Chapter 1 What is mentalizing?

Mentalization-based treatment (MBT) was originally developed in the 1990s and initially used to treat patients with borderline personality disorder (BPD) in a partial (day) hospital setting. More recently, MBT has grown into a more comprehensive approach to the understanding and treatment of personality disorders in a range of clinical contexts, including antisocial personality disorder (ASPD), the mentalizing-based treatment of which we are including in this new edition.

The mentalizing approach has changed – and, we hope, progressed considerably – over the past several years. Recent advances have in particular been influenced by new findings in developmental psychology, psychopathology and the neurosciences, and of course the lessons we have learned from our own clinical experiences in the practice and training of MBT.

In this chapter, we will explain the concept of mentalizing and describe the theory of mentalizing in its up-to-date and clinically relevant form. We will show how these developments in thinking on mentalizing have influenced both our understanding and clinical practice in relation to BPD and ASPD.

Mentalizing is the ability to understand actions by both other people and oneself in terms of thoughts, feelings, wishes and desires; it is a very human capability that underpins everyday interactions (see Box 1.1). Without mentalizing there can be no robust sense of self, no constructive social interaction, no mutuality in relationships and no sense of personal security (Fonagy, Gergely, Jurist, & Target, 2002). Mentalizing is a fundamental psychological process that has a role to play in all major mental disorders. Indeed, mentalizing techniques are now being used for the treatment of post-traumatic stress disorder, drug addiction, eating disorders, personality disorder in adolescents, particularly those who self-harm, and in work with families in crisis (much of this work is summarized in Bateman & Fonagy, 2012).

BOX 1.1 near here

Mentalizing involves an awareness of mental states in oneself or in other people, particularly when it comes to explaining behaviour. It is beyond question that mental states influence behaviour. Beliefs, wishes, feelings and thoughts, whether within or outside our awareness, always influence what we do. Mentalizing involves a whole spectrum of capacities: critically, this includes the ability to see one's *own* behaviour as coherently organized by mental states, and to differentiate oneself psychologically from others. These capacities often tend to be conspicuously absent in individuals with a personality disorder, particularly at moments of interpersonal stress.

Mentalizing is a uniquely human capacity – it can be seen as what defines humanity and separates us from other higher-order primates. However, this capacity is not an entirely stable, consistent or one-dimensional thing (see Box 1.2). We are not all able to mentalize to the same extent; many of us have strengths or weaknesses in particular aspects of mentalizing, and most of us are more likely to struggle to mentalize in moments of stress or anxiety. All of us have experienced mentalizing lapses to a greater or lesser extent. Trying to understand other people's behaviour in terms of mental states is almost always more difficult and more liable to go wrong than

explanations based on the impact of the physical environment – that is, the visibly contingent world of cause and effect. We can all act according to mistaken beliefs about others' mental states in particular situations, leading to everyday misunderstandings, difficulties and social faux pas or in situations of heightened threat of violence, leading to more tragic consequences.

BOX 1.2 near here

Mentalizing is a mostly preconscious, imaginative mental activity: we have to imagine what other people might be thinking or feeling. The ways in which different people mentalize can vary enormously, because each person's history and ability to imagine may lead them to different conclusions about the mental states of others. Sometimes, we may also need to make an imaginative leap to understand our *own* experiences, particularly when we are dealing with emotionally charged issues or find ourselves being overwhelmed by our own irrational, non-consciously driven reactions to situations. In essence, mentalizing is *seeing ourselves from the outside and others from the inside*. It helps us to understand misunderstandings by recapturing the mind states that led to misapprehensions. From a clinical perspective, at its core is 'mind-centredness' – a focus on acquiring a clear and coherent view of what our patient sees, having his/her mind in mind, being mind-minded and mindful of minds. Mentalizing is a key skill because our sense of personal continuity is dependent on envisioning the thoughts and feelings we had in the past and how these relate to our current experiences, and because how we envision ourselves in the future is rarely in terms of physical attributes (after middle age certainly) but rather in terms of projecting ourselves as a thinking and feeling person. Mentalizing, the representation of our mental states, is the spine of our sense of self and identity (Fonagy & Target, 1997b). Seeing oneself and others as agentive and intentional beings driven by mental states that are meaningful and understandable creates the psychological coherence about self and others that is essential for navigating a complex social world.

Central ideas in the mentalizing approach to personality disorders

The mentalizing approach aims to provide a comprehensive account of the phenomenology and origins of BPD and APSD from a developmental perspective. This fits with increasing interest, over recent years, in the emergence of BPD in childhood and adolescence, particularly as there is growing evidence to suggest that the disorder may have roots in genetic vulnerability and early development (Fonagy & Luyten, 2016).

A developmental and attachment-based approach

A developmental perspective is at the heart of the mentalizing approach to BPD and ASPD. The mentalizing model was first outlined in a large empirical study in which the security of infants' attachment to their parents proved to be strongly predicted not only by the parents' security of attachment during the pregnancy (Fonagy, Steele, & Steele, 1991), but even more by the parents' capacity to understand their childhood

relationships with their *own* parents in terms of states of mind (Fonagy, Steele, Steele, Moran, & Higgitt, 1991).

This study paved the way for a systematic programme of research demonstrating that the capacity to mentalize, which emerges in the context of early attachment relationships, may be a key determinant of self-organization and affect regulation. The concept of mentalization is based around the idea that one's understanding of others depends on whether one's own mental states were adequately understood by caring, attentive, non-threatening adults. We have particularly emphasized the central relevance of the 'marked mirroring' of the child's emotional reactions by an adult with the capacity to represent the child's affect in a manner that conveys understanding at the same time as communicating a sense of coping with, rather than merely reflecting back, the child's affect (Fonagy et al., 2002; Gergely & Watson, 1996). Problems in affect regulation, attentional control and self-control stemming from dysfunctional attachment relationships are thought to develop through a failure to acquire robust mentalizing skills. From this perspective, mental disorders in general can be seen as arising when the mind misinterprets its own experience of itself and of others, to the extent that a mental picture of others is inferred from one's experience of oneself (Bateman & Fonagy, 2010).

The capacity for automatic mentalizing seems to be an early emerging and possibly innate human characteristic, but the extent to which the potential for full mentalizing is achieved is unlikely to be genetically determined and appears to be highly responsive to environmental influences (Hughes et al., 2005). The development of mentalizing is thought to depend on the quality of the social learning environment, the child's family relationships and, in particular, his/her early attachments, as these reflect the extent to which his/her subjective experiences were adequately mirrored by a caregiver. The attachment figure's ability to respond with contingent and marked affective displays of their own experience in response to the infant's subjective experience makes possible the child's development of coherent second-order representations of these subjective experiences. A child whose mother makes proportionally more age-appropriate references to desires and emotions than to thoughts and knowledge when the child is 5 months old will have better explicit mind-reading performance at 24 months. If, at 24 months, the mother then changes to make more references to thoughts and knowledge than to desires and emotions, the child will have better explicit mind-reading skills at 33 months (Taumoepeau & Ruffman, 2006, 2008). We suggest that these developmental differences are driven by the mother's awareness of the child's needs, and that this awareness in turn drives the child's acquisition of mentalizing.

Specifically, we believe that the quality of affect mirroring by attachment figures plays a major role in the early development of affect regulative processes and self-control (including attention mechanisms and effortful control) as well as the capacity for mentalizing. Later development follows the same pattern. More generally, parents, in the role of 'expert mentalizers', have the task of communicating mental state concepts, and ways of representing these concepts, to their children. As the child acquires this competence and becomes an 'expert mentalizer', the knowledge and skill of mentalizing is passed on to the next generation. Thus, we see mentalizing as a *transactional* and *intergenerational* social process (Fonagy & Target, 1997a): it develops in the context of interactions with others, and its quality in relation to

understanding others is influenced by how well those around us mentalize us, as well as others around them. This experience of how other people mentalize is internalized, enhancing our own capacity for understanding ourselves and hence others and thus engaging better in interactive social processes; conversely, of course, early exposure to interactions characterized by poor mentalizing will lead the child to develop poor mentalizing too. Parents do not merely teach labels for mental states. The emotional and language environment they create conveys *concepts of mental state* (what does is it mean to 'think' something, how does it feel to 'feel' something, what is the meaning of being 'happy', how does a person behave when they are 'doubtful'?). The parent, through their interactions with and in the presence of the child, generates a format with which mental state concepts can be represented. In effect they pass on a set of processes that have evolved to represent mental states, (culturally) inherited primarily from their parents but also from others in their immediate social environment (O'Brien, Slaughter, & Peterson, 2011). We predict an association between the extent to which these mechanisms specialized for the representation of mental states are acquired and the quality of the relationship between members of the family (i.e. the individuals who undertake mental state-related discourse). The quality of adult–child relationships will influence the child's assumptions about the origin, location, and functioning of mental states. This in turn will lead individuals to attend to different aspects of observable behaviour, and, in addition, different appraisals of mind states will lead to different patterns of observable behaviour.

The multidimensional nature of mentalizing

Neuroscientists have identified four different components, or dimensions, to mentalizing (Lieberman, 2007), which are helpful to distinguish in clinical applications of the concept. These are:

1. automatic versus controlled mentalizing,
2. mentalizing the self versus others,
3. mentalizing with regard to internal versus external features, and
4. cognitive versus affective mentalizing.

To mentalize effectively requires the individual not only to be able to maintain a balance across these dimensions of social cognition but also to apply them appropriately according to context.

In an adult with personality disorder, imbalanced mentalizing on at least one of these four dimensions would be evident. From this perspective, different types of psychopathology can be distinguished on the basis of different combinations of impairments along the four dimensions (which we can refer to as different *mentalizing profiles*; see Figure 4.8 for an example of the mentalizing profile in BPD and ASPD).

Automatic versus controlled mentalizing

The most fundamental dimension to mentalizing is the spectrum between automatic (or implicit) and controlled (or explicit) mentalizing (see Box 1.3). *Controlled mentalizing* reflects a serial and relatively slow process, which is typically verbal and demands reflection, attention, awareness, intention and effort. The opposite pole of

this dimension, *automatic mentalizing*, involves much faster processing, tends to be reflexive, and requires little or no attention, intention, awareness or effort.

In day-to-day life and ordinary social interaction, most of our mentalizing tends to be automatic because most straightforward exchanges do not require more attention. Particularly in a secure attachment environment, when things are running smoothly on an interpersonal level, more deliberate or controlled mentalizing is not called for; in fact, the use of such a mentalizing style might hinder such interactions, making them feel unduly weighty or uncomfortably overwrought (*hypermentalized*). Both common-sense experience and neuroscience tell us that we relax controlled mentalizing and are less watchful of social intentions in a secure attachment environment; a parent playing with their child or close old friends reminiscing will conduct their exchanges along automatic, intuitive processes. However, when necessary, someone with normative, strong mentalizing abilities will be able to switch to controlled mentalizing if the situation demands it. For example, when a child starts to cry during play, the parent will respond by inquiring about the child's change in affect, or the friend in conversation may detect a change in tone and mood in their friend, and wonder whether the conversation has stumbled upon a difficult memory or association. In other words, well-functioning mentalizing involves the ability to switch flexibly and responsively from automatic to controlled mentalizing.

BOX 1.3 near here

Mentalizing difficulties arise when an individual relies exclusively on automatic assumptions about the mental states of the self or others, which tend to be oversimplistic, or when the situation makes it difficult for the individual to appropriately apply their automatic assumptions. In fact, it could be that any psychological intervention in essence involves challenging such automatic, distorted assumptions, and requires that the patient make these assumptions conscious and attempt to reflect upon these assumptions in partnership with the clinician. In other words, any effective treatment is, at that level, about getting the patient to mentalize (we will discuss this point further, in the section below, 'Reconceptualization of Treatment').

Most experts agree that two systems for mentalizing arise from different neurocognitive mechanisms, both specialized for thinking about mental states interpretation (Apperly, 2011). The automatic system develops early and tracks mental states in a fast and efficient way, while the explicit system develops later, operates more slowly and makes heavier demands on executive functions (working memory and inhibitory control). Explicit mentalizing allows us to explain and predict behaviour, and has a role in social regulation (McGeer, 2007). However, it is the balance of automatic and controlled mentalizing that is critical. Explicit reflection cannot feel real unless it is contextualized by an intuitive awareness of the mental states being reflected on.

Stress and arousal, especially in an attachment context, brings automatic mentalizing to the fore and inhibits the neural systems that are associated with controlled mentalizing (Nolte et al., 2013). This has important implications for clinical work: any intervention that calls for reflection, by asking for clarification or elaboration on a thought, is by its very nature asking the patient to engage in controlled mentalizing.

Many patients may perform relatively well (in terms of mentalizing) under low-stress conditions. But under higher level of stress, when automatic mentalizing naturally kicks in, the patient may find it much more difficult to activate the processes that underpin controlled mentalizing, and so will find it harder to understand and reflect on what might be happening.

Self versus others

This mentalizing dimension involves the capacity to mentalize one's own state – the *self* (including one's own physical experiences) or the state of *others* (see Box 1.4). The two are closely connected, and an imbalance signals vulnerability in mentalizing both others and/or the self. Individuals with mentalizing difficulties are likely to preferentially focus on one end of the spectrum, although they may be impaired at both.

It is a central tenet of our attachment-based approach that a sense of self and the capacity to mentalize both develop in the context of attachment relationships. The child observes, mirrors and then internalizes his or her attachment figures' ability to represent and reflect mental states. Hence, the self and others – and the capacity to reflect on the self and others – are inevitably closely intertwined. In line with these assumptions, neuroimaging studies suggest that the capacity to mentalize about others is closely related to the ability to reflect on oneself because the two capacities rely on common neural substrates (Lieberman, 2007). Therefore, it is not surprising that disorders that are characterized by severe impairments in feelings of self-identity – most notably, psychosis and BPD – are also characterized by severe deficits in the ability to reflect about others' mental states.

BOX 1.4 near here

However, this should not be taken to mean that an individual whose capacity to mentalize themselves is impaired will *always* show similar impairments in the ability to mentalize others. Some individuals may have fewer universal impairments in mentalizing in relation to the self and others, and have stronger skills at one end of this spectrum of mentalizing. For example, individuals with ASPD can often be surprisingly skilled in 'reading the mind' of others, but typically lack any real understanding of their own inner world.

Still, following the neuroimaging literature, we can identify two distinct neural networks used in self-knowing and knowing others (Lieberman, 2007). The first of these is a *shared representation* system, in which empathic processing relies on shared representations of others' mental states. This represents a kind of 'visceral recognition' that occurs while experiencing and observing others experiencing states of mind, which operates through a mirror-neurone motor-simulation mechanism (Lombardo et al., 2010). The second is the *mental state attribution* system, which relies more on symbolic and abstract processing (Ripoll, Snyder, Steele, & Siever, 2013). In line with our expectation of the way the dimensions of mentalizing function, these two systems may be mutually inhibitory (Brass, Ruby, & Spengler, 2009), in that the neural regions most often recruited in the inhibition of imitative behaviour are those involved in explicit mental state attributions.

Internal versus external mentalizing

Mentalizing can involve making inferences on the basis of the *external* indicators of a person's mental states (e.g. facial expressions) or figuring out someone's *internal* experience from what we know about them and the situation they are in (see Box 1.5). This dimension does not just refer to a process of focusing on the externally visible manifestations versus the internal mental state of others, it also applies to the self – it includes thinking about oneself and one's own internal and external states. From the perspective of clinical assessment, the internal–external distinction is particularly significant in helping us to understand why some patients appear to be seriously impaired in their capacity to 'read the mind' of others, yet they may be hypersensitive to facial expressions or bodily posture, giving the impression of being astute about others' states of mind. Someone who has poor access to and great uncertainty about their subjective experience may come to a conclusion about what they are feeling from observing their own behaviour as well as the reactions of others: their legs feel restless, therefore they must be feeling anxious. The external focus can make a person extremely vulnerable to the observable behaviour of others. The absence of confident knowledge about the internal creates a thirst for clues from others' reactions even when these are not direction at oneself. Seeing someone else anxiously fidget can stimulate an internal state of unease and worry to a greater extent than it might normally do if mentalizing was not imbalanced in favour of the external.

BOX 1.5 near here

Mentalizing difficulties may become apparent only when the balance of internal and external cues used to establish the mental states of others is considered. For example, BPD patients often tend to hypermentalize emotions in others, including the clinician. This is because they pay more attention to external indicators of mental states and their initial ideas are left unchecked by controlled/reflective mentalizing (which might limit the possibilities for attributing thoughts and feelings). For example, if the clinician leans back and opens his mouth even slightly, the patient may believe that this was a yawn indicating that the clinician is bored with them. Or if the clinician frowns, perhaps pensively, the patient may interpret this as looking angry or disgusted with them. There has been considerable research on BPD patients' hypersensitivity to facial cues; their performance in the 'Reading the Mind in the Eyes' test can be better than normal, creating an impression in clinicians that their patients are better than average mind-readers (sometimes called the 'borderline empathy paradox'; Dinsdale & Crespi, 2013). A focus on external features, in the absence of reflective mentalizing, makes an individual highly vulnerable in social contexts, as it generates the kind of interpersonal hypersensitivity well described by Gunderson and Lyons-Ruth (2008). In MBT, mentalizing interventions often need to start by examining the patient's interpretations of a person based on external cues and then consider possible plausible scenarios about what their internal states of mind may be – encouraging the patient to take into account the subtleties and complexities of people's internal worlds.

Cognitive versus affective mentalizing

Intense emotion appears to be incompatible with serious reflection on mental states. This point hardly needs to be made, but, as with much that is obvious, neuroimaging studies have provided biological confirmation. For example, emotional activation has been shown to limit people's ability to 'broaden and build' in the face of stress – that is, to open up their minds to new possibilities (broaden), and to build upon their personal resources that facilitate resilience and wellbeing. In an fMRI study of 30 healthy females, it was found that during a provocative confrontation, high emotional reactivity to threat suppresses recruitment of the mentalizing network (Beyer, Munte, Erdmann, & Kramer, 2014).

Cognitive mentalizing involves the ability to name, recognize and reason about mental states (again, in both oneself or others), whereas *affective* mentalizing involves the ability to understand the *feeling* of such states (again, in both oneself or others), which is necessary for any genuine experience of empathy or sense of self (see Box 1.6). Some individuals give undue weight to either cognitive or affective mentalizing. Studies have suggested that BPD patients have a deficit of cognitive empathy (Harari, Shamay-Tsoory, Ravid, & Levkovitz, 2010; Ritter et al., 2011), which is coupled with heightened sensitivity towards any kind of emotional cue (Lynch et al., 2006). This suggests that these patients may have an emotional processing advantage perhaps linked to a combination of amygdala overactivation and orbitofrontal cortex and PFC regulatory deficits (Domes, Schulze, & Herpertz, 2009).

BOX 1.6 near here

Context/relationship-specific nature of mentalizing

Mentalizing, then, is made up of different dimensions. All of us are likely to be more or less skilled at some of these dimensions, but individuals with personality pathology tend to have pronounced impairments along some of the dimensions, resulting in an imbalance in mentalizing and occasionally outright mentalizing failures. In this section we will discuss the situations that are more likely to trigger mentalizing failures or difficulties. As well as not being one single 'thing,' mentalizing changes over time, and particular situations and stimuli are more likely to lead to mentalizing difficulties. For instance, BPD patients may be able to perform mentalizing tasks relatively well in experimental settings, but when they become emotionally aroused (for example, in a difficult interpersonal situation), they may show considerable confusion as they become dominated by automatic assumptions about other people's internal states and find it challenging to reflect on and moderate these assumptions. In other words, when in a state of emotional arousal, they typically lose the ability for controlled mentalizing and are likely to struggle to imagine a rational scenario that might explain the states of mind of others.

Heightened psychological arousal tends to cause the capacity for controlled mentalizing to become increasingly difficult to access, and automatic and non-reflective mentalizing starts to dominate. Up to a point this is a normal 'fight or flight' response to stress, which has the advantage of allowing us to respond immediately to danger. However, in situations of social interpersonal stress, more complex, cognitive and reflective functioning may be more helpful, and an inability to use these more controlled and conscious skills can lead to real difficulties in dealing with other

people. We have all noticed that, given a certain amount of emotional arousal, it becomes hard to focus on someone else's point of view. When emotional, not only does it become much harder or even impossible to concern oneself with the other person's perspective; we can also be quick to make assumptions on the basis of flimsy observations. We can become convinced that our point of view is the only valid one, and ignore everything we know about the other person except what is relevant to support our point of view. Therefore, the degree to which an individual finds themselves affected by interpersonal stress may make a critical difference to their mentalizing skills across life experiences. It seems likely that the threshold for switching to an automatic (fight or flight) style of mentalizing will be lowered in people who have been exposed to stress or trauma in early life. There may also be a genetic influence on the ease with which people are likely to switch to this automatic, uncontrolled mentalizing mode.

There is also some evidence that the activation of the attachment system is linked with the deactivation of mentalizing. Imaging studies (for example, Nolte et al., 2013) have shown that the brain areas normally associated with maternal and romantic attachments appear to suppress activity in brain regions associated with different aspects of cognitive control, including those associated with making social judgements and mentalizing. Anything that stimulates the attachment system (beyond stress-induced arousal), therefore, seems to bring with it a general loss of mentalizing capacity. A traumatic experience will arouse the attachment, system and attachment trauma may do so chronically. The hyperactivation of the attachment system in people with a trauma history may account for the dramatic loss of mentalizing capacity experienced by some individuals in emotional situations that trigger their attachment-seeking instincts. Attachment trauma probably hyperactivates the attachment system because the person to whom the child needs to turn in a state of anxiety (their attachment figure, usually a parent) is the very person causing the fear in the first place. The quick-fire triggering of the attachment system in BPD may be a result of past trauma, and it shows itself in the tendency of BPD patients both to move to positions of intimacy with undue haste and to be vulnerable to the temporary loss of mentalizing skills when in interpersonally intense situations.

Such moments of mentalizing failure are significant because they make it difficult for someone to relate to others in the context of an attachment relationship. When mentalizing fails in this way, there tends to be a re-emergence of non-mentalizing modes of behaviour, which can lead to powerful complications and profound disturbances in relationships. We will discuss these non-mentalizing modes next.

The re-emergence of non-mentalizing modes

When mentalizing fails (as typically happens in individuals with BPD, particularly in high arousal contexts), individuals often fall back on non-mentalizing ways of thinking that have parallels with the ways in which young children behave before they have developed full mentalizing capacities (hence, they may also be termed *pre-mentalizing* modes). These modes of experiencing the self and others tend to re-emerge whenever we lose the ability to mentalize. The modes are termed *psychic equivalence mode*, *teleological mode* and *pretend mode*.

While the dimensions of mentalizing can reflect anomalies in terms of mechanisms, on the whole, that is not what the clinician sees. The whole-person perspective that clinicians are obliged to take must address the phenomenology or subjectivity of our patients. Their experience is not that of a single brain mechanism out of kilter with the rest, but of a whole system functioning sub-optimally. What the patient and the mentalizing clinician see is a product of a malfunctioning mentalizing system, driven by imbalances in the dimensions of mentalizing. We have grouped the outcomes of these malfunctions under three typical modes of non-mentalizing subjectivity for the purpose of clinical experience. These non-mentalizing modes are important for the clinician to recognize and understand, as they tend to emerge in the consulting room and refer to aspects of the patient's experience. It is important to address these, because they can cause considerable interpersonal difficulties and result in destructive behaviours.

In the *psychic equivalence mode*, thoughts and feelings become 'too real' to a point where it is extremely difficult for the individual to entertain possible alternative perspectives (see Box 1.7). When mentalization gives way to psychic equivalence, what is *thought* is experienced as being real and true, leading to what clinicians describe as 'concreteness of thought' in their patients. There is a suspension of doubt, and the individual increasingly believes that their own perspective is the only one possible. Psychic equivalence is normal in a child of around 20 months who has not yet developed full mentalizing skills. Young children, and patients with BPD who are in this mode, describe an overriding sense of certainty about their subjective experience, whether this is that *'there is a tiger under the bed'* or *'these drugs are harming me'*. Such a state of mind can be extremely frightening, adding a powerful sense of drama and risk to life experiences. The sometimes exaggerated reactions of patients are justified by the seriousness and 'realness' with which they can experience their own and others' thoughts and feelings. The vividness and bizarreness of subjective experience can appear as quasi-psychotic symptoms and are also manifest in the physically compelling memories associated with posttraumatic stress disorder.

BOX 1.7 near here

In the *teleological mode*, states of mind are recognized and believed only if their outcomes are physically observable (see Box 1.8). Hence, the individual can recognize the existence and potential importance of states of mind, but this recognition is limited to very concrete situations. For example, affection is perceived to be true only if it is accompanied by physical contact such as a touch or caress. A patient who experiences mentalizing failure and falls into the teleological mode may express this by 'acting out', by carrying out dramatic or inappropriate actions or behaviours in order to generate outcomes from others whose claims of subjective states (e.g. of being concerned about the patient) are not credible to them. The teleological mode shows itself in patients who are imbalanced towards the external pole of the internal–external mentalizing dimension – they are heavily biased towards understanding how people (and they themselves) behave and what their intentions may be in terms of what they physically do.

BOX 1.8 near here

In the *pretend mode,* thoughts and feelings become severed from reality (see Box 1.9). Taken to an extreme, this may lead to feelings of derealization and dissociation. A pre-mentalizing young child creates mental models and pretend worlds, which the child can maintain only as long as these are completely separate from the real world (for example, as long as an adult does not interrupt or spoil the game by 'getting it wrong'). Similarly, patients in pretend mode can discuss experiences without contextualizing these in any kind of physical or material reality, as if they were creating a pretend world. The patient may *hypermentalize* or *pseudomentalize*, a state in which they may say much about states of mind but with little true meaning or connection to reality. Attempting psychotherapy with patients who are in this mode can lead to lengthy but inconsequential discussions of internal experience that have no link to genuine experience. A patient who shows considerable cognitive understanding of mentalizing states but little affective understanding may often hypermentalize. This state can often be difficult to distinguish from genuine mentalizing, but it tends to involve excessively lengthy narratives, devoid of a real affective core or of any connection to reality. On first impressions, hypermentalizing can lead the clinician to believe that they are working with an individual with extraordinary mentalizing capacities, but after a little while they discover that they are unable to resonate with the feelings underlying their patient's mentalizing efforts (Allen, Fonagy, & Bateman, 2008). In addition, because in pretend mode there are no real feelings or emotional experiences providing the individual with constraints, he/she may misuse his/her cognitive capacity in self-serving ways (e.g., to get others to care for or feel compassion toward him or her, or to control or coerce others).

BOX 1.9 near here

As the astute reader will have noticed, imbalances within the dimensions of mentalizing predictably generate the non-mentalizing modes described above. Psychic equivalence is inevitable if emotion (affect) dominates cognition. Teleological mode follows from an exclusive focus on external features to the neglect of the internal. Pretend mode thinking and hypermentalizing are unavoidable if reflective, explicit, controlled mentalizing is not well established. Although we cannot go into detail here, the normal predominance of non-mentalizing in the early years can be predicted from what we know about the developmental unfolding of mentalizing capacities. For example, as affect-focused mental state thinking antedates more cognitive mentalizing (Harris, de Rosnay, & Pons, 2005), psychic equivalence (and the anxieties that accompany it) will almost inevitably be part of the life of a child 3 to 5 years of age.

These three pre-mentalizing modes are particularly important to recognize in patients as they are often accompanied by a pressure to externalize unmentalized aspects of the self (so-called *alien self* parts). This may be expressed in attempts to dominate the mind of others, self-harm, or other types of behaviour that in the teleological mode are expected to relieve tension and arousal, a feature typical of BPD (Fonagy & Target, 2000).

The concept of the 'alien self'

At moments of mentalizing failure, as well as falling back on pre-mentalizing modes, we also experience a pressure for what psychodynamic clinicians would recognize as

'projective identification'. This term has many meanings, and this has led us to talk about one aspect of this – the *externalization of the alien part of the self*.

Because mentalizing generates self-coherence, the faltering of mentalizing can signal a sense of fragmentation, a painful state from which we often seek shelter by extreme and even violent acts. Emotional interactions are painful in part because intense emotions undermine mentalizing and the natural and understandable reaction is to try to restore a sense of cohesion by dramatic action. When I find myself in an intense quarrel with a friend and I get 'emotional,' only a small part of this is the emotion I feel in relation to the argument; the lion's share is likely to be me trying to maintain my sense of self and identity. I may achieve this by:

1. Being excessively assertive (raising my voice)
2. Blinding myself to the potentially 'confusing' perspective of my friend,
3. Creating an image of him that is highly self-serving and confirms me in my position as coherent, accurate and, above all, beyond reproach
4. Forcing a reaction from him to affirm me still further or make me feel 'even more real'.

From a dispassionate, external perspective, the impression is as much of me trying to escape from a painful situation as trying to engage effectively in discussion or debate.

What am I so busy trying to protect myself from? To understand this we have to introduce the concept of the *alien self* (see Box 1.10). We assume, as suggested by Winnicott (1956), that when a child cannot develop a representation of his/her own experience via parental mirroring (of the psychological self), he/she internalizes the image of the caregiver for affirmation as part of his/her self-representation. While this is used to bolster the infantile self, it is not contingent with the self-state: it does not match it in quality, intensity, timing or tone. This discontinuity within the self is the 'alien self'. We understand the excessively controlling behaviour shown by children with a history of disorganized attachment as the persistence of a pattern analogous to projective identification, where the experience of incoherence within the self is reduced through externalization: that is, placing an aspect of the self on to another person by nudging them to behave according to the representation that requires externalization. As a young person one of us (PF) used to phone home in states of distress and talk about his situation in catastrophic terms until his parents were palpably panicked and then he would end the conversation feeling relieved. It was not until he received similar communications from his own children that he realized fully just what power this process had on parental wellbeing. If the alien self is an experience of vulnerability, the person creates this experience in his communication partner by generating chronic uncertainty; if it is aggression, he simply has to irritate him; if it is depression or disinterest and hopelessness, then he might force him to experience the potential of helping, only to dash his hopes again and again. In all these cases, the person resolves an internal incoherence, normally covered over by a capacity to create an illusion of coherence through mentalizing, by ridding oneself of its source – the alien self – on to someone in the external world.

BOX 1.10 near here

In people with personality disorder, the need for this externalizing can feel like a matter of life and death, not just a momentary relief from discomfort. This is because

the alien self can frequently become the vehicle for the experience of maltreatment, the host to a genuinely malevolent intentionality that has taken residence inside the self and expresses its malevolent intent from within through unmoderated self-destructiveness (see Box 1.11). This aspect of the alien self, too, is brought into relief by the disappearance of the 'self-generating' mentalizing narrative, which normally bridges cracks in the self-structure and prevents them from undermining self-coherence. Loss of mentalizing destabilizes the self, provoking an uncertainty – *'Who am I?'*; *'Who are they?'*; *'What do they want?'*; *'Who am I in relation to them?'* No answers are available to the individual and panic ensues. As it does so, the individual attempts to recapture a sense of self by schematic representation – *'I understand this if he does not like me – he is victimizing me and I am a victim'*. To manage this state of mind, individuals project aspects of themselves that are destabilizing, and see them in the other. The alien aspects of the self are most dangerous to the individual's integrity and narrative structure.

BOX 1.11 near here

Failures of mentalizing reveal discontinuities in the structure of the self. This happens simply because the narrative of intentionality that all of us continuously create for ourselves depends on mentalizing being available. When there is a break in mentalizing, discontinuities in our self-representation also become more prominent and threatening. At these points, coherence can be restored by attributing ownership to undesired aspects of oneself (those that are experienced as alien) to another person. In a personal quarrel, someone might accuse a friend of being controlling, inflexible, of caring nothing about other people's point of view, of being unable to listen to an argument, and so on. Non-mentalizing begets non-mentalizing. Relationships become rigid and fixed and the other has to be controlled, almost forced, to keep and not give up playing the roles of alien parts of the self. In fact, unfair accusations will only anger the friend and rile him exactly to the level of angry unreasonableness that the person finds hard to tolerate in himself. It is at odds with their normal self-representation, because it is a part of the self that a frequently tired and short-tempered mother, responding to my pleas for comforting, might have created there. The alternative to this successful externalization would be destructive non-mentalized self-criticism, experienced as truly persecutory in psychic equivalence mode. In a teleological mode, this state can represent a genuine risk – that is, a physical risk, through self-harm or ultimately suicide. The need for the other as a vehicle for the alien self can be overwhelming, as the patient experiences it as a matter for survival, and an adhesive, addictive pseudo-attachment to this individual can develop.

Ostensive cues and epistemic trust

The most recent theoretical developments in our thinking about mentalizing and therapeutic change have important implications for how we approach our clinical practice. This new thinking involves the theory of *epistemic trust*. In short, this theory emphasizes the social and emotional significance of the trust we place in the information about the social world that we receive from another person – that is, the extent and ways in which we are able to consider social knowledge as genuine and personally relevant to us (see Box 1.12). The theory builds on the ground-breaking work of the Hungarian psychologists Gergely and Csibra about the evolutionary

importance of human infants' capacity to learn from their primary caregivers. According to the theory, human beings have evolved to both teach and learn new and relevant cultural information, and to do this we have evolved particular sensitivity to forms of communication that indicate opportunities for this kind of learning. As part of this process of communication, a caregiver signals to the child that what they are conveying is relevant and can be considered useful and valid cultural knowledge (see Box 1.13). To do this, the caregiver uses what we term *ostensive cues*. Human infants are attuned to respond with particular attention to these signals (Csibra & Gergely, 2011). Ostensive cues include eye contact, turn-taking contingent reactivity, and the use of a special vocal tone (*motherese*), all of which appear to trigger a special mode of learning in the infant (see Box 1.13). We believe that this happens because the ostensive cues indicate to the infant that the caregiver recognizes the child as an individual, and as a mentalizing (thinking and feeling) 'agent'. In brief, sensitive responding to the child's need fosters not just a general confidence that he/she matters as a person, but also serves to open his/her mind more generally to receive new information as relevant and alter his/her beliefs and modify his/her future behaviour accordingly.

BOXES 1.12, 1.13 and 1.14 near here

Ostensive cues trigger *epistemic trust*: they signal that what the caregiver is trying to convey is relevant and significant, and should be remembered. A securely attached child is more likely to treat the caregiver as a reliable source of knowledge, and this trust is likely to generalize to other people in a position to teach and learn from. But what of individuals whose social experiences have led them to a state of chronic *epistemic mistrust*, in which (perhaps because of hypermentalizing) they imagine the motives of the communicator to be malign? Such individuals will appear to be resistant to new information, and might come across as rigid, stubborn or even bloody-minded, because they treat new knowledge from the communicator with deep suspicion and will not internalize it (i.e. they will not modify their internal mental structures to accommodate it). Their epistemic trust has been undermined by their previous experiences, and as a consequence an evolutionarily prepared channel for the acquisition of personally relevant information is partially blocked. We suspect that it is less likely to be the frank brutality of abuse that undermines epistemic trust (although of course it can do), and that genetic predisposition, in combination with neglect and emotional abuse, will play a larger role in making an individual excessively vulnerable to distrusting information from others (see Box 1.15).

BOX 1.15 near here

Everybody seeks social knowledge to help navigate the interpersonal world. We are all often insecure in relation to our own beliefs and intuitions, and seek input and reassurance from others. This, of course, is more likely to be the case for someone whose consistent insecurity has left them at the edge of the interpersonal lattice of social understanding. Yet, even though this individual's need for confirmation may be more intense than normal and anxiously sought, the content of such reassuring communications may be rejected, their meanings confused or they may even be misinterpreted as having hostile intent, leaving the person in a state of chronic uncertainty yet without means of meaningful redress. A person whose channels for learning about the social world have been disrupted – for example, one whose social

experiences with caregivers during childhood have caused a breakdown in epistemic trust – is stuck in a general state of uncertainty and permanent epistemic vigilance. An individual with a history of trauma has little reason to trust, and will reject information that is inconsistent with their existing beliefs. Precluding themselves from social information in this way will create an apparent rigidity, or reluctance to change. This rigidity is underpinned by epistemic mistrust and a state that may be characterized by 'hearing but not listening' (see Box 1.16).

As clinicians we may end up calling these individuals 'hard to reach', yet they are simply showing what may be a reasonable adaptation to a social environment where information from most attachment figures is 'tagged' as likely to be misleading (see Box 1.17). Notwithstanding the behaviour of a parent or a partner as faultlessly supportive and invariably acting in the patient's interest, or a clinician who consistently offers valuable and accurate advice, the patient apparently takes no notice, ignores the evidence of cooperativeness and support, and continues (from the point of view of others, 'persists') to feel abandoned, betrayed and unsupported. It is as if the patient is blind to the evidence, as it runs contrary to their belief. According to this perspective, we can see the destruction of trust in social knowledge as a key mechanism in pathological personality development. This has significant implications for how we understand how and why psychological therapies for BPD and ASPD work.

BOXES 1.16 and 1.17 near here

Reconceptualization of treatment: three systems

In the case of BPD, a considerable number of different therapies have now been found to be effective (Stoffers et al., 2012). What these treatments have in common is a clear theoretical framework and a reliable model for the delivery of treatment. Beyond this, though, it is not yet known whether there is a single factor, common to all these therapies, that makes them effective. Clearly, understanding what make interventions effective (or what renders them ineffective) would be of great value in the formulation of future interventions and the refinement of existing practice.

In the light of our argument above about epistemic trust, we suggest that successful treatments all involve three essential systems of communication relating to epistemic trust and social learning (see Box 1.18). MBT has been informed by these three principles of change. Over the past few years specific components have been increasingly emphasized to take into account our understanding of the processes underpinning effective treatment. In each section below we identify how MBT interventions relate to each component of the change system. Different techniques are emphasized at different stages of treatment and change, for example communication change system 1 being of greatest importance at the beginning of treatment although maintaining a place for the clinician and patient throughout treatment.

BOX 1.18 near here

Communication change system 1: the teaching and learning of content and the increase of epistemic openness

All evidence-based psychotherapies provide a coherent framework that enables the patient to examine the issues that are deemed to be central to him/her, according to a particular theoretical approach, in a safe and low-arousal context. These psychotherapies provide the patient with helpful skills or knowledge, such as strategies to handle emotional dysregulation or restructure thinking about interpersonal relationships. 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 for the patient to feel recognized and understood as an agent, empowered to make decisions and to alter the course of their path through life. The conceptual model of each treatment 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).

BOX 1.19 near here

MBT initially takes a more directive and informative approach, and we summarize some examples of how MBT addresses Communication System 1 here (see also Box 1.19). MBT requires the clinician and patient to:

(1) Develop a collaborative formulation with the patient early in the assessment process. This is written and shared with the patient and is constantly revised when new understanding develops

(2) Identify mentalizing vulnerabilities using examples personal to the patient. Pathways to the loss of mentalizing are identified and established as 'vulnerability points' to be monitored carefully;

(3) Discuss the diagnosis in terms of the patient's symptoms and history. The diagnosis is less important than agreeing a lens through which the variability of symptoms can be understood;

(4) Map attachment patterns and how they play out in current relationships. The identification of attachment strategies is essential if the patient and clinician are to recognize their deployment during treatment

(5) Engage the patient in an introductory phase which combines psychoeducation with some interpersonal process. The MBT-I group (see Chapter 10) offers the patient and clinician a shared framework for understanding BPD and the whole process of therapy

(6) Establish a developmental narrative of problems. The patient's background and context support a compassionate view of the problems

(7) Agree joint goals relevant to the patient so that therapy is about what is important to the patient.

In essence, we suggest that such explanations and suggestions may be seen as ostensive cues that signal to the patient the relevance to them of information that is being conveyed. These cues serve to trigger in the patient a feeling of being personally recognized by the clinician or the therapeutic situation. This process is important because it allows the patient to reduce his/her 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, as well as doubtless being useful in its own right, has the nonspecific effect of creating openness. This openness 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 content conveyed within the therapeutic model, which, because it is accurate and helpful, generates epistemic openness*. The growth of epistemic trust, in turn, allows the patient to take in further information that also serves to reassure and validate him/her. 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.

However, the fact that so many different therapies using widely differing theoretical models have been found to have considerable beneficial effects indicates that the significance of System 1 lies not so much in the essential truth of the wisdom conveyed by the clinician and the therapeutic model, but more importantly in the fact that it allows the patient to apply this new *received learning* in a more or less concrete way, changing the nature of the communication between patient and clinician in the direction of increased epistemic trust. This brings us to System 2.

Communication change system 2: the re-emergence of robust mentalizing

As noted above, through passing on knowledge and skills that feel appropriate and helpful to the patient, the clinician implicitly recognizes the patient's agency. The clinician's presentation of information that is personally relevant to the patient serves as a form of ostensive cueing that conveys the impression that the clinician seeks to understand the patient's perspective; this in turn enables the patient to listen to and hear the clinician's intended meaning. In effect, the clinician is demonstrating how he/she engages in mentalizing in relation to the patient. It is important that in this process both patient and clinician come to see each other more clearly as intentional agents (i.e. individuals seeking to mentalize). For example, when the clinician shows that his/her mind has been changed by the patient, the clinician gives agency to the patient and increases his/her faith in the value of social understanding. 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 a more trusting relationship to develop between clinician and patient. Ideally, the patient's feeling of having been sensitively responded to by the clinician opens a second virtuous cycle in interpersonal communication in which *the patient's own capacity to mentalize is regenerated* (see Box 1.20). This is the core of MBT.

BOX 1.20 near here

MBT recommends an authentic 'not-knowing' stance that forms the bedrock for exploration of the patient's perspective. Empathic validation and establishing a shared affective platform held between patient and clinician increases the patient's experience that he/she is not alone and indicates that another mind can be useful to

clarify mental states and increase a sense of agency. Increasing focus on affect and interpersonal interaction during a session and over time provides the context in which to explore ever more complex states of mind within an attachment context that would normally trigger loss of mentalizing. The mind of the clinician is open to the patient to the extent that the clinician actively demonstrates mentalizing about the patient stating what is in his mind and giving his perspective. Subjectivity is held to be of importance and not subjugated. The patient has to consider the clinician's viewpoint just as the clinician has to consider the patient's. Perspectives are expected to change when new information becomes available; minds change minds in a transactional manner.

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 that is increases 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 advanced 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 way their clinician thinks of them – it could be said that they 'find themselves in the mind of the clinician'. It is also vital to a further function of therapy: the rekindling of the patient's wish to learn about the world, including the social world. We believe that this is a complex and non-linear process, but it can be summarized briefly as follows: the insight obtained in therapy, whatever its content, creates or recreates the potential for the patient to have 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*.

Here we would like to emphasize a point that may seem initially puzzling, given our own declared commitment to mentalization-based psychotherapy: *mentalizing in itself is only an intermediate step, not the ultimate therapeutic objective*. Simply instructing the clinician to focus the patient on his/her 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 mind-set of the person undergoing treatment. However, the process of creating a more robust mentalizing function in therapy (System 2), although a likely necessary step, can no more assure enduring change in the patient than System 1. True and lasting improvement, we believe, rests on a third communication system: learning from experience beyond therapy.

Communication change system 3: the re-emergence of social learning with improved mentalizing

We hypothesize that rekindling epistemic trust through improved mentalizing, which permits the person to understand better and opens them up to feeling understood, in turn reopens the key evolutionarily determined route to information transmission and the possibility of taking in knowledge that is felt to be personally relevant and generalizable. Overcoming epistemic mistrust, so that positive social information that has previously been disavowed is now registered, enables the patient to alter his/her beliefs. This is the vital component of change; it is what brings about genuine alteration in previously rigidly held beliefs. In essence, the experience of feeling thought about enhances mentalizing, which in turn enables us to learn new things about our social world (see Box 1.21).

The therapeutic situation teaches about sources of knowledge. It provides a clear social illustration of trust, making the clinician a 'deferential source' of knowledge (Wilson & 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 experience feeling 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 they encounter these in day-to-day life.

BOX 1.21 near here

MBT recommends that early in therapy the patient's social context is stabilized. Change becomes impossible if housing, financial, employment, probation, and other stressors are dominant. The MBT clinician is an active advocate for the patient's link to the wider social system. Once treatment is stable and when mentalizing is established with greater constancy and less vulnerable to daily assaults, clinician and patient consistently work on interpersonal process both within and without the patient-clinician relationship. Exploration about attachment process in the therapy relationship is not seen as the end point but merely a stage to focus meaningfully on current relationships in the patient's life. How does the patient understand a negative response from an important person in her life, how does she respond to sensitive reactions from others? Too often epistemic hypervigilance interferes with getting what is good in an interaction and finding what might propel a joint relational endeavour forward.

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' behaviour. Social experiences that may have been positive but were in the past discounted as a result of the patient's epistemic hypervigilance and rigidity now have the potential to have a positive impact and be learned from. This is the third system of communication, which becomes available once the second system, tied to the therapeutic situation, has improved the patient's capacity to mentalize. As patients begin to experience social interactions as more benign and interpret social situations more accurately (e.g. being able to see an

experience of temporary social disappointment as simply this, rather than a total rejection of themselves), they update their knowledge of both themselves and others.

It is the recovery of the capacity for social information exchange that, we feel, may be at the heart of effective psychotherapies for BPD, of which MBT is one. 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 clinicians 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. Studies in which change was monitored session by session have suggested that the patient–clinician alliance in a given session predicts change in the next (Falkenstrom, Granstrom, & Holmqvist, 2013). This indicates that the change that occurs between sessions is a consequence of changed attitudes to learning engendered by therapy, influencing the patient's behaviour between sessions. The implication is that the extent to which a patient benefits from therapy depends partly on what he/she encounters in his/her social world during and after treatment. Because of this, 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. Clinical experience suggests that there is likely to be some validity to this assertion, although there is not yet evidence from research to support it.

This admittedly speculative model offers a way to integrate the specific and non-specific 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 non-specific pathway, the patient's capacity to mentalize is fostered. Both of these systems would be expected to lead to symptomatic improvement. Improved mentalizing and reduced symptomatology both improve the patient's experiences of social relationships. However, it is likely that these new and improved social experiences, rather than just what happens within therapy, serve to erode the epistemic hypervigilance that has previously prevented benign social interactions from changing the patient's experience of themselves and of the social world. *Meaningful change is thus possible only if the person can use their social environment in a positive way (and is the social environment is sufficiently supportive to allow this to happen).* For this to happen, recognition of self-agency is key, and this recognition is best achieved through the ostensive cues that are provided by feeling appropriately mentalized by another person. For the social environment to be accurately interpreted so that it can provide opportunities for new learning, mental state understanding of others' actions and reactions is critical – and only improved mentalizing will achieve this. For the benefit of social experience to be preserved through the maintenance of improved relationships, emotion regulation and good behavioural control are key – and, once again, only improved mentalizing will deliver these. This is essentially why MBT focuses on this capacity, and why its realization is the focus of this practical guide.

Allen, J. G., Fonagy, P., & Bateman, A. W. (2008). *Mentalizing in clinical practice*. Washington, DC: American Psychiatric Press.

Apperly, I. A. (2011). *Mindreaders: The cognitive basis of 'Theory of Mind'*. Hove, UK: Psychology Press.

Bateman, A., & Fonagy, P. (2010). Mentalization based treatment for borderline personality disorder. *World Psychiatry, 9*, 11-15.

Bateman, A. W., & Fonagy, P. (Eds.). (2012). *Handbook of mentalizing in mental health practice*. Washington, DC: American Psychiatric Publishing.

Beyer, F., Munte, T. F., Erdmann, C., & Kramer, U. M. (2014). Emotional reactivity to threat modulates activity in mentalizing network during aggression. *Social Cognitive and Affective Neuroscience, 9*, 1552-1560.

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences, 364*, 2359-2367.

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.

Dinsdale, N., & Crespi, B. J. (2013). The borderline empathy paradox: Evidence and conceptual models for empathic enhancements in borderline personality disorder. *Journal of Personality Disorders, 27*, 172-195.

Domes, G., Schulze, L., & Herpertz, S. C. (2009). Emotion recognition in borderline personality disorder – A review of the literature. *Journal of Personality Disorders, 23*, 6-19.

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

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., & Luyten, P. (2016). A multilevel perspective on the development of borderline personality disorder. In D. Cicchetti (Ed.), *Developmental psychopathology* (3rd ed.). New York, NY: John Wiley & Sons.

Fonagy, P., Steele, H., & Steele, M. (1991). Maternal representations of attachment during pregnancy predict the organization of infant-mother attachment at one year of age. *Child Development, 62*, 891-905.

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.

Fonagy, P., & Target, M. (1997a). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology, 9*, 679-700.

Fonagy, P., & Target, M. (1997b). Research on intensive psychotherapy with children and adolescents. *Child and Adolescent Psychiatric Clinics of North America, 6*, 39-51.

Fonagy, P., & Target, M. (2000). Playing with reality: III. The persistence of dual psychic reality in borderline patients. *International Journal of Psychoanalysis, 81*, 853-874.

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.

Gunderson, J. G., & Lyons-Ruth, K. (2008). BPD's interpersonal hypersensitivity phenotype: A gene-environment-developmental model. *Journal of Personality Disorders, 22*, 22-41.

Harari, H., Shamay-Tsoory, S. G., Ravid, M., & Levkovitz, Y. (2010). Double dissociation between cognitive and affective empathy in borderline personality disorder. *Psychiatry Research, 175*, 277-279.

Harris, P. L., de Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. . *Current Directions in Psychological Science, 14*, 69-73.

Hughes, C., Jaffee, S. R., Happe, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Origins of individual differences in theory of mind: From nature to nurture? *Child Development, 76*, 356-370.

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology, 58*, 259-289.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., . . . Baron-Cohen, S. (2010). Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience, 22*, 1623-1635.

Lynch, T. R., Rosenthal, M. Z., Kosson, D. S., Cheavens, J. S., Lejuez, C. W., & Blair, R. J. (2006). Heightened sensitivity to facial expressions of emotion in borderline personality disorder. *Emotion, 6*, 647-655.

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*.

McGeer, V. (2007). The regulative dimension of folk psychology. In D. Hutto & M. M. Ratcliffe (Eds.), *Folk psychology re-assessed* (pp. 137–156). Dordrecht, The Netherlands: Srpinger.

Nolte, T., Bolling, D. Z., Hudac, C. M., Fonagy, P., Mayes, L., & Pelphrey, K. A. (2013). Brain mechanisms underlying the impact of attachment-related stress on social cognition. *Frontiers in Human Neuroscience, 7*, 816.

O'Brien, K., Slaughter, V., & Peterson, C. C. (2011). Sibling influences on theory of mind development for children with ASD. *Journal of Child Psychology and Psychiatry, 52*, 713-719.

Ripoll, L. H., Snyder, R., Steele, H., & Siever, L. J. (2013). The neurobiology of empathy in borderline personality disorder. *Current Psychiatry Reports, 15*, 344.

Ritter, K., Dziobek, I., Preissler, S., Ruter, A., Vater, A., Fydrich, T., . . . Roepke, S. (2011). Lack of empathy in patients with narcissistic personality disorder. *Psychiatry Research, 187*, 241-247.

Stoffers, J. M., Vollm, B. A., Rucker, G., Timmer, A., Huband, N., & Lieb, K. (2012). Psychological therapies for people with borderline personality disorder. *Cochrane Database of Systematic Reviews, 8*, CD005652.

Taumoepeau, M., & Ruffman, T. (2006). Mother and infant talk about mental states relates to desire language and emotion understanding. *Child Development, 77*, 465-481.

Taumoepeau, M., & Ruffman, T. (2008). Stepping stones to others' minds: Maternal talk relates to child mental state language and emotion understanding at 15, 24, and 33 months. *Child Development, 79,* 284-302.

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

Winnicott, D. W. (1956). Mirror role of mother and family in child development. In D. W. Winnicott (Ed.), *Playing and reality* (pp. 111-118). London, UK: Tavistock.