

Significant Development Changes in Borderline Personality Disorder



What we have changed our minds about:

Part 1. Borderline personality disorder as a limitation of resilience

Abstract

This paper sets out a recent transition in our thinking in relation to psychopathology associated with personality disorder, in an approach that integrates our thinking about attachment, mentalizing (understanding ourselves and others in terms of intentional mental states) and epistemic trust (openness to the reception of social communication that is personally relevant and of generalizable significance) with recent findings on the structure of both adult and child psychopathology and resilience. In this paper – the first of two parts – we review evidence suggesting that a general psychopathology or p factor underlies vulnerability for psychopathology. We link this p factor to a lack of resilience using Kalisch and colleagues' positive appraisal style theory of resilience (PASTOR). We argue that vulnerability for (severe) psychopathology results from impairments in three central mechanisms underlying resilience – positive situation classification, retrospective reappraisal of threat, and inhibition of retraumatizing triggers – which in turn result from a lack of flexibility in terms of social communicative processes. We suggest that, from this perspective, personality disorders, and borderline personality disorder (BPD) in particular, can be considered to be the prototype of disorders characterized by a lack of resilience. Part 2 proposes an evolutionary developmental psychopathology account linking this inflexibility in social communication to problems with the development of epistemic trust – that is, an evolutionary pre-wired social communication system that normally facilitates resilience through salutogenesis, that is, the capacity to learn and derive benefit from the (social) environment.

Keywords: Borderline personality disorder, Resilience, Epistemic trust, Mentalizing, Attachment, Psychopathology

Background

A challenge for contemporary thinking about psychopathology arises from a general neglect by adult psychopathologists of the developmental psychopathology tradition established by Sroufe and Rutter [1] over 30 years ago. Specifically, the fact that when we consider an individual's psychiatric history over their life course, it rarely follows the discrete, symptom-led and time-limited categories that traditional models have used in conceptualizing mental disorder. This has increasingly come to be regarded as constituting a slow-burning crisis in the way we understand, and by extension treat, mental disorders. There is a heightened recognition of the salience of transdiagnostic features in

clinical presentations as well as across treatment protocols [2, 3]. Particularly in cases of more severe and persistent mental health difficulties, an individual's clinical presentation changes over time, one typical example being progression from conduct disorder to depression [4], or the extensive comorbidity between traditional 'symptom' disorders and personality disorders (PDs) (e.g. [5]).

Here we posit a reconceptualization of psychopathology associated with PD that speaks to these conceptual and diagnostic enigmas, in an approach that integrates our thinking about mentalizing (i.e. understanding ourselves and others in terms of intentional mental states) and epistemic trust (i.e. openness to the reception of social communication that is personally relevant and of generalizable significance) with recent findings on the structure of both adult and child psychopathology [3] and resilience [6].

At the core of the thinking set out here is an emphasis on the relationship between the social environment as a system on the one hand, and individual differences in the capacity for social cognition (as defined below) on the other. We argue that the presence or absence of resilience is the outcome of the dynamics of this relationship. Understanding the nature of resilience, we suggest, requires engagement at the level of the mechanism that channels the relationship between the social layer of communication and the individual's capacity for reorganizing mental processes. Attempts to intervene at the level of non-resilient responses, we suggest, can be of only limited effectiveness. This, we argue, explains the lack of clinical response of patients with BPD features to many traditional psychotherapeutic interventions.

A further informing principle is that the type of functioning associated with many forms of psychopathology might best be understood as an evolutionarily driven form of entrenched adaptation to stimuli from the social environment – often in interaction with genetic propensity [7] – rather than as a mere deficit. It is this adaptive imperative that underpins the enduring quality that is central to definitions of PD. The 'borderline mind', and related severe problems with social communication typically observed in what we commonly refer to as 'personality pathology', may therefore best be understood as a socially triggered outcome, a learned expectation about cultural context. Hence, while the processes we describe in this paper may be implicated in most, if not all, types of psychopathology, we consider severe PD, and BPD in particular, to be prototypical of the type of social communication problems that we now see as lying at the root of vulnerability for severe psychopathology.

Finally, in terms of clinical implications, we will indicate how this change in perspective drives a shift in clinical focus beyond the consulting room to the wider social systems that can promote resilience.

In the first part of this paper we review emerging evidence that a general psychopathology (or 'p') factor underlying psychopathology provides a comprehensive explanation for the extensive comorbidity among disorders, as well as many of the other features of individuals who we traditionally consider to be 'hard to reach'. We then argue that this p factor should not be primarily seen in terms of the presence of specific vulnerability factors (although these may well play an important role, and may be primarily responsible for the phenomenological heterogeneity observed among and within disorders), but in terms of the absence of resilience. We outline the recently formulated comprehensive positive appraisal style theory of resilience (PASTOR), and apply it to BPD as the prototype of disorders characterized by the absence of resilience. We argue that the absence of

resilience in BPD results from an inflexibility in the human capacity for social communication, and in problems with recalibrating the mind in the face of adverse experiences in interaction with others in particular.

In the second part of this paper, we will relate this lack of social communicative flexibility to impairments in epistemic trust from an evolutionary and developmental psychopathology perspective, and discuss the clinical implications of this shift in our views.

A general factor in psychopathology

Our starting point is the challenge presented to the traditional taxonomic structure of psychopathology by comorbidity (concurrent and sequential over time), recurrence and the unwieldy proliferation of diagnostic disorders. In our opinion, this challenge has been compellingly met by the suggestion that there is a general factor of psychopathology – in the words of Caspi and colleagues, 'one underlying dimension that summarized individuals' propensity to develop any and all forms of common psychopathologies' ([3], p. 131). In their analysis of the Dunedin longitudinal study, Caspi et al. examined the structure of psychopathology from adolescence to mid-life, considering dimensionality, persistence, co-occurrence and sequential comorbidity. They found that vulnerability to mental disorder was more convincingly described by one general psychopathology factor – labelled the 'p' (for pathology) factor – than by three high-order (spectral) factors (internalizing, externalizing and thought disorder). A higher p factor score was associated with 'more life impairment, greater familiarity, worse developmental histories, and more compromised early-life brain function' ([3], p. 131). In the meantime, several studies have replicated this higher-order p factor [8–11]. Importantly, the p factor concept may thus also explain why discovering isolated causes, consequences or biomarkers and specific, tailored treatments for psychiatric disorders has proved so elusive for the field [3].

This work on a general factor of psychopathology has recently also been extended to childhood and adolescence. A longitudinal study of 2450 girls aged 5–11 years, for instance, has further indicated the criterion validity of the p factor construct, and found it a significantly better fit than a correlated two-factor (internalizing and externalizing) model [9]. These findings weaken the argument that the p factor is a statistical artefact and reinforce the importance of further consideration of what the p factor might substantively represent [9]. In a large ($n = 23,477$) community-based sample aged 11–13.5 years, Patalay et al. investigated the traditional two-factor (internalizing and externalizing) model and a bi-factor model with a general psychopathology higher-order model [12]. Both models were found to fit the data well; however,

the general psychopathology however better predicted future psychopathology and academic attainment 3 years from the time of original assessment; with individuals with high *p* scores being 10 times as likely to have diagnosable disorder 3 years from assessment than individuals with lower *p* scores (see also [8]).

More specifically in relation to PDs, Sharp and colleagues have considered the question of whether a general factor for psychopathology exists in the context of PD diagnosis [13]. In a series of exploratory factor analyses based on a sample of 966 inpatients, only four of the six PDs (avoidant, schizotypal, narcissistic, and antisocial) examined formed factors with 75% of the criteria that mark their respective factors. Half the obsessive-compulsive PD criteria loaded with the narcissistic PD criteria, and the other half split across two other factors. However, Sharp et al. found that (a) a BPD factor included primary loadings from just over half (55.6%) of the BPD items, of which three had notable cross-loadings, each on a different factor; (b) nearly half (44.4%) of BPD items loaded most strongly on three non-BPD factors (although two had notable cross-loadings on the BPD factor); and (c) the BPD factor was also marked by a narcissistic PD item and had notable additional cross-loadings by other narcissistic as well as avoidant and schizotypal PD items. In the same study, Sharp et al. evaluated a bi-factor model of PD pathology in which a general factor and several specific factors of personality pathology account for the covariance among PD criteria. In the bi-factor model, it was found that all BPD criteria loaded only on to the general factor. Other PDs loaded either on to both the general and a specific factor or largely only on to a specific factor. The implication of this is that BPD criteria may capture the core of personality pathology, or may be most representative of all PDs. To compound more widely the salient status of BP traits, Caspi et al., in their work on the *p* factor, found that in terms of personality information, individuals who scored highly on the general psychopathology scale were characterized by 'three traits that compromise processes by which people maintain stability – low Agreeableness, low Conscientiousness and high Neuroticism; that is, high-*p* individuals experience difficulties in regulation/control when dealing with others, the environment and the self' ([3], p. 131). Such a profile appears to capture the core features of BPD – emotion dysregulation, impulsivity and social dysfunction – and speaks to trait profile approaches to PD [14]. Yet, to claim that such a profile in itself in some sense explains the developmental and life-course forecast that comes from '*p*' obviously risks approaching circularity.

The question that then remains is: what is the meaning of the general psychopathology factor at the level of mental mechanisms? Currently, we can only speculate

about the nature of this generic aetiological factor, but one association to be investigated may be childhood maltreatment. Studies indeed suggest that maltreatment, like *p*, increases the chance of most types of mental illness in adulthood [15] and worsens the course of mental illness [16]. It has been recently suggested that childhood maltreatment may be an ecophenotype associated with an earlier age at onset of psychopathology, greater symptom severity, higher levels of comorbidity, greater risk for suicide and, importantly, a poorer response to treatment [17].

In our opinion, research findings on maltreatment, although still too narrow, indeed point the way to understanding some of the mechanisms underlying the association between the *p*-factor and vulnerability to (severe) psychopathology. This emphasis on the role of adversity should not be associated with a narrowly environmental position on the relationship between adversity and BPD. Such a position would stand counter to growing evidence for a genetic determinant of BPD. Research showing the familial nature of BPD [18, 19], and classical twin studies that place heritability of BPD at around 40–50% [20–23], have been borne out further by more complex behaviour–genetic models that take into account siblings, spouses and twins [24]. Although a genetic anomaly associated with BPD has not so far been identified, it appears that an endophenotype for the disorder may be recognized. For example, impulsive aggression and suicidal behaviour have been linked to the tryptophan hydroxylase (TPH) gene, and patients with BPD have a higher frequency of two out of eight polymorphisms in one of the two known isoforms of the TPH gene [25].

Impulsive aggression has also been connected with reduced serotonergic responsiveness and the inefficient (short or '*s*') allele of 5-HTTLPR. This has been identified in patients with BPD [26] in some but not all accounts (e.g. [27]). There are suggestions that the *s* allele marks a vulnerability to stressful life events [28] on the one hand, and the positive influence of maternal sensitivity [29] on the other. Accumulating evidence supports the view that the *s* allele, in combination with secure attachment, increases agreeable yet autonomous social behaviour in adolescents [30]. In the context of attachment insecurity, this polymorphism is linked with poor self-regulation [31] and impulsiveness [30]. The implication may be that the *s* allele increases social sensitivity, making a child both more and less prosocial in response to different environmental stimuli.

Furthermore, the methylation of certain genes could mediate the long-term effects of adversity [32]. The glucocorticoid receptor gene promoter, for instance, has been shown to be more methylated in samples of brain tissue of individuals who had experienced adversity and suicide

[33]. The methylation of NR3C1 is associated with severity of maltreatment from DNA samples collected from peripheral blood leucocytes in bipolar disorder [34] and also in BPD [35]. In general, inherited differences in specific genes thus may moderate the effects of adversity and determine who is more resilient [36].

Interactional models of biological vulnerability combined with psychosocial risks are therefore being increasingly considered in relation to BPD (e.g. [37, 38]). The emphasis placed on social adversity in this paper should not be regarded as a statement of the exclusive pre-eminence of the environment in understanding the developmental origins of PD. Rather, the assumption that should be understood to underpin our discussion of the role of maltreatment and adversity is that such experiences in individuals who are biologically susceptible (and there may be different genetic routes that lead to this susceptibility) cumulatively strain the viability of resilience and, as we shall demonstrate, epistemic trust.

BPD as a limitation of psychological resilience

In further clarifying the relationship between BPD and the *p* factor, Kalisch and colleagues' [6] conceptual framework for the neurobiology of resilience is enlightening. Kalisch et al. [6] argue that psychological resilience is not an absence of disease processes, but a reflection of the work of active, biologically based mechanisms. In considering the relationship between PD and adversity, we have similarly tended to focus on identifying the characteristics of the patient who is experiencing mental health difficulties rather than attempting to delineate the competencies or capacities of the person who has remained functional and free of disorder despite substantial hardship. In fact, studies suggest that only a minority of individuals develop persisting trauma-related pathology as a result of experiencing or witnessing a single extreme or life-threatening event (e.g. Type I trauma). The majority of people have a remarkable capacity for resilience when faced with such events [39, 40].

Rather than searching for the clinical indicators of a transdiagnostic concept such as *p*, we may be wiser conceptualizing *p* as an indication of the absence of resilience and focusing on identifying mechanisms that 'normally' protect individuals from harsh conditions. Perhaps *p* may be more appropriately considered as pointing to protection (or rather the absence of protection).

Resilience has always been an important theme in discourse on mental health [41, 42] but recent concerns about healthcare costs have led to the concept increasingly occupying centre stage [43]. Work on the topic covers myriad different factors and explanations associated with psychological resilience, such as living in a stable and comfortable neighbourhood, family resources and family support, participating in community sporting or

extracurricular activities, racial or gender socialization, being securely attached, being able to regulate one's emotions, exposure to a sensitive style of parenting, or genetic factors. Many of these factors overlap conceptually as well as statistically. They are not explanations for resilience, but rather factors that predict the activation of psychological or biological mechanisms that produce resilience (the absence of pathology in the presence of adversity) as an outcome. Sadly, this conceptual clarity is often lacking in writings about resilience, especially those that concern interventions aimed at its promotion.

The diverse accounts of resilience, often advanced at radically different levels of explanation – from socioeconomic through to genetic – can be unified within the positive appraisal style theory of resilience (PASTOR) conceptual framework presented by Kalisch et al. [6]. According to this formulation, the process underlying resilience is driven by top-down processes in the form of the appraisal that is made of a stressful stimulus. The external and social factors that have been associated with resilience (such as social support or a secure attachment history) affect resilience either directly or indirectly in that they shape the individual's appraisal approach, or minimize exposure to stressors. This is not to deny the role of socio-environmental factors in determining an individual's resilience, or to deny the importance of interventions at a social or community level; it is to suggest that the mechanism by which these distal social factors affect individual resilience is via their impact on the individual's appraisal style.

Resilience and reappraisal

The appraisal theory of resilience is based on a specific understanding of the nature of higher-order cognition [44]. The theory is that the resilience process is as follows: a potentially stressful stimulus is perceived and mentally represented by the individual. The mental representation is then appraised using higher-order cognition, understood in terms of an ensemble of psychological mechanisms and phenomena, including executive function, attention, general intelligence and self-awareness. This in turn determines the emotional response of the individual – their resilience.

We consider this an important perspective but a narrow interpretation of what may be considered higher-order cognition. The outputs of neural processing intrinsically depend on the processing units that take input from the output of other units, perform specific functions, and generate output that in turn becomes the input of other processes. In most models of brain function, any psychological capacity is underpinned by a large number of such hypothetical processes [45, 46]. In this context, the nature of the organization of processing units, or indeed the system that determines their relative activation, may be either a simple

function of the efficiency of processing or, within a hierarchical system, determined by the functioning of a higher-order system. The higher-order meta-system monitors the performance of lower-order systems to ensure optimal performance within a particular context. These components of higher-order cognition are what constitute the core of a normal wakeful and wilful mind in the process of conscious perception, imagination, decision-making and action planning. These functions, taken together, create an opportunity for the internal reorganization of neural structures within the human brain. A consistently 'self-observing' process, which monitors the quality of outcome of neural processing units, enables the individual to reorganize the way neural structures subserve cognitive function. Mentalizing is a key facet of this self-observational process, and the extent to which intentionality fulfils expected behavioural outcomes is a critical indicator of the efficiency of neural processing and guides the way information processing is organized within available pertinent neural units. We assume that an efficiently functioning human brain representing a resilient system achieves such robustness because mentalizing provides a clear window on the efficiency of brain functioning. Multiple processing units cover similar functions in the brain. Some units, being more efficient than others, are more likely to be providers of output that is taken forward to other units. But circumstances change, and demands for adaption may reverse the hierarchy of efficient functioning of these processing units. Resilience is the appropriate appraisal and monitoring of the external social environment and internal functioning of processing units. Thus, as we will explain in more detail in Part 2 of this paper, higher-order cognition is the developmental capacity, based on early relationships and constantly renewed in changing social contexts, to appraise the efficiency of functioning, which in essence is intersubjective in its nature. The capacity to anticipate the reaction of another person, to regulate attention or to implement action plans are all shaped by the overarching need for survival in the context of social interaction. A failure of resilience arises when the individual is unable to change processing systems in a sufficiently flexible manner to maintain optimum outcome despite changed circumstances. When an individual cannot disengage a processing system that is no longer appropriate to the task – for example, a child whose perfectionistic attitudes serve them well during a period of knowledge acquisition and relatively simple tasks, but cause great problems when task complexity has increased to a point where perfection is impossible – the lack of flexibility is what creates vulnerability. Insensitivity does not create risk; the sensitivity of higher-order cognition is what provides protection through the appropriate appraisal of the functioning of neural structures relative to the environment. This is how the resilient brain functions; it is not a model that skirts

reification – it is a description of our assumptions of the nature of brain function.

Higher-order cognition appears to be more flexible within the brain than other, more specialized modal forms of cognition such as basic vision and hearing. For example, brains are able to preserve core aspects of the functional architecture of the information processing that sustains higher-order cognition in spite of substantial structural damage [47]. Higher-order cognition is a form of information processing, therefore, that does not completely rely on one single, static or fixed set of specialized brain regions and anatomical connections, within certain limits of course. It works by exploiting available neural resources and possible routes between them; it seems to use degenerated and pluripotent brain systems flexibly, enabling higher-order cognition to emerge as one of the most robust brain functions. In that sense, the mind does not exist in one physical location within the brain; rather, it is an abstraction, or code, and the brain is the code interpreter. Basic consciousness – the mechanism for the resiliency of cognitive and control systems – is thought to have evolved to be maximally resilient itself: 'consciousness itself can be interpreted as a general algorithm for resilience selected by evolution' ([47], p.22). This decoupling of higher-order cognition from a single location appears to be highly adaptive: its relatively abstract and algorithmic nature makes it more robust in the face of any localized damage or degeneration within the brain.

The algorithmic quality of consciousness may be regarded as a pinnacle of human evolution, but this should not detract from its highly pragmatic, adaptive purposes. This resilient framework is an essential condition for functioning autonomy and the capacity to adapt to the world's demands – particularly the highly complex demands of the human social world. As Paradiso and Rudrauf [48] have argued in their article on social cognition and social neuroscience, tellingly entitled 'Struggle for life, struggle for love and recognition: the neglected self in social cognitive neuroscience', the self, self-awareness and intersubjectivity are integral to social cognitions and actions. As described above, the appropriate functioning of higher-order cognition crucially depends on appropriate judgements about social contexts. In this sense, social cognition is part of the mechanism of higher-order cognition, although social cognition itself is made up of a set of processes that are monitored by the metacognitive evaluations that higher-order cognition performs: as in any feedback system, there is an inherent circularity in this conceptualization. This is inevitable given that we are describing the extent to which a system is capable of reorganizing its own functioning. Similarly, the modes of operationalizing the self and the identification of self-awareness are strongly shaped by

developmental contributions from the social environment – parents, sibling, peers and significant others. In other words, the abstract algorithm that creates personal consciousness cannot be separated from social interactions. This is what the algorithm was developed for, and what further shapes the algorithm of the self and its ongoing relationship with the outside world.

Although there are many factors at work in contributing to resilience, Kalisch describes the three underpinning appraisal mechanisms that determine resilient behaviour and responses [6], as follows:

- 1 *Positive situation classification.* This refers to the manner of immediate appraisal of a situation at the moment of encountering it (e.g. ‘What is the person approaching me carrying in their hand?’). In the case of an insignificant threat, a positive appraisal style enables the individual to view it in a manageable perspective. Clearly, in the context of an adverse event, a negative appraisal and stress response are called for. In such situations, resilience can be subsequently promoted through the second and third forms of appraisal.
- 2 *The retrospective reappraisal of threat.* Whether a traumatic event results in post-traumatic stress disorder, for example, is dependent on how it is retrospectively reappraised [49, 50]. This, as Kalisch et al. describe it, ‘shifts the emphasis from the external situation (or changes in the situation) to the individual’s ability to flexibly adjust current negative appraisal or to implement new, more positive appraisals and then to maintain those appraisals. Both processes have to occur in the face of interference from automatic and uncontrolled negative appraisals and the accompanying aversive emotional states’ ([6], p. 14).
- 3 *Inhibition of retraumatizing triggers.* This mechanism enables the individual to inhibit the threat-associated sensations that might be experienced when remembering a traumatic event and serve to reinforce, perpetuate and generalize the sense of threat.

BPD and the PASTOR model of resilience

To return to BPD, we can follow the PASTOR model by distinguishing between resilience factors and mechanisms. We suggest that a traditional clinical mistake in the treatment of BPD has been to intervene at the level of resilience factors rather than at the level of appraisal (i.e. mechanisms) – this in effect means that we have been working at the level of correlation rather than causation. In BPD, the appraisal mechanisms are at fault, in large part because of mentalizing difficulties (e.g. in the mistaken appraisal of threat at the moment of its

presentation) or a breakdown in epistemic trust, which damages the capacity to relearn different ways of mentalizing – or appraising – situations (i.e. the inability to change our understanding of the threat after the event). The outcome is the lack of resilience that is highly characteristic of BPD, regardless of its clinical presentation.

BPD and positive situation appraisal

Mentalizing has an interpretive role and allows us to explain and predict behaviour; in this sense it also has a social regulatory role [51]. Behaviour can be produced by rational interactions among beliefs and desires, which, when interpreted (appraised) according to specific culturally determined expectations, generate meaning (a meaning assigned to the observed action) in terms of putative mental states that could have engendered the perceived behaviours. Therefore, for our behaviour to be socially meaningful (predictable), it can and should obey these same conventions. Frequent behavioural deviations from these expectations may be considered as being core to PD. This is confusing and stressful for the observer because the normal process of reconstructing mental states from actions is disrupted.

The great importance of this process of meaning generation has been powerfully illustrated by studies in which participants were led to believe that deterministic neurological processes, rather than mental states, control behaviour: in other words, they were discouraged from believing in free will. Introducing an abstract disbelief in free will led to an observed weakening of neural signals associated with readiness planning; subjects became less prepared to act voluntarily [52]. Setting up a deterministic neurological bias also appeared to ‘free’ individuals from a sense of personal responsibility and generated more antisocial cheating and aggression [53].

If mentalizing is assumed to have such an interpretative and regulatory role, then individuals with BPD who have limited capacity to exercise this regulative function are at least partially deprived of the appraisal processes needed to reduce the stress of any social experience. This leaves them at times confused and vulnerable in both the interpretation and the convention-governed expression of mental states in behaviour. To put it plainly, they are frequently puzzled by others’ actions, and equally find themselves victims of misattributions by others. There is ample clinical evidence of limitations of appraisal in BPD (for examples, see [54–57]) although undoubtedly, as would be predicted by the p factor model, they are by no means the only clinical group to show concerning limitations in this area. Poor appraisal may be more severe and pervasive in BPD than, for instance, in major depressive disorder or generalized anxiety disorder without PD comorbidity.

Individuals with BPD tend to be very prone to automatic, non-reflective mentalizing; they often base their inferences on the immediate exterior features of others, and rely on affective rather than cognitive mentalizing. This has clear implications for the style in which they are likely to appraise social situations. As a result of their mentalizing tendencies, individuals with BPD tend to appraise situations and read others' expressions quite quickly: they may show a hypersensitivity to facial expressions [58, 59] and higher-than-normal sensitivity to non-verbal communication [60, 61]. For example, individuals with BPD have been found to outperform non-BPD comparisons on the Reading the Mind in the Eyes Test [62] or to be at least as good as normal controls on the same test [56, 63]. However, this emphasis on external and immediate cues in appraisal situations is accompanied by difficulties in making more reflective judgements based on what might be going on inside people's minds – so, for example, individuals with BPD tend to perform more poorly in social exchange tasks [55, 64]. They have also been found to be more likely to view characters/behaviours as negative or aggressive [65]; to have an impaired view of neutral faces in the context of anger or disgust [66]; and to react with hostility to neutral social interactions [67] – all suggestive of the negative appraisal style described by Kalisch and colleagues [6]. The emphasis on affective mentalizing also results in a heightened sensitivity to emotional cues [59], especially in cases of anger and fear [68, 69]. Furthermore, unbalanced mentalizing on the self–other dimension can cause individuals with BPD to experience severe difficulties in separating the self from the other [70–73] and to be unduly emotionally affected by others' affective states. This often leads to the experience of emotional contagion, which has clear implications for social appraisal situations [74, 75]: BPD individuals can feel forced to be rigid and highly controlling in order to maintain a subjective sense of coherence and integrity [76].

The mentalizing profile characteristic of an individual with BPD, in sum, results in an oversensitivity to possibly difficult social interactions (because distortions in mentalizing are more likely to result in mistaken interpretations of others' behaviour and motivation). In the aftermath of a challenging or stressful interaction, it is difficult for the individual to make sense of, contextualize or put aside potentially upsetting memories of experiences, leaving them more vulnerable to emotional storms. A capacity for explicit, reflective mentalizing in particular serves a dual interpretive (appraisal-strengthening) and self-regulatory role. The absence of this capacity deprives the individual of a fundamental tool in reducing stress.

However, one can see that in certain situations, for example, an emergency milieu characterized by high levels of interpersonal aggression, the heightened and immediate sensitivity and seemingly instinctive and physically

charged form of appraisal characteristic of BPD might in fact be adaptive, at least in the short term. In such an environment, extreme vigilance is a potential advantage, and similarly, the ability to form intense emotional relationships quickly might elicit resources or protection. The mentalizing profile associated with BPD and the appraisal style this generates is maladaptive in most stable social contexts, but we postulate that this mentalizing profile may be a response to cues suggestive of an unreliable and potentially threatening social environment. We thus should be wary of seeing apparent dysfunctions of the clinically 'hard to reach' as indicative of a deficit or any kind of sub-optimal functioning (as, indeed, we have done previously [77]). We would now consider that what may appear to us as dysfunction is an evolutionarily primed adaptation to specific environmental and social contexts. As a genetically triggered adaptation, the individual is biologically programmed to resist change in a behaviour pattern that signals increased chances of selection. We believe that enduring mental disorders (including BPD) are nested in the context of the evolutionary priorities of the human condition.

BPD and retrospective reappraisal

The mentalizing difficulties of BPD patients have also considerable implications for understanding the difficulties with retrospective reappraisal that may undermine resilience. Reappraisal can attenuate ongoing stress responses by appropriately adjusting negative appraisals and/or generating complementary positive appraisals. In strongly aversive situations the stress response is essentially unavoidable: the experience is automatically classified as negative and requires 'after the event' changes in the meaning of the stimuli. This is often achieved through reappraisal in terms of the mental states of the protagonists. To retrospectively appraise an event or situation in a way that promotes resilience, an individual needs to be able to reappraise it in a way that involves reflective, cognitive mentalizing. Such reappraisal will often also depend upon a capacity to mentalize the internal states of both the other and the self. In other words, the mentalizing strengths that this form of retrospective reappraisal requires are not congruent with the mentalizing profile typical of BPD, which is characterized by (a) a tendency to focus on the external rather than internal states of others; (b) the dominance of automatic, intuitive mentalizing over controlled, reflective, mentalizing that could help to put the potentially traumatic event into perspective; (c) an imbalance between affect and cognition in favour of the former, leading to self-perpetuating persistence of negative affect; and, finally, (d) difficulties in coherently representing the self independently of the other, undermining the potential to contextualize and make proportionate an event.

The mentalizing model for trauma has reappraisal of physical and psychological experience at its core [78, 79]. Similarly, trauma-focused cognitive-behavioural therapy and other exposure-based therapies (e.g. eye movement desensitization and reprocessing therapy) enhance mentalizing of the trauma experience, creating a second-order representation of the event in terms of greater coherence of the subjective experience of the victim and often also the perpetrator. Patients with BPD have a specific problem in relation to reappraisal proper because they find it challenging to generate second-order representations of mental states that might be modified to constitute more positive reappraisals of experiences or modify and thus mitigate (adjust) negative appraisals. In essence, this lies at the core of Gunderson and Lyons-Ruth's interpersonal hypersensitivity theory of BPD [80]. Interpersonal hypersensitivity is the likely consequence of a failure of reappraisal following stressful social interactions. In the absence of being able to mentalize in a balanced way, an event or a relationship can be endlessly discussed and dissected in an apparent attempt at reappraisal, but such attempts have an unreal quality. Complicated inferences about mental states are made, but they might have little connection with reality. We term this *pseudomentalizing*, or in extreme, *hypermentalizing*; it is a state of mind that can be clinically misleading in that it may present as a strong attempt at reflection and engagement, but it will ultimately be circular and unproductive. Hypermentalizing of trauma, the failure to move on from it, may be inevitable if individuals cannot reliably access and use social communication that could enable them to resolve or contain the sense of threat associated with a trauma (or if a perceived threat that has been misinterpreted as such, owing to problems in the first resilience mechanism). However, as our understanding of this state of 'petrification' has deepened [81], we also have come to recognize that mentalizing is not everything, or rather, that bodily experience has an important role in enabling access to further resilience strategies. This brings us to the importance of inhibition mechanisms.

BPD and the interference inhibition mechanism

According to Kalisch et al.'s conceptual framework [6], the final level of appraisal underpinning resilience is an inhibition mechanism based on interference. As mentioned above, a strongly aversive event naturally generates powerful negative appraisal responses. The ability to moderate and regulate such negative responses after the event can further determine the extent to which the event continues to cause difficulties in psychological functioning. This implies the inhibition of conflictive negative appraisals and acting deliberately to interfere with emotional reactions to information processing. The inhibition of negative and disruptive responses through

distraction or interference can enable the individual to begin the process of reappraisal proper, allowing a more resilient response to emerge. An individual's inhibitory capacity may to a large part be a trait-like characteristic, with some genetic basis. However, the extent to which the inhibition mechanism can be overwhelmed and how its restoration can be managed may be malleable to some degree.

Although much has been written on the nature of traumatic experiences, within the view outlined in this paper, an aversive event becomes traumatic in its aftermath when it is accompanied by a sense that one is not accompanied – that one's mental experience is not shared and the 'mind is alone' [78, 82]. Trauma obtains from a primitive, adaptive human terror of isolation. Here, again, we run into the key importance of social referencing to calibrate the mind. In the process of reappraisal, the social referencing provided by being able to access another mind enables us to frame and put into perspective an otherwise overwhelmingly frightening experience. This process, which drives a so-called broaden-and-build cycle [83], is far more available to individuals who are open to the benign social influence of other minds. As outlined in more detail in Part 2 of this paper, those who are able to manifest sufficient levels of epistemic trust to embark on the mutually mentalizing stance that is essential in soliciting other minds in support of one's own, are therefore more likely to be resilient. The commonly observed vicious cycle of BPD, comorbid trauma and the acute subjective experience of isolation captures the implications of the failure of this inhibition reappraisal mechanism.

Individuals with the diagnosis of BPD have been shown to have serious limitations in their capacity for the inhibition of conflictive negative appraisals and for interfering emotional reactions to information processing. They cannot cognitively inhibit retraumatizing triggers, leaving them vulnerable to threat-associated sensations that might be experienced when remembering a traumatic event, which serve to reinforce the sense of threat. It is not possible for these individuals to access mentalizing if the self is overwhelmed by negative interference that impairs normal cognitive function. This is congruent with the view that emotional dysregulation is the fundamental problem in BPD [84–86]. The idea of a failure of inhibition in BPD also echoes recently reported findings from Koenigsberg et al. concerning the failure of habituation in BPD [87, 88], which may have a genetic basis [89].

We have similarly (albeit not formulated in terms of the failure of interference or habituation) described the phenomenology of the unyielding nature of trauma-linked subjective experience in BPD [90] in terms of alien self experiences that consist of a sense of looming,

unmanageable anxieties that cannot be reappraised and contained, as the subjective outcome of incorporating an experience of overwhelming hostility into the self [91]. In this context, the focus is not on the development of this experience but rather how it is so persistently maintained despite intense and persistent efforts at reappraisal. This shift in perspective involves a recognition of the significance of the capacity for inhibition in the treatment of BPD. Individuals who are very poor at mentalizing may require more than cognitive interventions (talking) to bring about the inhibition of the stress response. Interventions may have to relate to the body more directly. We have always had a view that mentalizing was embodied [92], but we have not treated this fact seriously enough. We now see a role for physical activity in strengthening the capacity for inhibition at the same time as helping to restore mentalizing. In clinical terms, we suggest that physical activity has a role in strengthening the capacity for inhibition at the same time as, or as a precursor to, helping to restore mentalizing. Perhaps new areas for developing effective interventions may lie in this direction. For example, if an adolescent cannot communicate, activating interference to permit reappraisal via physical activity may well be more valuable than spending hours attempting to activate mentalizing via talking and reflection. The best initial approach may be a physical one: running with them, and discussing what the running was like. Such a simple focus on the embodied experience can be used to begin to rehearse the most basic principle of responding to and giving space to a stimulus outside the negative responses that normally overwhelm other forms of social cognition.

Conclusions

Although we still consider attachment and mentalizing to be key in our understanding of personality pathology, and in earlier formulations we have always emphasized the importance of the absence of resilience in BPD, there has been a notable shift in our views on the emergence and nature of BPD. Rather than seeing BPD primarily in terms of the presence of impairments in attachment and mentalizing, we consider the notable absence of resilience and the social communicative inflexibility that seems to underlie this absence as an adaptive strategy that individuals with BPD acquire within a social context where social inflexibility was often the only possible survival strategy and had considerable advantages in the short term.

We will further elaborate on these issues in Part 2 of this paper. Currently we are still faced with an important theoretical dilemma: from where does this absence of positive reappraisal mechanisms stem? How can we understand the inflexibility in social communicative processes in BPD and in all those suffering from serious

psychopathology, which seems to render these individuals so 'hard to reach'? How did this inflexibility develop over time? We believe that the answers to these questions lie in an evolutionarily informed developmental psychopathology account of BPD and related disorders that has considerable implications for prevention and intervention.

What we have changed our minds about:

Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication

Abstract

In Part 1 of this paper, we discussed emerging evidence suggesting that a general psychopathology or 'p' factor underlying the various forms of psychopathology should be conceptualized in terms of the absence of resilience, that is, the absence of positive reappraisal mechanisms when faced with adversity. These impairments in the capacity for positive reappraisal seem to provide a comprehensive explanation for the association between the p factor and comorbidity, future caseness, and the 'hard-to-reach' character of many patients with severe personality pathology, most notably borderline personality disorder (BPD). In this, the second part of the paper, we trace the development of the absence of resilience to disruptions in the emergence of human social communication, based on recent evolutionary and developmental psychopathology accounts. We argue that BPD and related disorders may be reconceptualized as a form of social understanding in which epistemic hypervigilance, distrust or outright epistemic freezing is an adaptive consequence of the social learning environment. Negative appraisal mechanisms become overriding, particularly in situations of attachment stress. This constitutes a shift towards a more socially oriented perspective on personality psychopathology in which the absence of psychological resilience is seen as a learned response to the transmission of social knowledge. This shift in our views has also forced us to reconsider the role of attachment in BPD. The implications for prevention and intervention of this novel approach are discussed.

Keywords: Borderline personality disorder, Resilience, Epistemic trust, Mentalizing, Attachment, Psychopathology

Background

Bringing together the threads of the argument we built in Part 1 of this paper, we propose that the common variance revealed by bi-factor studies of psychopathology indicates a shared variance in resisting socially expectable adversity. Moreover, persistent psychological distress associated with personality disorder (PD) has as a common element diagnostic criteria that we may particularly expect to see in BPD, making BPD features the core features linked to persistence of psychiatric problems. So far, we have outlined a model that inverts this

vulnerability from one focused on the common characteristics of the pathological condition to an alternative perspective that highlights the absence of resilience as the shared cause. Following Kalisch et al.'s persuasive model of resilience [1], we argued that the persistence of psychopathology, as observed prototypically in BPD, results from a pervasive limitation on the appraisal of stressful social experience, which could be linked to limitations in the capacity to mentalize.

What may explain this absence of capacity to reappraise stressful social experiences? Here, recent evolutionary and developmental accounts of the emergence of epistemic trust in humans may provide important answers. These views also, as we will demonstrate, necessitate a shift in our perspective on the role of

attachment in BPD. Put briefly, the theory of epistemic trust posits that the human infant – most usually first within the context of early attachment relationships – is instinctively inclined to develop openness to the reception of social communications from their primary caregivers. Stated otherwise, epistemic trust is an adaptation allowing the infant to receive social knowledge from their better-informed elders [2], enabling them to benefit from the complex edifice of human knowledge that their immediate culture has available to them.

There are two possible bases on which cultural knowledge can be accepted by a learner as credible: they can either work it out for themselves (which is time-consuming, difficult, and often impossible) or they can rely on the epistemic trust they have in the authority of the communicator [3, 4]. Trusting the communicator means that the learner does not have to go back to first principles each time they encounter novelty: a strange-looking tool without a self-evident purpose is accepted as being used as described by a trusted elder, because they have said so [5]. Being told in this way is enough, and saves an awful lot of time and effort, and indeed possibly allows the infant to grow up and build upon or revolutionize the use of the tool in question. This capacity to teach and learn social knowledge largely underpins the evolution of human culture [6]: it has been proposed that this form of cultural evolution, based on the transmission of knowledge via epistemically trusted communication, emerged during the late Pleistocene era [7].

The internalization of knowledge about the social world constitutes a particular kind of learning: it involves encoding the piece of knowledge as *significant*, *relevant* to the recipient and *socially generalizable* – that is, as an accepted and reusable piece of cultural currency. This specific form of learning is stimulated by ostensive cues generated by the communicator [8, 9]. Such cues trigger a pedagogic stance in the recipient, priming them to regard forthcoming communications as significant. Human infants display species-specific sensitivity and deference to non-verbal ostensive cues, such as eye contact, turn-taking contingent reactivity, being called by their name, and the use of a special tone of voice ('motherese') by the communicator [10, 11]. These ostensive cues have in common the quality that the recipient is recognized as a subjective, agentive self. Once epistemic trust is stimulated in this way, the channel for the transmission of knowledge is opened. Mimicry may be protected by human evolution because it generates epistemic trust, inevitably signalling recognition in the child by the imitating adult. A social smile (recognition of the self by the

other) probably increases the tendency for imitation because the smile generates epistemic trust and opens the communication channel to receive knowledge.

It has been argued that this mechanism for opening the epistemic channel exists because it cannot be left open by default: it is adaptive for humans to adopt a position of epistemic vigilance unless they are reassured that it is safe to do otherwise [4, 5]. The notion that children are promiscuously credulous to those around them has been disproved by ample evidence suggesting the ways in which dubious social signifiers and poor past performance may render a social communicator suspect and their assertions about the world regarded with scepticism [12, 13]. Epistemic vigilance is a necessary tool to protect against misinformation, whether as a consequence of malicious intent or incompetence on the part of the communicator [4]. Therefore, although the purpose of epistemic trust is the transmission of data, its application is a highly psychological and relational process, dependent on calculations about who is trustworthy, authoritative and knowledgeable – in other words, about whose information is worthy of being encoded as relevant and culturally significant to the self.

Epistemic mistrust and developmental psychopathology

In situations where a young learner's early environment is heavily populated by unreliable communicators, the opening of epistemic trust becomes problematic: it may be more adaptive to remain persistently vigilant about, or even closed off to, the communication of social knowledge. In the face of an abusive and hostile caregiver, whose intentions towards the infant or child are not benign, epistemic mistrust becomes entrenched as an appropriate adaptation that has been prepared by natural selection.

Consistent with these assumptions, an accumulating body of evidence indicates that childhood maltreatment, broadly defined, can have a negative impact on several aspects of social-cognitive competencies in individuals who have not yet been explicitly diagnosed with a mental disorder [14–17]. Young maltreated children display impairments with regard to several indices of mentalizing: (a) they engage in less symbolic and less child-initiated dyadic play [18, 19]; (b) they sometimes fail to show empathy when witnessing distress in other children [20]; (c) they have poor affect regulation, which contributes to psychopathology and peer rejection in later life [21–24]; importantly, (d) they make fewer references to their internal states [25]; and (e) they struggle to understand emotional expressions, particularly facial expressions [26, 27]; this latter feature has been observed even in studies that controlled for verbal IQ [28, 29].

The impact of maltreatment reaches into adulthood. A large-scale study of 5000 adults [30] found that maltreatment by parents in childhood was strongly associated with adult variations in theory of mind, or mental-state inferencing, as well as self-reported levels of social affiliation (social motivation and social support). Interestingly, this study found that face discrimination and face memory abilities in adulthood were relatively unaffected by early adversity. The findings confirm that social cognition may be the domain that it is particularly vulnerable to the effects of adverse childhood environments.

Impairments in epistemic trust are a further, and perhaps more damaging, long-term sequel of the experience of childhood maltreatment. Epistemic hypervigilance can manifest as the overinterpretation of motives, which can take the form of hypermentalizing [31, 32], or pseudomentalizing [33]. There is significant evidence that the quality of the relationship of a child to a given communicator determines the extent to which they acquire and generalize information from that communicator [34–36]. When in a state of epistemic hypervigilance, the recipient of a communication assumes that the communicator's intentions are other than those declared, and the information is therefore not treated as being from a deferential source. Most typically, epistemic mistrust manifests as the misattribution of intention and the assumption of malevolent motives behind another person's actions, and therefore treating them with epistemic hypervigilance (or conversely, in some instances, excessive inappropriate epistemic trust). There is evidence to suggest that a hypermentalizing stance is more characteristic of BPD in adolescence [31, 32]. It is possible that this hypermentalizing typically subsides into a flatter profile of outright epistemic mistrust as the individual matures. We speculate that this pattern may partially account for the common life-course history of BPD symptoms, which demonstrates a reduction in impulsive symptoms over time but no lessening of the affective and social symptoms associated with BPD.

In a state of epistemic mistrust, the recipient of social communication may well understand what is being expressed to him/her, but he/she cannot encode it as relevant, internalize it, and appropriately reapply it. The consequence is that the regular process of modifying one's stable beliefs about the world in response to social communication is closed down or disrupted. This generates the quality of rigidity and being 'hard to reach' that therapists have often described in their work in the field of PD [37]. Change cannot happen in the therapeutic setting because, although the patient can hear and understand the communications transmitted to them by the therapist, the information cannot be accepted as relevant to them and generalizable to other social contexts. The persistent distress and social dysfunction

associated with PDs is the result of the destruction of epistemic trust in social knowledge of most kinds.

PD may therefore be best understood as a failure of communication arising from a breakdown in the capacity to forge learning relationships. We believe that this quality underlies the painful sense of isolation that characterizes the subjective experience of an individual with BPD.

Reconsidering the role of attachment

The change of emphasis in relation to the role of attachment theory in the aetiology of PD we will consider in this section speaks to some of the long-standing criticisms of attachment theory that emerged from two directions: psychoanalysis and anthropology. The psychoanalytic criticism of attachment has tended to take the position that attachment theory is too mechanistic and reductionist; that its broad classifications leave attachment unable to engage with the subtlety and complexity of individual human subjectivity. These arguments have been well rehearsed [38, 39]. Meanwhile, anthropologists have suggested that attachment theory disallows other kinds of complexity: those that arrive from cultural differences and varying environmental imperatives. Varying contexts might indeed generate different family configurations and caregiving expectations and structure, for example, alloparenting [40]. As another example, the fluid capacity of caregivers to attach, disengage and reattach across their lives has been compellingly described by the anthropologist Scheper-Hughes in her work on mothering in an acutely impoverished milieu, where she observed mothers facing the death of their infants with apparently little sorrow, but become loving mothers to subsequent children or children who, having previously been given up on, went on to survive [41]. Similarly, historians have traced high rates of infanticide in many cultures (30–40% in early 19th century Milan, for example [42]). Indeed, early historians of childhood, such as Philippe Ariès [43] and Lawrence Stone [44], characterized it as a state of unremitting abuse and brutality. Stone argued that the high levels of infant and child mortality in the pre-industrial era precluded the investment of love and affection in children that we would now consider normative [44]. More recently, this depiction of the experiences of children in the past has been displaced by a more subtle and complex portrait of how parents have historically perceived and related to their children [45]. Ample examples have been found of the ways in which children were recognized, loved, protected and mourned for by their caregivers (e.g. [46, 47]). These academic skirmishes over the sameness and difference of being a parent and a child across time, and the co-existence of love and violence in human experience, should not

surprise us from a clinical point of view: they are in keeping with our understanding of attachment as a universal human (and indeed mammalian) instinct, while still allowing us to recognize, for example, the high rates of infanticide that historians have traced in some periods [42]. In all but the most cases extreme childrearing scenarios, attachments of some style do form; but it is possible that different social environments are likely to trigger different attachment styles as being more adaptive to each environment.

The attachment style to which the child is exposed may be protective of the child, even if it is harsh or cruel. We thus suggest that attachment styles are themselves one piece of social communication that the familial context is promoting about the most effective way to function in the prevailing culture. Attachment is part of a social signalling system telling the infant or young child to prioritize developing specific mentalizing capacities and particular patterns of behaviour. The family environment associated with BPD may entail triggering a particular style of adaptation to ensure survival to reproduction, albeit one that causes pain to the individual and is challenging to the immediately surrounding environment. For example, risky sexual behaviour in adolescents with a childhood history of neglect may be a way of increasing the likelihood that they will contribute to the gene pool. Such behaviours are resistant to change because the adaptation is triggered by natural selection; the individual's genes 'communicate' that this is most likely to ensure survival (of the genome) [48]. Lower levels of mentalizing, greater aggressiveness and higher sensitivity to perceived threats may be adaptive responses to certain cultural environments. Natural selection has charged families with psychologically enculturating their children to maximize their likelihood of survival. Social learning from the immediate family and culture can help us account for the relationship between individual behaviours and the culture that engenders them. Low levels of interpersonal understanding, or even frank attacks on the self-awareness of individual family members, may be biologically successful, evolutionarily selected strategies. A stance of dismissing attachment and non-mentalizing is not experienced as a deficit by the person adopting this stance, but rather as the most appropriate strategy to ensure their survival. It further follows that if mentalizing interventions are to succeed with children, they need to occur in the context of the family [33] and enhance the quality of mentalizing within the family system to which children are oriented to acquire social expectations.

At a theoretical level, this change in focus involves a certain reconfiguration of the role of attachment in developmental psychopathology. Like other authors [49], we have previously placed considerable weight on the

nature of attachment disorganization in our accounts of BPD based on the mentalizing model [50]. We maintain that the role of attachment is highly significant in the developmental origins of PD. However, we argue that its role might perhaps be best understood as only one (albeit very important) form of content learned from the social environment. This is congruent with recent work suggesting that the relationship between infant attachment status and later outcomes is more complicated than that suggested by early attachment studies [51]. Other findings have suggested limited evidence for linking childrearing environments to later outcomes and the fluctuating significance of infant attachment style across the life trajectory. For example, in infancy, the role of genes in determining security or insecurity of attachment is negligible and the childrearing environment is critical [52]; however, in adolescence, the impact of genetic factors rises considerably, such that they predict 38 and 35% of security and insecurity, respectively [51]. Meanwhile, parental sensitivity – previously considered key for the transmission of attachment security in infancy (see a major meta-analysis by Verhage et al. [53]) – may have other functions beyond ensuring secure attachment, although this function is, of course, an important one. The relationship between parental sensitivity and developmental outcomes, according to recent and highly compelling findings by Kok et al. [54], may be more general and structural than can be captured by infant attachment status: these findings indicate that normal variation in maternal sensitivity is related to markers of optimal brain development. This suggests that the parenting environment supports the neurobiological architecture of higher-order cognitive function upon which the capacity to mentalize depends.

We suggest that the relationship between parental sensitivity, attachment and epistemic trust lies in the way in which epistemic trust in most normal circumstances develops in the context of attachment relationships. Secure attachment, which provides mostly consistent contingent parental responses to the child, also provides mostly consistent ostensive cueing and therefore the most fertile ground in which epistemic trust can emerge and subsequently generalize to new relationships. This, of course, follows Bowlby's description of internal working models [55]. Attachment to a safe, sufficiently reliable and mentalizing caregiver provides the child with a sense of agency that allows the child to have some confidence both in their own interpretation of the social world, and in the good faith and general accuracy of their caregivers' communications [56].

The role of attachment in our conception of personality has shifted as we have increasingly come to regard the conceptualization of linear causation in psychopathology as unhelpful; instead, we conceptualize the

perpetuation of PD being driven by loosely coupled interacting systems working in a circular way. A linear approach would posit that the capacity for mentalizing is vulnerable because of the social-emotional quality of early attachment experiences; partial, erratic mentalizing turns into an interpersonal vulnerability whereby a person feels interpersonally brittle because they cannot reliably process the psychological meaning of social experience, and vulnerable because they cannot process their own emotional reactions to these experiences.

Evidence suggests that attachment stress derails mentalizing judgments [57]; working in the other direction, attachment schemas predict mentalizing in adolescence [58, 59]. According to this model, mentalizing and emotional regulation compete, and attachment insecurity has a catalytic role in disrupting the development of optimal mentalizing capacity.

Mentalizing difficulties lead to affect dysregulation, which in turn further disrupts mentalizing. Wherever this cycle starts, mentalizing problems lead to interpersonal conflict and social difficulties, which generate intense (social) affect such as shame, which is inadequately contextualized because of the failure of social cognition. This affect further undermines the capacity to mentalize, which can then create further social challenges, generating interpersonal conflict that will inevitably lead to higher emotional arousal. The emotional arousal is poorly modulated and causes further disruptions of social cognition as part of a recursive process, the final outcome of which is an individual lacking the higher-order cognitive capacity necessary to withstand even everyday social adversity.

The likely interaction between a history of adversity that challenges epistemic trust and mentalizing failure as both a cause and a consequence of emotion dysregulation culminates in a stance where the individual with limited mentalizing capacity cannot reliably detect ostensive cueing and adopts what is perceived to be a maladaptive pattern of rigidity – that is, inability to change. What emerges is an (implicit) attitude of mistrust in the social environment [60] and an incapacity to learn from social experience or to modify one's behaviour on the basis of social learning. In our view, these individuals are those with high 'p' scores whose disorders persist because of their inaccessibility to normalizing social influence. Their 'impermeability' to therapeutic influence comes not from the deep-seatedness of the pattern but its central manifestation of epistemic mistrust born of a dual core of a history of adversity and emotionally disrupted sensitivity to ostension. This is not a naive environmental theory promoting the quality of social interaction at the expense of biological factors: there is every reason to suspect that genetic predisposition, as well as the normal mixture of early environmental

determinants, makes an individual more or less receptive to ostensive cues. The fact that therapeutic interventions have the capacity to promote sensitivity to ostensive cues in no way prejudices the balance of biological versus psychosocial influences on sensitivity to social cues. Because clinicians have historically linked non-responsiveness to therapeutic intervention to *characteristics of their patient* rather than *features of their own relationship to the patient*, the pattern of epistemic mistrust/hypervigilance was regarded as a feature of the most stable system they could identify in their patient – their personality. As 'normal' personality is in fact far from stable, consistent or unmalleable in relation to social situations [61, 62], perhaps disorders of personality are so called because, unlike normal personality, individuals with PDs have in common an absence of flexibility and great difficulty in adapting to changing social situations. Hence, epistemic mistrust may have its roots in part in disturbed attachment experiences, but ultimately it is a disorder of social communication or social learning. Its core is a compromised capacity for appropriately interpreting social actions in terms of mental states, which is what normally bolsters resilience, leaving the individual with dysfunctional social learning systems that are inadequate to assure adaptation in the face of 'normal' adversity.

Although this perspective has considerable bearing on our understanding of the subjective experience of BPD, it is also one that is consistent with a conceptualization of the human mind as having evolved to be highly social and culturally responsive. Therefore, it is a theory that is relevant to how we think about the relationship between the individual and culture, and it is of relevance to a much broader and more interdisciplinary way of thinking than our previous position was. This rather more systemic, less intrapsychic approach involves a repositioning of the role of attachment in developmental psychopathology to accommodate the imperatives of the wider social environment within which the dyadic relationship is located. The anthropologist Thomas Weisner expressed it thus:

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. ([63], p. 263)

Our thinking has – albeit from a different direction – come to a similar conclusion.

The role of systems

If the lack of resilience we associate with BPD is to be understood as an inability to access positive appraisal and the inhibition mechanisms owing to imbalances in mentalizing and the associated compromise of epistemic trust, this also has implications for the system inhabited by that individual. As outlined earlier, we suggest that ‘personality’ dysfunction persists through the self-perpetuating cycle of social dysfunction and mentalizing difficulties. The resulting heightened affect disrupts the interpersonal environment, creating social challenges that derail mentalizing and in turn undermine social functioning.

A graphical display may help to illustrate these complex interactions (see Fig. 1). Emotion dysregulation, disrupted attachment histories and the disorganized insecure attachment system interact to generate social/interpersonal dysfunction, a shared characteristic of PDs [64, 65]. Such dysfunctions are best understood as communication failures rather than as properties or characteristics of the individual suffering from PD.

The failure of communication occurs at a number of levels. First, the social disruption associated with interpersonal conflict will itself compromise the processes of social learning and, in particular, of salutogenesis (the capacity to learn and benefit from the (social) environment). This is a systemic failure of communication that may characterize a family, the members of a social group such as a gang, a social subculture, or indeed an entire

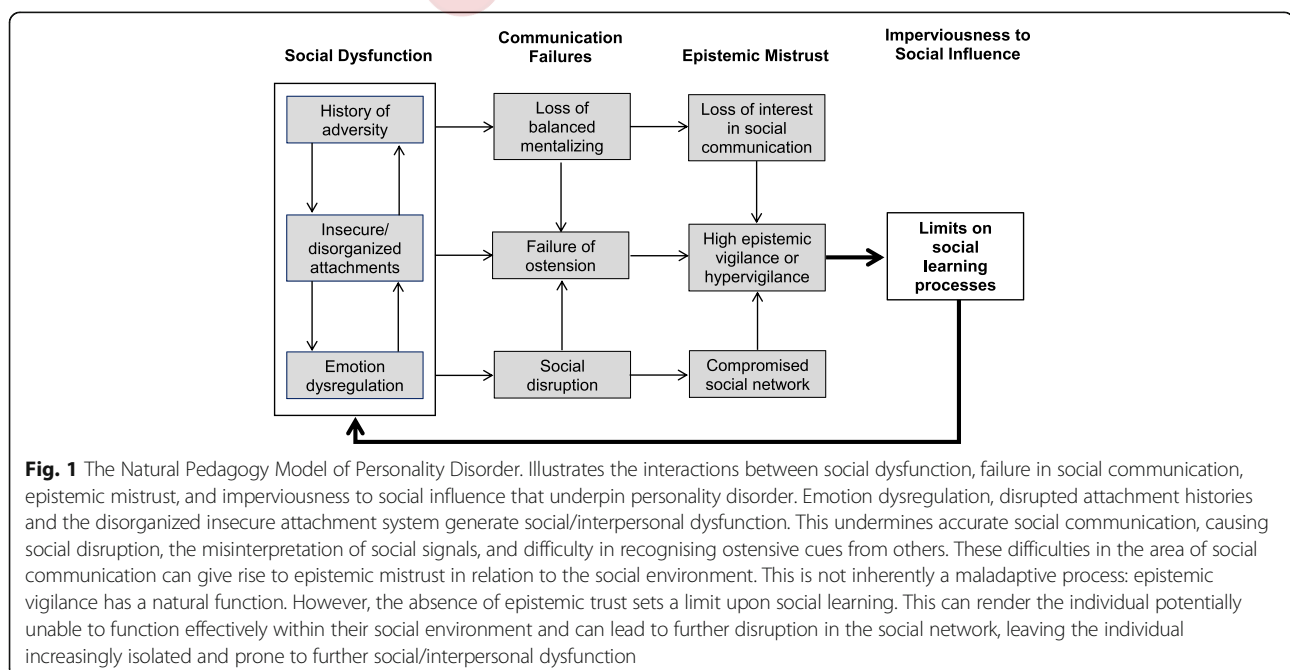
culture. We will discuss such systemic failures in more detail below in terms of their impact on the network of social influence within which all socialization occurs.

Second, the loss of balanced mentalizing triggered by interpersonal conflict generally lessens interest in the content of communication and social information exchange. There is a pervasive loss of interest in intentionality; observable outcomes are gradually prioritized as indicators of attitudes and the general tenor of verbal communication is perceived as meaningless ‘psychobabble’ with few or no substantive implications for the life of the individual.

Third, social dysfunction, as well as the misinterpretation of social signals associated with the loss of mentalizing, leads to a probable failure to appropriately identify ostension – the sense that a communication is of personal relevance.

These factors (and probably many others) contribute to the individual’s failure to develop epistemic trust in personally relevant communications. Again, we are keen to point out that this is not inherently a maladaptive process. The failure to develop epistemic trust leaves the natural function of epistemic vigilance in place. It is in fact an efficient adaptation and an indication that the individual is exercising appropriate caution in relation to social influence, which we see as manifesting in the undesirable persistence of antisocial expectations or schemata and the individual’s relative imperviousness to social influence.

However, the absence of epistemic trust sets a limit upon social learning. This can render the individual increasingly ill-suited to function effectively within their



social environment. Disruption of the social network within which the individual could (or perhaps should) function leaves them increasingly isolated and prone to further social/interpersonal dysfunction.

There are many levels at which systemic thinking applies to how we respond to PD. In terms of clinical work, a mentalizing team around the therapist is, we argue, essential for maintaining good practice. In the context of the persistent distress associated with PD, clinical encounters happen, by necessity, against the background of constant exposure to psychic equivalence and pretend or teleological modes [33, 66]. We suggest that it is the impact of non-mentalizing on the system of social communication, and not the unchangeability of non-mentalizing per se, that makes PDs clinically challenging conditions. One of the defining characteristics of PD is that the patterns of social dysfunction shown by the patient are enduring. Indeed, as mentioned above, BPD in particular has traditionally been regarded as an almost untreatable condition; this is one of the factors that have contributed to the stigma experienced by those receiving a PD diagnosis. However, effective therapies for BPD now exist: at least nine forms of treatment have been tested in at least 20 randomized controlled trials [67], and patients with BPD should no longer be regarded as 'unhelpable'. We would argue that the apparent inconsistency that a condition has long been believed to be untreatable, yet appears to be more responsive to therapy than most mental disorders, is to be found to lie in the way the non-mentalizing actions of BPD patients can create non-mentalizing social systems that sustain their condition – including in the consulting room. We suggest that it is unrealistic to expect a clinician working with such patients to themselves maintain an effective mentalizing stance in the medium to long term if they are not supported adequately to maintain their capacity to mentalize, ideally by a surrounding team that is not directly exposed to (and is thus protected from) the patient's dysfunctional social system.

Systemic interventions may be required to address these problems [68]. In principle, the patient and therapist are isolated in a room, albeit with bidirectional social influence – the therapist is, after all, in a position to enhance the patient's capacity to reflect, to question and to focus simultaneously on both other and self, inside and outside. But the reality is that the therapist becomes embedded within the patient's social survival mechanism, which subsumes the obliteration of balanced mentalizing (normally erring on the side of being unreflective, externally focused, emotional and dominated by resonance rather than reflectiveness). The clinician's mentalizing, even if exceptional, is unlikely to be sufficient to be able to deal with such highly intense emotional situations and conflicts. Therapists require their

own system of support relationships, primarily from other clinicians, in order to scaffold their capacity to mentalize and facilitate epistemic trust.

The self-perpetuating cycle of sustained dysfunction associated with BPD and a non-mentalizing social system reminds us of the international variability in the prevalence of BPD. It has been observed that BPD is less common in non-Western societies, possibly as a result of the fact that the lack of social capital and community support characteristic of many modern or modernizing societies leaves the individual more vulnerable to impulsivity and affective instability [69]. Available prevalence data suggest that Western countries with higher levels of inequality of wealth experience higher rates of BPD [70]. The anomie of modern life – that is, a lack of social connectiveness leading to dysregulation – described by Durkheim [71], and connected by other authors with the conditions that might account for national variations in BPD [69], can be read as a description of a systemic collapse of epistemic trust. This emphasis on the role of the social environment points to the value of thinking about ways in which a social climate can be encouraged to become more mentalizing to support a change process. Families are one obvious example of a systemic arena for the promotion of mentalizing that reinforces the learning of epistemic trust. Bateman and colleagues' development of the Families and Carers Training and Support programme (FACTS) for those supporting a family member with BPD is one example of a mentalizing intervention for the family [72].

The school is another system that seems ideal as the site for mentalizing interventions. Tellingly, evidence suggests that, of the many interventions that now exist to deal with bullying in schools, the most effective share the characteristic of involving the whole school [73]. A mentalizing-based approach, known as Creating a Peaceful School Environment (CAPSLE), is one of three bullying prevention strategies found by a large meta-analysis to be most effective [74] (the other two programmes were the Olweus Bully Prevention Program, whose generalizability has recently been questioned by Bradshaw [75], and Finland's KiVa national anti-bullying program [76]). The mentalizing approach of CAPSLE is a systemic one, which seeks to create a mentalizing climate and a group dynamic that can resist and limit the potency and currency carried by the individual acts of violence or aggression that are inevitable in a school [77–80].

AMBIT (adaptive mentalization-based integrative therapy) is a third example of a clinical approach that combines mentalizing with thinking about the systems that surround an individual [81, 82]. Originally developed for 'hard-to-reach' adolescents with complex needs, AMBIT is now being applied to younger and older client groups.

Such clients present with multifaceted difficulties and so tend to attract complicated multi-agency and multi-professional networks aiming to provide help. At the same time, these clients tend to be highly alienated from conventional social networks, while often forming personal relationships that carry further risks. AMBIT seeks to counter these difficulties by using a main keyworker to, where possible, simplify the individual's experience of the complex network that surrounds them. The keyworker simultaneously seeks to support and encourage the non-professional social networks that surround the individual (e.g. the family, friendship groups or extra-curricular/activity-based groups), while also serving as a secure attachment base from which the individual might explore the social opportunities their environment presents. A final crucial component of this approach is its emphasis on the need for a supportive mentalizing system around the keyworker, given the anxieties and pressures involved in such therapeutic work.

The systemic mentalization-based interventions outlined above have in common their view of the individual as being temporarily separated from their social network, and of their capacity to form bonds of trust being shaky and prone to disruption. Without intervention, the person loses their epistemic safety net; the socially defined network of meanings is under threat. The interventions address the *network*, not just the individual or the therapist. In AMBIT, the links between the keyworker and the 'dis-integrating' (the term used in AMBIT to indicate the frequency with which the various agencies around a client may pull in opposite directions in their various attempts to work with the client) social care system around the family are an important focus. In CAPSLE, the non-mentalizing bully-victim-bystander is focused on by everyone within the whole school. FACTS aims to address the non-mentalizing within the family system. Common to each of these approaches is its capacity to ensure that epistemic trust – the meaningful transfer of information from one person to the other – is ultimately assured and protected. It is evident in CAPSLE where the disruption of epistemic function makes the intervention necessary; indeed, one of the outcome measures for this intervention is the improvement of children's scores in standardized assessments of educational attainment [79]. In AMBIT, meaningful communication between different helping systems is resumed with the restoration of mentalizing. Similarly, in FACTS, with improved mentalizing the family can once again take up its function of social information transmission. It is in our opinion thus not mentalizing itself that is of direct benefit; it is the normal social functions that depend on mentalizing that bring the real therapeutic benefit.

Non-mentalizing social systems present a powerful cue that the individual is in an environment where social relations are not operating on the principle of shared goals, cooperation and interdependence. It is these behavioural imperatives that are, as Tomasello described, associated with our higher-order cognitive capacities [83]. When presented with cues that signify that we do not have access to collaborative social relations, we make cognitive adjustments, as evidenced by new research on Social Baseline Theory [84]. As a simple illustration, hills are judged to be less steep when one is standing next to a friend, and there is a dose-response effect: the longer the friendship, the less steep the hill appears to be [84, 85]. Coan et al. state that 'The human brain *expects* access to relationships characterized by interdependence, shared goals, and joint attention' ([84], p. 87). Violations of this increase stress and increase cognitive and physiological effort – 'social relationships decrease the predicted cost of the environment' ([84], p. 87). Social behaviour is so closely at the heart of the human evolutionary story that it is a fundamental instrument that humans use to 'mitigate risk and diminish the level of effort needed to accomplish goals' ([84], p. 87). In the absence of this social baseline, the environment is perceived to be more risky and costly in terms of effort. The accessibility of social support is one of the factors that humans – and other social animals – use in adjusting their appraisal bias.

Literature relating to research in non-human animals shows that the capacity of an organism to regulate its internal state according to evaluations of the external conditions (rather than through basic stimulus-response mechanisms) is fundamental to behavioural flexibility; it has been recently suggested that appraisal theory can be fruitfully brought into this thinking [86]. In particular, it has been suggested that cognitive biases arising from the interference of affective states, as well as genetic and environmental factors, can affect the appraisal of ambiguous situations, which subsequently shapes resilience to stressful events [86]. One example is Harding and colleagues' classic finding that rats exposed to unstable housing conditions made more pessimistic evaluations of ambiguous stimuli, in a way that is similar to how anxious or depressed people tend to make negative judgments about ambiguous stimuli [87]. Whereas previously, as attachment theorists, we may have made sense of the relationship between behavioural flexibility, social stimuli and appraisal in terms of internal working models, we now suggest that epistemic trust is the mechanism via which humans' behavioural flexibility arising from appraisals becomes compromised.

Implications for prevention and intervention

Different approaches to BPD from a theoretical and practical point of view appear to be embarrassingly similar in terms of outcome [88, 89] in BPD. Based on the considerations outlined in this paper, we suggest that all effective treatments of BPD involve the same structure, namely that the re-emergence of epistemic trust requires three initially sequentially implemented but, as treatments unfold, increasingly concurrent forms of communication.

Communication system 1

This entails the communication of therapeutic model-based content that indicates to the patient that the therapist has considerable knowledge as well as personal characteristics that may be highly valued by the patient. The knowledge communicated will naturally vary according to the treatment model (e.g. Transference Focused Psychotherapy will communicate information about primarily subtle intrapsychic relationships, while Dialectical Behavior Therapy will offer broader psychological constructs and coping strategies). Content analysis of all effective treatments reveal that the relationship of therapist and patient is supported by the former conveying a convincing understanding of the patient as an intentional agent which generates a sense of self-recognition. All evidence-based models of psychotherapy present models of mind, disorder and change that are accurate, helpful to patients and increase patients' capacity for understanding. However, they also need to overcome the epistemic hypervigilance ('not true', 'not relevant to me') presented by the patient. So, besides the content, this stage involves a subtle and rich process of ostensive cueing. Thus, even at this relatively early stage the therapist must present their information with mentalizing in mind, establishing collaboration with the patient, demonstrating that they see the patient's problems from their perspective, recognizing them as an agent, and with the attitude that the patient has things to teach the therapist. Through this, the therapist responds contingently to the patient. From the structural perspective we are presenting here, the therapist's attempt to apply his/her model to interactions with the patient serves as an ostensive cue, which increases the patient's epistemic trust and thus acts as a catalyst for therapeutic success. It does so to the extent that (a) the therapist is able to find and effectively transmit content that provides valuable ways for the patient to understand (mentalize) themselves and their reaction to others, and (b) the process of transmission involves the patient recognizing the truth and personal relevance of the content, so they become able to relax their epistemic mistrust.

Communication system 2

Mentalizing may be a common factor in effective psychotherapies, but not in the sense that we originally intended [90]. It is not that, regardless of the therapeutic model, patients learn the 'Esperanto' of mentalizing, or even the altogether more appealing discourse of 'plain old therapy' [91]. The constant engagement of the patient by the therapist has several key features that are relevant to the restoration of epistemic trust. First, the therapist consistently recognizes the patient's agentiveness, focuses on him/her as an actor and negotiates from the perspective of the patient's self. Second, by marking the patient's experiences, the therapist acknowledges the patient's emotional state. Third, the therapist makes extensive use of ostensive cues to denote the personal relevance of the information transmitted and its generalizable social value. By mentalizing the patient effectively, the therapist models mentalizing, creating an open and trustworthy environment with low arousal. Structurally, a 'virtuous cycle' is put into motion: the therapist responds sensitively to the patient, the patient takes a step back from epistemic isolation, and the patient gradually begins to exercise his/her mentalizing skills, which, step by step, extend from the confines of the therapeutic context and generalize to his/her wider social context. This elicits an emotional reaction by the patient to the social context, giving the therapist further opportunity to respond sensitively. This process involves a complex and non-linear progression. Improving mentalizing is not its main goal, but the improved mentalizing that results from it enables the patient to start to approach and learn from their wider social context. Answering the question of why patients with a better capacity for mentalizing improve more in psychotherapy than those whose mentalizing is poorer helps us to understand the process. Mentalizing moderates the impact of therapeutic communications: a poorly mentalizing patient will frequently interpret the therapist's ostensive cues erroneously, and epistemic trust is thus not established. With improved mentalizing, the therapist's communications are appreciated and interpreted as trustworthy – and have the intended influence on the patient. The experience of having one's subjectivity understood – of being mentalized – is a necessary trigger for being able to receive and learn from the social knowledge that has the potential to change one's perception of oneself and the social world. The 'gift' of a mentalizing process in psychotherapy is to open up or restore the patient's receptivity to broader social influence, which is a precondition for social learning and healthy development at any age.

Communication system 3

The greatest benefit from a therapeutic relationship comes from the generalizing of epistemic trust *beyond* therapy, such that the patient can continue to learn and

grow from other relationships. Social learning in the context of epistemic trust is (re)established, and this leads to salutogenesis. The third communication system is a process of opening the person's mind via establishing epistemic trust (collaboration) so he/she can once again trust the social world by changing his/her expectations of it. This means that it is not just *what* is taught in therapy that helps the patient, but that the patient's capacity for learning from social situations is rekindled. Enhanced mentalizing allows the patient to achieve improved social relationships and recognize who is a reliable and trustworthy source of information – that is, who one can 'be friends with'. The improved epistemic trust and abandonment of rigidity enables learning from experience once again. So, therapeutic change is probably a consequence of how the patient comes to use their social environment, and not to what happens in therapy per se. The benefits of therapy remain contingent on what is accessible to patients in their particular social world. Therapeutic interventions are effective because they open the patient to social learning experiences which feed back in a virtuous cycle. If the environment is at least partly benign, therapy will 'work'.

This third system – social learning in the context of epistemic trust – is, according to our thinking, the mechanism at work in the circular and self-perpetuating relationship between BPD and the social context. The conceptualization of the three communication systems outlined here involves an acknowledgment of the inherent limitations of clinical interventions in cases where the patient's wider social environment does not support mentalizing. The implication of this is that what happens in any therapeutic intervention cannot on its own be expected to be enough to bring about any lasting significant improvement in the patient's state. Indeed, in certain circumstances it would be maladaptive for the individual to develop epistemic trust and lower their social defences – for instance, in social environments characterized by high levels of aggression or violence, in which an external, non-reflective, rapidly responding affective focus on others as opposed to the self would be better prioritized as a survival strategy.

Conclusions

Several features of the theoretical approach presented in this paper await further empirical confirmation, but according to the theory of epistemic trust and social learning, the lack of resilience, or positive appraisal, characteristic of individuals with BPD may be, in a sense, mislabelling. It may be more accurate to characterize BPD as an 'emergency mode' form of social understanding in which epistemic hypervigilance, distrust, or outright epistemic freezing is an adaptive consequence to the individual's social environment. For various possible

reasons, the individual has adopted negative appraisal mechanisms as a default. This is a highly socially oriented perspective on personal psychopathology. The key argument is that BPD (or other manifestations of the absence of psychological resilience) is the outcome of the ways in which the individual has learned to respond to the transmission of social knowledge within their own social environment.

Future research is needed to investigate these assumptions in more detail. This may also lead to the development of new prevention and intervention strategies, which are urgently needed, particularly given the increasing recognition of the need for prevention strategies for BPD [92, 93].

"This course was developed and edited from the open access article: What we have changed our minds about:

Part 1. Borderline personality disorder as a limitation of resilience, Fonagy et al., Borderline Personality Disorder and Emotion Dysregulation (2017) 4:11 (DOI: 10.1186/s40479-017-0061-9), used under the Creative Commons Attribution License."

"This course was developed and edited from the open access article: What we have changed our minds about:

Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication, Fonagy et al., Borderline Personality Disorder and Emotion Dysregulation (2017) 4:9 (DOI: 10.1186/s40479-017-0062-8), used under the Creative Commons Attribution License."