Eli S. Neustadter,[1]
Aikaterini Fotopoulou,[2]
Matthew Steinfeld[1]
and Sarah K. Fineberg[1]

Mentalization and Embodied Selfhood in Borderline Personality Disorder

Abstract: *Aberrations of self-experience are considered a core feature of borderline personality disorder (BPD). While prominent aetiological accounts of BPD, such as the mentalization-based approach, appeal to the developmental constitution of self in early infant–caregiver environments, they often rely on a conception of self that is not explicitly articulated. Moreover, self-experience in BPD is often theorized at the level of narrative identity, thus minimizing the role of embodied experience. In this article, we present the hypothesis that disordered self and interpersonal functioning in BPD result, in part, from impairments in 'embodied mentalization' that manifest foundationally as alterations in minimal embodied selfhood, i.e. the first-person experience of being an individuated embodied subject. This account of BPD, which engages early intersubjective experiences, has the potential to integrate phenomenological, developmental, and*

Correspondence:
Email: eli.neustadter@yale.edu

1 Yale University School of Medicine, Department of Psychiatry, New Haven, CT, USA.
2 University College London, UK.

symptomatic findings in BPD, and is consistent with contemporary theories of brain function.

Keywords: borderline personality disorder; mentalization; minimal self; narrative self; interoception.

1. Introduction

Selfhood has been considered an organizing construct for theorizing borderline personality disorder (BPD) (Kerr *et al.*, 2015). Aberrant self-experience in BPD is characterized by dramatic changes in self-image, shifting goals and values, and feelings of emptiness, dissociation, and non-existence (Gunderson *et al.*, 2018). These experiences are distressing and dangerous: in a qualitative study, Brown, Comtois and Linehan (2002) found that more than 50% of interviewed women with BPD endorsed disturbances in self-experience as reasons for non-suicidal self-injury.

A prominent developmental account of BPD, the mentalization-based approach (MBA) (Fonagy and Luyten, 2009), appeals to the interpersonal constitution of 'self' structure in infancy. The caregiver's mentalization of the child's behaviour — i.e. the caregiver's ability to adopt an intentional stance and represent their child as having feelings, desires, and intentions — fosters the child's capacity to mentalize his or her own internal experience and sense of self. This occurs via the caregiver's re-presentation, or 'mirroring', of these states during infant–caregiver interactions (Slade, 2005). While the MBA seeks to explain 'pathologies of selfhood' in BPD, such as affect dysregulation, identity diffusion, and unstable self–other boundaries, it relies on a concept of self that is not explicitly articulated.

Philosophical approaches to selfhood in BPD have predominantly focused on the 'narrative self' (e.g. Gold and Kyratsous, 2017; Kernberg, 2006; Fuchs, 2007; Jørgensen, 2006). In this view, the self is a narrative constructed by the individual through the integration of memories, present experience, and future goals into a coherent story. Indeed, narrative coherence is lower in life story interviews of people with BPD compared to non-BPD narrators (Adler *et al.*, 2012). Fuchs (2007) goes a step further, writing that aberrations in self-experience are '*caused* by the inability to integrate past and future into the present and thus to establish a coherent sense of identity' (p. 379, italics ours). For some in the phenomenological tradition, BPD is distinguished by disturbances of the 'narrative self' while alterations of the more basic aspects of self are unique to psychotic-spectrum illnesses such as

schizophrenia (Parnas and Henriksen, 2014). And with regards to the MBA, the 'narrative self' may fully account for the implicit conception of self therein, given frequent appeals to 'identity', 'representation of self', and 'self-image' (e.g. Fonagy *et al.*, 2002).

However, while fragmented narrative identity is a feature of borderline pathology, anchoring self-disturbance in BPD only on narrative accounts risks minimizing the important role of embodied experience. Abnormal bodily experiences in BPD are common, including bodily dissociation, altered pain perception, and deficits in interoception (awareness and processing of bodily signals) (Löffler, Foell and Bekrater-Bodmann, 2018). Furthermore, recent empirical and theoretical advancements in mentalization theory highlight the *intercorporeality* of mentalizing activity, such that embodied subjectivity in early infancy develops through a 'co-construction of somatic experience within attuned bodily interactions with the caregiver' (Shai and Fonagy, 2014). This suggests that alterations of self in BPD present not just at the level of narrative identity, but in more basic aspects of embodiment.

In this paper, we present the hypothesis that disordered self and interpersonal functioning in BPD result, in part, from impairments in 'embodied mentalization' (Fotopoulou and Tsakiris, 2017) that manifest foundationally as alterations in minimal embodied selfhood, i.e. the first-person experience of being an individuated embodied subject. This theoretical expansion of the MBA seeks to further clarify how, mechanistically, early developmental processes may lead to embodied disturbances in BPD.

This account of BPD, which engages early intersubjective embodied experiences and links them to predictive coding models of brain function, has the potential to integrate phenomenological, developmental, and symptomatic findings in BPD. According to predictive coding accounts of brain function, the brain generates probabilistic and dynamic representations by integrating incoming sensory input (Clark, 2013). By this account, we can construe minimal embodied selfhood as the phenomenological manifestation of interoceptive inference, i.e. the generative representation of one's physiologic states (Seth and Tsakiris, 2018). Employing the concept of 'embodied mentalization', we will explain bodily self as shaped through infants' embodied interactions. Progressive integration of interoceptive and sensorimotor signals result in stable representations of self. This process may be impaired by known aetiological risk factors for BPD, including disorganized attachment and neglect (Fonagy, 2000).

In expanding this paper's main claims, we first review self-pathology as a core clinical dimension of BPD (Section 2). In Section 3, we elaborate the MBA and argue that it relies on an implicit conception of selfhood that is unarticulated. In Section 4, we consider two philosophical approaches to selfhood (i.e. the narrative and minimal self, respectively), and claim that, while minimal selfhood is under-theorized in BPD, it possesses explanatory utility in mentalization-based approaches to the disorder. In Section 5, we demonstrate how Fotopoulou and Tsakiris's (2017) concept of 'embodied mentalization' can provide a mechanistic account of alterations in minimal embodied selfhood in BPD that is coherent with contemporary predictive coding theories of brain function. In Section 6, we consider how this approach is consistent with empirical findings in BPD, and offers a framework to propose novel hypotheses regarding the entwinement of self and interpersonal pathology.

2. Self-Disturbance and Borderline Personality Disorder

BPD is a severe mental disorder associated with high rates of suicide and self-harm, marked functional impairment, and high cost to the individual and society (Leichsenring *et al.*, 2011). Clinically, BPD is characterized by four key symptom phenotypes: self-disturbance, interpersonal instability, emotion dysregulation, and behavioural dysregulation, such as impulsivity and self-harm. Self-disturbance is considered a core feature of BPD, and encompasses unstable or impoverished sense of self, instability of goals and values, chronic feelings of emptiness, and dissociation (Gunderson *et al.*, 2018).[3] First-person accounts attest to this self-disturbance in BPD:

> My mind was so murky, making everything seem like a dream. I remember walking around Newbury street, the foggy streetlights danced shadows around an old church. I saw myself in all of them. The hotel room had a TV in the bathroom mirror. I remember staring at myself.

[3] Self-disturbance in BPD is captured by the following DSM-V diagnostic criteria: identity disturbance (markedly and persistently unstable self-image or sense of self), chronic feelings of emptiness, and transient, stress-related paranoid ideation or severe dissociative symptoms. Other diagnostic criteria include: frantic efforts to avoid real or imagined abandonment, a pattern of unstable and intense interpersonal relationships, impulsivity, recurrent suicidal behaviour, affective instability due to a marked reactivity of mood, and inappropriate intense anger.

> My reflection felt as fake as the pixilated figure in the corner. For a moment, we fused. (Duffy, 2017)

This description of transitivism captures how self-instability impacts on the relationship between self and the non-self environment. This can be realized, as above, as a dissolution of the self–environment boundary, such that the self is felt as porous or undifferentiated. Self-pathology in BPD also extends into interpersonal settings. Individuals with BPD have a larger preferred social distance in interpersonal settings indicating an experientially enlarged sense of peripersonal space (i.e. the subjective sense of one's personal space immediately surrounding the body) (Fineberg *et al.*, 2018). (We discuss further alterations of 'minimal selfhood' in BPD in Section 4.2.) Furthermore, an instability in self-understanding extends to instability in thoughts related to others: individuals with BPD are less consistent in attributing personality traits to themselves and others, and more strongly endorse statements probing self–other undifferentiation (Beeney *et al*., 2016). (We discuss further alterations of 'narrative selfhood' in BPD in Section 4.1.) For Bender and Skodol (2007), 'disturbances in self and other mental representations are fundamental to borderline psychopathology' (p. 514). Similarly, Kerr *et al*. (2015) call for a 'reconceptualization' of BPD as 'a disorder of self and relationality' (p. 346). Interpersonal symptoms can include confusion about self–other boundaries with identity diffusion, projection of difficult affects onto interaction partners. One person with BPD evokes a bodily experience of this confusion (italics added):

> When I'm around other people, I can *feel* their energy. I can *feel* whether they're happy or unhappy. I can *walk into a room and feel* whether there's tension in the air or if everyone is getting along… Having my own identity issues makes it even harder to be around certain people. I feel other people's emotions so strongly that sometimes, I believe they are my own. (Mae, 2017)

Self, then, is a key concept to characterize symptoms (and symptom patterns) for the formulation of BPD. However, as Kerr *et al.* (2015) note, when used in the context of BPD research and formulation, 'a very wide range of concepts of the self has been employed by different writers, researchers and clinicians. [However], the concepts of self have only been defined to a limited extent, with many writers resorting to more "common parlance" or "folk psychological" definitions' (p. 340). They suggest that appeals to the self may cause considerable confusion if clinicians and scientists are working with

different implicit conceptions of the term. This concern is especially relevant for developmental accounts of BPD, across behavioural (e.g. Crowell, Beauchaine and Linehan, 2009) and psychodynamic traditions (e.g. Klein, 1946; Kernberg, 1985; Fonagy *et al.*, 2002), that rely on notions of self to articulate how early childhood experiences developmentally give rise to BPD symptomology. For example, in her biosocial developmental model of borderline personality, Linehan (1993) locates the origins of borderline self-disturbance in invalidating caregiver responses to a child's perceptions and beliefs about the world, resulting in the reliance of the developing individual on others to construct a sense of meaning. In Kernberg's (2006) object-relations theory, borderline pathology results from the 'pathological consolidation of the internalized world of object relations, reflected in a stable lack of integration of the concept of self and of significant others' (p. 980). To delimit the scope of this paper, we examine the articulation of self specifically in the mentalization-based account of BPD. This prominent developmental account of BPD, developed by Fonagy and others (e.g. Fonagy *et al*., 2002), grounds selfhood in the developmentally acquired capacity of an individual to 'mentalize' one's self and others. Furthermore, we believe it holds particular explanatory potential in elucidating alterations of embodiment in BPD (see Section 6).

3. Mentalization, Attachment, and the Development of the Self

According to the MBA, unstable sense of self is a core feature of BPD, for which symptoms are not isolated characteristics of the condition, but are aetiologically related to core processes in early self-development, i.e. the capacity to mentalize in the context of early infant–caregiver (attachment) relationships (Fonagy *et al.*, 2002). Slade (2005) characterizes mentalization 'narrowly as the capacity to understand one's own and others' behavior in terms of underlying mental states and intentions, and more broadly as a crucial human capacity that is intrinsic to affect regulation and productive social relationships'. Importantly, for the MBA, mentalization is not an inborn capability; rather, the ability of a developing infant to ascribe meaning to its own affective and behavioural states is a developmental achievement born within early attachment relationships. Infants learn about the mental world, and their own self-states, through the

caregiver's 'affective mirroring' of infant behaviour during infant–caregiver interactions:

> It is through… mothers' 'marking' of their very young infants' affect displays, producing an exaggerated version of realistic emotion expressions, in which the infant's state is reflected back to them as a 're-presentation' or proto-symbol, that the child first begins to organize his self experience. The 3-month-old first learns about mental states and about the mental world as he observes them in his caregiver, as representations of his self-state; only then can he begin to recognize them in himself. (*ibid.*)

During embodied interactions, caregivers are attuned to a rich repertoire of non-verbal communication signals from their infants, including hand gestures, facial expressions, and bodily movements that reflect aspects of their internal states and explorations of the physical environment. When a caregiver soothes a distressed infant, for example, they reflect the child's inferred mental state, with framing cues (e.g. raised eyebrows, titled head) and other affect displays (e.g. smiling, mocking) to communicate empathy while acknowledging the individuation of the infant's developing mind (Fonagy and Target, 1997).

In the MBA, it is posited that the unstable sense of self in BPD originates in 'an early psychosocial environment where [infants'] internal experiences were not adequately mirrored' (Fonagy and Luyten, 2009, p. 1359). Internal states then remain 'confusing, experienced as unsymbolized, and hard to regulate' (*ibid.*). Affective mirroring sculpts self–other boundaries and the individuation of the infant's mind because the infant learns that his caregiver can understand his internal state without sharing it (Allen and Fonagy, 2006). Furthermore, early infant–caregiver interactions are crucial in that they generalize to later relationships throughout development and adulthood (Cittern *et al.*, 2018). According to the MBA, secure caregiver attachment forms the foundation for fully developed mentalizing capacities and sense of self (Fonagy, 2000).

Secure attachment, in which caregivers form a secure base for infant exploration and discovery, is in part predicated on the caregiver's ability to reflect on their own and their child's mental experience, allowing them to accurately understand and predict behaviour. This capacity is known as parental reflective functioning (Luyten *et al.*, 2017). Importantly, parental reflective functioning supports secure attachment not only through cognizing infant mental states, but through promoting embodied interactions with infants that non-

verbally communicate parental attunement to their infant's thoughts and emotions (Shai and Meins, 2018).

Conversely, neglect, trauma, and insecure attachment — in which caregiver responses are inconsistent, ignoring, or aggressive — are associated with 'disrupt[ion of] the capacity for mentalization and, linked to this, the development of a coherent self-structure' (Fonagy and Luyten, 2009). Importantly, these adverse early experiences (Widom, Czaja and Paris, 2009; Korzekwa, Dell and Pain, 2009), and relatedly, insecure attachment style (Gunderson and Lyons-Ruth, 2008), are risk factors for BPD. Mentalization may be compromised in these settings because of the challenges of integrating the inconsistent beliefs and actions of abusive or poorly attuned caregivers who are less likely to consider, and thereby reflect, the internal states of their children (Ensink *et al.*, 2015). When affective mirroring is absent, or parental mentalization and mirroring are impaired (e.g. in the setting of parental substance use, parental psychopathology, or other psychosocial stressors), the child is left with a sense of self that is unknowable, and experientially empty (Slade, 2005).

3.1. Mentalization and BPD

The central premise of the MBA is that 'the failure of mentalizing, in combination with profound disorganization of self-structure, may account for the core features of borderline personality functioning' (Fonagy and Luyten, 2009, p. 1357). Unstable sense of self arises when infants internalize reflections from caregivers that do not align with their 'constitutional' inner states. Fonagy *et al.* (2002) describe the formation of an 'alien' self-representation within individuals with BPD that corresponds not to their constitutional states, but to the intentional, and often persecutory, mind of a maltreating attachment figure.

According to the MBA, this formation of an 'alien-self' underlies the development of canonical BPD defence mechanisms and symptoms. In efforts to stabilize the sense of self, individuals with BPD defensively project this misaligned representation onto the other. Projective identification functions to evoke thoughts and feelings which cannot be tolerated as arising from the self (Sharp and Vanwoerden, 2015). Reliance on splitting (i.e. the formation of distinct idealized and persecutory representations of a single mind) obviates the need to integrate opposing representations of self, and of self versus other. However, split representations of self and of

attachment figures can make it difficult to predict the behaviour and reactions of others, and engender a sense of chaos, and feelings of anxiety, hopelessness, and diminished agency (Fonagy, 2000).

According to the MBA, DSM-V social symptoms of BPD arise in part from the maladaptive defence mechanisms enacted to cope with an alien self-representation. For example, efforts to avoid real or imaged abandonment are bound up with the individual's attempt to stabilize the sense of self. Projecting painful and persecutory intentional states onto intimate partners and other attachment figures can also cause anger, resentment, and confusion in the other, and contributes to the volatile interpersonal relationships that are common in BPD. According to the MBA, childhood abuse, maltreatment, or neglect create a developmental environment in which 'recognition of the mental state of the other can be dangerous to the developing self' (*ibid.*, p. 1134). This inhibition of mentalization contributes to emotional dysregulation, such that high arousal states occur in 'the absence of symbolic representations of emotions', and are beyond self-control (Fonagy *et al.*, 2002, p. 360). People in this situation may focus on the immediate physical reality as opposed to understood intentions and affects of a social interaction partner, and impulsive behaviours can result, especially in interpersonal contexts (*ibid.*). Frantic efforts to avoid abandonment are tragic and common precipitants of self-harm and suicidal acts. Perceived abandonment entails facing and reinternalizing aspects of the alien self which are intolerable, and 'suicide represents the fantasized destruction of this alien other within the self' (*ibid.*, p. 363). Thus, the MBA seeks to elucidate the aetiology of a broad range of BPD symptoms.

4. Locating the Self in the MBA

While our brief exposition of the MBA demonstrates the conceptual centrality of the 'self', it also reveals the indeterminacy with which the concept is employed. According to the MBA, pathologies of selfhood in BPD are realized at the level of first-person experience, in the stability of 'self-representation', and in interpersonal contexts. In light of Kerr *et al.*'s (2015) concern regarding the utility of the concept in formulating BPD pathology, it appears that the MBA appeals to the 'self and its functions' without 'incorporat[ing] a formal concept of the self' (p. 340). Perhaps one benefit of relying on an implicit conception of self is that it enables the MBA to be consistent with a variety of self-conceptions across philosophical and psychological

approaches. However, even within philosophical traditions, the conceptual plurality surrounding selfhood has led some to claim that the self does not exist (e.g. Strawson, 1997), and that the concept should be abandoned (Metzinger, 2005). If this were the case, then contemporary formulations of BPD, such as that provided by the MBA, may be hindering our understanding of BPD by appealing to the concept of self.

One challenge to locating the self in the MBA is the inherent ambiguity of the concept. As Kyselo (2014) recently summarized, 'models and conceptions of the self are diverse. It is considered a substance or a thing, a concept, a narrative, a system, a process or a function… This list is not exhaustive but it makes a point: there is no unifying concept of *the* self' (p. 1). Conceptual clarification of the concept is not merely a philosophical concern, but has theoretical and empirical implications for scientific explorations of the self (*ibid.*). It also has the potential to support theoretical advances in the MBA and empirical investigations of BPD. In an influential paper, Gallagher (2000) divides philosophical approaches to the self that have relevance for the cognitive sciences into two important aspects: the 'narrative' self and the 'minimal' self. In the following sections, we will explore the application of these approaches to the MBA.

4.1. Narrative selfhood and BPD

According to the narrative-self approach, selfhood does not arise merely because an individual persists through time as the self-same physical being, or from an awareness of first-person experience. Rather, in narrative accounts, the self is constituted by the stories we tell about ourselves (Zahavi, 2005). Narrative selfhood arises from the capacity to mentally represent the continuity between past and future selves in a personal life narrative that is both intelligible and meaningful. Individuals behave in a way that demonstrates self-understanding and self-relating, such that the unity of narrative selfhood is enacted through lived experience: in the pursuit of goals, in making ethical decisions in accordance with one's values, and pursing a life of one's making (Fuchs, 2007). However, the narrative approach does not entail that selfhood is constructed by the individual alone: our lives are embedded in relationships and larger societal systems of meaning and language that anchor our identities. For Dennett (2017), the narrative self does not reside in the individual alone, but exists as an abstraction, a centre of 'narrative gravity' that specifies the locus of

intersection for the multiplicity of stories about ourselves that we and others enact.

Narrative selfhood has historically been the predominant approach by which self-pathology in BPD is theorized (Fuchs, 2007; Kernberg, 2006; Jørgensen, 2006). Fuchs conceives narrative identity as an achievement of integrating contradictory thoughts, experiences, and desires into 'a coherent, overarching sense and view of [one]self' (2007, p. 379). And, in BPD, it is the inability to construct a coherent self-concept that constitutes disordered selfhood (*ibid.*). The structure of narrative selfhood is also a central and distinguishing feature of BPD for Kernberg (2006), who conceptualizes it as a form of 'identity diffusion', characterized by inconsistencies in beliefs, values, and life goals, difficulties with interpersonal and professional commitments, and unstable representations of self and other.

Deficits in narrative selfhood have also been invoked to account for other clinical features of BPD. For example, Fuchs (2007) suggests that impulsivity in BPD occurs in the absence of 'enduring second-order volitions', such that individuals with BPD succumb to impulsive behaviours in the moment because they lack an overarching sense of their goals and values to resist in-the-moment urges (i.e. first-order volitions). Unable to integrate coherent mental representations of individuals across time, those with BPD rely on what Fuchs describes as 'temporal splitting' to align their present emotional experience with their representations of self and other. Moreover, it is this splitting of self and subsequent fragmentation of narrative identity that, for Fuchs, results in the chronic feelings of emptiness endorsed by individuals with BPD, as the present self, severed from its past and future, is experienced as empty and flat.

In considering our discussion above, it is *prima facie* plausible that the narrative account fully underlies the conception of self implicit in the MBA. For Jørgensen (2006), identity diffusion in BPD is a direct result of deficits in mentalization, as it underlies our capacity to understand and represent the emotions, intentions, and goals of ourselves and others. Likewise, for Fuchs (2007), the potential for stable narrative identity is realized in early child development: 'only if the child's experiences meet adequate mirroring… by others can they be integrated into a coherent understanding of what it means to be a self with… a basic temporal continuity' (p. 383). Conceiving unstable selfhood in BPD as a temporal instability is consistent with the narrative approach, as well as empirical findings regarding fluctuations in the representations of self and other over time (see Section 2). Cognitive

perspectives on the narrative approach suggest that the self exists as a dynamic set of mental representations that serve self-monitoring functions and are related to core processes in the formation and maintenance of autobiographical memory (Gallagher, 2000; Dennett, 2017; Prebble, 2013). Indeed, individuals with BPD demonstrate impairments in autobiographical memory, including memory gaps (Korzekwa, Dell and Pain, 2009) and over-generalization (Startup *et al.*, 2001), such that they are more likely to recall categories of events rather than specific details. Taken together, these theoretical and empirical considerations indicate profound alterations of narrative selfhood in BPD. However, we want to suggest that a reliance solely on the narrative approach risks minimizing self-disturbances in BPD related to embodied experience, and does not fully account for foundational aspects of self-development in the MBA.

4.2. The minimal self and embodied experience in BPD

The minimal approach to selfhood posits that the self exists independent to, and prior to, one's mental representations of self, and prior to self-reflection. For Zahavi (2005), the narrative approach tacitly relies on this more foundational conception of selfhood. Essential to the minimal approach is the commitment to the claim that selfhood is not merely a construction built over time but, rather, an intrinsic feature of conscious experience. Considered phenomenologically, the minimal self exists in first-person experience as a pre-reflective form of self-awareness such that whenever one has a conscious experience — whether it entails looking at a painting, smelling a flower, or reflecting on a conversation with a friend — it is experienced *as* one's own (Higgins, 2018). Importantly, minimal approaches do not locate the self in conscious experience abstracted from the physical world and body. Rather, minimal selfhood comprises essential embodied and experiential components, including a first-person perspective, self–other boundary, and sense of body ownership (Gallagher, 2013). The unreflective experience of having a body, what Seth and Tsakiris (2018) call 'being a body', is an essential component of minimal selfhood that 'describes a background experience of self… that shades into mood and emotion at one end and into experiences of body ownership at the other' (p. 8).

BPD is associated with disturbances in experiential embodiment. It is associated with high rates of body dysmorphia (Dyer *et al.*, 2013), altered bodily pain perception (Schmahl *et al.*, 2010), and dissociation

during periods of stress (Korzekwa, Dell and Pain, 2009). In particular, a phenomenologically reduced sense of body ownership is a frequent and specific manifestation of dissociation in BPD (Zanarini *et al.*, 2000) and, in interviews, individuals report a significantly reduced sense of belonging to their own bodies compared to those without mental illness (Kleindienst *et al.*, 2020).People diagnosed with BPD are also more susceptible to illusions of body ownership compared to individuals without mental illness, indicating decreased stability of their body schemas (Bekrater-Bodmann *et al.*, 2016; Neustadter *et al.*, 2019; Becker-Sadzio, 2019; but see Möller *et al.*, 2020, for only trend level differences). For example, in the rubber hand illusion paradigm, participants see a rubber hand touched while they feel their hidden hand touched in the same way. The extent to which they experience the illusion (endorsing that their touch sensation comes from the rubber hand, or, for some, even that the rubber hand is *their* hand) correlates with dissociative experiences (e.g. derealization, depersonalization) (Bekrater-Bodmann *et al.*, 2016).[4] People with BPD also have more difficulty identifying their own emotions than do those without mental illness or other personality disorders (New *et al.*, 2012), suggesting that the way they experience their own bodily sensations may be affected. More generally, individuals with BPD are known to present with alexithymia, a multidimensional failure of processing bodily signals (i.e. interoception) (Murphy, Catmur and Bird, 2018) that, in BPD, coincides with a painful felt-unknowingness of inner experience. These findings suggest that self-pathology in BPD manifests not only at the level of narrative identity, but also in fundamental aspects of embodied experience.

Given that the MBA purports to explain self-pathology in BPD, these findings raise the question whether the narrative approach fully captures the implicit conception of self in the theory. While the MBA appeals to the development of theory of mind, self-image, and mental representation, it also stresses the role of intersubjective embodiment as an essential component of parental reflective functioning and the

[4] Of note, increased susceptibility to body illusion paradigms is not specific to BPD. Rather, it is a common finding across several conditions that share symptomatic and clinical overlap with BPD, including schizophrenia (Thakkar *et al.*, 2011), body dysmorphia disorder (Kaplan *et al.* 2014), and eating disorders (Eskkevari *et al.*, 2011). In general, our aim is not to suggest that pathologies of embodiment are unique to BPD. Rather, we want to illuminate disturbances of embodiment in BPD, and provide an aetiological account of their development in individuals diagnosed with the condition.

development of mentalizing capacities in the infant. For example, Slade (2005) recognizes that, for pre-verbal infants, mental states are first re-presented by caregivers not through 'words and play', but through 'gesture and action'. Indeed, empirical research of moment-to-moment infant–caregiver interactions suggests that mirroring occurs through interactively contingent attunement of infant and caregiver with regards to facial expression, gaze direction, head orientation, and affect display (Beebe *et al.*, 2016). Some MBA theorists (Shai and Fonagy, 2014) also address features of self-development that correspond to primary components of minimal selfhood, including sense of body ownership and self-demarcation: 'a crucial element in the process of the development of a sense of self is establishing a sense of presence, volume, solidity, and mass. Furthermore, boundaries of the self need to be established to define limits between internal vs. external, me vs. not me, real vs. imaginary' (*ibid.*, p. 198). Thus, it appears that the narrative approach is too limited in scope to provide a conceptional foundation for the self in the MBA.

However, one may object to the idea that the MBA can also be utilized to account for the development of minimal selfhood and, therefore, the aetiology of minimal-self disturbance in BPD, in light of its central thesis: that the self is constituted *intersubjectively* in early development. For example, Zahavi (2014) claims that minimal selfhood exists prior to any form of socialization. Even within the MBA, references are made to a 'constitutional self' that is genetically programmed and upon which mental representations of self overlay (more or less) congruently (e.g. Fonagy, 2000; Bateman, 2006; Fonagy *et al.*, 2002). This tension is a reflection of what Kyselo calls the 'body-social problem': 'how bodily and social aspects figure in the individuation of the human individual self as a whole' (Kyselo, 2014, p. 2). For Kyselo, 'the self as a whole can be either embodied or social, but it cannot be both' (*ibid.*, p. 3). One reply to this problem is to appeal to Gallagher's (2013) 'pattern theory of self', that conceives different conceptions of self as compatible and related, instead of mutually exclusive. In this case, minimal and narrative approaches are not mutually exclusive; rather, narrative selfhood supervenes on more foundational (i.e. minimal) aspects of self (e.g. sense of body ownership and self–other boundary). However, this reply leaves unanswered whether early intersubjective experience shapes the development of minimal embodied selfhood. That sociality shapes minimal aspects of self, in addition to narrative aspects, is *prima facie* plausible. As Higgins (2018) observes, taking seriously the essential, and embodied,

infant–caregiver interactions at the earliest moments of life entails 'accepting the de facto equiprimordiality of minimal experientialism with a "minimal" form of relational selfhood, i.e. the co-constitution of experience through engagements with others' (p. 1). This observation is taken further in Fotopoulou and Tsakiris's (2017) account of 'embodied mentalization' (EM), which elaborates how minimal selfhood is constituted within embodied interactions, such that infants' generative models of sensorimotor signals and physiological states are co-constructed to form the foundations of minimal embodied selfhood.

5. Embodied Mentalization (EM): Mentalization and the Predictive Brain

Traditional accounts of mentalization conceive it as an inferential process related primarily to theory-of-mind processes and intention-attribution. According to Fotopoulou and Tsakiris (2017), however, these cognitive abilities actually reflect 'advanced forms of more primitive inferential processes of embodied perception and action' which they term 'embodied mentalization' (p. 5). EM reflects a reductionist extension of mentalization theory in so far as it grounds the development of self in the intersubjective process of schematizing bodily signals into neurocognitive representations that underlie experiential minimal selfhood. According to EM, proximal embodied interactions beginning in early infancy shape the felt experience of first-person experience along with progressively refined and elaborated representational distinctions between self and other, subject and object, and pleasure and pain (*ibid.*; Ciaunica and Fotopoulou, 2017). Importantly, EM relates mentalization theory to the most essential functions of early infant–caregiver attachment, i.e. the maintenance of survival and the regulation of bodily states (Bowlby, 2008), such that 'the constitution of the self is dependent upon the social mentalization of the body and particularly its homeostatic needs' (Fotopoulou and Tsakiris, 2017, p. 6). The special relevance of interoception for this process will be discussed in Section 5.1. The need to regulate homeostasis is also an organizing principle for predictive coding accounts of brain function, such as the free-energy principle (FEP), according to which the brain generates dynamic representations of our body and the world, in order to predict the causes of bodily states and the consequences of action for the purposes of survival (Friston, 2010). Thus, considering the application of EM to elucidate self-pathology in BPD

has the upshot of linking aetiological accounts of borderline pathology to contemporary theories of brain function.

Within predictive coding accounts of brain function, the brain is a dynamic organ that seeks to render the environment predictable by minimizing surprise and reducing uncertainty regarding the consequences of action (Clark, Watson and Friston, 2018). The brain 'discovers' the world by constructing generative models ('beliefs') about it that are represented as probability distributions regarding the causes of incoming sensory input. According to the FEP, the brain attempts to optimize its model of the world by minimizing 'free energy' (i.e. the upper boundary of surprise under certain simplifying assumptions) through the reduction of prediction error, i.e. the discrepancy between our brain's prediction ('prior') and actual sensory input (Bolis and Schilbach, 2018). The brain organizes expectations hierarchically, whereby predictions from higher cortical layers represent more abstract and integrated representations. Importantly, the precision associated with predictions relative to that of prediction errors — the confidence or uncertainty ascribed to prior beliefs or new information — modulates the integration of bottom-up and top-down information flow; e.g. imprecise prediction errors are less likely to update (top-down) precise prior representations. According to the FEP, free energy can be broadly minimized in two ways: brains can update representations about the world to reduce prediction error (i.e. 'perceptual inference'); second, action (via autonomic/motor reflexes or intentional behaviour) can change sensory input so as to better align with prior expectations (i.e. 'active inference') (Fotopoulou and Tsakiris, 2017).

Within a predictive coding framework, EM can be restated as an inferential brain process that maintains probabilistic 'beliefs' regarding embodied self-states by generating constantly updating models of the likely causes of sensory signals from both inside the body and from the environment (*ibid.*). Although the stability of one's sense of self, including sense of body ownership, is taken for granted in non-pathological situations, what predictive coding perspectives postulate is that representations of the bodily self are probabilistic and therefore malleable. This conceptualization of bodily self-representation is supported by evidence from behavioural paradigms that probe sense of body ownership experimentally. For example, body illusion tasks, such as the rubber hand illusion (RHI) (Botvinick and Cohen, 1998), manipulate the multisensory context in which a virtual or plastic body part is presented, so as to influence the inferences

participants make about whether that fake body part belongs to the self or not. Putatively, during RHI induction, while one visually observes a rubber hand being touched in temporal synchrony with the touch of one's real (but hidden) hand, bottom-up signals from visual, haptic, and proprioceptive modalities are temporarily integrated according to this particular context and this integration results in an updating of one's beliefs about the rubber hand belonging to the self.

5.1. Interoceptive inference: Linking physiology, emotion, and the minimal self

Importantly, according to the theory of EM, minimal selfhood results from the 'mentalization' of bodily signals across exteroceptive and interoceptive sensory modalities. Interoceptive processing in particular links the brain's role in regulating bodily states to the experience of emotion, and components of minimal embodied selfhood (Palmer and Tsakiris, 2018). Early definitions distinguished interoception from exteroception as the perception of stimuli originating from inside (e.g. visceral sensations) as opposed to outside the body (e.g. vision) (Murphy *et al.*, 2017). However, more contemporary approaches highlight the role of interoceptive signals in encoding information relevant for homeostatic needs, such as cardiac function and pain, as well as olfaction, taste, and affective touch (Fotopoulou and Tsakiris, 2017). Hierarchical models of interoceptive input enable allostatic regulation (i.e. the maintenance of stability through change) of homeostasis and the prediction of physiological consequences of future behaviour. For example, lower hierarchical levels minimize interoceptive prediction error through autonomic reflexes (e.g. the recruitment of vagal stimulation to control heart rate), while higher hierarchical models that integrate interoceptive and exteroceptive information contextualize interoceptive signals (e.g. hunger) to guide intentional action (e.g. opening the refrigerator) (Ondobaka, Kilner and Friston, 2017).

Theoretical (e.g. Lange and James, 1922) and empirical perspectives (Tsakiris and Critchley, 2016) also demonstrate how emotions arise from physiological changes within the body. Indeed, researchers have mapped the subjective quality of different emotions onto unique topographies of felt bodily sensation (Nummenmaa *et al.*, 2018). Moreover, the ability to perceive interoceptive signals (e.g. such as felt heart rate) is strongly linked with emotional experience (Badoud and Tsakiris, 2017), while alexithymia has been proposed as a marker of impaired interoceptive processing (Murphy *et al.*, 2017). Higher

hierarchical levels of interoceptive inference are thought to underlie emotional content, and, speculatively, the phenomenology of embodied experience. Seth and Tsakiris (2018) propose that exteroceptive modalities, such as vision, engender a phenomenology of objecthood given their role in generating models of the world (what they call 'epistemic inference'). However, because models generated by interoceptive inference are more relevant for controlling bodily states (i.e. predicting the consequence of autonomic/motor action on homeostasis), as opposed to 'epistemically discovering more about some particular external or internal state of affairs', they may uniquely relate to the affective and phenomenal character of first-person experience (*ibid.*, p. 8):

> Instead of delivering a phenomenology of objecthood… interoceptive inference plausibly underlies a phenomenology related to the evaluation of the allostatic consequences of regulatory actions. A non-localized, non-object-based phenomenology associated with both mood and emotion, and with the pre-reflective (i.e., non-reflexive) self-related experience of being an embodied organism. (*ibid.*, p. 9)

While the relationship between interoception and the phenomenal character of conscious experience remains speculative, interoceptive processing has been empirically linked to naturalized aspects of minimal selfhood, such as sense of body ownership. For example, those with high interoceptive accuracy (i.e. the ability to accurately perceive interoceptive signals) are less susceptible to illusions of body ownership (Tsakiris, Jiménez and Constantini, 2011). Conversely, synchronizing tactile stimulation of the rubber hand (in the case of the RHI) with interoceptive signals (such as heartbeat) enhances perceived ownership of it (Palmer and Tsakiris, 2018). What these findings suggest is that interoceptive processing underlies bodily-self representation, and importantly the stability of self-boundaries.

Through theoretical insights originating in the MBA, EM extends predictive coding accounts of self by describing how interoceptive inferences underlying minimal selfhood are sculpted by embodied infant–caregiver interactions in early development. Indeed, given the fact that humans are born without a fully developed motor system, infants rely wholly on caregivers for the regulation of bodily states and the functioning of interoceptive modalities (e.g. those underlying thirst, satiation, and thermoregulation). Furthermore, because infants rely on the *social* regulation of bodily homeostasis, it follows that 'the actions of their caregivers necessarily determine how they come to experience the affective core of their embodied selfhood' (Fotopoulou

and Tsakiris, 2017, p. 18). The integration of interoceptive and sensorimotor signals into progressively sophisticated models of bodily states, emotion, and body-schema are shaped by the responses of caregivers, who (ideally) provide continuous interactions that are rich in the synchronous and contingent sensorimotor attunement that underlie the embodied basis of attachment. Thus, EM provides a theoretical expansion of the MBA to consider how alterations in embodied experience in BPD may be related to core social processes underlying the constitution of minimal selfhood in early development.

We would like to note here that, by construing minimal selfhood as the phenomenal manifestation of multisensory integration, we follow a diverse tradition of thinkers that seek to mechanistically ground embodied consciousness in neurocomputational processes (Damasio, 2005; Seth, Suzuki and Critchley, 2012; Blanke, Slater and Serino, 2015; Craig, 2003; Apps and Tsakiris, 2014; Fotopoulou, 2014; Fotopoulou and Tsakiris, 2017). The upshot of this perspective is that it incorporates a naturalized conception of minimal selfhood to include empirically tractable phenomena such as sense of body ownership, peripersonal space, and interoception processing. We acknowledge that this mechanistic approach may diverge from strictly pre-reflective, phenomenological perspectives, and future specialist theoretical and empirical papers could explore similarities and differences between these approaches. One important difference to note here is the fact that, within predictive coding accounts of embodied selfhood, the distinction between the minimal self and narrative self is a matter of degree, not of kind, in that both are understood as manifestations of hierarchically organized neural representation (Allen and Friston, 2018). Thus, our approach differs from certain phenomenological perspectives that stipulate a categorical distinction between minimal and narrative selfhood (e.g. Parnas and Henriksen, 2014).

6. Embodied Mentalization and BPD

EM extends the scope of traditional mentalization approaches to theorize not only the dynamics of facial attunement and affective mirroring in the development of self-representation, but the multitude of proximal interactions between infant and caregiver (e.g. hugging, swaddling, feeding) that shape the very experience of bodily self. Consider how Fonagy conceptualizes the impact of maltreatment on self-development in BPD:

> The internalization of the caregiver's image of the child as an intentional being is central. If this is accurate, the child's emerging self-representation will map on to what could be called a primary or constitutional self (the child's experience of an actual state of being, the self as it is)… Maltreatment and difficulty in mentalizing preclude such an organic self-image. Internal experience is not met by external understanding; it remains unlabeled and confusing, and the uncontained affect generates further dysregulation. (Fonagy, 2000, p. 1137)

By this account, self-pathology results from the failure of caregivers to adequately model and reflect their child's intentional state. However, in considering the aetiology of BPD, we want to suggest that alterations in embodied mentalization may impact 'the child's experience of an actual state of being, the self as it is'. Indeed, the very distinction between a 'self-representation' and 'constitutional self' dissolves within the MBA once we establish that mentalization sculpts not only the narrative self, but also the foundational features of minimal selfhood that arise from integrated, hierarchical cortical representations of interoceptive and exteroceptive input. In addition, EM provides a framework to consider how the embodied interactions involved in the regulation of bodily and affective states contribute to aberrant interoceptive processing that could contribute to features of self-pathology in BPD.

That embodied mentalization may be implicated in the development of borderline pathology is suggested by recent predictive coding models of attachment security that demonstrate how distinct attachment styles emerge 'from a minimization of free energy — over interoceptive states relating to internal stress levels — when seeking proximity to caregivers who have a varying impact on these interoceptive states' (Cittern *et al.*, 2018, p. 1). The theoretical role of intersubjective regulation of bodily states in attachment security is also supported by evidence that the degree of non-verbal bodily attunement between infant and caregiver distinguishes the attachment styles of infants at eight months, over and above the cognitive reflective capacities of their caregivers (Shai and Meins, 2018). BPD is predominantly characterized by preoccupied attachment (60–100%), in which parents respond to the proximity-seeking behaviours of infants inconsistently (Korzekwa, Dell and Pain, 2009). According to EM, interoceptive predictions (including those used to regulate bodily and affective states) are formed in part by the actual and expected responses of caregivers to various bodily states (Fotopoulou and Tsakiris, 2017). However, inconsistent caregiver responses, as well as

maltreatment and neglect, may affect how those very interoceptive signals are integrated into models of bodily self and how they are subjectively experienced (Murphy *et al.*, 2017).

Several lines of evidence suggest impairments of interoception in BPD. At the physiological level, people with BPD have reduced vagal regulation of heart rate, which has been linked to emotional dysregulation and the perception of emotional cues (Koenig *et al.*, 2016). Neurologically, BPD is associated with reduced cortical representations of afferent bodily signals (i.e. the neural correlates of bodily self-awareness). Alexithymia, which is common in BPD, has also been linked to low interoceptive abilities (Goodman *et al.*, 2014), dissociation, and adverse childhood experiences (New *et al.*, 2012). Of note, one study (Hart *et al.*, 2012) found no differences in interoceptive accuracy between BPD and healthy controls as measured by heartbeat detection. Löffler, Foell and Bekrater-Bodmann (2018) hypothesized that, while objective interoceptive performance is unaffected in BPD, interoceptive *awareness* is impaired, such that interoceptive signals are experienced as less trustworthy and meaningfully integrated into cortical representations of self and other.

Within a predictive coding framework, the 'untrustworthiness' of interoceptive signals can be understood as a reduction in the precision of interoceptive predictions such that, in BPD, there is less stability in hierarchical models of bodily self. Within an embodied mentalization framework, interoceptive inference is sculpted by the dynamic physicality of the infant–caregiver dyad, wherein embodied interactions with caregivers putatively modulate the precision ascribed to interoceptive input, and subsequently affect how those signals are incorporated into more complex representations of emotion and self. Importantly, the emotional attunement of those bodily interactions are shaped by the extent to which caregivers accurately mentalize their children. For example, parental misattribution of their infant's mental states is associated with less emotionally-contingent tactile behaviours in their infants during dyadic interactions (Crucianelli *et al.*, 2019). In body illusion paradigms in adults, affective congruency of visuo-tactile stimuli in the rubber hand illusion leads to greater subjective embodiment of the rubber hand, suggesting that affective congruency of sensorimotor stimuli plays a role in stabilizing sense of embodied self (Filippetti *et al*., 2019). An ably mentalizing parent responds physically to the infant in an emotionally attuned manner that allows the infant to generate accurate and precise-enough expectations about their own body. Plausibly, known risk factors for BPD, including

neglect and disorganized attachment, contribute to the aetiology of self-pathology because the inconsistent responses of caregivers compromise the certainty ascribed to the developing predictive models of interoceptive needs and the exteroceptive (including social) experiences that will satisfy them, thus introducing a fundamental uncertainty about the very foundation of minimal selfhood.

Plausibly, disordered embodied mentalization in BPD contributes to significant alterations in subjective experience. For example, depersonalization (i.e. feeling disconnected from one's body) and derealization (i.e. feelings of unreality) are common and symptomatic manifestations of stress in BPD (Korzekwa, Dell and Pain, 2009). These dissociative states reflect aberrations in 'conscious presence' wherein the content of experience appears unreal or distant. Dissociation putatively arises from imprecise interoceptive prediction signal (Seth, Suzuki and Critchley, 2012) that disrupts the dynamic correspondence between interoceptive signal and top-down predictions that support typical embodied phenomenology (see Section 5.1) (Seth and Tsakiris, 2018). According to the MBA, BPD is associated with a low stress threshold for the activation of the attachment system in lieu of more controlled, cognitive mentalizing capacities (Fonagy and Luyten, 2009). And for those with BPD, activation of a disorganized attachment system in stressful contexts may be associated with a reduction in the precision (certainty) of interoceptive predictions that give rise to dissociative experiences (Seth, 2013). This process may also be mediated by attentional processes that suppress self-awareness (Sierra and David, 2011), which reflect the learned avoidance of expected negative consequences of caregivers who responded ineffectively to emotional reactions and further dysregulated them (Shai and Fonagy, 2014).

Alterations in embodied mentalization may also contribute to interpersonal difficulties. Bodily self-awareness depends on the integration of stimuli in peripersonal space: the physical space that immediately surrounds the body (Ardizzi and Ferri, 2018) and is central for establishing self–other boundaries in embodied social contexts. BPD is associated with a two-fold increase in preferred interpersonal distance in live dyadic contexts compared to healthy controls (Fineberg *et al.*, 2018). Importantly, interoceptive accuracy predicts a more narrow subjective sense of peripersonal space (Ardizzi and Ferri, 2018). In BPD, aberrations in interoceptive inference may result in the relative overweighting of exteroceptive signals. This overweighting also underlies increased plasticity of body ownership in illusory body

ownership paradigms (Palmer and Tsakiris, 2018). Müller *et al.* (2015) suggest that, developmentally, 'parental invalidation, insecurity, and abuse in childhood may result in an individual with BPD focusing on external cues instead of perceiving, integrating, and interpreting his or her own bodily signals and emotions' (p. 1082). Thus, people with BPD may prefer greater interpersonal distance in an effort to maintain self–other boundaries in the setting of a porous and unstable self-phenomenology.

Alterations in embodied mentalization may also contribute to deficits in empathy and theory of mind in BPD, and are traditionally understood in the MBA to result from impairments in the ability to model intentional states (Fonagy and Luyten, 2009). The capacity to correctly recognize emotions in others is related to the ability to identify interoceptive states in oneself. As Palmer and Tsakiris (2018) observe, 'understanding others… require[s] at least a "good enough" representation of one's own (interoceptive) state because the key element in representing other's states is how their states affect us' (p. 23). However, for people with BPD, who are more sensitive to the experience of others' emotions (Lynch *et al.*, 2006) and more susceptible to emotional contagion (Gunderson *et al.*, 2018), uncertainty regarding the origins of bodily and affective signals may contribute to the dissolution of self–other boundaries in their affective experience of the other as overwhelming and undifferentiated.

6.1. Remaining questions and future directions

While the role of embodied mentalization in the aetiology of BPD has yet to be empirically confirmed, EM provides a promising framework to investigate the bodily foundations of disordered selfhood and sociality in the disorder. For example, to clarify how early embodied interactions shape interoceptive processing, future research would be needed to examine the longitudinal association between bodily attunement during infant–caregiver interactions, measures of attachment style, and developing interoceptive sensitivity in infancy. Furthermore, investigations of embodiment in social and non-social paradigms should include self-report and interview measures, along with varied indices of interoceptive processing across interoceptive domains, to more precisely characterize how interoception mediates empirically measured and phenomenologically reported bodily disturbance in BPD.

Considering EM as an extension of the MBA also requires further clarification. Within the MBA, mentalization is not a unitary construct, nor is it considered as a stable trait: mentalizing abilities are thought to vary across time, relationships, and under periods of stress. Regarding BPD, future empirical work should consider how the stability of interoceptive inferences fluctuates as a function of stress and interpersonal context. In interrogating the implicit concept of self within the MBA, it was found that mentalization has implications for both narrative and minimal aspects of selfhood. Further work is needed to clarify how alterations in minimal embodiment are linked to identity disturbance, unstable goals and values, and fluctuating evaluations of self and other in BPD. For example, differences may be observed in children of parents that are somewhat more able to provide embodied mirroring without marking and vice versa, or parents who create environments of high volatility. In such environments, children may not only have difficulties understanding their own bodily and social needs, but they will particularly struggle to respond to unexpected changes in bodily or environmental conditions. That narrative identity is affected by alterations in minimal selfhood in BPD is plausible: inducing out-of-body illusions (in which experiences are had from a third-person rather than first-person perspective) in individuals engaged in social interactions impairs their ability to encode those interactions in memory (Bergouignan, Nyberg and Ehrsson, 2014). Thus, the relationship between dissociative experiences, impaired autobiographical memory, and incoherence of narrative identity in BPD may result from fundamental alterations in first-person experience during those episodes. As such, mentalization theory is consistent with a 'pattern theory of self' (Gallagher, 2013), whereby different aspects self (from minimal embodiment to identity) are thought to be related and interdependent. However, these relations require further elucidation.

Alterations of minimal selfhood have been implicated in several conditions that share clinical and symptomatic overlap with BPD, including schizophrenia (Parnas and Henriksen, 2014), post-traumatic stress disorder (Ataria, 2014), and depression (Thönes and Oberfeld, 2015). Further research is needed to clarify how the severity, quality, and aetiology of minimal-self disturbance vary across different diagnostic syndromes. For example, it is thought that schizophrenia is associated with more profound and extensive alterations of self-experience than BPD (Parnas and Henriksen, 2014). Moreover, the mechanistic bases of self-disturbance in depression, such as those

related to time perception (Thönes and Oberfeld, 2015) and sense of agency (Ratcliffe, 2013), are likely different from the early developmental processes described here. Validated phenomenological tools, such as the Examination of the Anomalous Self Experience (EASE) Scale (Parnas *et al.*, 2005), can help clarify phenomenological disturbances associated with psychopathology. Preliminary findings on the relationship between EASE ratings and borderline symptomology are mixed: one study found a correlation between the DSM-V BPD criterion 'identity disturbance' and mean total EASE score (Zandersen and Parnas, 2019), while another found no correlation between EASE score and BPD symptoms (Nelson *et al.*, 2013). However, these studies did not include samples of individuals diagnostically confirmed with BPD according to DSM criteria. Furthermore, interview scales and self-report measures may not capture subtle alterations of embodiment such as malleability of body ownership or enlargement of peripersonal space (or subjective time perception in the case of depression), which may not be conscious to the individual as such, but can be provoked under experimental conditions. Importantly, these alterations have meaningful experiential and clinical sequelae. Thus, investigations of alterations of selfhood in BPD and in psychopathology in general will benefit from a multimodal approach that synthesizes findings across different levels of analysis (e.g. phenomenological, behavioural, and neurobiological).

7. Conclusion

Borderline personality disorder is a severe mental illness associated with profound disturbances of selfhood that range from fragmentations in narrative identity to fundamental alterations in experiential embodiment. The mentalization-based approach (MBA) locates the origins of BPD in the interpersonal development of self-structure and the capacity to mentalize in early infancy. However, because traditional mentalization theory posits that impaired mentalization develops when formed self-representations misalign with the 'constitutional' experience of self, it fails to account for disturbances of minimal embodied selfhood in BPD. To address this limitation, we proposed the hypothesis that self-disturbance in BPD manifests foundationally as an alteration in 'embodied mentalization', i.e. the progressive integration of interoceptive and sensorimotor signals into relatively robust generative models of selfhood that underlie first-person embodied experience. The upshot of this proposal is that it

coheres mentalization theory to contemporary accounts of brain function, and links the development of the self to the necessary interpersonal regulation of bodily states during development. Additionally, it provides novel testable hypotheses regarding the role of interoception in the co-occurrence of self and social pathology in BPD.

Acknowledgments

E.S.N. was supported by the Yale School of Medicine Medical Student Research Fellowship. A.F. was supported first by a European Research Council (ERC) Starting Grant ERC-2012-STG GA313755 from the project 'Bodily Self' and then a Consolidator Grant ERC-2018-COG818070 for the project METABODY. S.K.F. was supported by the American Foundation for Suicide Prevention, young investigator grant ID: Y1G-1-045-16. This work was also funded in part by the State of Connecticut, Department of Mental Health and Addiction Services. This publication does not express the views of the Department of Mental Health and Addiction Services or the State of Connecticut. The views and opinions expressed are those of the authors. **Declaration of conflict of interests:** The authors declare that there is no conflict of interest.

References

Adler, J.M., Chin, E.D., Kolisetty, A.P. & Oltmanns, T.F. (2012) The distinguishing characteristics of narrative identity in adults with features of borderline personality disorder: An empirical investigation, *Journal of Personality Disorders*, **26**, pp. 498–512.

Allen, J.G. & Fonagy, P. (2006) *The Handbook of Mentalization-Based Treatment*, Hoboken, NJ: John Wiley & Sons.

Allen, M. & Friston, K.J. (2018) From cognitivism to autopoiesis: Towards a computational framework for the embodied mind, *Synthese*, **195**, pp. 2459–2482.

Apps, M.A.J. & Tsakiris, M. (2014) The free-energy self: A predictive coding account of self-recognition, *Neuroscience & Biobehavioral Reviews*, **41**, pp. 85–97.

Ardizzi, M. & Ferri, F. (2018) Interoceptive influences on peripersonal space boundary, *Cognition*, **177**, pp. 79–86.

Ataria, Y. (2014) Acute peritraumatic dissociation: In favor of a phenomenological inquiry, *Journal of Trauma & Dissociation*, **15**, pp. 332–347.

Badoud, D. & Tsakiris, M. (2017) From the body's viscera to the body's image: Is there a link between interoception and body image concerns?, *Neuroscience & Biobehavioral Reviews*, **77**, pp. 237–246.

Bateman, A. (2006) *Mentalization-Based Treatment For Borderline Personality Disorder: A Practical Guide*, Oxford: Oxford University Press.

Becker-Sadzio, J. (2019) *Die Somatosensorische Wahrnehmung Taktiler Vibrationsreize Und Ihre Beeinflussbarkeit Durch Die Rubber Hand Illusion Bei Patienten Mit Emotional Instabiler Persönlichkeitsstörung Vom Borderline-Typ*, PhD thesis, University of Tubingen, Germany.

Beebe, B., Messinger, D., Bahrick, L.E., Margolis, A., Buck, K.A. & Chen, H. (2016) A systems view of mother–infant face-to-face communication, *Developmental Psychology*, **52**, pp. 556–571.

Beeney, J.E., Hallquist, M.N., Ellison, W.D. & Levy, K.N. (2016) Self–other disturbance in borderline personality disorder: Neural, self-report, and performance-based evidence, *Personality Disorders*, **7**, pp. 28–39.

Bekrater-Bodmann, R., Chung, B.Y., Foell, J., Gescher, D.M., Bohus, M. & Flor, H. (2016) Body plasticity in borderline personality disorder: A link to dissociation, *Comprehensive Psychiatry*, **69**, pp. 36–44.

Bender, D.S. & Skodol, A.E. (2007) Borderline personality as a self–other representational disturbance, *Journal of Personality Disorders*, **21**, pp. 500–517.

Bergouignan, L., Nyberg, L. & Ehrsson, H.H. (2014) Out-of-body-induced hippocampal amnesia, *Proceedings of the National Academy of Sciences USA*, **111**, 4421.

Blanke, O., Slater, M. & Serino, A. (2015) Behavioral, neural, and computational principles of bodily self-consciousness, *Neuron*, **88**, pp. 145–166.

Bolis, D. & Schilbach, L. (2018) 'I interact therefore I am': The self as a historical product of dialectical attunement, *Topoi*, **39**, pp. 521–534.

Botvinick, M. & Cohen, J. (1998) Rubber hands 'feel' touch that eyes see, *Nature*, **391**, art. 756.

Bowlby, J. (2008) *Attachment*, New York: Basic Books.

Brown, M.Z., Comtois, K.A. & Linehan, M.M. (2002) Reasons for suicide attempts and nonsuicidal self-injury in women with borderline personality disorder, *Journal of Abnormal Psychology*, **111**, pp. 198–200.

Ciaunica, A. & Fotopoulou, A. (2017) The touched self: Psychological and philosophical perspectives on proximal intersubjectivity and the self, in Durt, C., Fuchs, T. & Tewes, C.(eds.) *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*, Cambridge, MA: MIT Press.

Cittern, D., Nolte, T., Friston, K. & Edalat, A. (2018) Intrinsic and extrinsic motivators of attachment under active inference, *PLoS One*, **13**, E0193955.

Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science, *Behavioral and Brain Sciences*, **36**, pp. 181–204.

Clark, J.E., Watson, S. & Friston, K.J. (2018) What is mood? A computational perspective, *Psychological Medicine*, **48**, pp. 2277–2284.

Craig, A.D. (2003) Interoception: The sense of the physiological condition of the body, *Current Opinion in Neurobiology*, **13**, pp. 500–505.

Crowell, S.E., Beauchaine, T.P. & Linehan, M.M. (2009) A biosocial developmental model of borderline personality: Elaborating and extending Linehan's theory, *Psychological Bulletin*, **135**, pp. 495–510.

Crucianelli, L., Wheatley, L., Filippetti, M.L., Jenkinson, P.M., Kirk, E. & Fotopoulou, A. (2019) The mindedness of maternal touch: An investigation of maternal mind-mindedness and mother–infant touch interactions, *Developmental Cognitive Neuroscience*, **35**, pp. 47–56.

Damasio, A. (2005) *Descartes' Error: Emotion, Reason, and the Human Brain*, London: Penguin.

Dennett, D.C. (2017) *Consciousness Explained*, Boston, MA: Little, Brown.

Duffy, C. (2017) What saved my life from borderline personality disorder, *TheMighty.com*, [Online], https://themighty.com/2017/07/My-Experience-Borderline-Personality-Disorder/ [12 November 2018].

Dyer, A., Borgmann, E., Feldmann, R.E., Kleindienst, N., Priebe, K., Bohus, M. & Vocks, S. (2013) Body image disturbance in patients with borderline personality disorder: Impact of eating disorders and perceived childhood sexual abuse, *Body Image*, **10**, pp. 220–225.

Ensink, K., Normandin, L., Target, M., Fonagy, P., Sabourin, S. & Berthelot, N. (2015) Mentalization in children and mothers in the context of trauma: An initial study of the validity of the child reflective functioning scale, *British Journal of Developmental Psychology*, **33**, pp. 203–217.

Eshkevari, E., Rieger, E., Longo, M.R., Haggard, P. & Treasure, J. (2012) Increased plasticity of the bodily self in eating disorders, *Psychological Medicine*, **42**, pp. 819–828.

Filippetti, M.L., Kirsch, L.P., Crucianelli, L. & Fotopoulou, A. (2019) Affective certainty and congruency of touch modulate the experience of the rubber hand illusion, *Scientific Reports*, **9**, 2635.

Fineberg, S.K., Leavitt, J., Landry, C.D., Neustadter, E.S., Lesser, R.E., Stahl, D.S., Deutsch-Link, S. & Corlett, P.R. (2018) Individuals with borderline personality disorder show larger preferred social distance in live dyadic interactions, *Psychiatry Research*, **260**, pp. 384–390.

Fonagy, P. (2000) Attachment and borderline personality disorder, *Journal of the American Psychoanalytic Association*, **48**, pp. 1129–1146.

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

Fonagy, P., Gergely, G., Jurist, E.L. & Target, M. (2002) *Affect Regulation, Mentalization, and the Development of the Self*, New York: Other Press.

Fonagy, P. & Luyten, P. (2009) A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder, *Development and Psychopathology*, **21**, pp. 1355–1381.

Fotopoulou, A. (2014) Time to get rid of the 'modular' in neuropsychology: A unified theory of anosognosia as aberrant predictive coding, *Journal of Neuropsychology*, **8**, pp. 1–19.

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

Friston, K. (2010) The free-energy principle: A unified brain theory?, *Nature Reviews Neuroscience*, **11**, art. 127.

Fuchs, T. (2007) Fragmented selves: Temporality and identity in borderline personality disorder, *Psychopathology*, **40**, pp. 379–387.

Gallagher, S. (2000) Philosophical conceptions of the self: Implications for cognitive science, *Trends in Cognitive Sciences*, **4**, pp. 14–21.

Gallagher, S. (2013) A pattern theory of self, *Frontiers in Human Neuroscience*, **7**, pp. 1–7.

Gold, N. & Kyratsous, M. (2017) Self and identity in borderline personality disorder: Agency and mental time travel, *Journal of Evaluation in Clinical Practice*, **23**, pp. 1020–1028.

Goodman, M., Carpenter, D., Tang, C.Y., Goldstein, K.E., Avedon, J., Fernandez, N., Mascitelli, K.A., Blair, N.J., New, A.S., Triebwasser, J., Siever, L.J. & Hazlett, E.A. (2014) Dialectical behavior therapy alters emotion regulation and

amygdala activity in patients with borderline personality disorder, *Journal of Psychiatric Research*, **57**, pp. 108–116.
Gunderson, J.G. & Lyons-Ruth, K. (2008) BPD's interpersonal hypersensitivity phenotype: A gene-environment-developmental model, *Journal of Personality Disorders*, **22**, pp. 22–41.
Gunderson, J.G., Herpertz, S.C., Skodol, A.E., Torgersen, S. & Zanarini, M.C. (2018) Borderline personality disorder, *Nature Reviews Disease Primers*, **4**, 18029.
Hart, N., Mcgowan, J., Minati, L. & Critchley, H.D. (2012) Emotional regulation and bodily sensation: Interoceptive awareness is intact in borderline personality disorder, *Journal of Personality Disorders*, **27**, pp. 506–518.
Higgins, J. (2018) The 'we' in 'me': An account of minimal relational selfhood, *Topoi*, **39**, pp. 535–546.
Jørgensen, C.R. (2006) Disturbed sense of identity in borderline personality disorder, *Journal of Personality Disorders*, **20**, pp. 618–644.
Kaplan, R.A., Enticott, P.G., Hohwy, J., Castle, D.J. & Rossell, S.L. (2014) Is body dysmorphic disorder associated with abnormal bodily self-awareness? A study using the rubber hand illusion, *PLoS ONE*, **9**, art. e99981.
Kernberg, O.F. (1985) *Borderline Conditions and Pathological Narcissism*, Lanham, MD: J. Aronson.
Kernberg, O.F. (2006) Identity: Recent findings and clinical implications, *The Psychoanalytic Quarterly*, **75**, pp. 969–1004.
Kerr, I.B., Finlayson-Short, L., Mccutcheon, L.K., Beard, H. & Chanen, A.M. (2015) The 'self' and borderline personality disorder: Conceptual and clinical considerations, *Psychopathology*, **48**, pp. 339–348.
Klein, M. (1946) Notes on some schizoid mechanisms, *The International Journal of Psychoanalysis*, **27**, pp. 99–110.
Kleindienst, N., Löffler, A., Herzig, M., Bertsch, K. & Bekrater-Bodmann, R. (2020) Evaluation of the own body in women with current and remitted borderline personality disorder: Evidence for long-lasting effects of childhood sexual abuse, *European Journal of Psychotraumatology*, **11**, pp. 264–277.
Koenig, J., Kemp, A.H., Feeling, N.R., Thayer, J.F. & Kaess, M. (2016) Resting state vagal tone in borderline personality disorder: A meta-analysis, *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, **64**, pp. 18–26.
Korzekwa, M.I., Dell, P.F. & Pain, C. (2009) Dissociation and borderline personality disorder: An update for clinicians, *Current Psychiatry Reports*, **11**, pp. 82–88.
Kyselo, M. (2014) The body social: An enactive approach to the self, *Frontiers in Psychology*, **5**, art. 986.
Lange, C.G. & James, W. (1922) *The Emotions*, Baltimore, MD: Williams & Wilkins.
Leichsenring, F., Leibing, E., Kruse, J., New, A.S. & Leweke, F. (2011) Borderline personality disorder, *The Lancet*, **377**, pp. 74–84.
Linehan, M. (1993) *Cognitive-Behavioral Treatment of Borderline Personality Disorder*, New York: Guilford Press.
Löffler, A., Foell, J. & Bekrater-Bodmann, R. (2018) Interoception and its interaction with self, other, and emotion processing: Implications for the understanding of psychosocial deficits in borderline personality disorder, *Current Psychiatry Reports*, **20**, art. 28.

Luyten, P., Mayes, L.C., Nijssens, L. & Fonagy, P. (2017) The parental reflective functioning questionnaire: Development and preliminary validation, *PLoS One*, **12**, E0176218.

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

Mae, K. (2017) Why BPD makes it hard for me to be around other people's emotions, *TheMighty.com*, [Online], https://themighty.com/2017/08/Borderline-Personality-Disorder-bpd-Identity-Emotions/.

Metzinger, T. (2005) Precis of Being No One, *Psyche: An Interdisciplinary Journal of Research on Consciousness*, **11**, pp. 1–35.

Möller, T.J., Braun, N., Thöne, A.-K., Herrmann, C.S. & Philipsen, A. (2020) The senses of agency and ownership in patients with borderline personality disorder, *Frontiers in Psychiatry*, **11**, art. 474.

Müller, L.E., Schulz, A., Andermann, M., Gäbel, A., Gescher, D.M., Spohn, A., Herpertz, S.C. & Bertsch, K. (2015) Cortical representation of afferent bodily signals in borderline personality disorder: Neural correlates and relationship to emotional dysregulation, *JAMA Psychiatry*, **72**, pp. 1077–1086.

Murphy, J., Brewer, R., Catmur, C. & Bird, G. (2017) Interoception and psychopathology: A developmental neuroscience perspective, *Developmental Cognitive Neuroscience*, **23**, pp. 45–56.

Murphy, J., Catmur, C. & Bird, G. (2018) Alexithymia is associated with a multidomain, multidimensional failure of interoception: Evidence from novel tests, *Journal of Experimental Psychology: General*, **147**, pp. 398–408.

Nelson, B., Thompson, A., Chanen, A.M., Amminger, G.P. & Yung, A.R. (2013) Is basic self-disturbance in ultra-high risk for psychosis ('prodromal') patients associated with borderline personality pathology?, *Early Intervention in Psychiatry*, **7**, pp. 306–310.

Neustadter, E.S., Fineberg, S.K., Leavitt, J., Carr, M.M. & Corlett, P.R. (2019) Induced illusory body ownership in borderline personality disorder, *Neuroscience of Consciousness*, **5**, art. 17.

New, A.S., Rot, M.A.H., Ripoll, L.H., Perez-Rodriguez, M.M., Lazarus, S., Zipursky, E., Weinstein, S.R., Koenigsberg, H.W., Hazlett, E.A., Goodman, M. & Siever, L.J. (2012) Empathy and alexithymia in borderline personality disorder: Clinical and laboratory measures, *Journal of Personality Disorders*, **26**, pp. 660–675.

Nummenmaa, L., Hari, R., Hietanen, J.K. & Glerean, E. (2018) Maps of subjective feelings, *Proceedings of the National Academy of Sciences USA*, **115**, 9198.

Ondobaka, S., Kilner, J. & Friston, K. (2017) The role of interoceptive inference in theory of mind, *Brain and Cognition*, **112**, pp. 64–68.

Palmer, C.E. & Tsakiris, M. (2018) Going at the heart of social cognition: Is there a role for interoception in self-other distinction?, *Current Opinion in Psychology*, **24**, pp. 21–26.

Parnas, J., Møller, P., Kircher, T., Thalbitzer, J., Jansson, L., Handest, P. & Zahavi, D. (2005) EASE-Scale (Examination of Anomalous Self-Experience), *Psychopathology*, **38**, pp. 236–258.

Parnas, J. & Henriksen, M.G. (2014) Disordered self in the schizophrenia spectrum: A clinical and research perspective, *Harvard Review of Psychiatry*, **22**, pp. 251–265.

Prebble, S.C. (2013) Autobiographical memory and sense of self, *Psychological Bulletin*, **139**, pp. 815–840.

Ratcliffe, M. (2013) Depression and the phenomenology of free will, in Fulford, K.W.M., Davies, M., Gipps, R.G.T., Graham, G., Stanghellini, G. & Thornton, T. (eds.) *The Oxford Handbook of Philosophy and Psychiatry*, pp. 574–591, Oxford: Oxford University Press.

Schmahl, C., Meinzer, M., Zeuch, A., Fichter, M., Cebulla, M., Kleindienst, N., Ludäscher, P., Steil, R. & Bohus, M. (2010) Pain sensitivity is reduced in borderline personality disorder, but not in posttraumatic stress disorder and bulimia nervosa, *The World Journal of Biological Psychiatry*, **11**, pp. 364–371.

Seth, A.K. (2013) Interoceptive inference, emotion, and the embodied self, *Trends in Cognitive Sciences*, **17**, pp. 565–573.

Seth, A.K., Suzuki, K. & Critchley, H. (2012) An interoceptive predictive coding model of conscious presence, *Frontiers in Psychology*, **2**, art. 395.

Seth, A.K. & Tsakiris, M. (2018) Being a beast machine: The somatic basis of selfhood, *Trends in Cognitive Sciences*, **22**, pp. 969–981.

Shai, D. & Fonagy, P. (2014) Beyond words: Parental embodied mentalizing and the parent–infant dance, in Mikulincer, M.S. (ed.) *Mechanisms of Social Connection: From Brain to Group*, Washington, DC: American Psychological Association.

Shai, D. & Meins, E. (2018) Parental embodied mentalizing and its relation to mind-mindedness, sensitivity, and attachment security, *Infancy*, **23**, pp. 857–872.

Sharp, C. & Vanwoerden, S. (2015) Hypermentalizing in borderline personality disorder: A model and data, *Journal of Infant, Child & Adolescent Psychotherapy*, **14**, pp. 33–45.

Sierra, M. & David, A.S. (2011) Depersonalization: A selective impairment of self-awareness, *Consciousness and Cognition*, **20**, pp. 99–108.

Slade, A. (2005) Parental reflective functioning: An introduction, *Attachment & Human Development*, **7**, pp. 269–281.

Startup, M., Heard, H., Swales, M., Jones, B., Williams, J.M.G. & Jones, R.S.P. (2001) Autobiographical memory and parasuicide in borderline personality disorder, *British Journal of Clinical Psychology*, **40**, art. 113.

Strawson, G. (1997) The self, *Journal of Consciousness Studies*, **4** (5–6), pp. 405–428.

Thakkar, K.N., Nichols, H.S., McIntosh, L.G. & Park, S. (2011) Disturbances in body ownership in schizophrenia: Evidence from the rubber hand illusion and case study of a spontaneous out-of-body experience, *PLoS ONE*, **6**, art. e27089

Thönes, S. & Oberfeld, D. (2015) Time perception in depression: A meta-analysis, *Journal of Affective Disorders*, **175**, pp. 359–372.

Tsakiris, M., Jiménez, A.T. & Costantini, M. (2011) Just a heartbeat away from one's body: Interoceptive sensitivity predicts malleability of body-representations, *Proceedings of the Royal Society B: Biological Sciences*, **278**, pp. 2470–2476.

Tsakiris, M. & Critchley, H. (2016) Interoception beyond homeostasis: Affect, cognition and mental health, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **371**, 20160002.

Widom, C.S., Czaja, S.J. & Paris, J. (2009) A prospective investigation of borderline personality disorder in abused and neglected children followed up into adulthood, *Journal of Personality Disorders*, **23**, pp. 433–446.

Zahavi, D. (2005) *Subjectivity and Selfhood: Investigating the First-Person Perspective*, Cambridge, MA: MIT Press.
Zahavi, D. (2014) *Self and Other: Exploring Subjectivity, Empathy, and Shame*, Oxford: Oxford University Press.
Zandersen, M. & Parnas, J. (2019) Exploring schizophrenia spectrum psychopathology in borderline personality disorder, *European Archives of Psychiatry and Clinical Neuroscience*, **270**, pp. 969–978.
Zanarini, M.C., Ruser, T., Frankenburg, F.R. & Hennen, J. (2000) The dissociative experiences of borderline patients, *Comprehensive Psychiatry*, **41**, pp. 223–227.

Paper received March 2020; revised July 2020.