How does psychotherapy work? The self and its disorders†

Gwen Adshead & Peter Fonagy

**Summary**

There is now extensive evidence that different types of psychological therapy are effective for various psychiatric disorders, both in terms of symptom reduction and in improving the patient’s capacity to function adaptively. In two articles, we set out some current ideas of how psychological therapies might 'work', drawing on neurobiological studies of brain plasticity and therapeutic efficacy. In this first article, we review the evidence that disorders of self-experience are common to many psychiatric disorders, and that psychological therapies 'work' on those disorders. The second article will appear in the next issue of *Advances*.

**Declaration of interest**

None.

Modern man (and woman) has been called ‘Homo psychologicus’ (Humphrey 1984): one who takes the experience of the self seriously. Historically, the self was understood principally in terms of its relationship with the body or with a divine creator, and the experiences of the self were described in stories, poems, songs and drama. However, the 19th century saw the self, and its disorders, become the object of medical discourse, which facilitated the development of psychological treatments.

Since the 20th century, pharmacological treatment of mental disorders has been expanding at an enormous rate, especially from the 1950s onwards (Herzberg 2008). However, the number of people who are referred for psychological therapies also continues to increase (Ofson 2002). In the UK, governmental policy has supported Improving Access to Psychological Therapies (IAPT). At the same time, senior psychotherapist posts are being abolished by local services on the grounds that long-term psychotherapies are not cost-effective compared with brief interventions that can be delivered by professionals who require less training (and therefore less pay).

In this first of two articles on psychological therapies, we draw on results from neuropsychiatry and animal studies to suggest that...
using very different imaging modalities (including 
fluodeoxyglucose positron emission tomography, 
single-photon emission computed tomography, 
functional magnetic resonance imaging and 
xenon-enhanced computed tomography) with 
patients presenting with a range of psychiatric 
diagnoses (including OCD, major depressive 
disorder, schizophrenia, phobias, post-traumatic 
stress disorder and personality disorder). The 
psychological therapies studied have included 
behavioural, cognitive–behavioural, interpersonal, 
group, cognitive rehabilitation, eye movement 
desensitisation and reprocessing (EMDR) and 
dynamic therapies. The areas of brain function 
shown to be altered by psychological therapies are 
listed in Table 1.

The evidence so far supports Roth & Fonagy’s 
(2004) conclusions that most major modalities 
can be effective in producing change in certain 
groups of patients, but that different techniques 
probably operate on different parts of the neu­ 
ronal system. Most neuroimaging research has 
involved study of cognitive–behavioural therapies 
for relatively discrete conditions, such as phobias 
or OCD, although Roffman and colleagues’ review 
included studies of the effect of interpersonal 
reflective therapies (Roffman 2005). As yet there 
are no published imaging studies of the major 
therapeutic orientations, such as transference­ 
focused therapy or mentalisation-based treatment, 
that have been shown to be effective with more 
complex disorders, although reports of such 
 studies are forthcoming.

The results of studies of the neuronal effect of 
psychological therapies are consistent with the 
increasing evidence of brain plasticity through­ 
n neuronal change after a change in memory 
status, i.e. that learning something new produced 
changes in synaptic function through a change in 
the expression of a gene for a synaptic peptide.

It seems obvious that learning produces change 
in junior developing brains, but research such as 
that of Kandel and Ramachandran & Blakeslee 
(2000), working with phantom limbs, suggest that 
plasticity is also present in adult brains. It may be 
harder, and take longer, but old dogs can learn 
new tricks, and their brains change when they do.

**Psychological distress and disorders of the self**

Patients who are referred to National Health 
Service psychotherapy services are often distressed 
and have significant levels of psychiatric morbidity 
(Chiesa 2007; Scott 2008). Compared with patients 
managed in primary care, they have more self­ 
harming behaviours and delusional symptoms and 
are more likely to have diagnoses of personality 
disorder. They also show higher levels of risk than 
primary care patients (Barkham 2005).

What might these patients have in common? 
As yet there has been no satisfactory answer to 
this question, although almost all agree that the 
essential characteristics of diagnoses such as 
personality disorder are not well articulated 
by the DSM diagnostic categories (Dahl 2008). 
Phenomenologically, perhaps the most critical 
feature is the disturbed sense of identity frequently 
described in a range of personality disorders (e.g. 
Blatt 1988). Fuchs (2007) linked the failure of 
these patients to establish a coherent self-concept 
to something akin to a ‘post-modernist’ stance that 
they appear to adopt in their lives, switching from 
one present to the next, always totally identified 
with their present state of affect. Jorgensen (2006) 
has linked this identity disturbance to social, 
organisational and cultural factors rather than 
individual history or personality structure. We 
have argued that a dysfunction or deficit in a sense 
of agency or self-directedness is a critical aspect 
of these problems (Fonagy 2009), which has been 
consistently identified in empirical clinical studies 
(e.g. Bender 2007).

The self is, among other things, a neuro­ 
psychiatric phenomenon, which can be studied 
using neuroimaging methods. Neuroimaging 
 studies have consistently supported the assumption 
that envisioning the mind of another is underpinned 
by the same brain systems as those that identify 
one’s own thoughts and feelings (Lieberman 2007; 
The common circuitry used in mentalising self 
and others may explain why it is that even the 
normally developing child will find it a struggle 
to acquire a sense of his own mind as separate 
and distinct from the minds of others, and can 
help us to understand the self–other confusions in 
disorders such as personality disorder, which may 
be associated with the disruption of these neural 

**Table 1** Brain areas affected by psychological 
therapies

<table>
<thead>
<tr>
<th>Therapy</th>
<th>Brain area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive therapy</td>
<td>Dorsal prefrontal cortex, medial prefrontal cortex</td>
</tr>
<tr>
<td>Reflective therapy</td>
<td>Anterior cingulate cortex, posterior cingulate cortex, precuneate nucleus, insula</td>
</tr>
<tr>
<td>Behavioural therapy</td>
<td>Amygdala</td>
</tr>
</tbody>
</table>
Different levels of complexity of thought:
- Memory and consistency of self-experience over time
- Self in groups or social self
- A theory of mind, including self-concept and self-recognition, and other concept and other recognition

**Domain of self-experience**

<table>
<thead>
<tr>
<th>Disorder of self-experience</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disorders of thinking</td>
<td>Rumination, cognitive distortion, delusions, obsessional activity</td>
</tr>
<tr>
<td>Disorders of memory</td>
<td>Post-traumatic stress disorder, amnésias, borderline states</td>
</tr>
<tr>
<td>Impaired theory of mind and self-recognition</td>
<td>Autism, Disorders of consistency and integrity of the self (e.g., borderline personality disorder), Disorders of the recognition of the bodily self (e.g., eating disorders, Capgras syndrome, attacks on the body), Disorders of self-esteem (e.g., after child maltreatment)</td>
</tr>
<tr>
<td>Impaired social self</td>
<td>Avoidant, paranoid and antisocial personality</td>
</tr>
<tr>
<td>Impaired relational self</td>
<td>Dyadic problems leading to repetitive relationship breakdown, Problems in parenting or caregiving relationships</td>
</tr>
</tbody>
</table>

**BOX 1 Phenomenology of self-experience**

- Different levels of complexity of thought:
  - First-order thinking (‘I am writing’)
  - Second-order thinking (‘I am wanting to write and think’)
  - Third-order thinking (‘I am thinking about wanting to think and write’)
- Memory and consistency of self-experience over time
- A theory of mind, including self-concept and self-recognition, and other concept and other recognition
- Self in groups or social self

(Kircher 2000)

**The self and the brain**

Self-representation involves the reptilian brain, the limbic cortex and the neocortex, and the different disorders of the self can be linked with specific neuronal areas. Executive function is known to involve the lateral prefrontal cortex, and self-experience has also been linked to the medial prefrontal cortex. Self-recognition involves the right limbic, left prefrontal and superior temporal cortex; the left prefrontal cortex has also been identified as being involved in integration of experience and a sense of agency. Memory functions are complex and are functionally organised in different parts of the brain. Damage to hippocampal and temporal regions produces obvious memory abnormalities, but more subtle pathological processes (such as uncontrollable memories that intrude into present consciousness or non-conscious personal semantic memories that affect behaviour) may be generated by different and additional areas.

**TABLE 2 Disorders of self-experience**

<table>
<thead>
<tr>
<th>Domain of self-experience</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disorders of thinking</td>
<td>Rumination, cognitive distortion, delusions, obsessional activity</td>
</tr>
<tr>
<td>Disorders of memory</td>
<td>Post-traumatic stress disorder, amnésias, borderline states</td>
</tr>
<tr>
<td>Impaired theory of mind and self-recognition</td>
<td>Autism, Disorders of consistency and integrity of the self (e.g., borderline personality disorder), Disorders of the recognition of the bodily self (e.g., eating disorders, Capgras syndrome, attacks on the body), Disorders of self-esteem (e.g., after child maltreatment)</td>
</tr>
<tr>
<td>Impaired social self</td>
<td>Avoidant, paranoid and antisocial personality</td>
</tr>
<tr>
<td>Impaired relational self</td>
<td>Dyadic problems leading to repetitive relationship breakdown, Problems in parenting or caregiving relationships</td>
</tr>
</tbody>
</table>

Reviews of the neuroimaging literature suggest that two distinct neural networks are shared by self-knowing and knowing others (Lieberman 2007; Uddin 2007). The first system involves a more body-based, frontoparietal mirror-neuron system which is involved in understanding the multimodal embodied self (e.g., face and body recognition) and understanding others through motor-simulation mechanisms (Gallese 2004; Rizzolatti 2004). This suggests that a fundamental process that allows us to appreciate the actions and emotions of others involves the activation of the mirror-neuron system for actions and the activation of visceromotor centres for the understanding of affect. This is thought to be one of the key evolutionary mechanisms underpinning social empathy – knowing from the inside, as it were, how someone else feels. This is an implicit, automatic system, providing physical other-to-self and self-to-other mapping, which is involved in the immediate understanding (or misunderstanding) of self and others.

The second, the cortical midline system, which consists of the medial prefrontal cortex, the anterior cingulate cortex and the temporal parietal junction in the lateral parietal cortex (Lieberman 2007; Uddin 2007), appears to play a central role in this process. This system is less body based, and processes information about the self and others in more abstract and symbolic ways (Frith 2007; Uddin 2007). Importantly, unlike the frontoparietal system, it appears to be mainly shaped across development by interpersonal relationships.

Kircher & David (2000) describe a phenomenology of self-experience (Box 1), which includes:

- different levels of thought (first-order thinking, second-order thinking about the first-order thoughts, and third-order thinking about the thinking process itself);
- memory and consistency of self over time;
- self-concept and self-recognition that involves having a theory of mind, not only of oneself, but by extension of others: this phenomenon has been widely studied in conditions such as autism and schizophrenia;
- self in groups or social self, which is a function of the evolutionary pressure on primates to have the capacity to make successful group attachments (Dunbar 2003).

**Disorders of the self**

This phenomenology of self-experience naturally suggests a phenomenology of disorders of self-experience (Table 2). Although the manifestations of such disorders may be complex and multifaceted,
one common thread across disorders of self-experience might be that they all represent different types of failures in the mind’s capacity to represent its own activities and contents. Manifestations of such failures might include disorders of thinking (including the classic thought disorder, which is manifest in speech made incoherent by odd words, grammar or syntax); disorders of memory; disorders of consistency and integrity of self (as commonly described in borderline personality disorder); disorders of self-concept and image (most notably found in eating disorders, but may also be one of the consequences of sexual abuse); and disorders of self-recognition (found in disorders such as Capgras syndrome, but may be one cause of self-harming behaviours). Disorders of a social or relational self may be localised to one form of relationship (usually intimate dyads) or may be more global (as is seen in antisocial personality disorder). All these instances have in common a confusion in the mind over the meaning, significance or value of its own activities.

There are other models of the self and personality which may be helpful when thinking about how change occurs in psychotherapy. McAdams (1996) describes three levels of mental organisation: the personality (which is a function of behavioural traits that are largely genetic and is comparatively stable); the self (which is individual and consists of values, beliefs and attitudes that can be modified); and identity (which is constructed in social narratives with others and therefore changes in relationships over time).

Duggan (2004) reviews the issue of change in personality and concludes similarly that there are some basic aspects of personality that do not change very much, but that characterological adaptations (McAdams’ “self”) are amenable to change.

From a bioethical point of view, it may also be helpful to think of psychiatric disorders as disorders of autonomy of will, thought and action, which can have a disastrous effect on the capacity to function effectively, and produce subjective distress (Gillon 1985; Adshead 2008).

Last, proper attention needs to be given to the affective self, which regulates and modulates emotions and is clearly disordered in a number of psychiatric conditions (Taylor 1997; Panksepp 1998). A common characteristic of patients referred for psychotherapy is that they are in psychological pain, as a result of a variety of negative affects that are dysregulated and seem overwhelming. Psychological pain, like physical pain, is modulated through a complex system of neuronal pathways, involving the hippocampus, thalamus and neocortex.

The self, mentalisation and its disorders

The process by which images of self or newly accessible memories are practised, considered and reflected on, is one aspect of what is called mentalisation (Bateman 2004) or metacognition (Dimaggio 2011). A key aspect of this process is self-reflective function, which operates at a number of levels. People with low levels of reflective function think in concrete ways and make poor-quality judgements about the mental states of both themselves and others. At even lower levels of reflective function, individuals cannot articulate their own self-experience at all, and other people’s experiences are not real to them.

Mentalising is an aspect of self-regulation; it helps an individual make sense of what is in their own mind and the mind of others, and facilitates recognition of what is self and what is not. Mentalising is a function of affective arousal, and the capacity to maintain certain types of mentalisation depends on the individual’s capacity to regulate arousal. This capacity, in turn, is affected by early attachment experiences and temperament that, as it were, ‘set’ the internal homeostatic regulator (Nolte 2011). It is important to emphasise that automatic mentalisation is ongoing at all times, and that it is both relationship and situation specific (Fonagy 2011). Thus, even the best mentalisers may have gaps or lapses in the process.

Advances in neuroscience, developmental, social and cognitive research, and accumulating clinical experience have enabled us (Fonagy 2009) to construct a more differentiated picture of mentalisation. This is based on four polarities (Box 2) which must be balanced appropriately to specific situations for mentalisation to be fit for purpose (the interpersonal and self-organisational function for which it was designed).

**Box 2** The four polarities of mentalisation

1. Implicit or automatic vs explicit or controlled mentalisation (Satpute 2008)
2. Mentalisation based on external vs internal cues about the internal state of self and others (Lieberman 2007)
3. Cognitive vs affective mentalisation (Shamay-Tsoory 2008)
4. A balance between the two distinct neural networks:
   - Self-knowing: located in the medial prefrontal cortex, the anterior cingulate cortex and the precuneus (Frith 2006, 2007)
   - Other knowing: located in the frontoparietal mirror-neuron system (Rizzolatti 2006)
Although the pattern of limitations in mentalising capacity may differ between individuals and across diagnostic conditions, we suggest that in most severe disorders which involve the personality, imbalanced mentalisation on one or other of the polarities mentioned in Box 2 would be evident in adults with clinical mental disorder. Mentalisation, as measured by reflective function, is low in patients with personality disorder, depression and severe mental illness (Levinson 2006; Levy 2006). Failure of the mentalisation process may be of childhood or adult onset, sustained or episodic, generalised or discrete to persons, situations or even topics. In all these contexts and levels, the imbalance may be mild or severe. Only rarely is mentalisation failure associated with threat, hostility or predatory activity.

**Why might poor mentalising arise?**

The answer is as complex as any ‘final common pathway’ in a neurodevelopmental process can be (Cicchetti 2006). We see mentalisation failure as the consequence of a series of risk events during development. Significant vulnerability risk factors will include the early attachment environment and the availability of a caregiver with an interest in the infant’s mental state, who through marked mirroring will establish robust self-state representations in the infant (Fonagy 2002). This process is of course moderated by constitutional factors such as infant temperament: an infant who is dominated by negative affect will be harder to respond to sensitively with the activation of the attachment system (Strathearn 2008) and will be less likely to elicit peripheral oxytocin (Stratheann 2009), which is known to enhance sensitivity to internal states (Domes 2007) and facial cues (Guastella 2008).

**Adverse experiences**

Adverse early experiences (which might include experiences of neglect, hostility, chronic fear, abandonment or actual pain infliction) result in a developmental environment which fails to stimulate proper dendritic growth or hyper-stimulates growth in immature neuronal networks in the executive cortex (Leckman 2004). There is evidence to suggest that mentalising failure associated with insecure attachment arises from subtle disorders of brain function. Ramachandran & Blakeslee (2000) argue that poor mentalising function, as revealed in immature psychiatric defences, is associated with abnormal right brain function. Functional magnetic resonance imaging studies of adults with insecure attachment show different patterns of activation compared with secure adults (Buchheim 2008; Vrticka 2008; Strathearn 2009). It is also possible that gene × environment interactions that increase the risk of developing psychiatric disorder are mediated by insecure attachment status (Caspers 2009).

**Risk factors**

The list of activating or provoking risk factors will undoubtedly include maltreatment, trauma and chronic stress, which undermine the language environment within which the child might acquire mentalisation (Lemche 2007; de Rosnay 2008). Their disruptive effects are maintained by mechanisms that can be more parsimoniously described in neurobiological terms, such as distortions in the functioning of arousal systems (Arnsten 1998; Mayes 2006), anomalous functioning of the prefrontal cortex (Cicchetti 2005) or even deficiencies in oxytocin levels (Heim 2008). Although an appropriate level of activation of the brain system mediating attachment appears to be both psychologically and biologically key to the development of mentalising (Fonagy 2007), imaging evidence and theorisation suggest that the hyperactivation of the attachment system undermines the capacity to think about mental states (Bartels 2004) (after all, love is known to be blind), along with other emotion-induced cognitive dysfunctions.

The so-called formation risk factors are likely to create the brain–behaviour environment within which mentalisation is difficult. The association between mentalisation, stress and attachment suggests that we should expect differences in the quality of mentalisation depending on the quality of the relationship within which it is observed (Allen 2008; Fonagy 2011). The specific condition which is likely to be triggered in the treatment of severe personality disorder is characterised by abnormally dense retrieval of negative memories and cognitions and an inhibition of judgements of social trustworthiness, paranoid thoughts and acute mentalising failure, as well as the re-emergence of pre-mentalising forms of subjectivity that are typical of the toddler.

However, a number of interacting systems are involved in the deterioration of psychological capacities, making it hard to discern causal sequences. For example, mentalisation (understanding the intent behind one’s own or others’ reactions) can exert influence over affect regulation, especially negative affects, so adults whose mentalising is lacking in one of its components frequently also have poor control over negative feelings. Emotion dysregulation may also be a key reason for mentalising failure. Momentary failures
Enhanced emotion regulation

Improved effortful control over maladaptive interpersonal schemata

Improved behavioural coping strategies in troubled interpersonal encounters

Limitation of the impact of negative self-experience by being able to take a third-person position

References


Levy KN, Meehan KB, Kelly KM, et al (2006) Change in attachment patterns and reflective function in a randomized control trial of


Linden DEJ (2006) How psychotherapy changes the brain: the contribu-


psychopharmacology 34: 2655–66.


### MCQs
Select the single best option for each question stem.

**1 Psychological therapies:**
- a. have been shown to be ineffective
- b. are all based on psychoanalytic theories of mind
- c. are equally effective for all conditions
- d. do not work on the brain
- e. need to be matched to the appropriate condition.

**2 Patients referred to psychotherapy services:**
- a. are best described as ‘the worried well’
- b. do not have significant psychiatric pathology
- c. rarely have personality disorder
- d. are at less risk of harm than primary care patients
- e. may struggle with a disturbed sense of self.

**3 The following brain areas have not been implicated in normal self-functioning:**
- a. medial prefrontal cortex
- b. lateral prefrontal cortex
- c. insula
- d. occipital cortex
- e. anterior cingulate cortex.

**4 Which of the following pairs of disorders is matched accurately with the disorder of self-experience?**
- a. anorexia nervosa and disorders of thought
- b. OCD and disintegration of self
- c. schizophrenia and attacks on the body
- d. post-traumatic stress disorder and disorders of memory
- e. depression and thought insertion.

**5 Enhanced mentalisation does not affect:**
- a. cognitive performance
- b. arousal regulation
- c. affect regulation
- d. behavioural control
- e. the ability to take a third-person perspective.