RESEARCH ARTICLE

The Role of Mentalizing and Epistemic Trust in the Therapeutic Relationship

Peter Fonagy and Elizabeth Allison

Research Department of Clinical, Educational and Health Psychology, University College London.

p.fonagy@ucl.ac.uk

ABSTRACT

Mentalizing—the capacity to understand others' and one's own behavior in terms of mental states—is a defining human social and psychological achievement. It involves a complex and demanding spectrum of capacities that are susceptible to different strengths, weakness and failings; personality disorders (PDs) are often associated with severe and consistent mentalizing difficulties (Fonagy & Bateman, 2008). In this paper we will argue for the role of mentalizing in the therapeutic relationship, suggesting that although Mentalization-Based Treatment (MBT) may be a specific and particular form of practice, the "mentalizing therapist" is a universal constituent of effective psychotherapeutic interventions.

The Role of Mentalizing and Epistemic Trust in the Therapeutic Relationship

Mentalizing theory was first elaborated in the context of formulating mentalization-based treatment (MBT) for the treatment of patients with Borderline Personality Disorder (BPD) in a partial hospital setting; MBT has more recently developed into a more comprehensive approach to the understanding and treatment of personality disorders in a range of clinical contexts. In this paper, we attempt to broaden our argument in relation to mentalizing and rather than explaining the particulars of the MBT approach (Bateman & Fonagy, 2006), we postulate that mentalizing might productively be conceptualised as the common factor across different forms of effective psychotherapy. Further, in understanding the mechanisms by which mentalizing works and how it can become disrupted, we broaden our argument to encompass the developmental significance of the transmission of epistemic trust in relation to social learning in the attachment context.

Mentalizing in therapy is a generic way of establishing epistemic trust (trust in the authenticity and personal relevance of interpersonally transmitted information) (D. B. Wilson & D. Sperber, 2012) between the patient and the therapist in a way that helps the patient to relinquish the rigidity that characterizes individuals with enduring personality pathology. The relearning of flexibility allows the patient to go on to learn, socially, from new experiences and achieve change in their understanding of their social relationships and their own behavior and actions. The very experience of having our subjectivity understood – of being mentalized – is a necessary trigger for us to be able to receive and learn from the social knowledge that has the potential to change our perception of ourselves and our social world.

Mentalizing in the Therapeutic Relationship and the Motion of Epistemic Trust

Mentalizing – despite in many ways being a defining human accomplishment – is not constitutionally guaranteed. It is, rather, a potential developmentally fulfilled in the early years of life; it is a developmental process that relies on good enough attachment relationships and early attachments in particular, as they reflect the extent to which our subjective experiences were adequately mirrored by a trusted other: i.e., the extent to which attachment figures have been able to respond with contingent and marked affective displays of their own experience in response to the infant's subjective experience, thus enabling the child to develop second-order representations of its own subjective experiences (Fonagy, 1998).

There is considerable evidence that a caregiver's capacity to mentalize predicts attachment in a child. Studies examining different ways in which caregivers' mentalizing is operationalized—including prenatal RF (Fonagy, Steele, Steele, Moran, & Higgitt, 1991), child-specific RF (Slade, Grienenberger, Bernbach, Levy, & Locker, 2005), mind-related comments (Meins, Fernyhough, Fradley, & Tuckey, 2001; Meins et al., 2002), and various other measures (Aber, Slade, Berger, Bresgi, & Kaplan, 1985; Koren-Karie, Oppenheim, Dolev, Sher, & Etzion-Carasso, 2002; Oppenheim, Koren-Karie, & Sagi, 2001; Sethna, Murray, & Ramchandani, 2012; Solomon & George, 1999) – have found that each of these facets of mentalizing capacity predicts attachment security in the child. To summarize this literature, it seems that Ainsworth's concept of *sensitivity* relates to the infant's sense of being recognized as an intentional agent, whether in the context of physical expressions of agency (Shai & Belsky, 2011a, 2011b) or more traditionally assessed indicators of sensitivity (Bretherton, 2013). Additionally, the caregiver's capacity to mentalize can offer protection from caregiver-related risk factors that are associated with

generating attachment insecurity, such as maternal trauma and disruptive maternal behaviors (e.g., Stronach, Toth, Rogosch, & Cicchetti, 2013). Furthermore, we know that the benefits of caregiver mentalizing extend beyond attachment outcomes: good mentalizing on the part of the caregiver is associated with higher performance of children in social cognition tasks (Laranjo, Bernier, Meins, & Carlson, 2010; Meins et al., 2002) and with general social cognitive development (Meins et al., 2003). In contrast, social environments characterized by adversity (e.g., child neglect or abuse) impair cognitive development (Ayoub et al., 2009; Fernald, Weber, Galasso, & Ratsifandrihamanana, 2011; G. S. Goodman, Quas, & Ogle, 2010; Rieder & Cicchetti, 1989). In brief, sensitivity to the young child's emerging intentionality, their nascent sense of subjective self inferred through parental mentalizing, increases the chance of secure attachment, enhances their resilience to adversity, and promotes cognitive, social-cognitive, and emotion-regulating capacity.

The related implied question is: by what mechanism does the child profit from the caregiver's mentalizing behavior? Answers to this may have powerful ramifications for our understanding of social development, as well as our understanding of how therapeutic interventions for BPD and other disorders characterized by interpersonal difficulties might work. We have argued elsewhere (Fonagy, Luyten, & Allison, 2013), building on pioneering work by Dan Sperber (Sperber et al., 2010; D. Wilson & D. Sperber, 2012) and accumulating evidence (e.g., Corriveau et al., 2009), that secure attachment experiences do not just pave the way for the acquisition of mentalizing, but that they are also key to the formation of *epistemic trust*—that is, an individual's willingness to consider new knowledge from another person as trustworthy, generalizable, and relevant to the self. In other words, attachment may mediate the reliable transmission of knowledge from one

generation to the next; secure attachment helps to create a benign condition for the relaxation of epistemic vigilance, sensitive and appropriate ostensive cueing (see below), is a key constituent element of sensitivity on the part of the primary caregiver. Attachment is a much older instinct, in evolutionary terms, than the imperative to generate epistemic trust; in that sense the two processes are distinct. In terms of the phenomenology of child development, however, they are closely interwoven. We shall now explore some recent evidence on this topic and consider its relevance to our understanding the role of mentalizing in the therapeutic relationship.

The Transmission of Culture: Natural Pedagogy and Epistemic Trust

The theory of natural pedagogy developed by Csibra and Gergely (2009) may offer a model to explain how an individual's attachment history could create distinct epistemic states. The theory posits a cue-driven social cognitive adaptation of mutual design dedicated to ensuring a highly effective and efficient transfer of culturally relevant knowledge between human beings. Csibra and Gergely argue that human communication is an evolutionary product of the requirement to transmit cognitively opaque cultural knowledge: generic knowledge that is robust to interference, is kind-generalizable, and becomes experienced as shared, in the sense that it immediately generates an expectation that others belonging to the same social group possess the same knowledge. Csibra and Gergely build on an idea first discussed by Bertrand Russell (1940) and extensively used by Sperber and Wilson (1995), suggesting that *an agent uses certain signals to prepare the addressee for the intent of the agent to communicate*. These cues may also serve to moderate natural *epistemic vigilance* (the self-protective suspicion towards information coming from others that may be potentially damaging, deceptive, or

inaccurate). Russell suggested that a process of *ostension*—the signaling of communicative intent—takes place as a part of communication. According to the theory of natural pedagogy, *ostensive cues* generate an attentional state in the addressee such that natural disbelief (what Sperber & Wilson, 1995 termed epistemic vigilance) is momentarily suspended and the addressee feels that the subsequent communication will contain information specifically relevant to them that should be remembered and encoded with other knowledge relevant to social situations. Such information could be about an object or the communicator's views and attitudes about the object, or about the beliefs communicated by the other (communicator) about the self (addressee) that are to be regarded as generalizable and relevant across situations. The information can be stored and used as part of procedural and semantic memory, not uniquely or as primarily episodic memory. The distinction between these memory systems is well established in neuroscience (Squire, 2004).

We have suggested that the concept of ostension, driving natural pedagogy, and maternal sensitivity, driving attachment security conceived in terms of mentalizing or sensitivity to intentional state, are loosely coupled and overlapping constructs (Fonagy, Gergely, & Target, 2007a). A securely attached child will believe their caregiver to be a reliable source of knowledge because the caregiver is more likely to have used ostension in the history of their relationship with the child. The predictors of secure attachment relationships are essentially also ostensive communication cues. The consistent emotional responses of a sensitive caregiver are clearly expressed to the child via ostensive cues such as making eye contact, accurate turn-taking, appropriate contingent (in time, tone, content) reactivity, and frequent use of a special communicational tone that addresses the child's

experiential world. All such cues appear to trigger a special mode of learning in the infant (G. Csibra & Gergely, 2011; Kiraly, Csibra, & Gergely, 2013). The biological predisposition of the caregiver to respond contingently to the infant's expressive displays creates a foundation for the infant to acquire further knowledge from that caregiver (Gergely, 2013).

During "marked mirroring interactions" (the caregiver's use of exaggerated facial displays and vocalizations in response to the infant's expressions of emotion to reflect how the infant is feeling back to the infant, but in a "play-acting" manner; Fonagy, Gergely, Jurist, & Target, 2002; Gergely & Watson, 1996), the caregiver "marks" his/her referential emotion displays to signal the generalizability of knowledge and effectively instruct the infant about the infant's subjective experience: "Look at me" (marked display/ostensive cue), "this is what you are feeling" (culturally transmitted self-knowledge; Fonagy, Gergely, & Target, 2007b). In other words, "marking" by the caregiver serves as an ostensive cue to the infant that the mirrored affect signals are relevant and generalizable. Babies display a sensitivity in relation to particular ostensive behavioral signals such as direct eye contact or being talked to with the special intonation of "motherese" (G Csibra & Gergely, 2006; G. Csibra & Gergely, 2009, 2011). They show attention preferentially to such signals, and the impact of these signals on their behavior is readily apparent.

To summarize, ostensive cues from the caregiver trigger *epistemic trust* at the same time as increasing the chance of a secure child–parent attachment (Fonagy et al., 2007b). Ostension sets aside the biological protection provided by epistemic vigilance and opens a channel of information exchange for transmitting and receiving knowledge about the social and personally relevant world, while encoding the information with the authority but not the person of the communicator, and helping to

ensure it will be remembered. Ostensive cues signal that we must go beyond a specific physical experience and acquire information that will be relevant across a range of settings. Epistemic trust is there to ensure that the individual can safely change their position; it triggers the opening of what we can think of as an "*epistemic superhighway*"—an evolutionarily protected mechanism that signals readiness to acquire knowledge.

Research has provided evidence of the links between attachment security and the ability to generate epistemic trust. In a longitudinal study of attachment, 147 children whose attachment was assessed in infancy were tested twice for epistemic trust at 50 and 61 months of age (Corriveau et al., 2009). For the test, the child's mother and a stranger made conflicting claims to the child concerning (a) the name of an unfamiliar object, (b) the name of a hybrid animal made up of 50% of each of two animals (e.g., an image made up of 50% horse and 50% cow; the mother might call it a cow, while the stranger says it is a horse), and (c) the name of a hybrid animal made up of 75% of one animal and 25% of another. In the latter case, the mother always made the improbable claim (e.g., that a picture made up of 75% squirrel and 25% rabbit was a rabbit), while the stranger gave the more likely answer ("squirrel"). The study aimed to gather data on which adult the child would spontaneously turn to for information and which they would believe, and whether this behavior was moderated by history of attachment security. The nature of a child's attachment relationship turned out to have a powerful effect on the child's trust in information imparted by the attachment figure (mother) and others (stranger). Children who were securely attached in infancy used a flexible strategy, showing a preference toward accepting claims made by their mother when they were plausible but trusting their own perception when the mother's claims appeared improbable.

Insecure-avoidant children appeared to withhold trust in their mother and preferred to attend to information from the stranger, while insecure-resistant children withheld trust in the stranger's claims even when their mother made improbable claims. Children with insecure-disorganized histories evidenced what we may call *chronic epistemic vigilance* or *epistemic hypervigilance*; they appeared to regard both information sources with suspicion. They had not much more trust in information given by the mother than by the stranger, but also showed little confidence in their own perception.

Attachment security, rooted in a history of feeling recognized, appears to increase the likelihood of trust in a source of communication when it is reasonably credible. A secure attachment history also generates confidence in one's own experience and belief and empowers one's judgment. In contrast, a history of attachment avoidance may generate epistemic mistrust; anxious attachment creates epistemic uncertainty through overreliance on the views of the attachment figure; and disorganized attachment can create epistemic hypervigilance, the mistrust of both the attachment figure and strangers as a source of information. In a child with a history of disorganized attachment, the unresolvable question "Who can I trust?" might contribute to this epistemic hypervigilance. As the study shows, children with a history of attachment disorganization mistrust information from attachment figures and strangers, and even their own experience. While an insecure attachment history may preclude complete confidence in one's subjective experience, an organized strategy biases either toward (insecure-resistant) or away from (insecure-avoidant) the attachment figure. But if neither source can be trusted, an unending epistemic search may ensue. The child seeks others to confirm or deny his/her own understanding, which he/she has little faith in, but, being unable to trust information

received from others, remains in a state of uncertainty and epistemic vigilance. This generates a state of interminable searching for validation of experience, coupled with the chronic lack of trust that we describe here as epistemic hypervigilance. This brings to mind many patients, but a young woman with BPD diagnosis in particular, who chronically lacked confidence in the accuracy of her interpersonal experiences (e.g. if X whom she met recently liked her or not) and would seek confirmation from friends and family only to find their response, whether confirmation or denial, unsatisfactory. But being thrown back on her own judgment failed to provide the desired certainty, leaving her to seek further independent verification in what appeared at times to be an interminable and frustrating process.

We suggest that while attachment may be a key mechanism for mediating epistemic trust, it is secondary to an underlying biological process preserved by evolution. In other words, secure attachment is unlikely to be *necessary* for generating epistemic trust but it may be *sufficient* to do so, and, further, it is the most *pervasive* mechanism in early childhood because it is a highly evolutionarily effective indicator of trustworthiness. Given that the infant needs to overcome the barrier created by natural epistemic vigilance and open their mind to acquiring the many pieces of culturally relevant information on which their survival will ultimately depend, it makes sense for humans to have evolved a mechanism to facilitate knowledge transmission between the teacher and the learner, based normally on a shared genetic inheritance (Hamilton, 1964).

Attachment, Mentalizing and Epistemic Trust

As outlined in the previous section, we believe that, through down-regulation of affect triggered by proximity-seeking in the distressed infant, attachment not only establishes a lasting bond between child and caregiver, but also opens a channel for

information to be used for knowledge transfer between generations. This is well demonstrated in studies of cognitive styles associated with patterns of attachment in adulthood.

Adult attachment insecurity is associated with a greater likelihood of cognitive closure, a lower tolerance for ambiguity, and a more pronounced tendency for dogmatic thinking (Mikulincer, 1997). Individuals with insecure attachment are also more likely to save intellectual effort and adopt stereotypes (Mikulincer, 1997). The same predisposition to knowledge inflexibility is apparent in the tendency of insecure individuals to make judgments based only on early information and to take subsequent information into account insufficiently (Green-Hennessy & Reis, 1998; Mikulincer, 1997). Insecure individuals, who fear the loss of their attachment figures, also anxiously hold on to their initial constructions. Kruglanski (Kruglanski, 1989; Kruglanski & Webster, 1996; Pierro & Kruglanski, 2008) proposed the concept of *epistemic freezing*, characterized by a tendency to defend existing knowledge structures even when they are incorrect or misleading (see also Fiske & Taylor, 1991). We consider that such a defensive strategy may be adaptive if an individual's self-esteem is vulnerable. Cognitive closure, dogmatism, and conservatism may simply be strategies to safeguard an inadequately individuated self (Bowlby, 1980). By contrast, the greater confidence of secure individuals enables them to be less defensive in relation to opening their minds to information that challenges their existing assumptions.

Mikulincer (1997) suggested that insecure individuals are more readily threatened by information that challenges their knowledge structures because their sense of self is vulnerable—in particular to being emotionally overwhelmed. If emotional dysregulation is experienced as a real and imminent threat, such

individuals may opt for knowledge stability, which temporarily serves to down-regulate arousal. Insecure individuals are less likely to revise their knowledge when faced with information that challenges their assumptions (Green-Hennessy & Reis, 1998; Green & Campbell, 2000; Mikulincer, 1997; Mikulincer & Arad, 1999), as if they not only have less confidence in the robustness of their bond to their attachment figure, but also fear the loss of epistemic trust. In sum, we assume that the epistemic superhighway that enables us to learn from others and from social experience is less efficient in those whose attachment representation in relation to their caregiver is insecure.

Developmental adversity, and particularly attachment trauma (Allen, 2012, 2013), may trigger a profound destruction of trust. There may be other reasons but, once epistemic trust has been lost, its absence creates an apparent *rigidity*. The rigidity is perceived by the communicator, who expects the recipient to modify his/her behavior on the basis of the information they received and apparently understood; yet in the absence of trust, the capacity for change is absent. The information given by the communicator is not used to update the recipient's understanding. In terms of the theory of natural pedagogy (G. Csibra & Gergely, 2009), the person has a (temporarily) reduced capacity to learn from "teachers". From a therapist's standpoint, he/she has become "hard to reach" and potentially interpersonally inaccessible. Looked at in another way, PD could be seen as a disorder of communication: chronic epistemic vigilance limits the capacity to internalize available knowledge as something that is "safe" to use to organize behavior.

Mentalizing as a Mediator of the Effectiveness of Psychotherapy

We have previously suggested, in a somewhat grandiose manner, that mentalizing provided an integrative framework that could bring together brain and

mind within a singular discourse and that a range of therapeutic modalities could be considered jointly, with mentalizing as their "common language" (Bateman & Fonagy, 2004; Fonagy & Bateman, 2006). In other words, we have suggested that mentalizing was a common factor in psychotherapy whilst also rather cheekily maintaining that it was specific to the approach of MBT (Bateman & Fonagy, 2006).

In an intriguing and unique study, Goldman and Gregory (2010) demonstrated that the therapeutic process of identifying, acknowledging, and sequencing emotional experiences correlated highly with the reduction of BPD symptoms in outpatients with the disorder. This is in line with our suggestion that the crux of the value of psychotherapy in BPD and a key factor with other clinical groups is the patient's experience of another person having the patient's mind in mind, and that therapy, regardless of the therapist's theoretical orientation, works by reviving the patient's capacity to interpret behavior as being motivated by mental states, both in themselves and in others (Bateman & Fonagy, 2004).

In a study that compared different psychotherapeutic modalities to assess the extent to which mentalization was a factor common to different therapies, by using the Psychotherapy Process Q-set (Ablon & Jones, 2002), Goodman (2013) showed that the prototype for Transference-Focused Psychotherapy (TFP) correlated with the psychodynamic psychotherapy prototype; the prototype for DBT correlated with the CBT prototype; and an MBT prototype (RF process) loaded on both TFP and DBT prototypes. Notably, the TFP prototype contained mentalizing items focused on the patient's mentalization of the therapist or other relationships, whilst the mentalizing elements of the DBT prototype focused on the patients' mentalization of themselves, perhaps in line with DBT's incorporation of mindfulness practice. Goodman argues that the RF process prototype encompasses the assertion that

enhancing mentalizing is central to therapy with BPD patients, and may be a unifying factor for effective approaches. Why should this be the case?

Mentalizing as a Means by which Epistemic Trust is Established

Mentalizing in the context of therapy must be distinguished from mentalizing in the social world. To "learn" to mentalize in treatment is not, in our view, an appropriate therapeutic aim. In fact, mentalizing is a key part of the therapeutic process because it enhances our general ability to learn in and from social situations and to generally benefit from interpersonal experience. Mentalizing in therapy is a generic way of establishing epistemic trust between the patient and the therapist with the aim of freeing the patient from rigidity, so that they can begin to learn from new experiences and achieve change in their understanding of their social relationships and their own behavior and actions. Having the experience of our subjectivity being understood is the necessary key to open us up to learning that has the potential to change our perception of our social world. Mentalizing our route to garnering knowledge relevant to us and being able to use it across contexts, independent of the learning experience. Put simply, *the experience of feeling thought about in therapy makes us feel safe enough to think about ourselves in relation to our world, and to learn something new about that world and how we operate in it*.

Mentalizing establishes a view of an individual as an agent, with a valid subjective experience that is worthy of engagement. Establishing epistemic trust in the creation of a collaboration between patient and therapist, through the explicit effort of seeing the world from the patients' standpoint, serves to open the patient's mind to the therapist's communication. The patient moves toward being able to trust the social world as a learning environment once again. This includes the therapeutic environment and all that is has to teach the individual about the nature of their

problems, the ways they have tried to cope with them in the past, and the options available to modify these strategies in the future. These are of course the "wisdoms" that the therapist has acquired to impart through their training. But as we shall explain in the next section, perhaps it is not what we learned to teach patients in therapy that matters most, but rather the potential of the therapeutic relationship to rekindle the capacity for learning from social situations.

The Psychotherapeutic Communication Systems: Why is Psychotherapy Effective?

We would like to argue that three sets of processes, which we will label "therapeutic communication systems", underpin the mechanism of change in the numerous forms of psychosocial treatment that have been found to be effective We suggest that the three systems relate to each other cumulatively to make change in revising an individual's experience of themselves as a consequence of therapy. From our point of view, a change in what Bowlby (1980) termed the 'internal working model' of attachment relationships (incorporating expectations about both self and other as well as expectations of the likely interaction between the two) is the critical change in therapy form which changes in symptoms and the quality of social adaptation follow. In individuals in a state of epistemic mistrust and hypervigilance, internal working models are impermeable to influence from social experience and such individuals are viewed as 'hard to reach' when considered outside a therapeutic context (and sometimes even inside it).

Communication System 1: The Teaching and Learning of Content

All evidence-based psychotherapies provide a coherent, consistent, and continuous framework that enables the patient to examine the issues that are

deemed to be central according to a particular theoretical approach (e.g., early schemas, invalidating experiences, object relations, current attachment experiences) in a safe and relatively low-arousal context. Thus, these psychotherapies provide the patient with helpful skills or knowledge, such as strategies to handle emotional dysregulation or restructured interpersonal relationship schemata. Perhaps more importantly however, all *evidence-based* psychotherapies implicitly provide for the patient a model of mind and an understanding of their disorder, as well as a hypothetical appreciation of the process of change, *that are accurate enough to enable the patient to feel recognized as an agent with intentionality*. The model contains considerable personally relevant information so the patient experiences feeling markedly mirrored or "understood". Helpful, directive approaches may be more likely to communicate a clear recognition of the patient's position than a generic exploratory style (McAleavey & Castonguay, 2013). The idea that psychotherapies have in common the creation of a sense of being understood while differing in the understandings they provide has been part of integrative approaches to psychotherapy since , common factor approaches were first proposed (e.g., Frank & Frank, 1991; Prochaska & Norcross, 2013; Rogers, 1951).

In essence, these (implicit or explicit) explanations may be seen as ostensive cues that signal to the patient the relevance to them of information that is being conveyed by triggering in the patient a feeling of being personally recognized by the therapist. This process is important because it allows the patient to reduce epistemic hypervigilance as he/she increasingly sees the model's relevance to his/her own state of mind. Thus, acquiring new skills and learning new and useful information about oneself has the nonspecific effect of creating openness, which makes it easier for the patient to learn the specific suggestions conveyed within the model. A

virtuous cycle is created: the patient "feels" the personal truth of the evidence-based content conveyed, which, because it is accurate and helpful, in turn generates epistemic openness; the growth of epistemic trust allows the patient to take in further information that also serves to reassure and validate them. The learning process is facilitated by the patient's experience of feeling mentalized by the "felt truth" of the content being communicated, either through its correspondence with phenomenology or through practical experience.

A wide range of explanations of experience will assist patients whose 'grip' on their own subjectivity is momentarily loose. Finding oneself prone to emotional outburst, an account in terms of a failure of emotion regulation is helpful. But so might be an account in terms of failures of appropriate soothing in early childhood, or the internalization of an aggressive self-critical voice. 'Felt truth' can come from biological as well as from social accounts but one may expect such explanations to be less compelling. One patient, for example, felt understood and validated by my hypothesis that their chronic suspicion when faced with the challenge of trusting others could be understood in terms of attachment experiences that were a catastrophic accumulation of disappointment. We know that without a coherent body of knowledge based on a systematically established set of principles, psychological therapy is of little value (Benish, Imel, & Wampold, 2008). Even in large cohort study meta-analyses, therapies without a credible and tight intellectual frame are observed to fail (Abbass, Rabung, Leichsenring, Refseth, & Midgley, 2013).

The fact that so many different therapies, using so many different theoretical models, have been found to have some beneficial effect indicates that the significance of communication system 1 lies perhaps not in the essential truth of the "wisdom" of the specific approach, but in the fact that it causes the patient to give

weight to communication from the social world (Ahn & Wampold, 2001; Paris, 2013). This brings us to the second communication system at work in psychotherapy.

Communication System 2: The Re-emergence of Robust Mentalizing

The experience of being provided with understanding creates a shift in the quality of interpersonal communication and generates a stepwise improvement in communication competence. As noted above, through passing on knowledge and skills that feel appropriate and helpful, the therapist implicitly recognizes the patient's agency. The therapist's presentation of information that is personally relevant to the patient serves as a form of ostensive cueing that conveys the impression that the therapist seeks to understand the patient's perspective; this in turn enables the patient to listen and to hear. In effect, the therapist is modeling (demonstrating) how he/she engages in mentalizing in relation to the patient. It is important that in this process both patient and therapist come to see each other more clearly as intentional agents. It is not sufficient for the therapist to present their "mentalizing wisdom" to the patient if they are not themselves clearly seen as an agentive actor whose actions are predictable given the principles of theoretical rationality (Kiraly et al., 2013). The context of an open and trustworthy social situation facilitates achievement of a better understanding of the beliefs, wishes, and desires underpinning the actions of others and of the self. This allows for a more trusting relationship in the consulting room. Ideally, the patient's feeling of having been sensitively responded to by the therapist opens a second virtuous cycle in interpersonal communication *in which the patient's own capacity to mentalize is regenerated*.

However, the mentalizing of patients—that is, acting in accordance with the patient's perspective—may be a common factor across psychotherapies not

because patients need to learn about the contents of their minds or those of others, but because mentalizing may be a generic way of increasing epistemic trust and therefore achieving change in mental *function*. We would maintain that the patient's capacity to mentalize improves in all effective therapies. This is likely to have generic benefits in increasing the patient's self-control and sense of self-coherence; it increases the accuracy of their social understanding, reduces their experience of mental pain, and improves their ability to think coherently in the context of attachment relationships. This has been a key part of our understanding of the mechanisms of change since we devised the MBT model (Fonagy & Bateman, 2006). Understanding the patient's subjectivity is vital to this process, as the patient's self-discovery as an active agent occurs through the social interchange where they experience themselves as an agent in the mind of their therapist—they "find themselves in the mind of the therapist". It is also vital to a further function of therapy, which we wish to note separately: the (re-)kindling of the patient's wish to learn about the world, including the social world. In brief, and to simplify what we believe is a complex and non-linear process, the insight obtained in therapy, whatever its content, creates or re-creates the potential for a learning experience, which in turn makes other similar learning experiences more productive because it *enables the patient to adopt a stance of learning from experience by increasing their capacity to mentalize*.[1]

[1] In using this phrase we are leaning heavily on Wilfred Bion's discoveries (Bion, 1962). We are doing this intentionally to explicitly acknowledge the intellectual indebtedness we feel, although we are equally aware that those who see themselves as maintaining his tradition may well be dismayed by our claim of intellectual communion.

We would like to underline a point that may seem initially puzzling given our own declared commitment to mentalization-based psychotherapy: *mentalizing in itself is not the therapeutic objective*. Simply instructing the therapist to focus the patient on their own thoughts and feelings, or the thoughts and feelings of those around them, will not achieve change by itself. It may, along with other techniques, initiate change by changing the mindset of the person undergoing treatment. However, the process of creating a more robust mentalizing function in therapy (communication system 2) can no more assure enduring alteration in the patient than does communication system 1. True and lasting improvement rests on communication system 3: learning from experience beyond therapy.

Communication System 3: The Re-emergence of Social Learning

We hypothesize that feeling understood opens a key biological route to information transmission and the possibility of taking in knowledge that is felt to be personally relevant and generalizable; this is what brings about change in previously rigidly held beliefs. In essence, the experience of feeling thought about enables us to learn new things about our social world.

The therapeutic situation teaches about sources of knowledge. It provides a clear social illustration of trust, making the therapist a “deferential source” of knowledge (D. Wilson & D. Sperber, 2012) with the capacity to undo previously rigidly held beliefs about the self and about others, and to reduce the patient’s experience of epistemic isolation, which is embodied in the rigidity of their subjective experience. This initiates a third virtuous cycle. Improved understanding of social situations through improved mentalizing leads to better understanding of significant others in the patient’s life, which in turn creates potential for the person to notice a sensitive response and feel understood. Reopening the potential to feel sensitively

responded to—both within and outside the therapeutic setting—may in itself initiate more trusting interpersonal relationships, and thus open the patient up to new understandings of specific social situations as these arise.

We hypothesize that, as the patient's state of epistemic hypervigilance relaxes, his/her capacity for trust increases and he/she can discover new ways of learning about others. This facilitates an increase in the patient's willingness to modify his/her cognitive structures for interpreting others' behavior. Social experiences that may have been positive but were in the past discounted as a result of the patient's epistemic hypervigilance now have the potential to have a positive impact. This is the third system of communication, which becomes available once the second system, tied to the therapeutic situation, has enhanced the patient's capacity to mentalize. As patients begin to experience social interactions in a more benign way and view their social situations more accurately (e.g., not seeing an experience of temporary social disappointment as an outright rejection), they update their knowledge of both themselves and others.

It is the recovery of capacity for social information exchange that, we feel, is at the heart of effective psychotherapies. They impart an ability to benefit from benign social intentions, and to update and build on knowledge about the self and others in social situations. The improved sense of epistemic trust derived from mentalizing enables learning from social experience; in this way the third virtuous cycle is maintained beyond therapy.

As therapists we often assume that the process in the consulting room is the primary driver of change, but experience shows us that change is also brought about by what happens beyond therapy, in the person's social environment. Empirical evidence from studies where change was monitored session by session suggests

that the therapeutic alliance in a given session foretells change in the next (Falkenstrom, Granstrom, & Holmqvist, 2013; Tasca & Lampard, 2012). This suggests that the change that occurs in between sessions is a consequence of changed attitudes to learning engendered by therapy. The implication is that the extent of the benefit a patient derives from therapy may depend on what he/she encounters in his/her particular social world. We predict that psychotherapy for BPD is much more likely to succeed if the individual's social environment at the time of treatment is largely benign. Although we do not know of any systematic studies that have explored this moderator, clinical experience suggests that there is likely to be some validity to this assertion.

This admittedly speculative model offers a way to integrate the specific and nonspecific factors in effective psychotherapy. Specific factors associated with "therapies that work" create experiences of truth, which in turn encourage the patient to learn more. In this process, via a nonspecific channel, the patient's capacity to mentalize is fostered. Both these systems would be expected to lead to symptomatic improvement. Improved mentalizing and reduced symptomatology both improve the patient's experiences of social relationships. It is likely that these new social experiences, rather than only what happens within therapy, serve to erode the epistemic hypervigilance that has hitherto prevented benign social interactions from changing an individual's experience of themselves and of the social world. *Change is thus likely due to transformations of the ways a person uses their social environment, not just to what happens in therapy.*

Summary

Mentalizing in the therapeutic relationship is a complex process. Patients with the diagnosis of BPD often struggle in therapy because their disorder serves to undermine their capacity to benefit from the therapeutic process, whatever its modality. Individual or group therapy, particularly of an unstructured kind, can serve to activate the attachment system, causing what may be a catastrophic collapse of mentalizing. It is hardly surprising that in such situations patients often abandon therapy against the advice of their therapist. Awareness of the nature of the mentalizing problems that BPD patients face can help in managing these challenging situations. The MBT technique was designed to guide therapists to avoid this type of iatrogenesis in the course of treatment.

MBT is not the only effective therapy for BPD. In fact, many therapies can consider themselves “evidence-based”. In this paper we have speculated that mentalizing may play a role in the change process regardless of modality. Based on modern formulations of the communication and learning process, and in line with most formulations of “common factors” in psychotherapy, we have suggested that limitations in patients’ capacity to learn from experience (i.e., being “hard to reach”) are generically overcome by specific interventions that make patients feel understood. Mentalizing is a common tool for achieving this sense of being individually responded to. Feeling understood in therapy restores trust in learning from social experience (epistemic trust) but at the same time also serves to regenerate a capacity for social understanding (mentalizing). Improved social understanding alongside increased epistemic trust makes life outside therapy a setting in which new information about oneself and about the world can be acquired and internalized. Ultimately, it may be that therapeutic change is not due to new

skills or new insights gained in the consulting room, but rather to the capacity of the therapeutic relationship to create a potential for learning about oneself and others in the world outside of therapy. In the past, modifications of the patient's social world were felt to fall outside the concerns of psychotherapy. It is possible, however, that effective treatments depend as much on ensuring that the patient's social environment is benign as on ensuring a similar emotional tone in the consulting room.

References

Abbass, A. A., Rabung, S., Leichsenring, F., Refseth, J. S., & Midgley, N. (2013). Psychodynamic psychotherapy for children and adolescents: A meta-analysis of short-term psychodynamic models. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*, 863-875.

Aber, J., Slade, A., Berger, B., Bresgi, I., & Kaplan, M. (1985). *The Parent Development Interview. Unpublished manuscript*. Barnard College, Columbia University. New York, NY.

Ablon, J. S., & Jones, E. E. (2002). Validity of controlled clinical trials of psychotherapy: Findings from the NIMH Treatment of Depression Collaborative Research Program. *American Journal of Psychiatry, 159*, 775-783.

Ahn, H., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. . *Journal of Counseling Psychology, 48*, 251−257.

Allen, J. G. (2012). *Restoring mentalizing in attachment relationships: Treating trauma with plain old therapy*. Washington, DC: American Psychiatric Press.

Allen, J. G. (2013). *Mentalizing in the development and treatment of attachment trauma*. London, UK: Karnac Books.

Ayoub, C., O'Connor, E., Rappolt-Schlictmann, G., Vallotton, C., Raikes, H., & Chazan-Cohen, R. (2009). Cognitive skill performance among young children living in poverty: Risk, change, and the promotive effects of Early Head Start. *Early Child Research Quarterly, 24*, 289-305.

Bateman, A. W., & Fonagy, P. (2004). Mentalization-based treatment of BPD. *Journal of Personality Disorders, 18*, 36-51.

Bateman, A. W., & Fonagy, P. (2006). *Mentalization based treatment for borderline personality disorder: A practical guide*. Oxford, UK: Oxford University Press.

Benish, S. G., Imel, Z. E., & Wampold, B. E. (2008). The relative efficacy of bona fide psychotherapies for treating post-traumatic stress disorder: A meta-analysis of direct comparisons. *Clinical Psychology Review, 28*, 746-758.

Bion, W. R. (1962). *Learning from experience*. London, UK: Heinemann.

Bowlby, J. (1980). *Attachment and loss, Vol. 3: Loss: Sadness and depression*. London, UK: Hogarth Press and Institute of Psycho-Analysis.

Bretherton, I. (2013). Revisiting Mary Ainsworth's conceptualization and assessments of maternal sensitivity-insensitivity. *Attachment and Human Development, 15*, 460-484.

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., . . . de Rosnay, M. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development, 80*, 750-761.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In M. H. Johnson & Y. Munakata (Eds.), *Processes of change in brain and cognitive development. Attention and Performance XXI* (pp. 249-274). Oxford, UK: Oxford University Press.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148-153.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 366*, 1149-1157.

Falkenstrom, F., Granstrom, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology, 60*, 317-328.

Fernald, L. C., Weber, A., Galasso, E., & Ratsifandrihamanana, L. (2011). Socioeconomic gradients and child development in a very low income population: Evidence from Madagascar. *Developmental Science, 14*, 832-847.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York, NY: McGraw-Hill.

Fonagy, P. (1998). An attachment theory approach to treatment of the difficult patient. *Bulletin of the Menninger Clinic, 62*, 147-169.

Fonagy, P., & Bateman, A. (2008). The development of borderline personality disorder: A mentalizing model. *Journal of Personality Disorders, 22*, 4-21.

Fonagy, P., & Bateman, A. W. (2006). Mechanisms of change in mentalization-based treatment of BPD. *Journal of Clinical Psychology, 62*, 411-430.

Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). *Affect regulation, mentalization and the development of the self*. New York, NY: Other Press.

Fonagy, P., Gergely, G., & Target, M. (2007a). The parent-infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry, 48*, 288-328.

Fonagy, P., Gergely, G., & Target, M. (2007b). The parent–infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry, 48*, 288-328.

Fonagy, P., Luyten, P., & Allison, E. (2013). Teaching to learn from experience: Epistemic mistrust at the heart of BPD and its psychosocial treatment. *Manuscript in preparation*.

Fonagy, P., Steele, M., Steele, H., Moran, G. S., & Higgitt, A. C. (1991). The capacity for understanding mental states: The reflective self in parent and child and its significance for security of attachment. *Infant Mental Health Journal, 12*, 201-218.

Frank, J. D., & Frank, J. B. (1991). *Persuasion and healing: A compatative study of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.

Gergely, G. (2013). Ostensive communication and cultural learning: The natural pedagogy hypothesis. In J. Metcalfe & H. S. Terrace (Eds.), *Agency and Joint Attention* (pp. 139). Oxford, UK: Oxford University Press.

Gergely, G., & Watson, J. S. (1996). The social biofeedback theory of parental affect-mirroring: The development of emotional self-awareness and self-control in infancy. *International Journal of Psycho-Analysis, 77 (Pt 6)*, 1181-1212.

Goldman, G. A., & Gregory, R. J. (2010). Relationships between techniques and outcomes for borderline personality disorder. *American Journal of Psychotherapy, 64*, 359-371.

Goodman, G. (2013). Is mentalization a common process factor in transference-focused psychotherapy and dialectical behavior therapy sessions? *Journal of Psychotherapy Integration, 23*, 179-192.

Goodman, G. S., Quas, J. A., & Ogle, C. M. (2010). Child maltreatment and memory. *Annual Review of Psychology, 61*, 325-351.

Green-Hennessy, S., & Reis, H. T. (1998). Openness in processing social information among attachment types. *Personal Relationships, 5*, 449-466.

Green, J. D., & Campbell, W. K. (2000). Attachment and exploration in adults: Chronic and contextual accessibility. *Personality and Social Psychology Bulletin, 26*, 452-461.

Hamilton, W. D. (1964). The genetic evolution of social behaviour. *Journal of Theoretical Biology, 7*, 1-52.

Kiraly, I., Csibra, G., & Gergely, G. (2013). Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations. *Journal of Experimental Child Psychology, 116*, 471-486.

Koren-Karie, N., Oppenheim, D., Dolev, S., Sher, E., & Etzion-Carasso, A. (2002). Mothers' insightfulness regarding their infants' internal experience: Relations with maternal sensitivity and infant attachment. *Developmental Psychology, 38*, 534-542.

Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases*. New York, NY: Plenum Press.

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: "Seizing" and "freezing". *Psychological Review, 103*, 263-283.

Laranjo, J., Bernier, A., Meins, E., & Carlson, S. M. (2010). Early manifestations of children's theory of mind: The roles of maternal mind-mindedness and infant security of attachment. *Infancy, 15*, 300-323.

McAleavey, A. A., & Castonguay, L. G. (2013). Insight as a common and specific impact of psychotherapy: Therapist-reported exploratory, directive, and common factor interventions. *Psychotherapy*.

Meins, E., Fernyhough, C., Fradley, E., & Tuckey, M. (2001). Rethinking maternal sensitivity: Mothers' comments on infants' mental processes predict security of attachment at 12 months. *Journal of Child Psychology and Psychiatry, 42*, 637-648.

Meins, E., Fernyhough, C., Wainwright, R., Clark-Carter, D., Das Gupta, M., Fradley, E., & Tuckey, M. (2003). Pathways to understanding mind: Construct validity and predictive validity of maternal mind-mindedness. *Child Development Perspectives, 74*, 1194-1211.

Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Development, 73*, 1715-1726.

Mikulincer, M. (1997). Adult attachment style and information processing: Individual differences in curiosity and cognitive closure. *Journal of Personality and Social Psychology, 72*, 1217-1230.

Mikulincer, M., & Arad, D. (1999). Attachment working models and cognitive openness in close relationships: A test of chronic and temporary accessibility effects. *Journal of Personality and Social Psychology, 77*, 710-725.

Oppenheim, D., Koren-Karie, N., & Sagi, A. (2001). Mothers' empathic understanding of their preschoolers' internal experience: Relations with early attachment. *International Journal of Behavioral Development, 25*, 16-26.

Paris, J. (2013). How the history of psychotherapy interferes with integration. *Journal of Psychotherapy Integration, 23*, 99-106.

Pierro, A., & Kruglanski, A. W. (2008). "Seizing and freezing" on a significant-person schema: need for closure and the transference effect in social judgment. *Personality and Social Psychology Bulletin, 34*, 1492-1503.

Prochaska, J. O., & Norcross, J. C. (2013). *Systems of psychotherapy: A transtheoretical analysis* (8th ed.). Belmont, CA: Brooks/Cole Cengage Advantage Books.

Rieder, C., & Cicchetti, D. (1989). Organizational perspective on cognitive control functioning and cognitive-affective balance in maltreated children. *Developmental Psychology, 25*, 382-393.

Rogers, C. R. (1951). *Client-centered therapy*. Boston, MA: Houghton Mifflin.

Russell, B. (1940). *An inquiry into meaning and truth*. London, UK: Allen & Unwin.

Sethna, V., Murray, L., & Ramchandani, P. G. (2012). Depressed fathers' speech to their 3-month-old infants: A study of cognitive and mentalizing features in paternal speech. *Psychological Medicine, 42*, 2361-2371.

Shai, D., & Belsky, J. (2011a). Parental embodied mentalizing: Let's be explicit about what we mean by implicit. *Child Development Perspectives, 5*, 187-188.

Shai, D., & Belsky, J. (2011b). When words just won't do: Introducing parental embodied mentalizing. *Child Development Perspectives, 5*, 173-180.

Slade, A., Grienenberger, J., Bernbach, E., Levy, D., & Locker, A. (2005). Maternal reflective functioning, attachment, and the transmission gap: A preliminary study. *Attachment and Human Development, 7*, 283-298.

Solomon, J., & George, C. (1999). The caregiving system in mothers of infants: A comparison of divorcing and married mothers. *Attachment and Human Development, 1*, 171-190.

Sperber, D., Clement, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*, 359-393.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Malden, MA: Blackwell.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82*, 171-177.

Stronach, E. P., Toth, S. L., Rogosch, F., & Cicchetti, D. (2013). Preventive interventions and sustained attachment security in maltreated children. *Development and Psychopathology, 25*, 919-930.

Tasca, G. A., & Lampard, A. M. (2012). Reciprocal influence of alliance to the group and outcome in day treatment for eating disorders. *Journal of Counseling Psychology, 59*, 507-517.

Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge, UK: Cambridge University Press.

Wilson, D. B., & Sperber, D. (2012). *Meaning and Relevance*. Cambridge, UK: Cambridge University Press.