Chapter 5

The paradoxical self

Vilayanur Ramachandran and William Hirstein

Summary

We consider a number of syndromes incorporating paradoxical phenomena that lie at the boundary between neurology and psychiatry. Amongst the phenomena we examine are Cotard's Syndrome (belief that one is dead/dying), Capgras Syndrome (belief that a personally familiar person has been replaced by an imposter), and Apotemnophilia (desire to have a limb amputated). We use these phenomena to speculate on the manner in which the brain constructs a sense of self. We propose that, despite the extraordinary variety of paradoxical symptoms encountered in neuropsychiatry, certain key assumptions can help explain most of these self-related phenomena. (1) Discrepancies and conflict between the information dominating different brain systems. (2) Disturbance of Me/Other distinctions caused by dysfunctional interactions between the mirror neuron system, frontal lobe structures and sensory input. (3) Misattribution of symptoms to spurious causes, so as to minimize internal discrepancies. (4) The existence of three functionally distinct visual systems, as opposed to the two conventionally accepted ones, with selective damage or uncoupling between them. (5) Recruiting one neural map for another unrelated function, or one neural structure serving as a template for transcribing on to another neural structure. We suggest that, paradoxically, the mechanisms that give rise to psychiatric delusions and illusions may themselves sometimes have adaptive value in evolutionary terms.

Introduction

An approach we have pursued in our laboratory involves exploring precisely those phenomena that have long been regarded as paradoxical or anomalous, that do not fit the overall framework of science as currently practised, or that appear to violate established conventional assumptions. Neurology and psychiatry are full of such examples. What can be more paradoxical than a person with Cotard's Syndrome denying his own existence, when the very denial implies existence? Conditions such as Cotard's Syndrome, and many others, have shown that the self is not the monolithic entity it often believes itself to be. Our unitary sense of self may well be an illusion that incorporates distinct components, each of which may be studied separately. If it is an illusion, it is not enough to merely state that fact; we need to explain how it arises. An interesting question is, for example, whether it is an adaptation acquired through natural selection (cf. McKay and Dennett, 2009).

Our emphasis in this chapter will be on 'borderline' syndromes that straddle the boundary between neurology and psychiatry, with a focus on delusions directly or indirectly

The Paradoxical Brain, ed. Narinder Kapur. Published by Cambridge University Press.

involving the concept of self (Hirstein and Ramachandran, 2009). Consider the following disorders that illustrate different aspects of self.

1. Cotard's Syndrome – a patient claiming he is dead or that he does not exist (Pearn and Gardner-Thorpe, 2002).
2. Capgras Delusion – a patient claims that his mother looks like his mother but is in fact an imposter (Sinkman, 2008).
3. Apotemnophilia – an otherwise apparently normal person develops an intense desire to have his arm or leg amputated (Money *et al.*, 1977).
4. A person with temporal lobe epilepsy (TLE) claims to see or feel God (Devinsky and Lai, 2008).
5. Somatoparphrenia – a patient with right superior parietal lobule (SPL) damage and insula damage claims his left arm doesn't belong to him (Berlucchi and Aglioti, 1997; Bisiach and Geminiani, 1991).
6. Tactile hyperempathy in amputees – a patient who has had his arm amputated watches another's arm being touched. Astonishingly, he feels the touch on his own phantom arm, an extreme form of 'tactile empathy' (Ramachandran and Brang 2009; Case and Ramachandran, 2010) – the barrier between him and others has been dissolved.
7. Out-of-body experiences – following a right hemisphere stroke, a patient reports floating out into space and being able to see her own body. Paradoxically, she says that she feels the sensation of pain 'in that other body down below' but not in herself (Blanke *et al.*, 2004).

Can we resolve these paradoxes of consciousness and self using our knowledge of neuroanatomy? Here we offer proposals, some still speculative, to help to resolve these paradoxes.

Three visual pathways

Our scheme, introducing three parallel visual pathways (Figure 5.1), emerged initially from our attempt to deal with the Capgras Syndrome. In this disorder, an apparently lucid patient claims, for example, that his mother is an imposter – a woman who *looks* identical to his mother but is not (Capgras and Reboul-Lachaux, 1923; Young *et al.*, 1994; Devinsky, 2009).

To understand this disorder we will start with the notion that there are three functionally distinct, quasi-autonomous intra-cortical visual pathways in man (see Jeannerod and Jacob, 2005 for a discussion of the two-pathway model). We will mainly focus on the feed-forward features of these pathways, recognizing that they may also have important feed-backward components – e.g. the amygdala feeds back to the temporal cortex (Amaral and Price, 1984), allowing, for example, for the response of the fusiform gyrus to be modified by emotional faces (Hung *et al.*, 2010).

Pathway 1 ('dorsal stream') is well known – it goes from the superior colliculus to the parietal lobe via the pulvinar and is involved in navigation and spatial vision (e.g. avoiding obstacles, reaching for objects, stepping over branches, dodging missiles). It is not involved in identifying objects, this is achieved by pathways 2 and 3.

Focusing on the most salient of objects, the human face is initially processed in the fusiform gyrus. Here, faces are discriminated, dropped into 'bins' (as being different faces) and tagged; in much the same way that an entomologist might classify and label

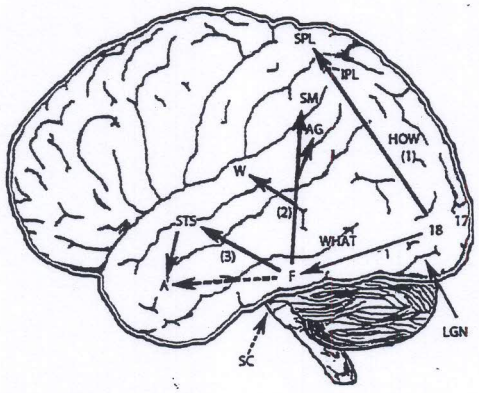

Figure 5.1 Diagram of postulated pathways (1), (2) and (3), and related neural structures. In this schematic representation, locations are approximate and readers should be aware that structures such as the amygdala, lateral geniculate nucleus and superior colliculus are deep within the brain and not on its surface. The superior colliculus has direct connections with the parietal lobe. The HOW pathway (1) is also likely to represent WHERE information in relation to any particular stimulus. 17 and 18 refer to Brodmann areas. A, amygdala; AG, angular gyrus; F, fusiform gyrus; IPL, inferior parietal lobule; LGN, lateral geniculate nucleus; SC, superior colliculus; SM, supramarginal gyrus; SPL, superior parietal lobule; STS, superior temporal sulcus; W, Wernicke's area.

hundreds of different butterflies without knowing anything about them. The fusiform face area projects to dozens of higher areas but, crudely speaking, the output segregates into two pathways (which we call 2 and 3). One of these, *Pathway 2* (the semantic pathway), evokes a halo of memories associated with the face, and has an inferior, ventral route that terminates in the anterior temporal lobe. *Pathway 3* (the emotional pathway) is involved in the early detection of biologically salient (e.g. terrifying, sexually evocative) objects. Pathway 3 gives you that jolt of familiarity when you recognize important people, such as your mother or spouse. It projects mainly via the superior temporal sulcus (STS) and amygdala to evoke both subjective manifestations (fear, joy, familiarity) and physiological manifestations (sweating, heart racing) of the emotional response, mediated through the hypothalamic output. This superior temporal sulcus is rich in mirror neurons (Aziz-Zadeh *et al.*, 2006; Noordzij *et al.*, 2009), cells that are active during the planning of an action and when that action is observed in others (Di Pellegrino *et al.*, 1992; see further below). Here they respond to changing facial expressions and biological motion. Pathway 3 is functionally distinct from Pathway 2, although perhaps with some overlap. The degree to which Pathway 3 is anatomically segregated from Pathway 2 remains to be seen.

Pathway 3 is also connected (via STS) to the insular cortex, which has cells that both *respond* to changing facial expressions of disgust and also fire when the individual *produces* the corresponding expression. These could be regarded as 'emotional expression' mirror neurons that require cross-modal abstraction of information from the motor system about one's own face-muscle twitches and visually perceived facial features of others. We suggest that the insula's proximity to left hemisphere language areas makes it ideally suited for transforming literal expressions of disgust (e.g. facial expression of disgust to faeces) to metaphorical ones ('he is disgusting' or 'that was a tasteless remark'). Our bold prediction is that such verbal expressions would be incomprehensible to someone with insular damage.

The insula also receives information about the relative locations of our limbs, our orientation with respect to gravity, and from our internal viscera – in other words, somatic data that help keep the 'self' anchored in the body. If these data are discrepant with those from other systems, parts of our body may simply not feel 'right'. Perhaps for this reason, somatoparaphrenia (the delusional denial that a body part is one's own) has sometimes been observed following insular damage (Vallar and Ronchi, 2009).

Interactions between Pathways 2 and 3

If the amygdala's role in Pathway 3 is to imbue familiarity and emotional salience to faces and other objects, what is its role in Pathway 2? We can speculate that, together with the hippocampus, it may draw on accumulating semantic and contextual associations to direct attention to particular faces and objects marked as salient (familiar, attractive, distorted). The amygdala is a highly differentiated structure and it's also possible that Pathways 2 and 3 initially activate different portions of it before salient inputs get 'transferred' from Pathway 2 to Pathway 3.

This scheme requires the additional, somewhat radical, postulate that a neural algorithm that is initially set up in one brain region can, with repeated stimulation, serve as a template to entrain another brain region which initially has only crude responses. Thus, non-automatic and initially slow responses to complex stimuli in Pathway 2 become transcribed onto Pathway 3, which would then start responding automatically and rapidly to the trained stimuli – thereby making it emotionally salient and familiar.

Summary of pathways

In summary, the amygdala receives both direct projections from fusiform (Pathway 2) and indirect ones via STS (Pathway 3) (Amaral and Price, 1984; Freese and Amaral, 2006; Smith *et al.*, 2009). Pathway 2 is responsible for evoking emotions, including familiarity. From a phylogenetic standpoint, it evolved as an early warning system with circuits 'hardwired' to quickly detect eyes, faces, etc. (in the generic sense), snakes, bananas, water (Rolls, 1999) and sexually attractive characteristics – in short, biologically salient stimuli. This system is characterized by rapid processing of a small but important class of specific inputs (visual and auditory) leading to rapid – almost reflexive – action. Response to breasts, for example, is hardwired in this pathway, manifesting as the mammo-ocular reflex in humans (S. M. Anstis, personal communication). More complex objects – e.g. mother or father – are initially processed by circuits in Pathway 2 (acquired by associative learning). The result is sophisticated recognition but slower response. However, with repeated stimulation, according to our speculation, responses to such complex stimuli get 'transferred' to Pathway 3; so 'wife' becomes as salient as water, breasts or snakes, in addition to being imbued with familiarity. Whether originating from hardwired circuits in Pathway 3 or acquired initially in Pathway 2 and 'transferred' to Pathway 1, the emotions cascade from the basolateral amygdala via the amygdalofugal pathways and stria terminalis to generate hypothalamic outflow – influencing a range of responses such as fighting, fleeing, feeding, reproduction, hormonal regulation, subjective emotions, etc.

Obviously these anatomical flow diagrams are a gross oversimplification. However, a simplified picture is not a bad place to start, especially given that it has already proved to be of heuristic value in explaining aspects of psychopathology such as misidentification syndromes (see below).

Capgras Syndrome and abhorrence of discrepancy

We suggested that the Capgras Syndrome, where a familiar person has been replaced by an imposter, arises because the fusiform area becomes disconnected from the STS and amygdala, but remains connected to relevant semantic processes (as represented in structures such as Wernicke's area and the left angular gyrus): a face is recognized and a penumbra of associated memories activated, but it does not evoke emotion – there is no jolt of

familiarity. Our patient, David, on recognizing his mother, is therefore forced to rationalize away his curious sensations by concluding that she is an imposter. We tested this idea by measuring changes in skin sweating (skin conductance response, SCR) when David looked at photographs. For healthy volunteers, photos of their mothers typically evoke reliable SCR in a way that pictures of strangers do not. In David, there was no difference, supporting the disconnection hypothesis (Hirstein and Ramachandran, 1997; cf. Ellis *et al.*, 1997). It is worth noting similarities with a theory proposed by Young *et al.* (1993). A difference is that, in their scheme, the disconnection was between Pathways 1 and 2 (in our terms), rather than between Pathways 2 and 3.

One might ask: why doesn't a patient with Capgras Syndrome merely say, 'I know it *is* my mother, but she doesn't feel familiar for some reason', instead of saying she is an imposter or even (in some cases) developing paranoid ideation about the imposter's intent? To account for this, we introduce one of our key postulates – abhorrence of discrepancies between outputs of brain modules and/or a lack of consistency between expected and actual input (cf. Coltheart *et al.*, 2009). In Capgras Syndrome, the patient is confronted with a peculiar dilemma. He 'recognizes' his mother from a semantic standpoint, but fails to do so from an emotional standpoint. This creates a paradox, a peculiar discrepancy in his brain that he is unable to interpret or convey and which produces distress or even paranoia. Additionally, there may also be some damage to frontal (especially right frontal) structures, which makes individuals more prone to delusions (Ramachandran and Blakeslee, 1998; Hirstein and Ramachandran, 1997).

David also had difficulty abstracting across successive encounters of a new person seen in different contexts. Without the flash of recognition he ought to have experienced in the second, third or *n*th exposure, we hypothesize that he was unable to bind these experiences together to form an enduring identity for the other person. Even more remarkably, this may also have applied to his experience of himself across successive episodes. He would, for example, refer to 'The other David who is on vacation' as if, unable to bind these experiences together, his conclusion was that he existed in a form of temporal multiplicity.

Prosopagnosia and Capgras

A specific difficulty in recognizing faces despite intact basic visual processes is called prosopagnosia. It seems at first peculiar that a person with such 'face blindness' can nevertheless show a relatively normal emotional response to familiar faces, as indexed, for example, by changes in skin conductance (Bauer, 1984). However, we have all had an inkling of it when we run into someone completely out of context, someone we 'know' well (e.g. an ex-student, dentist, etc.). He looks completely familiar yet his identity eludes you – you do not know who he is and cannot quite place him. As we shall see later, the fact that such a dissociation can occur at all in healthy people suggests some separation between recognition and familiarity, although it could be that familiarity is a necessary precursor for recognition. Capgras Syndrome, in which people recognize others but do not experience familiarity, provides the crucial 'double dissociation' suggesting complete independence of process (Young *et al.*, 1994). In our scheme, this double dissociation is explained by the partial segregation between the semantic and emotional Pathways (2 and 3): damage to Pathway 2 leads to no semantic recognition but normal emotional responses (prosopagnosia), whereas damage to Pathway 3 leads to preserved recognition but anomalous emotional responses (Capgras).

A strange feature of the Capgras Syndrome is that it can be modality-specific. A patient may claim that his mother is an imposter when he sees her, but has no delusion when he talks to her on the phone (Ramachandran and Hirstein, 1998). We suggested that this might be caused by selective damage to Pathway 3 with simultaneous preservation of input from the auditory cortex to the amygdala.

Apotemnophilia

Apotemnophilia refers to a condition in which an apparently neurologically normal person has had – from early childhood – an intense desire to have a limb amputated, often going to the extreme of actually having it removed (Money *et al.*, 1977; Bayne and Levy, 2005; First, 2005). What causes this?

Based on several lines of evidence we proposed that the disorder has a neurological rather than a purely 'psychological' basis. We will mention two sets of evidence. First, the patient can use a pen to precisely draw the line along which he desires amputation. Second, as in somatoparaphrenia (limb ownership denial), the left limb is more commonly involved than the right. We therefore suggested and provided the first experimental evidence that the right superior parietal lobe (SPL) is implicated in the condition (Brang *et al.*, 2008). An identical theory, based on the same line of reasoning, was later proposed by Blanke *et al.* (2009).

The SPL in the right hemisphere has a complete representation of the body (body image or 'body schema'). It is probably topographically organized and receives its input from several primary sensory areas. Although malleable, through experience, the map in SPL is probably hard-wired (for example, those who have never had a limb due to congenital malformation may nevertheless experience a 'phantom' of that limb; Ramachandran and Blakeslee, 1998). We recently postulated (Ramachandran and McGeoch, 2007) and found evidence that, in apotemnophiliacs, the arm (or leg) portion is congenitally absent from the SPL body map (Brang *et al.*, 2008). Since the sensory input from the leg is intact but there is no corresponding recipient zone in SPL, the result is once again a discrepancy that leads to stressful aversion. Indeed, we have found that touching the affected limb produces a much higher SCR than touching the normal limb (McGeoch *et al.*, 2007) and substantially *lower* activation in SPL, as measured by MEG (magnetoencephalography; Ramachandran *et al.*, 2008). On the other hand, if the sensory input itself is lost (as in brachial plexus avulsion), then there is no activation of either S1 or SPL so there is no discrepancy and no desire for amputation. The same holds for a limb that has had deafferentation from stroke. Finally, in somatoparaphrenia, both S1/S2 and SPL are damaged in the right hemisphere. Since there is no discrepancy between S1 and SPL, no desire for amputation emerges.

One might predict, therefore, that if the limb is anaesthetized in an apotemnophiliac (and the arm made to appear to be visually missing using an optical trick) there might be a temporary alleviation of the desire for amputation. Merely occluding the limb may not suffice – one may need to make it *look* amputated. (For absence of evidence for the limb is not the same thing as evidence of absence.)

We recently invoked a similar theory to account for the intense desire that male to female transsexual men develop for amputation of the penis (Ramachandran and McGeoch, 2007). We suggested that these men have no representation of the penis in SPL, so they develop a form of specific 'apotemnophilia' caused again by the discrepancy between body image and afferent input. Conversely, we have seen (Ramachandran and McGeoch, 2007) that a majority of female to male transsexuals who have been asked, report having had

a phantom penis ever since childhood and request genital reassignment surgery to correct the mismatch between external anatomy and the map in SPL.

Apotemnophila and sexual preferences

One curious aspect of apotemnophilia that is unexplained by our model is the associated sexual inclinations in some individuals, namely a desire for intimacy with another amputee. These sexual overtones are probably what misled people to propose a Freudian view of the disorder.

We postulate that one's sexual 'aesthetic preference' for certain body morphology, even across the obvious sexual differences between men and women, is dictated in part by the shape of one's own body image, hardwired into the right superior parietal, and possibly insular cortex (Ramachandran *et al.*, 2009). Expanding on this, we suggest that there is a genetically specified mechanism that allows a template of one's body image to act on limbic connections thereby determining aesthetic visual preference for one's own body image 'type'. Consequently, if a person with apotemnophilia has an arm missing in his internal (genetically hardwired) body image, then that would affect his limbic circuits and explain his sexual affinity for amputees. The pathways that enable limbic structures to determine visual aesthetic preference remain obscure, but there are known back-projections from basolateral amygdala to almost every stage of extrastriate visual processing. (These pathways may also be involved in art appreciation; patients with Capgras Syndrome sometimes lose interest in art and natural beauty.)

This notion of the circuitry in one brain region serving as a template for being transcribed onto another region may have more general applicability in helping us understand brain development and function.

Mirror neurons in psychopathology

We will now consider two aspects of self that are considered almost axiomatic. First, its essentially private nature. You can empathize with someone, but never to the point of experiencing his sensations or dissolving into him (except in pathological states like *folie à deux* and romantic love). Second, it is aware of its own existence. A self that negates itself is an oxymoron. Yet both these axioms can collapse in disease, without affecting other aspects of self. For example, an amputee can literally feel his phantom being touched when he merely watches a normal person being touched (Ramachandran and Brang, 2009). A person with Cotard's Syndrome will deny that he exists, claiming that his body is merely an empty shell. Explaining these disorders in neural terms can help illuminate how the *normal* self is constructed, especially its peculiar recursive quality.

To account for some of these syndromes we need to again invoke mirror neurons (Di Pellegrino *et al.*, 1992; see reviews by Rizzolatti and Craighero, 2004 and Iacoboni, 2009). Neurons in the prefrontal cortex send signals down the spinal cord that orchestrate skilled and semi-skilled movements such as putting food in your mouth, pulling a lever, pushing a button, etc. These are 'ordinary' motor command neurons, but some of them, known as mirror neurons, also fire when you merely *watch* another person perform a similar act. It's as if the neuron (more strictly the network of which the neuron is part) was using the visual input to do a sort of 'virtual reality simulation' of the other person's actions, allowing you to empathize with her and view the world from her point of view. We have previously speculated that these neurons cannot only help simulate other people's behaviour, but

can also be turned 'inward' – as it were – to create second-order representations or 'meta-representations' of your *own* earlier first-order brain processes. This could be the neural basis of introspection, and of the reciprocity of self-awareness and other awareness. This complements Humphrey's ingenious hypothesis (Humphrey, 1978) that the selection pressure for the emergence of introspection in humans came from the need to model and predict other people's behaviour. There is obviously a chicken-or-egg question here as to which evolved first, but that is tangential to our main argument. The main point is that the two co-evolved, mutually enriching each other to create the mature representation of self that characterizes modern humans. Our ordinary language illustrates this, as when we say, 'I feel a bit self-conscious', when we really mean that I am conscious of others being conscious of me. Or when I speak of being self-critical or experiencing self-pity. It is arguable whether an ape can experience pity (as when responding to a beggar), but it is almost certainly incapable of self-pity, let alone *knowing* that it is engaging in self-pity.

We also suggest that, although these neurons initially emerged in our ancestors to adopt another's allocentric *visual* point of view, they evolved further in humans to enable the adoption of another's *metaphorical* point of view ('I see the idea from his point of view', etc.). Our idea that metaphorical thinking is parasitic on, and has evolved from, early brain processes is similar to Lakoff and Johnson's (1980) ingenious speculations. We are merely trying to flesh it out by referring it to actual brain structures. Just as we might model another's metaphorical point of view via structures developed for the prediction of motor acts, we suggest that many gestures produced by the motor system may be 'metaphorical' echoes of concepts. For example, when trying to convey a conceptually precise point we oppose thumb and index finger in a 'precision grip' unique to humans. Or we clench our fist literally to metaphorically convey that we have 'come to grips' with the situation.

As we have seen, there are also 'touch mirror neurons' that fire not only when your skin is touched, but when you watch someone else being touched (Keysers *et al.*, 2004). This raises an interesting question – how does the neuron know what the stimulus is? Why doesn't the activity of these neurons lead you to literally experience the touch delivered to another person? There are two answers. First, the tactile receptors in your skin tell the other touch neurons in the cortex (the non-mirror neurons) that they are *not* being touched and this null signal selectively vetoes some of the outputs of mirror neurons. This would explain why our amputee experienced touch sensations when he watched our student being touched, a phenomenon that we have dubbed 'tactile hyperempathy' – the amputation had removed the vetoing (Ramachandran and Brang, 2009). Astonishingly, the same kind of referral is seen immediately following a brachial plexus block (Case and Ramachandran, 2010).

A second reason why your mirror neurons do not lead you to mime everyone you watch or to literally experience their tactile sensations might be that your frontal lobes send feedback signals to partially inhibit the mirror neurons' output (cf. Brass *et al.*, 2009). It cannot completely inhibit them, otherwise there would be no point having mirror neurons in the first place. As expected, if the frontal lobes are damaged, you *do* start imitating people ('echopraxia'). Recent evidence suggests that there may also be mirror neurons for pain, disgust, facial expression – perhaps for all outwardly visible expression of emotions. Some of these are in the anterior cingulate, others in the insula.

We suggest that many otherwise inexplicable neuropsychiatric symptoms may arise from flaws in the three-way interactions between the mirror neuron system, frontal structures and external sensory input, systems that are normally in dynamic equilibrium. These circuits allow you to have deep empathy while at the same time maintaining your

distinctiveness – your brain holds on to two parallel representations. Where this goes wrong, 'you–me' confusion and impoverished ego-differentiation would occur. Our group has seen strong preliminary hints that autistic children have a paucity of mirror neurons (Altschuler *et al.*, 1997; Oberman *et al.*, 2005; Oberman and Ramachandran, 2007) which would not only explain their poor imitation, empathy and 'pretend play' (which requires role-playing), but also why they sometimes confuse the pronouns 'I' and 'You', and have difficulty with introspection.

At this juncture, we should point out that some scholars have recently raised questions about the significance of mirror neurons for cognitive functions like theory of mind, and indeed about their very existence in humans (Hickok, 2008; Lotto *et al.*, 2009; Lingnau *et al.*, 2009). Some of this, we would argue, may be a manifestation of 'neuron envy' – a deep-seated fear of reductionism. This is irrational because mirror neurons are not meant to replace the theory of mind any more than DNA replaces heredity. On the contrary, the mirror neuron system (MNS) provides a mechanism by which theory of mind can be instantiated in the brain and might have emerged in evolution. A second criticism is that there is too much hype surrounding it, but the existence of media hype in itself does not invalidate a theory. A third criticism is that although the MNS exists in monkeys, it has not been unequivocally shown to exist in humans; the brain imaging techniques are inherently unreliable. However, absence of evidence is not evidence of absence – to a biologist it would be strange to suggest that a system of neurons demonstrated in monkeys would have suddenly disappeared in humans; in fact, the default position is the opposite (we never went through a phase when people claimed that Hubel and Wiesels' discovery of orientation selective cells is applicable only to monkeys). A fourth criticism is that the neurons with MNS-like properties are simply the result of Hebbian associative conditioning, that there's nothing special about them. There are two answers to this. First, even if the statement were true, it would not detract from their importance for understanding brain function. The whole brain could be set up through associative conditioning, but no one would argue from this that studying the actual circuitry in the brain is useless for our understanding of how the brain works. The second answer is that if it is all a matter of associative learning then why are only a *subset* of V5 neurons (or S2 neurons) – about 20% – capable of this? Why not the other 80%? You cannot argue that only 20% are wired up genetically to acquire MNS-like properties – that just takes you back to square one. If we are right about MNS deficiency in autism spectrum disorder (ASD) then one might be able to devise new therapies to tap into any residual MNS function and enhance or rejuvenate their function. For example, one could create multiple mirror images of a 'trainer' dancing to a rhythm and have the child synchronize with the multiple images. This might revive MNS function and its benefits might spill over into other more cognitive domains.

Cotard's Syndrome

Let us return to Cotard's Syndrome – the ultimate paradox of the self-negating its own existence (sometimes claiming 'I am dead'). We postulate that this arises from a combination of two lesions, resulting in something akin to derealization. First, a lesion that is analogous to that which causes the Capgras familiarity impairment, but which is far more pervasive. Instead of emotions being disconnected just from visual centres, they are disconnected from all sensations and even memories of sensations. So, rather than just familiar people, the entire world becomes an imposter, it feels unreal. Second, there may be

dysfunctional interaction between the mirror neurons and frontal inhibitory structures leading to a dissolution of the sense of self as being distinct from others (or indeed from the world). Lose the world and lose yourself – and it's as close to death as you can get.

Geschwind Syndrome and hyper-religiosity

Now, imagine if these same circuits that were outlined above became *hyper*active. This can happen during seizures originating in the temporal lobes (TLE or temporal lobe epilepsy). The result would be an intense heightening of the patient's sensory appreciation of the world (Ramachandran *et al.*, 1997) and intense empathy for all beings, to the extent of perceiving no barriers between himself and the cosmos. In that this may be interpreted as a religious experience, it is perhaps not surprising that hyper-religiosity can occur in a disorder particularly associated with TLE, Geschwind Syndrome (Benson, 1991). Other features of the syndrome can include hypergraphia (excessive written output), altered sexuality (often reduced sexual drive), and a degree of viscosity/'stickiness' in interpersonal interactions (inappropriate over-attachment). Whilst retrospective diagnosis from selective contemporary accounts is always dangerous, arguments have been made that some of history's religious leaders have had TLE. We hasten to add that the fact that fervent belief can arise in the context of TLE neither refutes nor supports the existence of God(s).

Let us turn now to out-of-body experiences (Easton *et al.*, 2009). Even a normal person, perhaps such as yourself, can at times adopt a 'detached' allocentric stance, but this does not become a full-blown delusion because other neural systems keep you anchored. However, damage to the right fronto-parietal regions or ketamine anaesthesia (which may influence the same circuits) removes the inhibition and you can feel that 'you' have left your body, even to the extent of not feeling your own pain (Muetzelfeldt *et al.*, 2008). You see your pain 'objectively' as if someone else was experiencing it. Out-of-body experiences can also be produced by using a system of parallel multiple-reflecting mirrors (Altschuler and Ramachandran, 2007) or video cameras (Ehrsson, 2007). This raises the possibility of using such systems to allow patients with chronic pain – such as fibromyalgia – to dissociate themselves from their pain (Ramachandran and Rogers-Ramachandran, 2008), mimicking ketamine anaesthesia.

Unity and discrepancy

The purported unity or internal consistency of self is also a myth, or highly tenuous at best. Most patients with left arm paralysis caused by right hemisphere stroke complain about it as, indeed, they should. However, a subset of patients who have additional damage to the 'body image' representation in the right SPL (superior parietal lobule; and possibly also the insula) claim that their paralysed left arm does not belong to them (Critchley, 1953). The patient may assert that it belongs to his father or spouse (as if he had a selective 'Capgras' for his arm). Such syndromes challenge even basic assumptions such as 'I am anchored in this body' or 'This is my arm'. They suggest that 'belongingness' is a primal brain function hardwired through natural selection because of its obvious selective advantage to our hominid ancestors. It makes one wonder if a Californian with this disorder would deny ownership of (or damage to) the left fender of his car and ascribe it to his mother's car.

There appears to be almost no limit to this. An intelligent and lucid patient we saw recently claimed that her own left arm was not paralysed and that the lifeless left arm on her lap belonged to her father who was 'hiding under the table'. Yet when we asked her to touch

her nose with her left hand she used her intact right hand to grab and raise the paralysed left hand – using the latter as a 'tool' to touch her nose! Clearly somebody in there knew that her left arm was paralysed and that the arm on her lap was her own, but 'she' – the person we were talking to – did not.

Finally, we have speculated that some aspects of pain in phantom limbs may arise from discrepancies between visual input and corollary discharge to the (now) missing limb. Restoring visual/motor congruence using visual feedback ('mirror visual feedback', MVF) reduces pain in many patients possibly by restoring the congruence (Ramachandran *et al.*, 1995; Ramachandran and Hirstein, 1998; Ramachandran nd Altschuler, 2009). Even more intriguingly, we (Altschuler *et al.*, 1999) found that MVF could produce substantial recovery of function in the paralysed arm, following stroke.

Discrepancy between sensory and motor signals may also be involved in producing the excruciating, intractable pain that arises in reflex sympathetic dystrophy (RSD) or complex regional pain syndrome (CRPS). In 1996, we proposed (Ramachandran, 1996) using mirrors for treating this condition and recent double-blind placebo-controlled clinical trials have shown striking reduction of pain (Cacchio *et al.*, 2009).

Hemispheric specialization

Confabulation refers to a memory (or report of a memory) for events that did not happen, but which feel to the reporter to be true. There is evidence from a number of sources that language centres in the left hemisphere tend to 'patch up' discrepancies in sensory input (or indeed between inputs and pre-existing stable beliefs) by engaging in confabulation (Gazzaniga and LeDoux, 1978). In fact, we have suggested (Ramachandran, 1998; Ramachandran and Blakeslee, 1998; Pettigrew and Miller 1998) that many of the so-called Freudian defences (of which confabulation is one example; others include 'projection' rationalization, 'reaction formation', etc.) are mainly attempts by the more 'action'-oriented left hemisphere to confer stability and coherence on behaviour – you do not want to orient to every small discrepancy. On the other hand, if this is overdone, it can lead to delusions even of manic proportions (it is hardly ever adaptive to deny poverty and overuse a credit card). We therefore postulated, albeit a gross over-simplification, a 'devil's advocate' in the right hemisphere that constantly questions the status quo instead of clinging to it (Ramachandran, 1998). Thus, damage to the right frontoparietal region destroys the devil's advocate leading to florid delusions such as denial of paralysis and even denial of arm ownership (when the STS and right insula are also damaged) tinged with aversion because of the discrepancy, as noted above.

In the final section of this chapter, we consider how the patterns that we have seen in neurological illness, including paradoxical denial of the self, may illuminate general human responses to extreme conditions, anxiety problems such as panic attacks, and depression.

Panic attacks, depersonalization and derealization

In emergency situations, such as during an assault, people can experience significant alterations in state, including a sense of depersonalization, in which they feel apart or separated from the outside world. These responses can sometimes be set off by no obvious external threat, and the result is a panic attack. For example, if the neural circuitry responsible for a reflexive response to an emergency is triggered accidentally – caused by a mini TLE-type 'seizure' – it sets in motion an autonomic storm: your heart rate and blood

pressure go up and you start sweating. You continue to be vigilant as well, and the anterior cingulate mediates this. But this time there is no tangible external danger to ascribe it to. Once again, an inexplicable discrepancy is set up leading to acute distress. So you ascribe the physiological changes to an impending heart attack or some other inexplicable danger – a foreboding of death that is vague yet deeply disturbing.

If this theory is correct, it immediately suggests a therapy for panic attacks. Since the main problem is a discrepancy arising from the absence of a target, one could substitute a false target. During the premonition of the attack, the patient could view a short horrifying video clip on, say, his iPhone. On the assumption that a tangible threat (anxiety that can be attributed to a specific external trigger) is less disturbing than an intangible one, the film clip may abort the attack. The fact that you know 'It's only a film clip' may not matter much – no more than knowing you are merely watching horror movie prevents you from vicariously 'enjoying' the horror. Now consider what would happen if the same circuitry starts malfunctioning on an extended timescale rather than just a few seconds. There is an increase in vigilance combined with no external cause and (possibly) lack of emotions as well (Sierra and Berrios, 2001). Additionally, there may be dysfunctional MNS causing poor ego differentiation and recursive representation of self (shades of Cotard's Syndrome here). Once again the organism – the brain – is confronted with a peculiar set of discrepancies that it would never ordinarily encounter; hence the patient concludes that the world does not exist or is not real (derealization), or that he does not exist or is not real (depersonalization). In desperate attempts to regain their sense of being anchored in reality, or indeed in their bodies, such patients may deliberately self-harm.

Self/other confusion

One of our propositions is that many neuropsychiatric disturbances arise from dysfunction of self/other differentiation, which in turn is based on deranged interactions between MNS and frontal inhibitory circuits. The result is a range of behaviour spanning the gradient from normalcy to pathology; from (1) *lack of empathy* – seen in some autism spectrum disorders; (2) *sympathy* – I sympathize with your plight, but maintain the clinical detachment of a surgeon; (3) *empathy* – I actually adopt your world view, to gain the deeper appreciation towards a close friend; (4) *hyperempathy* – a profound sense of feeling the other's emotional responses; (5) *pathological empathy* – experiencing the other's qualia as your own (e.g. a patient feeling his phantom being touched when watching someone else being touched); (6) *cosmic union* – a sense of relatedness with all beings and with the cosmos (God delusions).

Future challenges and questions

Some of the ideas and formulations outlined in this chapter are speculative, but we hope nevertheless that it will help generate testable predictions and new therapies and provide a starting point for a more comprehensive explanation of self-related paradoxes in neuropsychiatry.

One issue that needs to be addressed in future research is the role of individual differences in the manifestation of these syndromes. If there is an explanation in terms of particular networks and pathways, why do we not see these syndromes more often? Could it be the case that the syndromes will only become apparent in individuals with a particular set of past experiences and repositories of knowledge, and/or individuals who

happen to have particular goals that they need to achieve, and that the syndromes emerge due to pressures from these two factors?

In the last decade there has been a tremendous resurgence of interest among neuroscientists in the nature of consciousness and self. The goal is to explain specific details of certain complex mental capacities in terms of equally specific activity of specialized neural structures. Just as the functional logic of heredity (pigs give birth to pigs, not donkeys) was mapped precisely on to the structural logic of the DNA molecule (the complementarity of the two strands), we need to map the mental phenomena on to specialized neural structures and their interactions.

As we have seen, one way to achieve this might be to explore inherently paradoxical syndromes that lie at the interface between neurology and psychiatry, using them to illuminate both normal and abnormal mental functions. We have used this strategy to provide a conceptual framework for thinking about neuropsychiatric disturbances. Many of the ideas have been proposed by others, but the particular synthesis we present may have some novel features.

Acknowledgements

The ideas we have expressed in this article are an attempt to provide a novel synthesis, but rely a great deal on the work of many researchers – most notably Francis Crick, Pat Churchland, Antonio Damasio, Nick Humphrey, Joe LeDoux, Orrin Devinsky, Haydn Ellis, Andrew Young, Mauricio Berrios, M. Sierra, Jack Pettigrew and many others.

References

Altschuler, E. L., & Ramachandran, V. S. (2007). A simple method to stand outside oneself. *Perception*, **36**: 632–4.

Altschuler, E. L., Vankov, A., Wang, V., Ramachandran, V. S. & Pineda, J. A. (1997). *Person see, person do: human cortical electrophysiological correlates of monkey see monkey do cells* (poster session presented at the 27th annual meeting of the society for neuroscience, New Orleans, LA).

Altschuler, E. L., Wisdom, S. B., Stone, L., *et al.* (1999). Rehabilitation of hemiparesis after stroke with a mirror. *Lancet*, **353**: 2035–6.

Amaral, D. G., & Price, J. L. (1984). Amygdalo-cortical projections in the monkey (*Macaca fascicularis*). *The Journal of Comparative Neurology*, **230**: 465–96.

Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J., & Iacoboni, M. (2006). Lateralization of the human mirror neuron system. *Journal of Neuroscience*, **26**: 2964–70.

Bauer, R. (1984). Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the Guilty Knowledge Test. *Neuropsychologia*, **22**: 457–69.

Bayne, T., & Levy, N. (2005). Amputees by choice: body integrity disorder and the ethics of amputation. *Journal of Applied Philosophy*, **22**: 75–86.

Benson, D. (1991). The Geschwind Syndrome. *Advances in Neurology*, **55**: 411–21.

Berlucchi, G., & Aglioti, S. (1997). The body in the brain: neural bases of corporeal awareness. *Trends in Neuroscience*, **20**: 560–4.

Bisiach, E., & Geminiani, G. (1991). Anosognosia related to hemiplegia and hemianopia. In: Priganto, G. P. and Schacter, D. L. (Eds.). *Awareness of Deficit After Brain Injury: Clinical and Theoretical Issues*. Oxford: Oxford University Press.

Blanke, O., Landis, T., Spinelli, L., & Seeck, M. (2004). Out-of-body experience and autoscopy of neurological origin. *Brain*, **127**: 243–58.

Blanke, O., Morgenthaler, F. D., Brugger, P., & Overney, L. S. (2009). Preliminary evidence for a fronto-parietal dysfunction in able-bodied participants with a desire for a limb

amputation. *Journal of Neuropsychology*, **3**: 181–200.

Brang, D., McGeoch, P., & Ramachandran, V. S. (2008). Apotemnophilia: a neurological disorder. *Neuroreport*, **19**: 1305–06.

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **364**: 2359–67.

Cacchio, A., De Blasis, E., Necozione, S., di Orio, F., & Santilli, V. (2009). Mirror therapy for chronic complex regional pain syndrome type 1 and stroke. *New England Journal of Medicine*, **361**: 634–6.

Capgras, J., & Reboul-Lachaux, J. (1923). L'illusion des 'sosies' dans un délire systématisé chronique. *Bulletin de la Société Clinique de Medicine Mentael*, **11**: 6–16. Reprinted in H. D. Ellis, J. Whitley and J. P. Luauté (Eds.). (1994) Delusional misidentification: the three original papers on the Capgras, Fregoli and intermetamorphosis delusions. *History of Psychiatry*, **5**: 117–46.

Case, L., & Ramachandran, V. S (2010). Immediate interpersonal and intermanual referral of tactile sensation following anesthetic block of the brachial plexus. *Archives of Neurology*, **57**: 1521–3.

Coltheart, M., Menzies, P., & Sutton, J. (2009). Abductive inference and delusional belief. *Cognitive Neuropsychiatry*, **15**: 1–27.

Critchley, M. (1953). *The Parietal Lobes*. London: Edward Arnold.

Devinsky, O. (2009). Delusional misidentifications and duplications: right brain lesions, left brain delusions. *Neurology*, **72**: 80–7.

Devinsky, O., & Lai, G. (2008). Spirituality and religion in epilepsy. *Epilepsy and Behavior*, **12**: 636–43.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, **91**: 176–80.

Easton, S., Blanke, O., & Mohr, C. (2009). A putative implication for fronto-parietal connectivity in out-of-body experiences. *Cortex*, **45**: 216–27.

Ehrsson, H. H. (2007). The experimental induction of out-of-body experience. *Science*, **317**: 1048.

Ellis, H., Young, A., Quayle, A., & De Pauw, K. (1997). Reduced autonomic responses to faces in Capgras Delusion. *Proceedings of the Royal Society, Biological Sciences*, **264**: 1085–92.

First, M. B. (2005). Desire for an amputation of a limb: paraphilia, psychosis, or a new type of identity disorder. *Psychological Medicine*, **35**: 919–28.

Freese, J. L., & Amaral, D. G. (2006). Synaptic organization of projections from the amygdala to visual cortical areas TE and V1 in the Macaque monkey. *The Journal of Comparative Neurology*, **496**: 655–67.

Gazzaniga, M., & LeDoux, J. (1978). *The Integrated Mind*. New York, NY: Springer.

Hickok, G. (2008). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, **21**: 1229–43.

Hirstein, W., & Ramachandran, V. (2009). He is not my father, and that is not my arm: accounting for misidentification of people and limbs. In: Hirstein, W. (Ed.). *Confabulation. Views from Neuroscience, Psychiatry, Psychology and Philosophy*. Oxford: Oxford University Press, pp. 109–38.

Hirstein, W., & Ramachandran, V. S. (1997). Capgras syndrome: a novel probe for understanding the neural representation and familiarity of persons. *Proceedings of the Royal Society of London*, **264**: 437–44.

Humphrey, N. (1978). Nature's psychologists. In Ramachandran, V. S. and Josephson, B. (Eds.). *Consciousness and the Physical World*. Oxford: Pergamon Press.

Hung, Y., Smith, M. L., Bayle, D. J., *et al.* (2010). Unattended emotional faces elicit early lateralized amygdala–frontal and fusiform activations. *Neuroimage*, **50**: 727–33.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, **60**: 653–70.

Jeannerod, M., & Jacob, P. (2005). Visual cognition: a new look at the two-visual systems model. *Neuropsychologia*, **43**: 301–12.

Keysers, C., Wicker, B., Gazzola, V., Anton, J. L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, **42**: 335–46.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Lingnau, A., Gesierich, B., & Caramazza, A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences*, **106**: 9925–30.

Lotto, A., Hickok, G. S., & Holt, L. L. (2009). Reflections of mirror neurons and speech perception. *Trends in Cognitive Sciences*, **13**: 110–4.

McGeoch, P. D., Brang, D., & Ramachandran, V. S. (2007). Apraxia, metaphor and mirror neurons. *Medical Hypotheses*, **69**:1165–8.

McKay, R., & Dennett, D. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, **32**: 493–510.

Money, J., Jobaris, R., & Furth, G. (1977). Apotemnophilia: two cases of self-demand amputation as a paraphilia. *Journal of Sex Research*, **13**: 115–25.

Muetzelfeldt, L., Kamboj, S. K., Rees, H., *et al.* (2008). Journey through the K-hole: phenomenological aspects of ketamine use. *Drug and Alcohol Dependence*, **95**: 219–29.

Noordzij, M. L., Newman-Norlund, S. E., de Ruiter, J. P., *et al.* (2009). Brain mechanisms underlying human communication. *Frontiers in Human Neuroscience*, **3**: 14.

Oberman, L. M., & Ramachandran, V. S. (2007). The stimulating social mind: the role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychological Bulletin*, **133**: 310–27.

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., & Ramachandran, V. S. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, **24**: 190–8.

Pearn, J., & Gardner-Thorpe, C. (2002). Jules Cotard (1840–1899): his life and the unique syndrome that bears his name. *Neurology*, **58**: 1400–03.

Pettigrew, J. D., & Miller, S. M. (1998). A 'sticky' interhemispheric switch in bipolar disorder? *Proceeding of the Royal Society of London, Series B, Biological Sciences*, **265**: 2141–8.

Ramachandran, V. S. (1996). What neurological syndromes can tell us about human nature: some lessons from phantom limbs, Capgras syndrome, and anosognosia. *Cold Spring Harbor Symposia on Quantitative Biology*, **61**: 115–34.

Ramachandran, V. S. (1998). Consciousness and body image. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **353**: 1851–9.

Ramachandran, V. S., & Altschuler, E. L. (2009). The use of visual feedback, in particular mirror visual feedback, in restoring brain function. *Brain*, **132**: 1693–710.

Ramachandran, V. S., & Blakeslee, S. (1998). *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. New York, NY: William Morrow.

Ramachandran, V. S., & Brang, D. (2009). Sensations evoked in patients with amputation from watching an individual whose corresponding intact limb is being touched. *Archives of Neurology*, **66**: 1281–4.

Ramachandran, V. S., & Hirstein, W. (1998). The perception of phantom limbs. The D. O. Hebb lecture. *Brain*, **121**: 1603–30.

Ramachandran, V. S., & McGeoch, P. D. (2007). Occurrence of phantom limb genitalia after gender reassignment surgery. *Medical Hypotheses*, **69**: 1001–03.

Ramachandran, V. S., & Rogers-Ramachandran, D. (2008). Sensations referred to a patient's phantom arm from another subject's intact arm: perceptual correlates of mirror neurons. *Medical Hypotheses*, **70**: 1233–4.

Ramachandran, V. S., Brang, D., McGeoch, P. D., & Rosar, W. (2009). Sexual and food preference in apotemnophilia and anorexia: interactions between 'beliefs' and 'needs'

regulated by two-way connections between body image and limbic structures. *Perception*, **38**: 775–7.

Ramachandran, V. S., McGeoch, P. D., & Brang, D. (2008). *Apotemnophilia: a neurological disorder with somatotopic alterations in SCR and MEG activation.* Society for Neuroscience Meeting, Washington, D.C.

Ramachandran, V. S., Rogers-Ramachandran, D., & Cobb, S. (1995). Touching the phantom limb. *Nature*, **377**: 489–90.

Ramachandran, V. S., Vilayanur, S., Hirstein, W. S., Armel, K. C., Tecoma, E., & Iragul, V. (1997). The neural basis of religious experience (paper presented at the 27th Annual Meeting of the Society for Neuroscience, New Orleans, LA).

Rizzolatti, G., & Craighero, L. (2004). The mirror neuron system. *Annual Reviews of Neuroscience*, **27**: 169–92.

Rolls, E. T. (1999). Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, **9**: 467–80.

Sierra, M., & Berrios, G. E. (2001). The phenomenological stability of depersonalisation: comparing the old with the new. *Journal of Nervous and Mental Disorders*, **189**: 629–36.

Sinkman, A. (2008). The syndrome of Capgras. *Psychiatry*, **71**: 371–8.

Smith, C. D., Lori, N. F., Akbudak, E., *et al.* (2009). MRI diffusion tensor tracking of a new amygdalo-fusiform and hippocampo-fusiform pathway system in humans. *Journal of Magnetic Resonance Imaging*, **29**: 1248–61.

Vallar, G., & Ronchi, R. (2009). Somatoparaphrenia: a body delusion. A review of the neuropsychological literature. *Experimental Brain Research*, **192**: 533–1.

Young, A. W., Leafhead, K. M., & Szulecka, T. K. (1994). The Capgras and Cotard delusions. *Psychopathology*, **27**: 226–31.

Young, A. W., Reid, I., Wright, S., & Hellawell, D. J. (1993). Face-processing impairments and the Capgras delusion. *British Journal of Psychiatry*, **162**: 695–8.