

A Mentalization-Based and Neuroscience-Informed Model of Severe and Persistent Psychopathology

Tobias Nolte, Chloe Campbell & Peter Fonagy

Background

This chapter describes a recent reconceptualization of psychopathology associated with personality disorders (PD) that speaks to the conceptual and diagnostic challenges currently encountered in the field (Fonagy, Luyten, Allison, & Campbell, 2017a, 2017b¹). It is 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 (Caspi et al., 2014) and resilience (Kalisch, Müller & Tuscher, 2015). This chapter seeks to bring an additional dimension to this thinking through the elaboration of the neuroscientific developments that underpin this theoretical framework. In particular, we will highlight how advances in computational neurosciences can be harnessed to contribute a mechanistic understanding of the relationships between higher order social cognition, mentalizing and epistemic trust in generating psychopathology.

In the first part of this chapter we will discuss an emerging body of work on a general psychopathology (or "p") factor, which may provide a compelling explanation for the extensive comorbidity among disorders, as well as many of the other features associated with chronic or severe mental disorder. We will then suggest that a general factor for psychopathology might be most usefully described as an absence of resilience, rather than as reflecting 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). We outline the theory that factor "p" may reflect an absence of resilience resulting from difficulties in social communication emerging from disruptions in epistemic trust, and in problems imagining one's own or other people's thoughts in the face of interpersonal stressors, particularly when they arise in the context of attachment relationships.

A general factor in psychopathology

As has been observed by many in the field, the utility and meaning of the traditional categorical structure of psychopathology are significantly undermined by the overweening

¹ This chapter is based upon previous work by Fonagy, Luyten, Allison & Campbell (2017) as well as by Debanné and Nolte (2018)

presence of comorbidity (concurrent and sequential over time), recurrence and the noisy overabundance of diagnostic disorders. The promise of potential clarity in the face of this clinical and conceptual confusion has been suggested by recent work on a general factor of psychopathology (p factor) – defined by Caspi and colleagues as ‘one underlying dimension that summarized individuals’ propensity to develop any and all forms of common psychopathologies’ (Caspi et al., 2014, p. 131). Caspi and colleagues scrutinised the mental health trajectory of over a thousand individuals in the New Zealand city of Dunedin, across twenty years, from adolescence to mid-life, examining the persistence, co-occurrence and comorbidity of mental health disorder. They found that vulnerability to mental disorder was more accurately described by one general psychopathology factor than by three high-order (spectral) factors delineating 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’ (Caspi et al., 2014, p. 131). Several studies have replicated this higher-order p factor (e.g. Laceulle, Vollebergh & Ormel, 2015; Lahey, Rathouz, Keenan, Stepp, Loeber, & Hipwell, 2015; Murray, Eisner, & Ribeaud, 2016; Del Giudice, 2016, Gibbon, Nolte, Constantinou, Montague & Fonagy, in preparation). Importantly, the p factor concept may thus also explain why there has been such limited success for the field in locating precise causes, consequences or biomarkers and developing tailored treatments for specific psychiatric disorders (Caspi et al., 2014).

More specifically in relation to personality disorders (PD), Sharp and colleagues have considered the question of whether a general factor for psychopathology exists in the context of PD diagnosis (Sharp et al., 2015). In a series of exploratory factor analyses based on a large 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 Borderline PD (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 BPD 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’ (Caspi et al., 2014, p. 131).

The question that follows from this is what does the p factor represent, beyond a statistical construct? Currently, we can only speculate about the nature of this generic aetiological factor, but one association to be investigated may be childhood adversity. Studies indeed suggest that maltreatment, like p, increases the chance of most types of mental illness in adulthood (Scott, McLaughlin, Smith, & Ellis, 2012) and worsens the course of mental illness (Nanni, Uher & Danese, 2012). 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 (Teicher & Samson, 2013). In our opinion, research findings on maltreatment, although still too narrow, 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 (to take a commonly used example in our work). Such a position would stand counter to the growing evidence for a substantial genetic determinant of BPD. Research showing the familial nature of BPD (White, Gunderson, Zanarini & Hudson, 2003; Zanarini, Barison, Frankenburg, Reich & Hudson, 2009), and classical twin studies that place heritability of BPD at around 40–50% (Bornoalova, Hicks, Iacono & McGue, 2009; Distel et al., 2008; Kendler et al., 2008; Torgersen et al., 2000) have been borne out further by more complex behaviour–genetic models that take into account siblings, spouses and twins (Torgersen et al., 2008). We would rather position the role of adversity as one possible – and not uncommon – way in which the psychological “immune system” of an individual becomes compromised. To extend the metaphor, difficulties in social communication arising from early adversity can limit the individual’s capacity to mobilise defences (through recursive cooperative access to one’s own and other people’s thoughts) to stave off potentially destabilising incursions from the environment.

Resilience, mentalizing and reappraisal

In illuminating the nature of the relationship between mentalizing and the p factor, Kalisch and colleagues’ (2015) conceptual framework for the neurobiology of resilience is valuable. 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 this group. According to the formulation, the mechanism underlying resilience is driven by top-down processes in the form of the representation of adverse experience with subsequent 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 (including defence mechanisms), or minimize exposure to stressors. Kalisch et al. (2015) argue that psychological resilience is not an absence of disease processes, but a reflection of the work of active, biologically based mechanisms. The appraisal theory of resilience is based on a specific understanding of the nature of higher-order cognition (Folkman, Lazarus, Dunkel-Schetter, DeLongis, & Gruen, 1986).

Kalisch describes the three underpinning appraisal mechanisms that determine resilient behaviour and responses (Kalisch et al., 2015), 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 appropriately to view it as non-menacing. 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. The mentalizing profile typically associated with individuals with BPD tends to lean towards the automatic, non-reflective, emotional poles on the mentalizing dimension (Fonagy & Luyten, 2009). They have also been found to be more likely to view characters/behaviours as negative or aggressive (Barnow et al., 2009); to have an impaired view of neutral faces in the context of anger or disgust (Daros, Zakzanis, & Ruocco, 2013); and to react with hostility to neutral social interactions (Domes, Czeschnek, Weidler, Berger, Fast, & Herpertz, 2008) – all suggestive of the negative appraisal style described by Kalisch and colleagues (2015). The mentalizing profile characteristic of an individual with BPD may therefore lead to an oversensitivity to possibly difficult/threatening social interactions (because distortions in mentalizing are more likely to result in mistaken interpretations of others' behaviour and motivation). Further, 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.

- 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 (Gross, 1998; Lazarus & Folkman, 1984). 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. Both processes have to occur in the face of interference from automatic and uncontrolled negative appraisals and the accompanying aversive emotional states' (Kalisch et al., 2015, p. 14). Patients with BPD or high on 'p' 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 to 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 (Gunderson & Lyons-Ruth, 2008).

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 (Fonagy & Luyten, 2009), 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*.

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. An incapacity to inhibit these responses can serve to reinforce, perpetuate and generalize the sense of threat or distress. Individuals with the diagnosis of BPD may often find it difficult to limit the incursion of conflictive negative appraisals and find themselves overwhelmed by 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 or of being overwhelmed. It is not possible for these individuals to access mentalizing if the self is flooded by negative interference that impairs normal cognitive function. This is congruent with the view that emotional dysregulation is the fundamental problem in BPD (Siever & Davis, 1991; Linehan, 1993; Crowell, Beauchaine, & Linehan, 2009).

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 and the neural updating

processes that occur at the intersection between external reality, interoception and predictive capacities (see further below).

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 (Rumelhart & McClelland, 1986; Marr, 1982). We are still faced with an important theoretical dilemma: where does this absence of positive reappraisal mechanisms originate from? How can we understand the inflexibility in social communicative processes, which seems to render these individuals with severe personality pathology so ‘hard to reach’? How did this inflexibility develop over time? We believe that the answers to these questions lie in an evolutionarily informed account of developmental psychopathology.

Epistemic trust and social communication

Bringing together the threads of the argument we have built up so far, 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 has as a common element diagnostic criteria that we may particularly expect to see in BPD, thereby making BPD characteristics (namely compromised mentalising of the self in relation to others) 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 (2015), 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. It thus emerges that rather than searching for the clinical indicators of a transdiagnostic concept such as *p*, we may be wiser to conceptualize *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 thereof).

An important question remains unanswered: What may explain this absence of a protective capacity to reappraise stressful social experiences? A core component of the model outlined here 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 (this thinking is very much informed by the work of, among others, Belsky et al., 2012, Del Giudice, etc). It is the heft of these evolutionary imperatives that creates the

quality of entrenchment that is typically regarded as characteristic of PD: the severe problems with social communication typically associated with such pathology, should be, according to this perspective, considered to be a socially generated outcome of a particular context or developmental niche.

This is where we suggest that evolutionary and developmental accounts of how the capacity for epistemic trust evolves in humans may provide some important answers. To summarise, the notion of epistemic trust posits that the human infant – most usually first within the context of early attachment relationships – is instinctively inclined to develop an openness to the reception of social communications from their primary caregivers. Epistemic trust is thus an adaptation allowing the infant to receive social knowledge from their better-informed elders (Gergely, Egyed & Kiraly, 2007), enabling them to benefit from the complex edifice of human knowledge that their immediate culture (the attachment dyad and the wider caregiving network) has available to them.

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 (Recanati, 1997). Being told in this way is enough, and saves time and effort (and thereby processing capacity), 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 (Wilson, 1976): it has been proposed that this form of cultural evolution, based on the transmission of knowledge via epistemically trusted communication channels, emerged during the late Pleistocene era (Wilson & Wilson, 2007).

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 (Sperber & Wilson, 1995; Russell, 1940). Such cues trigger a pedagogic stance in the recipient, priming them to regard forthcoming communications as significant and as worthwhile to be paid attention to. 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 (Csibra & Gergely, 2009; Csibra & Gergely, 2011). These ostensive cues have in common the quality that the recipient is recognized as a subjective, agentive self – (s)he is being mentalized. What is required, in other words, is a second order capacity by the communicator: he or she must take into account which cues will indicate to the receiver the intention of communicating to this specific person. 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 (Sperber et al., 2010; Recanati, 1997). 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 (Brosseau-Liard, Cassels, & Birch, 2014; Koenig & Harris, 2005; Shneidman & Woodward, 2016). 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 (Sperber et al., 2010). 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 experiences early caregivers as unreliable communicators, the opening of epistemic trust becomes problematic. 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. It may be more adaptive to remain persistently vigilant about, or even closed off to, the communication of social knowledge (under which we also include the “learning” about mental states when being contingently mirrored in a marked way.

The resulting impairments in epistemic trust are a further, and perhaps more damaging, long-term sequela of the experience of childhood maltreatment. Epistemic hypervigilance can manifest as the overinterpretation of motives, which can take the form of hypermentalizing (Sharp et al., 2011; Sharp et al., 2013), or pseudomentalizing (Asen & Fonagy, 2012). 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 (Lane & Harris, 2015; Shafto, Eaves, Navarro, & Perfors, 2012; Mascaró & Sperber, 2009). 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).

The consequence is that the regular process of modifying one's stable beliefs about the world in response to social communication (and in the light of new data) 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 (Fonagy, Luyten & Allison, 2015). 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. PD may therefore be best understood as a failure of communication arising from a breakdown in the capacity to forge trust-based learning relationships.

Furthermore, 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 (Nolte, Bolling, Hudac, Fonagy, Mayes, & Pelphey, 2013); working in the other direction, attachment schemas predict mentalizing in adolescence (Dykas & Cassidy, 2011; Sharp, Fonagy & Allen, 2012). According to this model, mentalizing and affect regulation compete, and attachment insecurity or disorganization have a catalytic role in disrupting the development of optimal mentalizing capacity.

Wherever this cycle starts, mentalizing problems lead to interpersonal conflict and social difficulties, which generate intense (social) affect such as shame or rage, which is inadequately contextualized because of the failure of social cognition. This affect further undermines the capacity to mentalize, which in turn leads to social challenges, generating interpersonal conflict that will inevitably lead to higher emotional arousal. The upregulated 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 (e.g. perspective-taking leading to reappraisal) necessary to withstand even everyday social challenge.

The failure of communication described 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.

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 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 and imbued with the potential for meaning making.

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 equipped to function effectively within their social environment as the higher order cognition underpinning resilience, we believe, develops within and is iteratively sustained by the social context. 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.

Based on the above considerations, we propose that there are three distinct processes of communication that cumulatively make psychotherapy effective:

- *Communication System 1: The teaching and learning of content*
The different therapeutic schools belong to this system. They may be effective primarily because they involve the therapist conveying to the patient a model for understanding the mind that the patient can understand as involving a convincing recognition and identification of his/her own state. This may in itself lower the patient's epistemic vigilance and create a minimal interest in psychic reality.
- *Communication System 2: The re-emergence of robust mentalizing*

When the patient is once again open to social communication in contexts that had previously been blighted by epistemic hypervigilance, he/she shows increased interest in the therapist's mind and the therapist's use and modulation of thoughts and feelings, which stimulates and strengthens the patient's capacity for mentalizing. Improvements in mentalizing or social cognition may thus be a common factor across different interventions.

- *Communication System 3: The re-emergence of social learning*

The relaxation of the patient's hypervigilance via the first two systems of communication enables the patient to become open to social learning beyond the dyad. This allows the patient to apply his/her new mentalizing and communicative capabilities to wider social learning, outside the consulting room. This final part of the process depends upon the patient having a sufficiently benign social environment to allow him/her to gain the necessary experiences to validate and bolster the improved mentalizing, and to continue to facilitate relaxation of epistemic mistrust, in the wider social world.

Basic Principles in Computational Neuroscience

As shown elsewhere (Debbané & Nolte, 2018), it is our view that experience-dependent reorganisation of the brain's structure-function relationships is critical for lifelong adaptation. Resilience, from that perspective, is governed by higher-order cognition (HOC) comprising a range of psychological processes such as perspective taking, metacognition, executive functions, attention, memory, general intelligence, and self-awareness. In the brain, HOC can be seen as an overarching re-routing of structural and functional connections between ensembles of neuronal populations or networks (e.g. Rudrauf 2014, Fonagy & Bateman, 2016; Fonagy, Luyten, Alison & Campbell, 2017) – also described as a particular form of neural plasticity. As a result, new routes for information processing between regions are created (for instance after structural damage due to a neurological event). The resulting optimised neural architecture seeks to compensate for the compromised structure by generating an alternative set of brain connections to sustain *functional* resilience in the face of partial organ failure.

Mentalizing in a broader sense, we postulate, is a prime candidate in order to dynamically preserve or recover adaptive functioning, i.e. it supports psychological resilience and robustness in the face of increased allostatic load impacting the brain. This can arise from either external stressors or changes in the environment, or through interpersonal and intrapsychic conflict. When both neural integrity as well as that of the felt self are under threat (for instance, during childhood adversity, illness, prolonged distress in maladaptive relationships or traumatic experiences), HOC can act as a compensatory cognitive reserve (Stern, 2009) to rearrange computations in the brain in order to ensure “business as usual” (Fonagy & Bateman, 2016; Fonagy, Luyten, Alison & Campbell, 2017).

Heuristically speaking, following our notion of resilience, it is the human capacity for imaginative activity or perspective taking that constitutes the crucial mechanism enabling abstraction, separation of thoughts from feelings, and reappraisal which is key to flexible and resilient thinking.

A partial failure of resilience arises when the individual is unable to change processing systems in a sufficiently flexible manner so that an optimal outcome under changed circumstances can be maintained. 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 – this inflexibility can generate 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 in interaction with the environment.

Recently, computational psychiatry, which focuses on understanding the mechanistic aspects of psychopathology, has developed as a compelling possible approach to investigating these phenomena of the self in exchange its social context in a rigorous empirical way. As an emerging discipline, computational psychiatry is located at the intersection between computational neuroscience, psychology and psychiatry. It seeks to provide insight into the key mechanisms underpinning mental functioning by examining the associations between brain function, interplay with the environment the brain is trying to model or act upon, and psychiatric symptoms that arise from adaptational failures via formalised computational processes (Montague et al., 2012, Friston 2015; Stephan & Mathys, 2015 and Adams, Huys & Roiser, 2016).

The notion of *active inference* is one of the main principles underpinning this novel approach. It posits that the brain performs so-called *Bayesian* inference processes. In the heuristics of this framework, the brain is viewed not just as a passive filter of its input, but rather as a ‘statistical organ that generates hypotheses or fantasies that are tested against sensory evidence’ (Friston, Stephan, Montague & Dolan 2014, p. 148). This is of particular importance as navigating the social world requires constant modelling of other people, as well as approximating to what intentional states might underpin their behaviours or our own. As a result of such *recursive modelling* coherent self-with-other experiences are generated.

Recognising discrepancies between one’s assumptions (our prior beliefs about the causes of what we perceive) and reality generates ‘surprise’ or ‘unbound energy’. Such a mismatch then leads to a neural alert signals, the so-called *prediction error*, which in turn incline the neural system to update the relevant working hypotheses as one accumulates more and new evidence or information’ (Debbané & Nolte, 2018). Through pre-existing mental structures, our generative models of self and other – the brain’s working hypotheses

– are afforded greater or lesser precision via top-down neuromodulatory gain control mechanisms. These prescribe how prone the revisions of our belief are to rigidity or indeed flexibility. We may draw a bridge to the notion of *learning from experience* here, and the accompanying capacity to hold alternative viewpoints – a core capacity required for effective reappraisal as postulated in the PASTOR framework.

The key notion of this specific approach in computational neuroscience is that neural functioning shapes sensory input into experience according to Bayesian probabilistic calculus. As captured by Holmes (2017), in healthy minds, ‘discrepancies between prediction and input are: (a) identified, via ‘tolerating uncertainty’, (b) explored, by action aimed at reducing uncertainty, (c) leading, finally, to cognitive restructuring and updating of ‘prior’ probabilities’. A mentalizing therapist supports patients in bearing the inevitable uncertainty or opaqueness that is intrinsic to any knowledge about mental states. He or she, furthermore, models affect regulation, can stimulate reflection, and thereby scaffolds the exploration of alternative perspectives so that the process of thinking and imagining can be rekindled, constantly reshaped and dynamically updated through new experiences.

Developmentally speaking, the generating of meaning in infancy within the biobehavioral (a-)synchronies with caregivers and the physical environment, the pressures to regulate allostatic challenges (both continuous life stressors as well as particular traumatic events) and the reappraisal, i.e. mentalising of caregiving experiences – all can be captured from the point of view of Bayesian principles. The accompanying basic generative models or beliefs, in the attachment relationship, undergo constant revision and refinement. In other words ‘a baby’s brain is in the game of minimizing surprise, while acting epistemically to reduce expected surprise (i.e. uncertainty)’ about the novel events it is constantly encountering (Friston, 2017). We can understand the baby’s need to adapt to its specific developmental niche as ‘conforming to one imperative; namely, to maximize the evidence for its own existence’ (Friston, 2017). This is captured by the notion of a self-evidencing brain that strives to make sense of what is going on inside and around her/him.

In early development, it is the HOCs ‘borrowed’ from the parents’ minds, through their caregiving disposition, that provide most of the infant’s adjustment and the accompanying affect regulation. Shai & Belsky postulate that parental *embodied* mentalizing is initially non-verbal and implicit (Shai & Belsky, 2011). Critically, these embodied gestures present a model of the infant from the caregiver’s perspective, which at both the neural and experiential level, assist the infant to integrate the primary sensory and multisensory signals (Fotopoukou & Tsakiris, 2017). Predictive models will spatially, affectively, and cognitively orient the infant to explore the environment and the mind of others. This is echoed in computational neuroscience: if the brain is designed to elaborate hierarchical generative models of the world, then the infant brain will fundamentally seek to forge and to differentiate between self versus non-self causes of sensations. ‘In other words, the self-evidencing baby critically requires predictable input from the caregiver to integrate self-

other boundaries. As a result of contingent and marked mirroring a continuity of selfhood is then established. When this input is missing or when it is conveyed in a affectively overwhelming or un-“marked” fashion, it carries the opposite effect: that of blurring the self-other boundary’ (Debbané & Nolte, 2018) with the resulting risk for developmental disorders.

Following this line of thinking, psychopathology and compromised resilience, in a contemporary neuroscience framework, can be understood in terms of atypical and impaired developmental patterns underpinning *computational* aspects of such inference and prediction of caregiving behaviour.

The modelling of computations sustaining ongoing experience-dependent adaptation in life can generate empirical tools that will inform the mentalizing framework to better conceptualize processes that contribute to both health and to psychopathology. Significant conceptual contributions from this approach all bear relevance for a deepened understanding of *how* aspects of mentalizing are implemented neurally. In other words; they extend beyond identifying which network of brain areas is involved when a particular process is being called into action, but rather they seek to uncover the algorithms underlying how neural populations perform their computations. Recent advances in this area are, for instance, concerned with the inference on self and other representations (Moutoussis et al., 2014) including the development of caregiving-attachment contingencies (Cittern, Nolte, Friston & Edalat, in press), interoception and homeostasis of the self when engaged with the outside world (Gu et al., 2013; Seth, 2014), compulsivity (Hauser et al., 2016), and deep cognitive phenotyping (Xiang et al., 2012; Hula, Dayan & Montague, 2015). They can further be complemented by tangible applications focused on social learning (Diaconescu et al., 2017) and theory of mind in two party interactions, for instance in the multi-round Trust Game (Hula, Vilares, Dayan & Montague, 2017;).

Conclusion: P Factor and the Brain

In line with the contemporary developmental account of severe personality impairment outlined above, computational neuroscience may shed light on both environmental and constitutional contributions to the general vulnerability or propensity of this persistent psychopathology. Furthermore, it can help identify key computational processes and brain-behaviour interactions of compromised functioning, for instance sub-optimal recruitment of HOCs. With initial evidence emerging that casts “p” as strongly associated with high levels of childhood maltreatment and low mentalising capacities (Caspi et al., 2014; Gibbon, Nolte, Constantinou, Montague & Fonagy, in preparation) the neural underpinnings of mental state processing will illuminate transdiagnostic brain-behaviour

relationships that may facilitate an understanding of why, for instance, certain symptoms cluster together or follow a sequential order.

With regards to psychopathology, such computational models can be informative for identifying where a single patient is located along a continuum of (interpersonal) problems (Stephan et al., 2017). They can also drive the identification of endophenotypes associated with higher-order mental processes that cut across traditional diagnoses. Most importantly, these new approaches may shed light on aspects of brain structure-function relationships and plasticity underpinning resilient adaptation (Kalisch, Müller & Tüscher, 2015) such as the epistemic trust-modulated broaden and build cycles – i.e. the individual's capacity to use social support and to learn from other mentalizing minds (Fonagy, Luyten & Allison, 2014). Thereby, computational neuroscience may further our still limited knowledge on the three therapeutic communication systems outlined above. Most importantly it may help to illuminate how insights gained in the consultation room are generalized in the patient's wider social environment, and how the patient's 'inferential organ' can be made more receptive to external influence and thus more amenable to updating on the basis of new interpersonal experience.

A better mechanistic understanding of how hierarchically organised microcircuits of the mentalising neural network in the brain orchestrate and update beliefs about relational experiences (including those schemata that are recruited without a person's awareness, i.e. unconsciously) will help to identify computational building blocks of self-other inference and HOC-related processes. This includes technological advances that will allow for new insights into how two or more mentalising brains influence one another e.g. (Bolis & Schilbach, 2017). 'As we know, entraining flexible thinking (balanced mentalising) and thereby an openness to psychological knowledge through epistemic trust results in learning from a social environment experienced as benign and offering support. Thus, a final extension to such between-brain processes and their neural machinery (with parent-infant or therapist-patient as paradigmatic dyads) has to be the investigation of how broader social networks and the wider social environment are represented and experienced as well as how the communicative processes outlined above are implemented neurally (Debbané & Nolte, 2018).

References

References

Debbané & Nolte (2018). The Neurobiology of Mentalising (in press). In Fonagy, P. & Anthony Bateman (Eds): Handbook of Mentalizing in Mental Health practice, 2nd edn.

Chapter 2. Mentalization and Psychopathology

- Fonagy, P. and Luyten, P., 2009. A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and psychopathology*, 21(4), pp.1355-1381.
- Gibbon, L., Nolte, T. Constantinou, M, Montague, R. & Fonagy, P. (in preparation). Modelling axis I and personality disorder symptomatology and its associations with childhood trauma and mentalizing.
- Gibbon, L., Fonagy, P. & Nolte, T. (in preparation). A review of the general psychopathology factor and its clinical implications.
- Shneidman, L. and Woodward, A.L., 2016. Are child-directed interactions the cradle of social learning?. *Psychological bulletin*, 142(1), p.1.
- Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017a). What we have changed our minds about: Part 1. Borderline personality disorder as a limitation of resilience. *Borderline Personal Disord Emot Dysregul*, 4, 11. doi:10.1186/s40479-017-0061-9
- Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017b). What we have changed our minds about: Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication. *Borderline Personal Disord Emot Dysregul*, 4, 9. doi:10.1186/s40479-017-0062-8
- Teicher, M.H. and Samson, J.A., 2013. Childhood maltreatment and psychopathology: a case for ecophenotypic variants as clinically and neurobiologically distinct subtypes. *American journal of psychiatry*, 170(10), pp.1114-1133.
- Widiger, T. A., Lynam, D. R., Miller, J. D., & Oltmanns, T. F. (2012). Measures to assess maladaptive variants of the five-factor model. *Journal of Personality Assessment*, 94(5), 450-455. doi:10.1080/00223891.2012.677887

Chapter 2. Mentalization-Based Therapy and Neuroscience

Arnsten AF, Mathew R, Ubriani R, et al: α -1 noradrenergic receptor stimulation impairs prefrontal cortical cognitive function. *Biol Psychiatry* 45:26-31, 1999

Asen E, Fonagy P. Mentalization-based therapeutic interventions for families. *J Fam Ther.* 2012;34:347-70.

Barnow S, Stopsack M, Grabe HJ, Meinke C, Spitzer C, Kronmuller K, Sieswerda S. Interpersonal evaluation bias in borderline personality disorder. *Behav Res Ther.* 2009;47:359-65.

Bateman AW, Fonagy P: *Psychotherapy of Borderline Personality Disorder: Mentalization-Based Treatment.* Oxford, Oxford University Press, 2004

Bateman AW, Fonagy P: *Mentalization-Based Treatment for Borderline Personality Disorder: A Practical Guide.* Oxford, Oxford University Press, 2006

Bateman AW, Fonagy P: *Mentalization-Based Treatment for Personality Disorders: A Practical Guide.* Oxford, Oxford University Press, 2016

Belsky DW, Caspi A, Arseneault L, Bleidorn W, Fonagy P, Goodman M, Houts R, Moffitt TE. Etiological features of borderline personality related characteristics in a birth cohort of 12-year-old children. *Dev Psychopathol.* 2012;24:251-65.

Bolis D, Schilbach L: Beyond one Bayesian brain: Modeling intra- and inter-personal processes during social interaction: Commentary on "Mentalizing homeostasis: The social origins of interoceptive inference" by Fotopoulou & Tsakiris. *Neuropsychanalysis* 19:35-38, 2017

Bonanno GA, Diminich ED. Annual research review: Positive adjustment to adversity –Trajectories of minimal-impact resilience and emergent resilience. *J Child Psychol Psychiatry.* 2013;54:378-401.

Bornoalova MA, Hicks BM, Iacono WG, McGue M. Stability, change, and heritability of borderline personality disorder traits from adolescence to adulthood: a longitudinal twin study. *Dev Psychopathol.* 2009;21:1335-53.

Borsboom D, Cramer AO: Network analysis: An integrative approach to the structure of psychopathology. *Annu Rev Clin Psychol* 9:91-121, 2013

Brosseau-Liard P, Cassels T, Birch S. You seem certain but you were wrong before: developmental change in preschoolers' relative trust in accurate versus confident speakers. *PLoS One.* 2014;9:e108308.

Carter CS, Porges SW: The biochemistry of love: an oxytocin hypothesis. *EMBO Rep* 14:12-16, 2013

Caspi A, Houts RM, Belsky DW, et al: The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin Psychol Sci* 2:119-137, 2014

Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, Meier MH, Ramrakha S, Shalev I, Poulton R, Moffitt TE. The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin Psychol Sci.* 2014;2:119-37.

Chapter 2. Mentalization-Based Therapy and Neuroscience

Csibra G, Gergely G. Natural pedagogy. *Trends Cogn Sci*. 2009;13:148-53.

Csibra G, Gergely G. Natural pedagogy as evolutionary adaptation. *Philos Trans R Soc Lond B Biol Sci*. 2011;366:1149-57.

Cittern D, Nolte T, Friston K, Edalat A: Intrinsic and extrinsic motivators of attachment under active inference. Manuscript under review, 2018

Crowell SE, Beauchaine TP, Linehan MM. A biosocial developmental model of borderline personality: elaborating and extending Linehan's theory. *Psychol Bull*. 2009;135:495-510.

Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med*. 2013;11:126.

Daros AR, Zakzanis KK, Ruocco AC. Facial emotion recognition in borderline personality disorder. *Psychol Med*. 2013;43:1953-63.

Debbané M, Badoud D, Sander D, et al: Brain activity underlying negative self- and other-perception in adolescents: The role of attachment-derived self-representations. *Cogn Affect Behav Neurosci* 17:554-576, 2017

Debbané & Nolte (2018). The Neurobiology of Mentalising (in press). In Fonagy, P. & Anthony Bateman (Eds): *Handbook of Mentalizing in Mental Health practice*, 2nd edn.

Debbané M, Salaminios G, Luyten P, et al: Attachment, neurobiology, and mentalizing along the psychosis continuum. *Front Hum Neurosci* 10:406, 2016

Del Giudice M. The life history model of psychopathology explains the structure of psychiatric disorders and the emergence of the p factor. *Clin Psychol Sci*. 2016;4:299-311.

Diaconescu AO, Mathys C, Weber LAE, et al: Hierarchical prediction errors in midbrain and septum during social learning. *Soc Cogn Affect Neurosci* 12:618-634, 2017

Distel MA, Trull TJ, Derom CA, Thiery EW, Grimmer MA, Martin NG, Willemsen G, Boomsma DI. Heritability of borderline personality disorder features is similar across three countries. *Psychol Med*. 2008;38:1219-29.

Domes G, Czeschnek D, Weidler F, Berger C, Fast K, Herpertz SC. Recognition of facial affect in borderline personality disorder. *J Pers Disord*. 2008;22:135-47.

Dykas MJ, Cassidy J. Attachment and the processing of social information across the life span: theory and evidence. *Psychol Bull*. 2011;137:19-46.

Ebert A, Brüne M: Oxytocin and social cognition, in *Current Topics in Behavioral Neurosciences*. Berlin, Springer, 2017, pp 1-14

Feldman R: The neurobiology of human attachments. *Trends Cogn Sci* 21:80-99, 2017

Chapter 2. Mentalization-Based Therapy and Neuroscience

Feldman R, Weller A, Zagoory-Sharon O, Levine A: Evidence for a neuroendocrinological foundation of human affiliation: Plasma oxytocin levels across pregnancy and the postpartum period predict mother-infant bonding. *Psychol Sci* 18:965-970, 2007

Folkman S, Lazarus RS, Dunkel-Schetter C, DeLongis A, Gruen RJ. Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes. *J Pers Soc Psychol.* 1986;50:992-1003.

Fonagy P, Bateman AW: Adversity, attachment, and mentalizing. *Compr Psychiatry* 64:59-66, 2016

Fonagy P, Campbell C: What touch can communicate: Commentary on “Mentalizing homeostasis: the social origins of interoceptive inference” by Fotopoulou and Tsakiris. *Neuropsychoanalysis* 19:39-42, 2017

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

Fonagy P, Luyten P: A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Dev Psychopathol* 21:1355-1381, 2009

Fonagy P, Luyten P, Allison E: Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *J Pers Disord* 29:575-609, 2015

Fonagy P, Luyten P, Allison E. Epistemic petrification and the restoration of epistemic trust: a new conceptualization of borderline personality disorder and its psychosocial treatment. *J Pers Disord.* 2015;29:575-609.

Fonagy P, Luyten P, Allison E, Campbell 2017

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

Friston KJ, 2015

Friston KJ: Self-evidencing babies: Commentary on “Mentalizing homeostasis: The social origins of interoceptive inference” by Fotopoulou & Tsakiris. *Neuropsychoanalysis* 19:43-47, 2017

Friston KJ, Stephan KE, Montague R, Dolan RJ: Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry* 1:148-158, 2014

Gergely G, Egyed K, Kiraly I. On pedagogy. *Dev Sci.* 2007;10:139-46.

Gibbon L, Nolte T, Fonagy P: Modelling Axis I and personality disorder symptomatology and its associations with childhood trauma and mentalizing. Manuscript in preparation, 2018

Gibbon, L., Fonagy, P. & Nolte, T. (in preparation). A review of the general psychopathology factor and its clinical implications.

Gross JJ. The emerging field of emotion regulation: an integrative review. *Rev Gen Psychol.* 1998;2:271-99.

Chapter 2. Mentalization-Based Therapy and Neuroscience

- Gu X, Hof PR, Friston KJ, Fan J: Anterior insular cortex and emotional awareness. *J Comp Neurol* 521:3371-3388, 2013
- Gunderson JG, Lyons-Ruth K. BPD's interpersonal hypersensitivity phenotype: a gene-environment-developmental model. *J Pers Disord.* 2008;22:22-41.
- Gunnar M, Quevedo K: The neurobiology of stress and development. *Annu Rev Psychol* 58:145-173, 2007
- Hauser TU, Fiore VG, Moutoussis M, Dolan RJ: Computational psychiatry of ADHD: Neural gain impairments across Marrian levels of analysis. *Trends Neurosci* 39:63-73, 2016
- Heinrichs M, Domes G: Neuropeptides and social behaviour: Effects of oxytocin and vasopressin in humans. *Prog Brain Res* 170:337-350, 2008
- Holmes, 2017
- Holmes J, Slade A: *Attachment in Therapeutic Practice*. London, Sage, 2018
- Hula A, Montague PR, Dayan P: Monte Carlo planning method estimates planning horizons during interactive social exchange. *PLoS Comput Biol* 11:e1004254, 2015
- Hula A, Vilares I, Dayan P, Montague PR: A model of risk and mental state shifts during social interaction. Manuscript in preparation, 2018
- Jones W, Klin A: Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature* 504:427-431, 2013
- Kalisch R, Müller MB, Tüscher O: A conceptual framework for the neurobiological study of resilience. *Behav Brain Sci* 38:e92, 2015
- Kalisch R, Muller MB, Tuscher O. A conceptual framework for the neurobiological study of resilience. *Behav Brain Sci.* 2015;38:e92.
- Kendler KS, Aggen SH, Czajkowski N, Roysamb E, Tambs K, Torgersen S, Neale MC, Reichborn-Kjennerud T. The structure of genetic and environmental risk factors for DSM-IV personality disorders: a multivariate twin study. *Archives of general psychiatry.* 2008;65:1438-46.
- Kim P, Strathearn L, Swain JE: The maternal brain and its plasticity in humans. *Horm Behav* 77:113-123, 2016
- Koenig MA, Harris PL. The role of social cognition in early trust. *Trends Cogn Sci.* 2005;9:457-9.
- Laceulle OM, Vollebergh WAM, Ormel J. The structure of psychopathology in adolescence: replication of a general psychopathology factor in the TRAILS study. *Clin Psychol Sci.* 2015;3:850-60.
- Lahey BB, Rathouz PJ, Keenan K, Stepp SD, Loeber R, Hipwell AE. Criterion validity of the general factor of psychopathology in a prospective study of girls. *J Child Psychol Psychiatry.* 2015;56:415-22.
- Lane JD, Harris PL. The roles of intuition and informants' expertise in children's epistemic trust. *Child Dev.* 2015;86:919-26.

Chapter 2. Mentalization-Based Therapy and Neuroscience

Lazarus RS, Folkman S. Stress, appraisal and coping. New York, NY: Springer; 1984.

Lecours S, Bouchard MA: Dimensions of mentalisation: outlining levels of psychic transformation. *Int J Psychoanal* 78:855-875, 1997

Lieberman MD: Social cognitive neuroscience: A review of core processes. *Annu Rev Psychol* 58:259-289, 2007

Linehan MM. Cognitive-behavioral treatment of borderline personality disorder. New York, NY: Guilford Press; 1993.

Lupien SJ, McEwen BS, Gunnar MR, Heim C: Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci* 10:434-445, 2009

Luyten P, Fonagy P: The neurobiology of mentalizing. *Personal Disord* 6:366-379, 2015

Luyten P, Fonagy P. Psychodynamic treatment for borderline personality disorder and mood disorders: a mentalizing perspective. In: Choi-Kain LW, G. GJ, editors. *Borderline personality and mood disorders: Comorbidity and controversy*. New York, NY: Springer; 2015. p 223-51.

Luyten P, van Houdenhove B, Lemma A, et al: A mentalization-based approach to the understanding and treatment of functional somatic disorders. *Psychoanal Psychother* 26:121-140, 2012

Marr D. Vision: a computational investigation into the human representation and processing of visual information. San Francisco, CA: W. H. Freeman; 1982.

Marty P: *Mentalisation et Psychosomatique*. Paris, Synthélabo, 1991

Mascaro O, Sperber D. The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*. 2009;112:367-80.

Mayes LC: A developmental perspective on the regulation of arousal states. *Semin Perinatol* 24:267-279, 2000

Meaney MJ: Epigenetics and the biological definition of gene x environment interactions. *Child Dev* 81:41-79, 2010

Montague PR, Berns GS, Cohen JD, et al: Hyperscanning: Simultaneous fMRI during linked social interactions. *Neuroimage* 16:1159-1164, 2002

Montague PR, Dolan RJ, Friston KJ, Dayan P: Computational psychiatry. *Trends Cogn Sci* 16:72-80, 2012

Moutoussis M, Fearon P, El-Deredy W, et al: Bayesian inferences about the self (and others): A review. *Conscious Cogn* 25:67-76, 2014

Moutoussis M, Shahar N, Hauser TU, Dolan RJ: Computation in psychotherapy, or how computational psychiatry can aid learning-based psychological therapies. *Comp Psych*;10.1162/CPSY_a_00014 2017

Chapter 2. Mentalization-Based Therapy and Neuroscience

Murray AL, Eisner M, Ribeaud D. The development of the general factor of psychopathology 'p factor' through childhood and adolescence. *J Abnorm Child Psychol*. 2016;44:1573-86.

Nanni V, Uher R, Danese A. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *Am J Psychiatry*. 2012;169:141-51.

Nolte T, Bolling DZ, Hudac CM, et al: Brain mechanisms underlying the impact of attachment-related stress on social cognition. *Front Hum Neurosci* 7:816, 2013

Nolte T, Bolling DZ, Hudac CM, Fonagy P, Mayes L, Pelphrey KA. Brain mechanisms underlying the impact of attachment-related stress on social cognition. *Front Hum Neurosci*. 2013;7:816.

Powers SI, Pietromonaco PR, Gunlicks M, Sayer A: Dating couples' attachment styles and patterns of cortisol reactivity and recovery in response to a relationship conflict. *J Pers Soc Psychol* 90:613-628, 2006

Recanati F. Can we believe what we do not understand? *Mind Lang*. 1997;12:84-100.

Rudrauf D: Structure-function relationships behind the phenomenon of cognitive resilience in neurology: Insights for neuroscience and medicine. *Adv Neurosci* 2014:1-28, 2014

Rudrauf D, Bennequin D, Granic I, et al: A mathematical model of embodied consciousness. *J Theor Biol* 428:106-131, 2017

Rumelhart DE, McClelland JL. *Parallel distributed processing*. Cambridge, MA: MIT Press; 1986.

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

Rutherford HJ, Wallace NS, Laurent HK, Mayes LC: Emotion regulation in parenthood. *Dev Rev* 36:1-14, 2015

Sander D, Scherer KR: *The Oxford Companion to Emotion and the Affective Sciences*. Oxford, Oxford University Press, 2009

Schulkin J: Social allostasis: Anticipatory regulation of the internal milieu. *Front Evol Neurosci* 2:111, 2011

Scott KM, McLaughlin KA, Smith DA, Ellis PM. Childhood maltreatment and DSM-IV adult mental disorders: comparison of prospective and retrospective findings. *Br J Psychiatry*. 2012;200:469-75.

Seth AK: Interoceptive inference, emotion, and the embodied self. *Trends Cogn Sci* 17:565-573, 2013
Seth, 2014

Shafto P, Eaves B, Navarro DJ, Perfors A. Epistemic trust: modeling children's reasoning about others' knowledge and intent. *Dev Sci*. 2012;15:436-47.

Shai D, Belsky J: When words just won't do: Introducing parental embodied mentalizing. *Child Dev Perspect* 5:173-180, 2011

Chapter 2. Mentalization-Based Therapy and Neuroscience

Sharp C, Fonagy P, Allen JG. Posttraumatic stress disorder: a social-cognitive perspective. *Clin Psychol Sci Pract.* 2012;19:229-40.

Sharp C, Ha C, Carbone C, Kim S, Perry K, Williams L, Fonagy P. Hypermentalizing in adolescent inpatients: treatment effects and association with borderline traits. *J Pers Disord.* 2013;27:3-18.

Sharp C, Pane H, Ha C, Venta A, Patel AB, Sturek J, Fonagy P. Theory of mind and emotion regulation difficulties in adolescents with borderline traits. *J Am Acad Child Adolesc Psychiatry.* 2011;50:563-73.

Sharp C, Wright AGC, Fowler JC, Frueh BC, Allen JG, Oldham J, Clark L. Borderline personality pathology as the 'g' factor of personality disorder. *J Abnorm Psychol.* 2015;124:387-98.

Siever LJ, Davis KL. A psychobiological perspective on the personality disorders. *Am J Psychiatry.* 1991;148:1647-58.

Simic M, Fombonne E. Depressive conduct disorder: Symptom patterns and correlates in referred children and adolescents. *J Affect Dis.* 2001;62:175-85.

Southwick SM, Bonanno GA, Masten AS, Panter-Brick C, Yehuda R. Resilience definitions, theory, and challenges: interdisciplinary perspectives. *Eur J Psychotraumatol.* 2014;5.

Sperber D, Wilson D. *Relevance: communication and cognition.* 2nd ed. Malden, MA: Blackwell; 1995.

Sperber D, Clement F, Heintz C, Mascaro O, Mercier H, Origg G, Wilson D. Epistemic vigilance. *Mind Lang.* 2010;25:359-93.

Sroufe LA, Rutter M. The domain of developmental psychopathology. *Child Dev.* 1984;55:17-29.

Stephan KE, Mathys C: Computational approaches to psychiatry. *Curr Opin Neurobiol* 25:85-92, 2014

Stephan KE, Mathys C, 2015

Stephan KE, Schlagenhaut F, Huys QJM, et al: Computational neuroimaging strategies for single patient predictions. *Neuroimage* 145:180-199, 2017

Stern Y: Cognitive reserve. *Neuropsychologia* 47:2015-2028, 2009

Teicher MH, Samson JA. Childhood maltreatment and psychopathology: a case for ecophenotypic variants as clinically and neurobiologically distinct subtypes. *Am J Psychiatry.* 2013;170:1114-33.

Torgersen S, Lygren S, Oien PA, Skre I, Onstad S, Edvardsen J, Tambs K, Kringlen E. A twin study of personality disorders. *Compr Psychiatry.* 2000;41:416-25.

Torgersen S, Czajkowski N, Jacobson K, Reichborn-Kjennerud T, Roysamb E, Neale MC, Kendler KS. Dimensional representations of DSM-IV cluster B personality disorders in a population-based sample of Norwegian twins: a multivariate study. *Psychol Med.* 2008;38:1617-25.

Vrticka P, Sander D, Anderson B, et al: Social feedback processing from early to late adolescence: Influence of sex, age, and attachment style. *Brain Behav* 4:703-720, 2014

Chapter 2. Mentalization-Based Therapy and Neuroscience

Wagner S, Baskaya O, Dahmen N, Lieb K, Tadic A. Modulatory role of the brain-derived neurotrophic factor Val66Met polymorphism on the effects of serious life events on impulsive aggression in borderline personality disorder. *Genes Brain Behav.* 2010;9:97-102.

Wagner S, Baskaya O, Lieb K, Dahmen N, Tadic A. The 5-HTTLPR polymorphism modulates the association of serious life events (SLE) and impulsivity in patients with borderline personality disorder. *J Psychiatr Res.* 2009;43:1067-72.

White CN, Gunderson JG, Zanarini MC, Hudson JI. Family studies of borderline personality disorder: a review. *Harv Rev Psychiatry.* 2003;11:8-19.

Widiger TA, Lynam DR, Miller JD, Oltmanns TF. Measures to assess maladaptive variants of the five-factor model. *J Pers Assess.* 2012;94:450-5.

Wilson DS. Evolution on the level of communities. *Science.* 1976;192:1358-60.

Wilson DS, Wilson EO. Rethinking the theoretical foundation of sociobiology. *Q Rev Biol.* 2007;82:327-48.

Winnicott DW: The theory of the parent-infant relationship. *Int J Psychoanal* 41:585-595, 1960

Xiang T, Ray D, Lohrenz T, et al: Computational phenotyping of two-person interactions reveals differential neural response to depth-of-thought. *PLoS Comput Biol* 8:e1002841, 2012

Zanarini MC, Barison LK, Frankenburg FR, Reich DB, Hudson JI. Family history study of the familial coaggregation of borderline personality disorder with axis I and nonborderline dramatic cluster axis II disorders. *J Pers Disord.* 2009;23:357-69.

Debbane & Nolte (2018). The Neurobiology of Mentalising (in press). In Fonagy, P. & Anthony Bateman (Eds): *Handbook of Mentalizing in Mental Health practice*, 2nd edn.

Fonagy, P. and Luyten, P., 2009. A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and psychopathology*, 21(4), pp.1355-1381.

Gibbon, L., Nolte, T. & Fonagy, P. (in preparation). Modelling axis I and personality disorder symptomatology and its associations with childhood trauma and mentalizing.

Gibbon, L., Fonagy, P. & Nolte, T. (in preparation). A review of the general psychopathology factor and its clinical implications.

Chapter 2. Mentalization-Based Therapy and Neuroscience

Shneidman, L. and Woodward, A.L., 2016. Are child-directed interactions the cradle of social learning?. *Psychological bulletin*, 142(1), p.1.