and computer-based neurocognitive tasks, the Continuous Performance Task (CPT) and the Wisconsin Card Sorting Test (WCST). RF was unrelated to gender, ethnicity, age, or intelligence quotient (IQ).

**Impulsivity.** On the CPT, RF was not correlated with $d'$, a measure of sustained attention. However, there was a significant inverse relationship between RF and $\beta$, a measure of impulsivity. These findings were similar in the BPD and the comparison samples. Given that sustained attention was maintained, the relationship between RF and impulsivity is likely not a function of poor attention.

**Concept Formation.** On the Wisconsin Card Sorting Test, there was no relationship between RF and nonperseverative errors, categories completed, or trials needed to complete the task. However, there was a significant inverse relationship between RF and both perseverative errors (i.e., when a participant persists with the wrong answer despite feedback that it is incorrect) and failure to maintain the set (i.e., when a participant changes a correct answer to an incorrect response despite feedback that the initial answer was correct). Higher RF was correlated with fewer errors of both types. These findings were significant in both the BPD and comparison samples. Perseverative errors are not uncommon and are characteristic of a number of other psychiatric conditions. Failure to maintain the set is a relatively rare response, and a review of the literature suggests that this type of error is characteristic of only one other psychiatric condition, schizotypy. The literature regarding schizotypes suggests that they make these types of errors because of attentional problems (Lenzenweger, personal communication, March 17, 2004). Given that sustained attention was maintained on the CPT, the relationship between RF and failure to maintain the set is likely not a function of poor attention. We suggest that the relationship is a function of a lack of contingency in the mind of patients who have BPD. We believe that this lack of contingency is a function of fluctuating mental states and unintegrated representations of experience.
Attention Networks. We (Posner et al., 2002) examined the attentional control system in more detail by utilizing the Attentional Network Task (ANT), developed by Posner and colleagues (Fan, McCandliss, Sommer, Raz, & Posner, 2002). Participants were 39 patients who had BPD, 20 control subjects who were matched to the patients in having very low self-reported effortful control and very high negative emotionality, and 30 control subjects who were average in these two temperamental functions. In the ANT, participants are to determine whether the central arrow in the target display points left or right. Before the target either no cue, a double cue, a single cue at the location of the upcoming target, or a single central cue is given. These three conditions are used to assess the efficiency of each network. Each subject was given a total of 256 trials—one-fourth in each of the four cue conditions. Each trial began with either no cue or one of the three cues. The cue was followed after a variable interval with a mean of 1 second by either a congruent, incongruent, or neutral target with equal frequency. Three aspects of attention are assessed: alerting, orientating, and conflict. Subtraction of the double-cue from the no-cue condition RTs provides a measure of the ability of subjects to maintain alertness on trials in which they are not cued and to take advantage of warning signal. Subtracting RTs to a cue at the target location (either above or below fixation) from a central cue (where no targets are ever presented) provides information on the skill of orienting to the target location. Finally by subtracting the congruent RTs from the incongruent RTs we have a measure of the ability of subjects to resolve conflict introduced by the flankers when they are in the opposite direction from the target. We found that the patients exhibited significantly greater difficulty in their ability to resolve conflict among stimulus dimensions in a purely cognitive task than did average control subjects. Temperamental control subjects also had elevated conflict scores, although they were not significantly different from those of the average control subjects. No other attentional network appeared to be impaired in these patients. Consistently with the idea of a constitutional bias, we conclude that temperament may play a role in the disorder, possibly in predisposing children to acquire it, but some other environmental or temperament factor must be involved. We also have some preliminary data suggesting that RF is correlated with the conflict score but not the alerting or orienting score.

Affect. Kernberg’s (1984) concept of identity consolidation, which is characterized by the integration of mental representations of self, others, and affective experience, also appears to be strongly related to patterns of affect and self-regulation. Using Blatt’s Object Representation Inventory, Levy (2000) found that less differentiated and integrated representations of self and other were significantly related to the self-reported use of more maladaptive strategies (e.g., self-injurious behaviors, promiscuous sex, illicit drug use, and violent fantasies and behaviors) to regulate negative affective states. These findings taken together suggest, as Kernberg and Fonagy contend, that deficits in reflective function may result in difficulties with impulsivity and affective instability.

Summary and Conclusions

TFP, similarly to most prominent theories of borderline personality disorder, hypothesizes an interaction between a constitutional emotional vulnerability and environmental experiences. The exact nature of the constitutional vulnerability can vary but is generally thought to involve an abundance of negative affect, particularly aggression, in ratio to positive affect. Research has shown that positive affect acts as a buffer to negative experiences, including one’s own internal negative mental states (Fredrickson, 1998). Without the buffer of positive affect, negative affect colors the perceptions of interactions and
distorts internal representations. Despite support from neurobiological and social personality research, some have criticized Kernberg’s hypothesis of a constitutional aspect of BPD. Some have argued that the negative affect and aggression seen in BPD patients are solely the result of real traumatic experiences. Although a large and significant portion of patients who have borderline personality disorder have experienced varying types of traumatic experiences, including sexual abuse, which may result in increased levels of aggression, many, and in some samples most, BPD patients have not experienced such traumatic events. In addition, many other disorders are characterized by equivalent types and levels of abuse but do not show the same symptom pattern or pervasive disruptions that are seen in BPD. Thus, although a significant proportion of BPD patients may have achieved their level of anger because of real experience, we need a theory that is broader than the simple “abuse equals aggression in BPD.” Within psychoanalytic circles, some have criticized Kernberg’s focus on constitutional factors, suggesting that they are not necessary for his conceptualization of BPD but instead are a vestige of his loyalty to Freudian theory (e.g., Mitchell, 1988). These critics ignore the growing evidence from neuroscience suggesting endogenous temperamental inputs that a child (and by extension, parents) must grapple with and metabolize. Nevertheless, there are probably multiple pathways for development of BPD, one of which involves serious, prolonged trauma, particularly sexual abuse, that, regardless of constitutional factors, would overwhelm almost anyone’s capacity to integrate the experience (Pynoos, 1993). Other pathways might include disruptions in psychological coherence resulting from high levels of aggression that cannot be metabolized normally.

Similar to Kernberg, Bateman and Fonagy contend that emotional instability in borderline personality disorder is secondary to unstable internal states. Bateman and Fonagy contend that either a deficiency or an inhibition of the capacity to mentalize leads to instability in one’s sense of self. Hence, the putative mechanism of change within patients who have BPD, according to Bateman and Fonagy, is an increase in the capacity to mentalize. This capacity is thought to increase emotional stability in BPD patients by allowing them to shift their attention when experiencing negative emotional states and to find more contextualized meaning in their own and other people’s behavior. Both Bateman and Fonagy (2003) and Kernberg (1984) hypothesize this capacity is a developmental achievement that occurs in the context of a secure attachment relationship to a caregiver, and that it is integrally related to one’s sense of self. Consistently with the hypothesis of a relation between a deficit in mentalization and borderline pathology, there is now empirical evidence to suggest that mentalization is an important mechanism of change within BPD patients who are treated with TFP (Levy et al., 2005).

The concept of mindfulness, which Linehan (1993) integrated into DBT treatment, bears a striking similarity to Bateman and Fonagy’s (2003) concept of mentalization and, as can mentalization, can be seen as a product of integrated representations (Levy, 2004). Mindfulness is becoming increasingly central in Linehan’s conceptualization of the treatment of BPD (Brodsky & Stanley, 2002; Dimidjian & Linehan, 2003; Robins, 2002). Mindfulness involves the ability to observe, reflect, and describe emotional states while developing focused attention, and it is thought to help BPD patients to develop perspective and tolerate and regulate negative affective experiences without being overwhelmed. According to Linehan, and similarly to Kernberg’s conceptualization, the patient begins to understand a separation between the observer and the observed, and thoughts are not taken as literally “true” and to be acted on. Although there are subtle differences, the concepts of mindfulness, mentalization, and integrated representation share important conceptual overlaps. Fonagy (Fonagy et al., 1997) notes that RF is intimately related to the representation of the self (Fonagy & Target, 1995, 1996; Target & Fonagy, 1996).
In sum, we have been able to show that during the course of TFP, consistently with hypotheses, there are changes in integration and reflectiveness, as assessed by the coherence and reflective function scores on the AAI, respectively. These changes are specific to TFP and parallel significant changes in outcome (Clarkin et al., 2005). Future research will examine in-therapy therapists’ technique as it relates to changes in coherence and RF (Levy, 2005).

References


guideline for the treatment of patients with borderline personality disorder. Journal of Personality Disorders, 16(2), 122–129.


