Manuscript Title:

The Role of Epistemic Trust in Mentalization-Based Treatment of Borderline Psychopathology

Tobias Nolte[1,2], Joost Hutsebaut[3,4], Carla Sharp[5], Chloe Campbell[1,2], Peter Fonagy[1,2] & Anthony Bateman[2]

Author Affiliations:

[1]Anna Freud National Centre for Children and Families, London, UK

[2]Research Department of Clinical, Educational and Health Psychology, University College London, London, UK

[3]Viersprong Institute for Studies on Personality Disorders (VISPD), P.O. Box 7, 4660 AA, Halsteren, The Netherlands

[4]Tilburg university, The Netherlands
[5]Department of Psychology, University of Houston, USA

Corresponding author: Tobias Nolte,

Abstract

Building on the recently developed notion of epistemic trust as facilitating social learning, in this paper we explore and clarify how interventions from Mentalization Based Treatment (MBT) for severe psychopathology such as Borderline Personality Disorder potentially generate this process in adults with this diagnosis.

We suggest first that being mentalized may be a critical cue in social interactions to establish epistemic trust, the individual's willingness to consider new knowledge as trustworthy and relevant and therefore worth integrating into their lives, and second that epistemic mistrust may represent a final common pathway through which aversive relational experiences in the past may exert their influence on psychosocial treatments – both as a disposition of the patient and as a characteristic of the therapist-patient encounter. We argue that our developmental, interpersonal view on the stimulation of epistemic trust in the context of MBT creates a new perspective on the role of the therapeutic relationship, especially in the work with patients with personality disorders in whom the capacity to internalize new information through social learning is undermined by the absence of epistemic trust.

By charting the interventions and building blocks of Mentalization-Based Therapy (MBT) from the initial assessment and formulation, through individual and group therapy sessions, to re-engaging with the wider social environment, this paper examines how each of these can potentially establish a “we-mode”, or an interpersonal experience associated with being mentalized. This, in turn, can unlock the barrier posed by epistemic vigilance. In addition, implications for relational mentalizing and rupture and repair within the therapeutic

relationship are discussed in terms of fostering the patient's sense of being recognized and better regulating affect.

Keywords: mentalizing, epistemic trust, epistemic stance, borderline personality disorder, attachment, we-mode, childhood adversity, Mentalization-Based Therapy (MBT)

Introduction

In this paper, we seek to clarify the role of various treatment components and clinical interventions in Mentalization-Based Therapy (MBT) for Borderline Personality Disorder (BPD) in facilitating social learning in patients. We will begin by describing how the particular experiences and difficulties associated with BPD are understood within MBT thinking. We will then consider recent conceptualizations of BPD and epistemic trust, conceptualizing BPD – and aspects of other personality disorders more generally – as expressing a breakdown in social communication, driven by disruptions in the associated social cognitive processes of mentalizing, the capacity for joint attention and the achievement of what is called the "we-mode", and openness to social learning facilitated by a context-appropriate reduction of epistemic hypervigilance. We will then discuss the clinical implications of these ideas, in particular, the task within MBT of restoring openness to epistemic trust and the salutogenic opportunities this restoration can provide.

The mentalization-based developmental framework to understand BPD

We have described mentalizing as both an implicit and explicit process by which we make sense of actions by others as well as our own behaviours by attributing intentional mental states. Mentalizing is conceptualized as a multidimensional and complex social cognitive

process which entails forming beliefs about the internal states of those with whom we interact and creating representations of such interactions that include the mental states underpinning our own actions (Bateman et al., Ventura et al., 2020). The interdependence of self and other mental state understanding suggests that the capacity to mentalize is a developmental achievement, and its attainment is dependent on the quality of early social relationships. This reflects the extent to which the child's subjective experiences are adequately mirrored and contingently responded to by trusted figures in the child's life (Fonagy, Gergely, Jurist, & Target, 2002; Fonagy & Luyten, 2009). There is a substantial body of evidence demonstrating that adversity and complex trauma have a detrimental effect on the development of mentalizing abilities (Luyten & Fonagy, 2019) (Luyten, Campbell, & Fonagy, 2019). Further, it appears effective caregiver mentalizing and in particular the capacity to mentalize in relation to traumatic experiences (trauma-RF) may be an important protective factor in moderating the relationship between childhood adversity and later outcomes (Berthelot et al., 2015; Ensink, Begin, Normandin, & Fonagy, 2017). As we have further argued, this notion needed to be extended so that attention is also given to the processes that enable or hinder learning from the wider social environment (Fonagy, Luyten, Allison, & Campbell, 2017a, 2017b; Luyten et al., 2020). Individuals with a diagnosis of BPD often employ ways of thinking and appraising interpersonal events and internal states that reflect inadequate or ineffective mentalizing. (Fonagy, Luyten, Allison, & Campbell, 2017a, 2017b; Luyten et al., 2020). The primary attributes of Borderline Personality Disorder (BPD) from a Mentalization-Based Therapy (MBT) perspective are: (a) a lack of stability in mentalization when in emotionally charged relationships; (b) a reversion to pre-mentalizing thought patterns during periods of stress; (c) an inclination to externalize internal states, which has been interpreted as the projection of painful and disorganized or unbearable self-

states; (d) a tendency to disrupt and obstruct efforts to restore interpersonal relationships, including the therapeutic relationship and (e) a disrupted process of social learning from reliable/trusted others (Bateman, Campbell & Fonagy, 2021). Indeed, MBT for BPD was developed on the understanding that susceptibility to loss of mentalizing, particularly during emotionally charged interpersonal interactions, is the underlying pathology that brings about the characteristic interpersonal, emotional and impulse symptoms of BPD (Fonagy & Luyten, 2009; Euler et al., 2021). Therefore, a key mechanism that underpins symptomatic improvement, such as the reduction of self-harm or suicidality, is the enhancement of effective mentalizing. More recently, it has become clearer (Fonagy & Allison, 2014 and Fonagy, Luyten and Allison, 2015) that a challenge to addressing these features therapeutically is to be found in the patient's epistemic stance. What is common to the recourse to pre-mentalizing modes under heightened arousal, when the attachment system is activated, is not only a momentary and sometimes prolonged shutdown of mentalizing (Nolte et al., 2013) and with this a lowering of the individual's capacity to reappraise self-states and the internal states of significant others but also, critically, to learn effectively (in a broad sense) from the social environment. This significant failure in learning from others includes, but is of course by no means limited to, the therapeutic relationship.

Epistemic trust and social learning in the We-mode

The work to support more robust and effective mentalizing is done not with the aim of the emergence of this engagement with mental states as an end goal, but rather as a process. We postulate that all psychotherapies in some way or another 're-appraise' (events, self, others etc). What we would like to stress is that openness to re-appraising requires a we-mode in order to generate mutual exchange of perspectives. From that perspective, the creation of

we-mode experiences is a pre-requisite for such reappraising processes to take place and to have impact. Through the therapeutic process, patient and therapist jointly focus on mental states, engaging in joint attention and co-mentalizing, resulting in more complex representations of what is occurring both within and outside the therapeutic relationship. This is achieved by activating, separating and differentiating the different representational levels of self and other experience within affective experience. The result is a form of social cognition known as "we-mode" or "we-ness" which is processed within the therapeutic relationship as "us" rather than "self" and "other" (Gallotti & Frith, 2013; O'Madagain & Tomasello, 2019; M. Tomasello, 2016) (see Fig. 1 for the three types of experience that underpin MBT thinking; Bateman et al., 2023). Frith (2012) argues that such metacognition is especially relevant to cooperative or we-mode procedures, which occupy a great deal of human waking life. There is well documented neuroscientific evidence of the brain processes that underpin the addressing functioning in the receiver of information – a state of particular attention that creates openness for joint attention and thus subsequent information transfer. (e.g. Kampe et al., 2003; Leong et al., 2017). Frith futhermore cites a range of experimental evidence showing how inaccurate unmodulated self-appraisal can be – we cannot easily see ourselves as others see us. He has shown experimentally how two heads are better than one: "through discussions of our perceptual experiences with others, we can detect sensory signals more accurately." (Frith, 2012).

I-Mode

- Creation of subjective experience of core self - continuous over time and not referenced through others.

  Sameness across time
- Interoceptive (the ability to identify, access, understand, and respond appropriately to the patterns of internal signals)
- Exteroceptive information (sensitivity to stimuli that are outside the body) contribute to aspects of I-mode

  Interacts with Me-mode at first order representation – others have different beliefs (e.g. false beliefs) and *may do so about me*
- Mismatch creates instability and experience of self and influence of alien self
  - Alien self embedded in self-image (this is a 'not me' experience)
  - Control others perception of self - protect I-mode forcefully
  - Become what I am seen as – defer and relegate I-mode
  - Distrust and avoidance – protect I-mode by deactivating social mentalizing

Me-mode

- Self as Object/Self in Social context
- Capacity to infer others mental states efficiently for smooth social interaction which include second order representations (recognition of how other person is thinking about/experiencing me)

- Components of self are referenced and calibrated through others in social interaction – self is an object that is described or narrated (Me-mode), rather than an entity that is validated by its coherence and action (I-Mode)
- Me-mode personalised
  - Ability and **confidence** in differentiating their mental states about me - how much insight we think that others have into our thoughts and intentions (i.e., 'do they know what I'm thinking?')
  - **Confidence** in hiding mental states – 'they don't know what I am thinking'

We-Mode

- Significance of 'joint attention' in human social cognition
    - Joint attention refers to the ability to focus with another on both external objects, and on mental content – of particular significance is the ability to understand how and why mental states might differ
    - Infant who benefits from being effectively mentalized is rich in experiences of joint attention to mental states. As well as feeling intensely rewarding for an infant, such moments of "we-ness" confer the powerful benefit of stimulating epistemic trust and creating an openness to collaborative social learning
- Coordination of perspectives
- Appreciate the distinction between the subjective (one's own view) and the objective (actual physical reality "out there")
- Coordinate knowledge (content) of another individual's mental state: quite a complex triangulation
- Cooperation is immeasurably advanced by being able to compare and coordinate different perspectives on the same situation
- Experience of being part of a set of thoughts and feelings that are beyond their own.

Figure 1. Modes of experience.

In MBT, 'We-ness' in the clinical interaction is distinct from either patient or clinician individual self/other perspectives. As an inherently relational representation, it is a state actively developed and promoted in MBT. The we-mode, it should be emphasised, does not involve the abdication of one's own agency or a fusion of thoughts. On the contrary, the value of the we-mode is that it is predicated on different perspectives on a shared object, allowing for the generation of perspectives and (self)-knowledge that could not be reached separately by any one mind – it is for this reason that, we think, the we-mode is such a powerful tool for harnessing social cooperation to achieve complex or difficult objectives (M. Tomasello, 2020) (Colle et al., 2020). The epistemic closure and rejection sensitivity often seen in individuals with a diagnosis of BPD make it difficult to achieve and sustain the "we-mode" experience;

this difficulty may be reflected in the subject's painful isolation from others and their intense need at times to become lost in other minds. We consider it key to recognizing one's agency and experiencing the salutogenic effects of cooperative intersubjectivity, which can provide a buffer when faced with difficulties or challenges. The phenomenon of social thinning (Goemans, Viding, & McCrory, 2021; McCrory et al., 2022) often associated with BPD might be understood as the traumatized individual's inability to experience the we-mode, which is rooted in the difficulties in mentalizing found in such individuals. Thus, we argue that the regulation of affect and the restoration of balanced mentalizing are necessary within the therapeutic relationship prior to the activation of the relational process. Without a certain degree of regulated, balanced mentalizing, the patient is unable to take into consideration the therapist's mind. As a result, the patient is likely to project their own narrative, thereby precluding the expereince of "we-ness". We have elsewhere described the clinical process which can overcome this state of petrification thus: (1) the patient's *imagined* sense of self (their personal narrative) (2) is imagined by the therapist in full establishing a prospect for the we-mode in which the complexity of the patient's inner experience, that is not only their dominant narrative but also the sub-dominant (pre-conscious narrative) experience associated with it, is recognised and communicated so that (3) this image is perceived by the patient creating the potential we-mode when (4) this communicated and perceived image is compared with the patient's personal narrative and (5) in case of a match the shared or co-representation has been created triggering the we-mode which in turn (6) removes the I-mode's self-protection from from extrinsic change and (7) the channel for more rapid, efficient knowledge transfer is opened (Fonagy et al., 2021). "Putting it at its simplest, if I feel that I am understood, I will be disposed to learn from the person who understood me. This will include learning about myself but also about others and about the entire world I live in"

(Wurman et al., 2021, p. 36). A brief vignette of a therapist's response may convey the potential that empathic validation has for generating we-mode experiences: a patient could present herself as over-adapted, friendly, and always accommodating (her dominant narrative). Only if the therapist can detect, mentalize and represent for her how humiliating this may be for her, because she has to constantly negate her own needs, and what enormous anger she feels about this (subdominant narrative) and how disappointing it may additionally be that this dynamic is now also repeated in therapy (empathic validation), an experience of epistemic match in we-mode is established.

Developmental origins - epistemic (mis)trust, hypervigilance and maltreatment

As human beings we are evolutionarally determined to seek cultural knowledge. Trust in knowledge (which we call epistemic trust, ET, following Sperber et al., 2012) is at the heart of what it means to be human. Once trust has been established, we are open to internalize and assimilate the knowledge offered by the trusted or deferentially regarded source in order to optimize our future interactions with our social world [Sperber, 2010; Recanati, 1997; Wilson, 2012]. Thus in order to ensure effective cultural knowledge-transfer via teaching (including the various forms this can take within a broader notion of "teaching" in psychotherapy), humans needed to evolve a reliable way of distinguishing trustworthy sources of knowledge. This gives the instructor a special deferential status and allows them access to modify the learner's ideas and beliefs, which are normally safeguarded by our naturally conservative vigilant stance to novelty. Feeling recognised as an agent serves as an ostensive cue and thus opens someone up for learning– ideas postulated a natural pedgagogy (Csibra & Gergely, 2009 and 2011).

The theory of natural pedagogy (Csibra & Gergely, 2009) underlines the highly interpersonal nature of the process by means of which epistemic trust is generated in infancy. Overcoming adaptive epistemic vigilance and appropriately identifying whom to trust can be regarded an imperative for adaptive cooperative functioning. As the anthropologist Thomas Weisner suggests, teaching and learning such discrimination is a central task of parenting and early development:

> The question that is important for many, if not most, parents and communities is not, "Is [this individual] child 'securely attached?'" but rather, "How can I ensure that my child knows whom to trust and how to share appropriate social connections to others? How can I be sure my child is with others and situations where he or she will be safe." Parents are concerned that the child learns culturally appropriate social behaviours that display proper social and emotional comportment and also show trust in appropriate other people. (Weisner, 2014, p.263)

We have suggested that one of the advantages that derives from a secure attachment, is that it creates a template for the child that helps them recognise who they can trust (Fonagy et al., 2017b). Individuals who are securely attached appear to be able to take an appropriately agentive epistemic stance: they are better able to resist misinformation, but are also able to display adaptive epistemic trust when it appears warranted (Corriveau et al., 2009) (Campbell et al., 2021). Those who have, for example, been exposed to higher levels of adversity in childhood appear to show higher levels of both epistemic mistrust – showing excessive levels of resistance or impermeability to social learning (e.g. McGuire & Jackson, 2018; Cowell, Cicchetti, Rogosch, & Toth, 2015; Fry et al., 2018; Jay & Mc Grath-Lone, 2019) – and/or epistemic credulity – characterized by excessive openness and a lack of discrimination and agency in relation to new information. (Campbell et al., 2021). Because an indivdual's epistemic stance shapes their capacity to respond to and adapt to interpersonal communications, it appears to be highly implicated in the quality of social functioning. We also suggest that it may be highly pertinent to any psychotherapy outcome; and we have

suggested that the treatment resistance associated with some chronic and severe forms of pathology such as BPD arises out of a failure of the treatment to overcome the patient's difficulties in relation to social communication (Fonagy, Campbell, & Luyten, 2018) and the resulting difficulties in identity coherence, relatedness and affect regulation that characterize BPD. Whilst we consider the facilitation of epistemic trust a mechanism to affect common factors across all treatments, there is some indication that patients with BPD benefit more from manualized treatments (Cristea et al., 2017) such as MBT and others that more explicitly address the quality of the therapeutic relationship, including potential ruptures. The implication of this thinking is that the epistemic stance is a trait-like feature of individual functioning, analogous to the internal working model concept in attachment theory but also prone to state-driven changes. We have increasingly emphasized the role of the wider social environment in supporting the emergence of epistemic trust, and the value of creating mentalizing social systems around children and young people (Campbell & Allison, Forthcoming; Fonagy et al., 2021; Talia et al., 2021). Fisher, Zilcha-Mano & Fonagy (2022) suggest: "A range of interpersonal contexts can offer differential exposure to reliable sources of information in which social learning can occur. Teachers, peers, social media, and psychotherapy may change people's general expectations of trustworthiness". MBT aims to support patients with BPD to be able to make use of reliable relational resources (again).

If the caregivers around the infant are not reliably responsive, not benign and/or not able to recognise what is meaningful and relevant to the infant's self, this can undermine the development of epistemic trust leading to a variety of epistemic disruptions. The first, and arguably the most damaging way in which this disruption to social learning can happen is via early adversity, severe neglect or maltreatment, perhaps the most generally agreed transdiagnostic cause of mental disorder (Coughlan et al., 2022) - in interplay with

constitutional and neurobiological factors (e.g. Fonagy & Luyten, 2009). In such circumstances, a child might "switch off" their instinct for social learning, for entirely understandable reasons: a negligent, hostile or abusive caregiver cannot be trusted to be a reliable source of information who is invested in helping the child get the most from their environment generally but particularly so when learning about the nature of mental states (in self and other, i.e. their intentions and feelings) are concerned. In this context this "switch off" may be seen as a potentially effective adaptation to an inherently untrustworthy social environment. In such circumstances it may be a better strategy to regard what others try to convey to us as irrelevant, suspect, or misleading. However, as many of those writing about the impact of trauma point out, a strategy that may have survival value in the short term, may turn out to generate significant difficulties later [McCrory, 2015; Huang et al., 2020].

The problem with such a strategy is that in the longer-term, the child is unable to reap the full benefits of social learning. Child maltreatment has been consistently linked with poor academic and school-related problems [Jonson-Reid, 2004][Romano, 2015][Trout, 2008].. Of course, other qualities of the social environment that normally support psychological resilience in the face of challenges [Brown, 1989] – being able to seek and receive help, having a social network, being open to change – are also potentially lost.

Depending on the severity of the adverse environment, the developing individual will be more prone to persistent and pervasive epistemic *mis*trust. This can take the form of epistemic hypervigilance, a state in which epistemic trust is rare and hard to generate. In other cases this tendency may result in prolonged moments of "epistemic freeze" – a state in which epistemic trust is impossible to generate and all forms of social learning become inaccessible regardless of their source. Epistemic disruption (mistrust, credulity or a combination of both) then prohibits the generation of moments of experience in we-mode (e.g. an experience of an

understanding and validating other as in the above example of a therapist) as the other is not imbued with the capacity to consider and represent the communicator's (patient's) mind with benevolent intention.

There is accumulating evidence that suggests the capacity for orienting effectively to the mental state of others is dramatically reduced by adversity and early life stress, particularly neglect or abuse, and in combination with other disadvantages that can be best described from an intersectionality perspective. Whilst we consider the developmental impact of such experiences to affect most types of mental disorder, there is sufficient evidence they are key factors assoiated with or contributing to the development of BPD. In a recent study, we explored the mediating role of personality functioning between childhood adversity and symptoms of PTSD and complex PTSD in a representative German sample of adults, some of whom had reached clinical diagnosis levels. We found that including epistemic disruption as a predictor added to this mediating role (Kampling et al., 2022). We could demonstrate that the clinical profiles of patients with features of BPD with and without PTSD and cPTSD (complex PTSD) are highly correlated with compromised mentalizing, dissociation scores, and childhood trauma. Further, mentalizing appears to mediate the relationship between BPD symptoms and post-traumatic stress symptoms (PCL-5) and cPTSD diagnoses (Bateman et al., under review).

In providing a model for vulnerability to mental disorder we wish to establish a link between openness to social learning and the feeling of being acknowledged or recognised as an agent. A sense of personal agency has been a focus for us, and other researchers, as offering an important link between the concept of a singular general vulnerability to mental disorder and and various sources of vulnerability, such as genetics, personal history, emotional

dysregulation, and executive dysfunction [Ryan, 2016][Roth, 2019]. For individuals in a state of epistemic disruption, cultural knowledge may become inaccessible, be ignored, misinterpreted, or seen as hostile, and new information and experience may not be able to replace existing knowledge structures.

In summary, we postulate that epistemic trust or its disruption, is the key factor - a common final pathway - that accounts for the liability to developing borderline personality disorder and other psychopathologies. This is a biopsychosocial framework that suggests that these result from a combination of constitutional factors and early caregiving, which can "create brain and personality structures that are, in turn, shaped by and will further shape later experiences" (Knapen et al., 2020). A combination of unfavorable constitutional factors and early aversive experiences may render an individual prone to developing psychopathology through the effects that epistemic disruption has on social functioning and the propensity for help-seeking. This transdiagnostic notion could explain downstream effects as recently highlighted by hierarchical empirical models, which suggest that psychopathology is best captured by a general psychopathology factor (e.g. Caspi et al., 2014; Wendt et al., 2022) and some broad personality-related spectra (e.g. Kotov et al., 2017).

Three communications systems and the role of epistemic trust and ostension: implications for the therapeutic alliance

In an attempt to conceptualise what actually happens in effective psychotherapeutic help for those suffering from BPD (and other more severe psychopathology), we have described three communication systems, three interlocking and over-lapping processes at work across interventions . We suggest that these systems do not just apply to mentalization-based treatment; rather we suggest that any form of meaningful help tends to involve the

communication, internalization and reapplication of new forms of "learning" about oneself and oneself in relation to other people. As we will demonstrate later, MBT (as well as some other manualized treatments) explicitely addresses how the epistemic stance and learning can be improved via the three systems described above and thus makes explicit what are perhaps more implicit mechanism of change in other therapies for BPD.

Communication System 1: The teaching and learning of content

All different therapeutic schools activate this system when the therapist conveys to the patient a model for understanding the mind that feels relevant to the patient and makes them feel recognized and understood. The experience of being recognized as an independent agent reduces the patient's epistemic vigilance and begins to prime the patient to social learning. Crucial to this system is the therapist's capacity to mentalize the patient as it requires the therapist to apply and communicate their therapeutic model in such a way that it is experienced as meaningful by the patient, creating an epistemic match – with specific adaptation of ostension required for remote/online interventions (Wurman et al., 2021; Fisher et al., 2021: Aisbitt, Nolte & Fonagy; 2022).

Communication System 2: The re-emergence of robust mentalizing

When the patient is once again open to social communication in contexts that had previously been blighted by epistemic hypervigilance, they show increased interest in the therapist's mind and use of thoughts and feelings, which stimulates and strengthens the patient's capacity for mentalizing – 'how does this person see me as they do'? The not-knowing stance of the clinician with assiduous focus on the patient experience kick starts this. The emergence of mentalizing in the patient develops a virtuous circle in which curiosity about mental states

and social learning through greater epistemic openness support each other within the therapeutic relationship and allow for the integration of different levels of experience and affective states. The patient and therapist join in the we-mode, with the patient's state of mind and response to what is happening to them the subject of their joint attention.

Communication System 3: The re-emergence of social learning

Applying social learning in the wider environment. Being mentalized by another person frees the patient from their state of temporary or chronic social isolation, and (re-)activates the capacity to learn: this frees the person to grow in the context of relationships outside therapy. This view thus implies that it is not just the content, techniques or insight acquired through treatment that are key to its success, but perhaps primarily, it is the patient's capacity for social learning and thinking about mental states that improve their functioning as the patient gradually becomes able to "use" their environment in a different way: Being less hypervigilant to the social environment may also enable positive events in a person's environment to have implications for the person's experience of self and others A further implication is, of course, that psychological interventions may need to also intervene at the level of the social environment when needed or appropriate. The bringing in of this emphasis on the patient's wider social ecology as a key factor in making change possible acknowledges the limitation of what can be achieved in the consulting room alone, and opens up a clear clinical and theoretical challenge for thinking about psychopathology – that social systems matter and the clinician cannot achieve all alone. [1]

[1] As one participant in a mentalizing group at a UK prison so articulately conveyed to one of the authors (PF): 'This group and mentalizing is really good and I like it a lot. But when I get back on the wing, it's ****ing useless'

Although we describe these processes numerically, we should not expect them to unfold along a neatly linear trajectory. There are inevitable disruptions, ruptures and work on repair across treatment, which may involve the activation of different communication systems, or their overlapping activation. For example, a patient may, half-way through treatment, enter a session feeling distressed and angry about a difficult exchange they recently had with a family member. In a state of psychic equivalence (i.e. when what is thought or felt is experienced as completely real and true), the patient may regard anything the therapist says as meaningless, useless or provocative: communication breaks down and any attempt to expect the patient to take in or apply social learning is likely to fail. Instead, communication 1 needs to be reactivated, in which the patient's state of mind is understood and recognized, and the capacity for thinking and learning gently reinstated.

The MBT approach – general features related to epistemic trust

The attitude and stance of the MBT clinician is one of authentic interest in mental states and how they play out in managing emotions and impulses and in understanding social and interpersonal interactions. Employing this not-knowing stance, the clinician follows a hierarchy of interventions in both individual and group therapy with the overall aim to stop non-mentalizing and to engender a process that brings mentalizing back. In addition the not knowing stance requires the attitude of the clinician to be actively focused on seeing things from the patient's perspective, that is elaborating their reality rather than imposing a significantly different one. This ensures the initial MBT processes of psychoeducation and formulation are personalised, worked on jointly, and are tailored to the individual patient, their history and current life. Also, during this phase of MBT, psychoeducation is provided not

as a lesson to be learned but as a frame acting as a scaffold for the patient to attach their own experiences to, to make better sense of themselves for future consideration.

In order to engender effective mentalizing, openness to social learning needs to be established or ET restored. Recent conceptualiastions have spelled out how ET plays a key role in this spanning across and being interlinked with all three communication systems. Fisher and colleages have introduced a triadic model that conceptualises ET as a mechanism of change (both its trait and state-like characteristics, thus viewing it as interpersonally and dynamically embedded). They suggest three aspects of ET that capture its expression in psychotherapy: sharing, the we-mode, and learning (Fisher et al., 2020): "When individuals put their overall experience into perspective and calibrate their mind to those of others (sharing) while establishing a mutual discourse for the processing of ideas (we-mode), new information pertinent to social adaptation (learning) can be acquired. However, these three postulated components can manifest differently within psychotherapy, depending on the patient's trait-like ET characteristics" (Fisher, Zilcha-Mano & Fonagy, 2022, p. 32). Again, we agree with Fisher that this triad is relevant for the delivery of all forms of treatment; we here focus on how MBT explicitly fosters ET in order to effect greater use of the three aspects.

One potential mechanism to bring about state-like changes in ET is associated with therapists' recognition and articulation of their patients' more fine- grained self-experiences (Fisher et al., 2022; Fonagy et al., 2019). "Feeling accurately reflected in the therapist's mind may pave the way to social learning and the restoration of ET in the patient" (Fisher et al., 2022; see also Sharp et al., 2020). The patient's experience of having agentive selfhood recognized and in particular, of achieving joint attention between therapist and patient on the subject of the patient's mental state involves working in the we-mode. Generating the we-mode in this way activates cooperative social learning, enabling the patient to become more open to new

forms of learning from their therapist. This includes the therapist presenting his own experience and responsiveness in an authentic way. Thus, the patient can take in different ways of thinking about themselves and about how they relate to and impact others that in effect constitute (in attachment terms) an adjustment of their internal working model. The unfolding of this process is dependent on the therapist communicating their recognition of the patient's mental state in a regulated and validating way. It must be borne in mind that for individuals who have been avoiding engaging in the we-mode, often out of an adaptive aversion to the experience of joining with other minds, thinking together should not be assumed to be a non-arousing or non-threatning experience for individuals who are vulnerable to being overwhelmed, or swallowed up by other people's minds (a process associated with epistemic credulity) or who rigidly avoid engagement with others (epistemic mistrust), or indeed who veer between these two positions in disorganized fashion (perhaps particularly associated with borderline functioning) or to whom the experience of being with a benign thinking and mentalizing mind is both desired but deeply disturbing at the same time. As a further complication, there can also be epistemic exhaustion of the interpersonal system that 'palpates' the social environment in such a way that credulity manifests in one of the following ways: i) self-representation is too diffuse whereby all things feel as though they fit (indiscriminate trust), ii) representations of self are distorted (creation of an illusory fit when none exists ("I am accurately seen as bad person") and iii) inaccurate view of the self (defensively generated – very robust: "I am being unfairly treated") whereby the perception of the personal narrative is calculated by a manipulative instructor to be experienced as a match by the learner (manipulation of the match).

MBT treatment components and specific interventions with regards to their relevance to restoring epistemic trust and reducing disruption

Inspired by Kamphuis and Finn's work on the ET-facilitating effects of Therapeutic Assessment in personality disorders (2019) in what follows (table 1), we provide an in-depth proposal that explores and clarifies all aspects of MBT treatment delivery with regards to ET and epistemic disruption including sharing, we-mode exploration and social learning. In our view, this helps with rupture and repair cycles and, although empirical support is still lacking, may reduce treatment drop out. We propose that it, furthermore, creates a resilient alliance and potential for extra-therapeutic change.

While it is beyond the scope of this paper to define in detail MBT interventions (e.g. "contrary moves" or "stop, explore and rewind") that have been introduced and illustrated elsewhere (e.g. Bateman & Fonagy, 2006; Bateman & Fonagy, 2020) it is important to emphasize that some of the aspects or MBT "ingredients" listed below under individual or group sessions also apply to the respective other setting: the flexible integration of appropriate ET and epistemic vigilance (balancing of the epistemic stance), exploration of explicit and implicit patient beliefs etc., do not only take place in either individual or group sessions but extend across all treatment components.

Treatment component	Intervention/What therapist does	Purpose	relevance to ET and epistemic disruption
Assessment	Comprehensive exploration of personal history including trauma history and developmental vulnerabilities, past and present relationships, positive achievements, previous treatment, impulsivity and risk, symptom profile, sextratherapeutic support systems.	To aid formulation linking 'symptoms and behaviours' to changes in mental states	Creating mind as the subject of focus – Identifying mentalizing profile the importance of a stable 'I' and opening up questioning of an exclusive overly inclusive or 'me-mode' frame of reference and experience.

			considering over and under use of mentalizing dimensions and ineffective mentalizing modes. Patients also "learn" that their symptoms/functioning/reasons for referral are related to changes in mental states and that it becomes helpful to focus on these to improve (communication system 1).
MBT-I	Psychoeducation about treatment, mentalizing, emotions, interpersonal processes and attachment strategies, and psychopathology. Exercises to tailor the information to each patient; supports use of topics to patient life; Identifies practise points for each patient.	Creates facilitating atmosphere for hearing and listening; Examples all topics using own everyday experience; Normalise mentalizing failures; Focus explicitly on engaging patients in treatment and keeping them motivated for change.	Framework which scaffolds the forthcoming work. Creates a shared model in communication system 1 to reduce EV – there is a way of understanding the non-understandable. Reframing so two or more perspectives can be taken. Culture of curiosity is created. Experiences as something that can be shared, reduction of sense of alienation and loneliness.
Individual Sessions			
Formulation	Initial draft written or drawn as picture/flow diagram/circle of problem by therapist with the patient present (later all from team involved in care can contribute); non-jargon, experience-near, shared with patient throughout treatment, updated as treatment progresses. Reviewed every 3 months. Patient in group 'presents' formulation to group when starting and all current patients review their formulation with the new patient.	-provides a shared construction of the interplay between developmental history, sensitivity to specific types of triggers and interpersonal functioning, highlights mental states and processes and the resulting individual symptoms -Engenders sense that personality is not fixed but can change and develop reflected upon together and revised after patient's input. Provides focus for treatment.	Patients sees himself and his relationships from the outside, through mind of the therapist/team and begins to feel understood from within. Recognises that experiential 'self' is a construction from I-mode and Me-mode functioning. Opportunity to calibrate understanding by co-creating aspects hitherto not fully or accurately grasped by team (or patient) – creates agency. Enlisting clients' curiosity and sense of agency via co-constructing understanding, especially through peer interaction in group. Transparency and collaborative communication (communication system 2).
Mentalising as key to understanding mental and social function/diagnosis	Consider borderline and other aspects of personality function in discussion. Discuss 'diagnosis' if appropriate. Demonstrate that problems as described by patient are understood and change is possible.	Increase coherence of narratives; Specific problems related to imbalances/problems with contextual use of mentalizing. connecting different pieces to create understanding and foster hope that this is a well-known condition that can be treated.	Potential recognition of not being alone. There are others out there like me. Emphasis on diagnosis as not fixed but something that can be explored and tracked over time in we-mode (communication system 1)..

Co-defining treatment goals	Formulation leads to clinician and patient jointly agreeing and specifying short term and long term goals.	Creation of joint task and focus on motivation	Necessary for development of we-mode – we both look at the same goal and try to consider our individual perspectives together (communication systems 1 and 2)
Mentalising Process			(communication system 2).
	Active stance in exploring Authenticity Equality Collaborative	Not Knowing Stance and Perspective-Taking are engendered	Seeing it from patient's perspective and not imposing other perspective -> epistemic matached and foundation for shared intentional structures for opening learning systems
	Empathic Validation - affect and effect; Apply to past and present experience. Basic emotion and its effects in present moment. Example: "I'm just imagining how difficult it is to be experiencing me as looking upon you in a judgemental, critical way, or even disappointed with you, when you've been putting so much effort into resisting urges to hurt yourself. And it seems to leave you in such a hopeless place where not only you feel all the progress you've made is lost, but you've also lost me as someone in your mind who might be on your side in supporting you with your efforts, and so you are left alone with all this..."	Conveying Emotional Attunement; Being with - 'two-getherness'.	Joining with but not being in the same state as the other. Other mind is given credibility. Creates opening of ET - triggers patient experience of being mentalized in their current emotion - mentalized affectivity; identifies the effects the emotion is having. Affect experienced as valid and shareable so that a sense of security and recognition is created for further exploration potential to be found in the therapist's mind -> epistemic match. Appraisal of historical need of certain non-mentalizing defenses: Kamphuis and Finn pay particular attention to shame as an affect and PD-related symptoms as defenses against shame (or humiliation). (this, at times, may include marking bodily sensations or dissociative states)
	Marking	Defines different perspectives Increase agency and self-definition	Separation of representations of self and other mind states on which to begin generation of we-mode
	Contrary Moves	Creates recognition of over or under activation use of mentalizing dimensions; increases breadth and complexity of reflection	A necessary pre-cursor to we-mode potential. Creates some flexibility in mentalizing and so frees the mind from being stuck in a single perspective
	Stop, explore and rewind	Prevents collapse into ineffective mentalizing modes	Attenuates the tendency for mind to continue without reflection
	Titration of affective arousal	Therapist monitors arousal (own and patient's) and intervenes when too high/too low	Required if mentalizing is to be kick-started, subsequent micro-slicing of reported events support nascent naming and representation of affect that has hitherto only been experienced in pre-mentalizing modes.
	Attentiveness to Therapist contributions	Therapist becomes a person to the patient. Mind of therapist	Generation of me-mode – patient considers themselves through the mind of therapist

		is of interest to learn about self	
	Attentiveness to Misunderstandings	Separation of mind states – is this me or you? And to what effect? Therapist as "real" and authentic Indicate early in treatment as important moments to reflect on.	Rupture and repair and normative interpersonal cycles with an experience that those can be survived, an experience of "betrayal" can be mentalized
	Role of ostension from therapist (leaning forward, raising eyebrow etc.)	Increase readiness for communication;	Authentic and appropriate use of ostensive cues creates attentional state of attuned listening
	Identifying and Stopping Non-Mentalising Modes	Creating potential to (re)-engender mentalizing	The 'How' - Regaining mentalizing is required before any possibility of we-mode developing unless other mind is also non-mentalizing mode
	Challenges	Disrupt ineffective mentalizing particularly hypermentalizing; Establish clinician as a person of relevance - 'Get back in the room'	Collision of mind states - to address complete loss of capacity for we-mode and also lack of self-other (I-mode:Me-mode) representation
	MBT Loop	Name and Note ineffective mentalizing; disrupt ineffective mentalizing and rebuild mentalizing; create a new perspective	Not joining with low mentalizing and avoiding iatrogenic process in which we-mode 'fake news'. We-mode can masquerade as joint false beliefs and even be 'I-Mode' looking like we-mode (narcissism) Looping avoids these harmful processes and allows re-instatement of higher levels of mentalizing.
	Exploring when and with whom trust and mistrust are appropriate	Cautioning against Credulity	The individual may appear to be highly open and receptive to content and ideas but a lack of clarity in the patient's mind about their own mental states make it hard for these ideas to gain meaningful traction. High epistemic credulity might be associated with the pretend mode of functioning. Using gentle, compassionate challenge and linking of affect to reality are important tools here.
	Mentalising Affective Narrative	Affective trajectory exploration – there and then, here and now, re-appraisal from current affective state; identification of patterns of behaviour and action linked to mentalizing process.	The 'what' – reflection on the content of mind with integration of memory past and present and balanced mentalizing process e.g. cognitive and affective. Linking affect focus and affect clarification with sense of being understood via epistemic matches

		Affect focus of sub-dominant themes	
	Identifying and Reinforcing Positive Mentalizing	Alight on positive affect and experience; effects that has on experiential self	
	Relational Mentalising	Identify common attachment strategies with patient; go beyond self-other interactional components; Create an 'us' to be looked at through sharing responsibility of interaction; generalise to patient life; consider current relationships from new perspective	Attempt to intensify we-mode reflection within relationship as well as 'on' the relationship (distinguish in and on here for full we-mode – together and representing something that is neither of us but both of us together. Taking meaning from interaction that is useful to 'me'
		Understanding Misunderstandings	Being misunderstood as opportunity, role of ruptures in providing understanding if curiosity in them can be maintained, acknowledgment of therapist's involvement in how misunderstanding came about – opportunity for patient to learn from modelled curiosity – sense of shared intentionality and of being understood when both parties arrive at joint understanding of "what went wrong"
Crisis plan	Jointly creating a document about what helped and did not in crisis in the past and how to respond in and prevent future crises		Patient and clinician develop a future self in crisis and both take perspectives on how to manage that crisis should it occur. Also write or draw out how to manage mental pain when it starts. This is the triangulation aspect of we-mode
Crisis intervention	Implement agreement about response; focus on crisis and managing it with pre-agreed plan. Mentalizing of pre-cursors of crisis and context. Link to formulation and revisit crisis plan for future. Rehearse mentalizing over next 24 hours. Implement safety procedures	Create mentalizing in anxiety through calibration with other.	Working on mentalizing self and calibration of emotional experience through another person. Immediate triangulation in we-mode with self as object of scrutiny and safety/protection.
Group Sessions			Communication systems 1 and 2 with sharing of experiences in system 3.
		Generation of group values	Practice of we-mode with the group itself as an object of exploration. Group is more than the sum of the individuals Organise ourselves according to values we all share

			Develop a culture that is recognisable and transferable to new participants and contributed to by all
		Presentation of personal formulation to group members	Presenting story of oneself and some problems and strengths to others and having others present themselves increases prosocial interaction and forms an initial platform for we-mode – 'getting to know me and getting to know you'.
		Thinking collaboratively about form of relationships.	group psychotherapy as an environment in which individuals can practice, participate in, and contribute to a mentalizing social environment.
		Reduction of epistemic isolation	Shared understanding of patients' core narratives and then help "each client to get the feeling of being an 'author' of his or her new story, a story which is more compassionate, useful, emotionally viable, and coherent than the previous one" (Aschieri, Finn, & Bevilacqua, 2010, p. 257) – other can feel similar to me. Restoration of social learning beyond the therapeutic dyad.
		Anchoring of experience and social feedback via link between group and individual	thinking collaboratively and across context.
	Triangulation	Prevent collapse into low mentalizing modes and restricted we-mode interaction	We-mode in group has to have the whole group attending to the problems – ' we all are in this together and we need to see to it'.
	Parking	Seeks to increase patient capacity for attentional control to supress a dominant desire in the service of joining with others to explore a theme that is sub-dominant for the individual	Attenuates the tendency for mind to continue without reflection When picked up later: conveys interest in patient's experience and potential for we-mode instead of high-arousal narrative trumping reflection
	Siding with	Supports an individual following sudden collapse of a sense of belonging or rupture in relationship	Repair is possible. Change and learning occurs in context of mending relational rupture; doing so with a trusted source gives experience of vicarious mentalizing as a way of increasing mentalizing of others.
Treatment review with patient	Often with individual and group therapists jointly present. Recently: with extra clinician present who is not directly involved with treatment delivery.	Where is patient at in treatment, what works, what doesn't? How can the team	Benign curiosity that also opens discourse about therapist/team contribution: potential for exploting progress and areas that are not working in we-mode.

		help better? Revision of treatment goals etc. Modeling curiosity of team in relation to extra clinician's perspective. Focus explicitly on engaging patients in treatment and keeping them motivated for change whilst also identifying aspects of treatment "failure". Can also introduce focus on coming end of treatment.	Review in its totality is potentially a we-mode process (communication system 2).
Across all therapeutic interactions		Attentiveness to the patient's interpersonal experiences in their lived reality outside treatment	Does ET gained or restored in therapy translate into applying social learning in patients community (romantic relationships, those with children and parents, with friends but also the wider social network including at work and with institutions and authorities) Establishment of shared important that "practicing out there" is important (as well as exploring what helps and prohibits it) (communication system 3).
Outcome data collection	Combination of symptom-based outcome and process instruments with suggested shift to more patient goal-based or co-created mesures.	Creation of meaningful monitoring of patient experience – 'this is for me', not for the clinician or research.	Additional self-feedback in relation to treatment goals and success. If instruments used that capture patient perspective, more fine-grained sense of areas of change or stuckness and deterioration can be evoked.

Table 1 depicts the progression of both general trust-related dynamics and those specific to epistemic trust organised according to the domains of MBT treatments. These dynamics reflect common clinical observations in treatment for all patients. However, there are individual variations to these patterns rooted in the individual's maltreatment experiences and their predominant attachment strategy (hypo- or hyperactivating, or disorganized).

Implication for MBT treatment

Via the generation of a mentalizing and validating processes in MBT, the three aspects of Fisher's tripartite model – sharing, we-mode experiences, and learning – can begin to mutally reinforce each other. The co-creation by patient and therapist of 'we-mode' moments (solidified through successfully negotioaded – "survived"- rupture and repair cycles) enables

learning about the self, others, and the self-with-others and subsequently an increased viability of sharing. As these processes do not evolve in a linear fashion, a continued not knowing stance can ensure that therapist and patient maintain a shared understanding of interpersonal dynamics.

Importantly, mutual sharing of mental states enshrined in the not-knowing stance and in counter-relational mentalizing facilitate we-mode: “Communicating meaningful experiences can create a sense of companionship or allyship, which can mitigate a person’s feelings of isolation” (Fisher et al., 2022). In MBT, patients are invited to share – often for the first time – with the hope that these can be accurately represented by a therapist. This is common to most therapies but MBT emphasizes the importance of explicit use of mentalizing the counter-relationship. In this process the clinician accepts that their mental states are part of the joint process and shares their experience in dynamic relationship to the patient’s mental states. The clinician firstly validates the experience of the patient and secondly both states of mind become available for scrutiny with consideration of how the interaction has been generated. With time, the effects of what is being shared on both the patient, the therapist themselves and also others can be examined (Fonagy & Campbell, 2017). The therapist’s role will be, in particular when working with those who are less able to “tune into their social environment and perceive the therapist’s intentions as beneficial or at least benign, which prompts them to believe that sharing their conflicts is worthwhile” (Fisher et al., 2022), to help mentalize the patient’s experiences and expectations (sometimes openly shared, something kept covertly) to be faced with an malevolent environment. Feeling misunderstood or being exploited, being subject to criticism, and being made to feel guilt or shame, according to Fisher and colleagues, may lead to patients abandoning the possibility of sharing altogether. The philosopher Patrizio conceives of trust as “constitutively involving

vulnerability to betrayal. When trusting rather than merely relying, one subjects themselves to the possibility that one will be betrayed, rather than merely let down" (Patrizio, 2022) – a particular kind or experience of relational rupture that requires acknowledgment. Validating such expectations is critical in earlier stages of MBT (see table 2) and the potentially ensuing alliance ruptures need to be addressed and understood with an emphasis not only on the patient's experience but also on the therapist's or team's contributions to ruptures in order to reduce epistemic vigilance (Allen, 2022). One but not the only way through which a disposition to epistemic mistrust is manifested specifically in younger patients, is through the patient's tendency to hypermentalize (Sharp et al., 2011). This manifestation of imbalanced mentalizing denotes the "overattribution of mental states far beyond what there is evidence for (…), accordingly, hypermentalizing is described as making excessively convoluted inferences on the basis of others' social cues" (McLaren et al., 2022). Hypermentalizing may in this context be reformulated as an attempt of the patient to detect trustworthy others who can be learnt from in a world characterized by insecurity, isolation and further intersectional burden. Hypermentalizing may make the patient extra sensitive to a lack of authenticity as a core characteristic of untrustworthiness in others, as if 'their authenticity sensors are on full blast to protect themselves from betrayal and deception' by others (de Wit, de Jong, & Mulder, 2019). Experienced inauthenticity may interfere with the therapeutic alliance, enhancing epistemic mistrust in treatment, and resulting in disengagement from treatment. Interestingly, this tendency to be mistrustful towards many others, may be complemented by a tendency to be epistemically naïve towards and thus inappropriately credulous towards some selected others. Exploration of trust-related interpersonal beliefs in MBT includes the mentalizing of when, with whom and why sharing was easier and appropriate and when and why it became impossible. Following the elaboration of ET as a core process for change this

process of exploration has become a structured process in MBT groups - the MBT clinician and all participants work together to agree values, some suggested by the clinician and some by the participants, to be followed by the group; current group members introduce themselves to new participants by discussing their ability to share with others in the group and in their outside lives, outlining the benefits and disappointments experienced; in MBT-trauma (Bateman et al., under review), once an individual's management of anxiety and dissociation is stabilised, participants share some of their trauma experiences with the group and group members are tasked with helping them do this without making any judgements. This is followed by them talking about how sharing has changed their perception of the person presenting and how it has changed how they think about themselves; in the final phase the group look to the future and consider how they use what they have learned in their lives. These processes are deemed to overcome epistemic isolation, i.e. the relational experience that no other processing system or only a dysfunctional one underpinned by non-mentalizing is available to process trauma.

With progression of treatment sufficient attention should be given towards the process of sharing in extra-therapeutic relationships where "interpersonal relationships can be reconfigured and scaffold important adjustments in self-perception" (Fisher et al., 2022). It is here that there are further implications for MBT. The generation of ET within therapy itself needs to be seen as a stepping stone and greater emphasis placed on generalizing a changed attitude and new perspectives to the patient's current and future life. Looking to the future needs to become part of the early stage of therapy, so often necessarily taken over by stabilizing risk behaviours and crisis management, and not left by default to the final phase. When sharing experiences becomes possible, " a co-creation of mutual communication is enabled" (Fisher et al., 2022) which creates a joint looking at the experiences (including those

unfolding within the therapeutic relationship) in the we-mode. The ability to do so may both depend on the "dispositional (mis)trust as resulted from the attachment history of the patient, and on the capacity of the therapist to overcome this dispositional mistrust" (Knapen et al., 2020) and trigger or restore (epistemic) trust within this specific relational context. Exploring and understanding facets of the patient's epistemic stance can help identify aspects of the interpersonal dysfunction that matter to the issue of affective states and social learning, which is a "pre-requisite for the transfer of skills, advice, insight, and other modes of information within a therapeutic encounter" (Knapen et al., 2020) including reflective and regulating capacites. Achieving this in the we-mode requires acknowledging the subjective reality of patient and therapist with joint attention being established, at least temporarily, on the subject matter of the emotional experience in the here and now of the therapeutic relationship. In line with Fisher and colleagues we propose that it is this "sense-making process that gives conscious meaning to inner narratives and may allow for an internal dia-logue to take place in the patient's mind and with others" (2022).

Interventions in MBT create a potential for positive change to epistemic mistrust so that permeability to new informants and information can be increased which, in turn, is thought to aid social adaptation and the calibration of minds. But the questions remain of which aspects might be focused on more and which less, what new interventions are there that might increase ET through improving mentalizing, and will making such changes improve outcomes.

Implications for ET-informed research

Knapen and colleagues argue that ET/EM holds promise for being a psycho-marker to identify "patients 'at risk' for treatment failure as it provides us with an accessible feature of the

person that may help to determine his or her eligibility for interpersonally driven help" (Knapen et al., 2021).

There is a growing interest in understanding the contribution of ET to successful treatment outcomes and patient experience (e.g., Bo et al., 2017; Folmo et al., 2019; Li et al., 2022; Sprecher et al., 2022). Yet there is no research to date on whether symptom changes within MBT (or any indeed other treatment) are supplemented with stable changes in ET. If an increase in trust as engendered by MBT is a mechanism of change, then MBT and other treatments should be evaluated in terms of change in ET processes including the reduction of epistemic disruption as demonstrated by Riedl et al. for psychosomatic rehabilitation (2023). Empirical assessments (e.g. the Epistemic Trust, Mistrust and Credulity Questionnaire, ETMCQ; Campbell et al., 2021 or experimental procedures: Schröder-Pfeifer et al., 2022) could provide predictive validity of individual differences in pre-treatment ET and how ET unfolds in therapy for therapy outcome, process and treatment completion. Focusing on individual differences can help identify how psychotherapy works for certain subpopulations (Zilcha-Mano, 2021) depending on their trait-like ET and the shaping of the alliance and rupture and repair phenomena. Future research should thus focus on identifying the specific interventions by the therapist and their interplay with ET-related patient factors (e.g. during MBT-I or in group vs individual therapy) that may lead to sustained changes in ET as well as on what stage and context of therapy they should occur (see Folmo et al., 2022). We thus envisage a number of desiderata and related challenges:

- Design of instruments to assess ET that may also show incremental validity compared to existing measures assessing traits like suspiciousness, negative affectivity or abilities

such as mentalizing or personality functioning. This requires establishing the specificity of the construct and its measurement as well as easy administration.

- Similarly, for process research (e.g. based upon video recodings or transcripts of sessions as in MBT), development of an assessment tool that operationalizes quality and appropriateness of ostention and ET-facilitating interventions
- ET may inform treatment assignment. If the hypothesis is correct that epistemically mistrustful persons may benefit less from social learning opportunities (like therapy), we may predict that high versus low EM may predict efficacy of manualized treatments (medication / psychosocial interventions / guideline-approved therapies
- ET may inform treatment change trajectories: we may assume that increase in ET may be required for change to occur (whilst also distinguishing between ET in relation to the therapist/team/unit and ET in relation to group members)
- We may delineate a set of techniques that are specifically designed to address EM, e.g. at the first 3-5 sessions in (any) treatment and investigate the value of this intervention in predicting outcome of the subsequent intervention.

Finally, in addition to the need for outcome variables beyond symptom-reduction and those co-created with patients, a meaningful assessment of therapy success should include changes in relation to goals set by the patient as well as a focus on assessing change in areas such as loneliness/epistemic isolation, social learning, and generalisation in ET to extra-therapeutic relationships. The keys question here is whether improved trust within the therapueitc relationship therapy generalise to other relationships? Or, as demonstrated for learning in children: Does improved attachment security lead to an increase in social learning? Recent advances in methodology could facilitate research programms to operationalize and assess

changes and dynamics in communication system 3, e.g. via the us of Social Network Analysis (Bevington et al., in press), Ecological momentary assessment (EMA; Ellison et al., 2021) or changes assessed by third observers.

References:

Aisbitt, G. M., Nolte, T., & Fonagy, P. (2022). Editorial Perspective: The digital divide–inequalities in remote therapy for children and adolescents. *Child and Adolescent Mental Health*.

Allen, J. G. (2021). *Trusting in psychotherapy*. American Psychiatric Pub.

Bateman, A., Campbell, C., & Fonagy, P. (2021). Rupture and repair in mentalization-based group psychotherapy. International Journal of Group Psychotherapy, 71(2), 371-392.

Bateman, A., Constantinou, M. P., Fonagy, P., & Holzer, S. (2021). Eight-year prospective follow-up of mentalization-based treatment versus structured clinical management for people with borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment, 12*(4), 291-299. doi:10.1037/per0000422

Bateman, A. et al., (under review). Mentalization based treatment – trauma focused (MBT-TF): rationale of a mentalizing model for treatment of people who have experienced developmental trauma.

Bateman, A., Fonagy, P., Campbell, C., Luyten, P., Debbané, M., & Allison, E. (2023). Cambridge Guide to Mentalization-Based Treatment (MBT) (Cambridge Guides to the Psychological Therapies). Cambridge: Cambridge University Press.

Berthelot, N., Ensink, K., Bernazzani, O., Normandin, L., Luyten, P., & Fonagy, P. (2015). Intergenerational transmission of attachment in abused and neglected mothers: The role of trauma-specific reflective functioning. *Infant Mental Health Journal, 36*(2), 200-212. doi:10.1002/imhj.21499

Campbell, C., & Allison, E. (Forthcoming). Mentalizing the modern world. *Psychoanalytic Psychotherapy*.

Campbell, C., Tanzer, M., Saunders, R., Booker, T., Allison, E., Li, E., . . . Fonagy, P. (2021). Development and validation of a self-report measure of epistemic trust. *PLoS One*. doi:10.1371/journal.pone.0250264

Colle, L., Dimaggio, G., Carcione, A., Giuseppe, N., Semerari, A., & Chiavarino, C. (2020). Do competitive contexts affect mindreading performance? *Frontiers in Psychology* doi:10.3389/fpsyg.2020.01284

Corriveau, K., Harris, P., Meins, E., C., F., 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*(3), 750-761.

Coughlan, B., Duschinsky, R., Bakermans-Kranenburg, M., Bakkum, L., Skinner, G., Markham, A., ... & van IJzendoorn, M. H. (2022). Sequelae of Child Maltreatment: Umbrella synthesis on mental health correlates in over 11 million participants. PsyArXiv. December, 19.

Cristea, I. A., Gentili, C., Cotet, C. D., Palomba, D., Barbui, C., & Cuijpers, P. (2017). Efficacy of psychotherapies for borderline personality disorder: a systematic review and meta-analysis. Jama psychiatry, 74(4), 319-328.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *13*(4), 148-153.

Ensink, K., Begin, M., Normandin, L., & Fonagy, P. (2017). Parental reflective functioning as a moderator of child internalizing difficulties in the context of child sexual abuse. *Psychiatry Research, 257*, 361-366. doi:10.1016/j.psychres.2017.07.051

Euler, S., Nolte, T., Constantinou, M., Griem, J., Montague, P. R., Fonagy, P., & Personality and Mood Disorders Research Network. (2021). Interpersonal problems in borderline personality disorder: associations with mentalizing, emotion regulation, and impulsiveness. Journal of Personality Disorders, 35(2), 177-193.

Fertuck, E. (2018). Presentation at 5th International Congress on Borderline Personality Disorder and Allied Disorders, Barcelona. Enhancing interpersonal trust in Transference-Focused Psy- chotherapy: Theory, research, and principle interventions

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. Psychotherapy, 51(3), 372.

Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. Journal of personality disorders, 29(5), 575-609.

Fonagy, P., Campbell, C., Constantinou, M., Higgitt, A., Allison, E., & Luyten, P. (2021). Culture and Psychopathology. *Development and Psychopathology*, 1-16. doi:10.1017/S0954579421000092

Fonagy, P., Campbell, C., & Luyten, P. (2018). It's all in the eye of the beholder: Personality and personality disorder as interpersonal phenomena. *European Journal of Personality, 32*(5), 548-549.

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017a). What we have changed our minds about: Part 1. Borderline personality disorder as a limitation of resilience. *Borderline*

Personality Disorder and Emotion Dysregulation, 4, 11. doi:10.1186/s40479-017-0061-9

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017b). What we have changed our minds about: Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication. *Borderline Personality Disorder and Emotion Dysregulation, 4*, 9. doi:10.1186/s40479-017-0062-8

Fisher, S., Guralnik, T., Fonagy, P., & Zilcha-Mano, S. (2021). Let's face it: video conferencing psychotherapy requires the extensive use of ostensive cues. Counselling Psychology Quarterly, 34(3-4), 508-524.

Fisher, S., Zilcha-Mano, S., & Fonagy, P. (2022). Conceptualizing epistemic trust in psychotherapy: A triadic model. Psychotherapy Bulletin, 57(2), 29-35.

Folmo, E. J., Karterud, S. W., Kongerslev, M. T., Kvarstein, E. H., & Stänicke, E. (2019). Battles of the comfort zone: modelling therapeutic strategy, alliance, and epistemic trust—a qualitative study of mentalization-based therapy for borderline personality disorder. Journal of Contemporary Psychotherapy, 49(3), 141-151.

Folmo, E. J., Langjord, T., Myhrvold, N. C., Stänicke, E., Lind, M., & Kvarstein, E. H. Pedagogical stance in mentalization-based treatment. *Journal of clinical psychology*.

Fotopoulou, A., & Tsakiris, M. (2017). Mentalizing homeostasis: The social origins of interoceptive inference. Neuropsychoanalysis, 19(1), 3-28.

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences, 17*, 160-165. doi:10.1016/j.tics.2013.02.002

Goemans, A., Viding, E., & McCrory, E. (2021). Child Maltreatment, Peer Victimization, and Mental Health: Neurocognitive Perspectives on the Cycle of Victimization. *Trauma Violence Abuse*, 15248380211036393. doi:10.1177/15248380211036393

Huang, Y. L., Fonagy, P., Feigenbaum, J., Montague, P. R., Nolte, T., & London Personality and Mood Disorder Research Consortium. (2020). Multidirectional pathways between attachment, mentalizing, and posttraumatic stress symptomatology in the context of childhood trauma. Psychopathology, 53(1), 48-58.

Kamphuis, J. H., & Finn, S. E. (2018). Therapeutic assessment in personality disorders: Toward the restoration of epistemic trust. Journal of personality assessment.

Kampe, K. K., Frith, C. D., & Frith, U. (2003). "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. Journal of Neuroscience, 23(12), 5258-5263.

Kampling, H., Kruse, J., Lampe, A., Nolte, T., Hettich, N., Brähler, E., ... & Riedl, D. (2022). Epistemic trust and personality functioning mediate the association between adverse childhood experiences and posttraumatic stress disorder and complex posttraumatic stress disorder in adulthood. Frontiers in Psychiatry.

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., & Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. Proceedings of the National Academy of Sciences, 114(50), 13290-13295.

Knapen, S., Hutsebaut, J., van Diemen, R., & Beekman, A. (2020). Epistemic trust as a psycho-marker for outcome in psychosocial interventions. *Journal of Infant, Child, and Adolescent Psychotherapy*, *19*(4), 417-426.

Li, E. T., Midgley, N., Luyten, P., Sprecher, E. A., & Campbell, C. (2022). Mapping the journey from epistemic mistrust in depressed adolescents receiving psychotherapy. *Journal of Counseling Psychology*.

Luyten, P., Campbell, C., Allison, E., & Fonagy, P. (2020). The mentalizing approach to psychopathology: State of the art and future directions. *Annual Review of Clinical Psychology, 16*(1). doi: 10.1146/annurev-clinpsy-071919-015355

Luyten, P., Campbell, C., & Fonagy, P. (2019). Borderline personality disorder, complex trauma, and problems with self and identity: A social-communicative approach. *Journal of Personality*, doi: 10.1111/jopy.12483. doi:10.1111/jopy.12483 [epub ahead of print]

Luyten, P., & Fonagy, P. (2019). Mentalizing and trauma. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (2nd ed., pp. 79-99). Washington, DC: American Psychiatric Publishing.

Neil, L., Viding, E., Armbruster-Genc, D., Lisi, M., Mareschal, I., Rankin, G., ... & McCrory, E. (2022). Trust and childhood maltreatment: evidence of bias in appraisal of unfamiliar faces. *Journal of child psychology and psychiatry*, *63*(6), 655-662.

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.

McLaren, V., Gallagher, M., Hopwood, C. J., & Sharp, C. (2022). Hypermentalizing and borderline personality disorder: A meta-analytic review. American Journal of Psychotherapy, 75(1), 21-31.

McCrory, E., Foulkes, L., & Viding, E. (2022). Social thinning and stress generation after childhood maltreatment: a neurocognitive social transactional model of psychiatric vulnerability. The Lancet Psychiatry.

O'Madagain, C., & Tomasello, M. (2019). Joint attention to mental content and the social origin of reasoning. *Synthese*. doi:10.1007/s11229-019-02327-1

Patrizio, R. F. (2023). Trust's meno problem: Can the doxastic view account for the value of trust?. *Philosophical Psychology*, 1-20.

Riedl, D., Rothmund, M. S., Grote, V., Fischer, M. J., Kampling, H., Kruse, J., ... & Lampe, A. (2023). Mentalizing and epistemic trust as critical success factors in psychosomatic rehabilitation: results of a single center longitudinal observational study. Frontiers in Psychiatry, 14.

Schröder-Pfeifer, P., Georg, A. K., Talia, A., Volkert, J., Ditzen, B., & Taubner, S. (2022). The Epistemic Trust Assessment—An experimental measure of epistemic trust. Psychoanalytic Psychology, 39(1), 50.

Sprecher, E. A., Li, E., Sleed, M., & Midgley, N. (2022). 'Trust me, we can sort this out': a theory-testing case study of the role of epistemic trust in fostering relationships. *Qualitative Research in Psychology*, 1-26.

Talia, A., Duschinsky, R., Mazzarella, D., Hauschild, S., & Taubner, S. (2021). Epistemic trust and the emergence of conduct problems: aggression in the service of communication. *Frontiers in psychiatry*, *12*.

Tomasello, M. (2016). *A Natural History of Human Morality*. Cambridge, MA: Harvard Univ Press.

Tomasello, M. (2020). The role of roles in uniquely human cognition and sociality. *Journal for the theory of social behaviour, 50*(1). doi:org/10.1111/jtsb.12223

Weisner, T. S. (2014). The socialization of trust: plural caregiving and diverse pathways in human development across cultures. In H. Otto & H. Keller (Eds.), *Different Faces of Attachment: Cultural Variations on a Universal Human Need* (pp. 263-277). Cambridge: Cambridge University Press.

Ventura Wurman, T., Lee, T., Bateman, A., Fonagy, P., & Nolte, T. (2021). Clinical management of common presentations of patients diagnosed with BPD during the COVID-19 pandemic: The contribution of the MBT framework. Counselling Psychology Quarterly, 34(3-4), 744-770.