Introduction: the self and neuroscience

Tilo Kircher[1] and Anthony S. David[2]

[1] Department of Psychiatry, University of Tübingen, Germany
[2] Institute of Psychiatry and Maudsley Hospital, London, UK

Who are we and what makes us who we are? Like our world, our self is a construction of our minds. But we do not live in isolation. The self is also a construction of our relations with other selves. And most intriguingly, the self is a construction of its relation with itself. One question is, how does the mind construct this world and ourselves in it? Constantly we think, feel, decide, perceive. Understanding how these things happen is central to our grasp of what kind of being we are. The way our mental life is constituted is also important to our understanding of who we are individually, because the variation of our mental lives constitutes our feeling of differentiation between our fellow humans. Mental states, unlike most other things of our everyday experience, have no spatial characteristics and they do not seem to belong to a world constituted by physical things. How to place our mental experience in the physical universe is therefore perplexing. Mental phenomena also interest us because we infer from ourselves that others have similar mental experiences. Social interactions require us to understand each other's thoughts and feelings. And language would not exist as a medium of expressing our inner world without our elaborate cognitive abilities. We seem to understand the content of our mind readily from our own experience. The problem arises when we try to know objectively, independently from ourselves, what we experience. From this arises the general problem of how the study of mind should proceed.

Much of our knowledge about our mind is immediate, and seems to have some sort of privileged status. That is, only we ourselves really know what is going on in our mind, and nobody else can know exactly what we feel or experience. Nothing seems to mediate between our mental states and knowledge we have of those states; such knowledge seems both direct and automatic. The privilege may not be absolute, it may not mean we are usually correct about our own mental states, nor that what we know about them is all there is to know. But our automatic and immediate access to our own mental states leads to a natural presumption that our beliefs about ourselves are correct. It is tempting to regard this special access as superior to

any other sort of knowledge we could have about mind. This tantalizing problem has so far been widely neglected by scientific studies.

What we have said so far brings about two main topics of interest: firstly, the relation between *mind and body*, and secondly the nature of the *self* or *self-consciousness*. The latter is what we are concerned with in this book. What we mean by *self* here is as a first approximation the commonly shared experience, that we know we are the same person across time, that we are the author of our thoughts/actions, and that we are distinct from the environment. It is the immediate, pervasive, automatic feeling of being a whole person, different from others, constant over time, with a physical boundary, the centre of all our experience. These feelings are so fundamental to our human experience that we hardly ever think about them.

These are exciting times for the closer examination of self-consciousness. For many years, the topic has been studied primarily at a philosophical level (Rosenthal, 1991; Metzinger, 1995; Block et al., 1997; Gallagher & Shear, 1999; Zahavi, 2000). More recently, however, progress has been made by linking theories and experimental procedures from psychology to the results of neuroscience. This has allowed us to begin to understand how processing of self-relevant material is taking place on a cognitive and neural level, and how models of self-processing can explain some pathological states of mind.

For every author who has written about the *self* there are as many concepts, so the literature is full of diverse definitions and overcrowded by misunderstandings. The most basic thing to keep in mind is the level of description and study:

1. On a philosophical level, we can distinguish basically between two different schools, phenomenology and philosophy of mind. Phenomenology, in its broadest sense, describes the *essence*, the content and *feel* of a mental state. Philosophy of mind, based on concepts of analytical philosophy, is, for our purposes, mainly concerned with the logical connection and systematization of our knowledge of the mind.
2. Social science, social and personality psychology are concerned with how people regard themselves ('What kind of person am I? How do others see me?'), the different roles one person can have in society (researcher, mother, amateur musician . . .), and how these things interact and change over time.
3. Cognitive science tries to build models of how the mind works, derived from computer simulations as well as experimental data on healthy subjects and patients with brain lesions or mental disorders.
4. The neurosciences try to correlate mental phenomena with brain states and structures, using brain imaging or electrophysiological techniques.
5. The clinical sciences: descriptive psychopathology describes and classifies pathological mental states. Neurology and classical neuropsychology try to relate mental faculties to distinct brain areas by examining patients with cerebral lesions.

When we talk about mental or brain states we always have to be aware which level of investigation we are talking about. Confusion often arises when concepts of different fields are mixed, particularly points 1, 3 and 4, described above. However, often the goal of an enquiry is an explanation of a phenomenon in one field with concepts or results from another. Here it is particularly important to remember that a model from one field (e.g. in cognitive sciences *attention*) does not necessarily have a clearcut correlation with findings from other areas (e.g. a particular brain area). It becomes even more difficult when we cross borders between philosophy and sciences. However, this is what we have tried in this volume for a particular purpose. The neuroscientific study of the self and self-consciousness is in its infancy. There are no established models, very little data and not even the vocabulary to describe neuroscientific notions on these topics. For a start it is therefore necessary to draw from as many sources as possible to form a basis of enquiry. This volume brings together contributions from different fields, but focuses on the cognitive and neurosciences, and particularly pathological states of the self in schizophrenia.

We know that we are the same person across time, that we are the author of our thoughts/actions, and that we are distinct from the environment. This means there is a fundamental, affective tone of mental, emotional and bodily unity, which is so basic to our experience that it is very difficult to grasp. However, there are neuropsychiatric conditions such as schizophrenia where this basic tone of selfhood loses its natural givenness, with subsequent changes in the perception of oneself and the environment. This makes it possible to interview and test patients with impairments in self-experience. We can describe their experiences and compare results from experiments with those of healthy controls and thus generate tentative models of the underlying neurocognitive structure, correlating with the experience.

In this volume, we focus on schizophrenia, because the core pathology of the disorder is a disentanglement of the normal unity of body, thoughts and emotions. Schizophrenia is one of the most interesting and tantalizing of all human diseases, because what is most central to our existence, the mind, is lost or distorted (at least in the severe and acute phases). While it is commonplace in the cognitive neurosciences to draw inferences from the loss or disturbances of functions seen in patients, this is less common in psychiatric patients. However, it is only in disorders such as schizophrenia that the mental architecture underlying the self is so cruelly exposed.

The clinical presentation of schizophrenia varies both between individuals and within the same individual at different stages of the illness. But there are some prevalent features which most frequently comprise acoustic hallucinations (hearing voices) and delusions (false, uncorrectable beliefs, e.g. the ability to control the weather). Other common symptoms are thought withdrawal and insertion (patients

have the feeling that their own thoughts and emotions are introduced from outside), formal thought disorder (language disorder), incongruous affect and negative symptoms (Crow, 1985), such as apathy, social withdrawal and flattened affect (McKenna, 1994; Crow, 1985; Cutting, 1985; American Psychiatric Association, 2000; Gelder et al., 2000). Schizophrenia is a disorder with a worldwide incidence of 2–4 cases per 10 000/year. The lifetime risk is about 1% in the general population; the disease usually starts before the age of 30. It is therefore a common disorder which leads to chronic disability in about one-third of affected patients.

In this volume, we have brought together scholars from different fields of study to present their views on consciousness and self-consciousness, which are probably the most complex phenomena we know of. We focus on the cognitive and neurosciences and give special weight to the pathological self-states in schizophrenia. The book is divided into three parts. In the first part (chapters 1–4), some important theoretical and conceptual foundations are laid out. In the second part the cognitive and neurosciences present empirical data and models about the self from its different aspects (chapters 5–10). In the third part (chapters 11–21), concepts, models and data are presented to explain normal and disturbed self-states, focusing on schizophrenia. In the final chapter, our own integrative view, encompassing most of the aspects dealt with in the previous contributions, is attempted. Most authors agree that there is a feeling of *self* (thoughts, emotions, body, across time), that this is a mental state, that mental states are represented in the brain and that these states can be investigated scientifically.

Berrios and Marková, in their challenging contribution (chapter 1), regard the *self* as a mere construct of western thought that can be traced back to Greek philosophy. Consequently, they argue that it is not a natural entity, that it cannot be investigated scientifically, and *self-pathology*, such as passivity phenomena, are pure metaphors. In contrast, Northoff and Heinzel (chapter 2) elaborate the idea put forward above, that it depends on the level of description (or perspective, as they put it) what kind of model is to be applied to describe the self. These different perspectives lead to particular implications regarding notions of the self in the different scientific disciplines, that are not mutually interchangeable. This idea is developed further by Zahavi (chapter 3), who focuses on one particular perspective, the *first-person perspective*, from a phenomenological point of view. He goes on to criticize the higher-order representation theory put forward by the philosophy of mind school, which is the basis for the investigations described mainly in chapters 17–21. Stamenov (chapter 4) dissects the relationship between self-awareness, self-consciousness, auditory hallucinations and linguistics. The symbolic representation of the self in language is the personal pronoun 'I'. Based on Chomsky's generative grammar, the smallest enactment of the self in the world is a simple sentence.

In chapter 5, O'Brian and Opie, as a linking contribution between philosophy of mind and cognitive science, develop their multitrack model of consciousness. They argue that although we have the feeling of being a mental unity, in fact the underlying mechanism is an aggregate of phenomenal elements (units of experience), each of which is the product of a distinct consciousness-making mechanism in the brain. Barnard (chapter 6) offers a cognitive multilevel theory and argues that the mind has a modular architecture with specialized subsystems to process information. Meaning, such as the representation of self-related material, is created by an emotionally charged interaction of different system levels. Gallup, Anderson and Platek (chapter 7) open up the view to incorporate ontogenetic and phylogenetic aspects of self-awareness. They demonstrate that mirror self-recognition in infants and primates must go hand in hand with a sense of self and with theory of mind abilities. Keenan, Wheeler and Ewers (chapter 8) extend this view to human adults and present experiments on facial self-recognition using different techniques. The memory aspect of our own past, as a constituent of a coherent feeling of self in time, is discussed by Markowitsch (chapter 9). He presents a neuropsychologically based theory of autobiographical memory and integrates levels of psychology, cognitive science, clinical neurology and functional imaging. Yet another important aspect of self-representation is discussed by Panksepp (chapter 10): the emotional aspect, and how this might be implemented on a neural level.

Chapters 11–21 mainly focus on psychopathological states of self-disturbances with an emphasis on schizophrenia. From a phenomenologically oriented position, Parnas (chapter 11) presents detailed clinical descriptions of the disorders of self-experience in patients with schizophrenia spectrum disorders. Further, data on prepsychotic stages demonstrate anomalies of self-aspects already present in the early stages of the disorder. From a similar conceptual background, Sass (chapter 12) argues that some schizophrenic symptoms can be understood as phenomena that would normally be taking place naturally and unnoticed which are instead taken as objects of one's awareness. Similar to Parnas and Sass, Scharfetter (chapter 13) regards a self-disturbance as the core symptomatology of schizophrenia. He presents data on a newly developed rating scale, based on Karl Jaspers' ego-pathology, for the description and classification of severe psychotic experiences. Chapters 14–16 are based on models derived from social and personality psychology to explain some of the symptoms and life courses of patients with schizophrenia. Bentall (chapter 14) introduces a model of self-attribution ('How am I, compared to other people, compared to how I would like to be?') and applies it to paranoid ideation. The way we see ourselves over time and make a coherent story out of it for ourselves is called self-narrative. Phillips (chapter 15) discusses this notion in a broader context of philosophy and the social sciences and introduces histories of patients to show how the disorder might have influenced their course of life. From a more

cognitive point of view, Gallagher (chapter 16) defines four capacities that are necessary to construct a coherent self-narrative and tries to link them to cerebral structures. In another overarching attempt, Vogeley (chapter 17) introduces his definition of self-consciousness and applies it in more detail to the symptom of auditory hallucinations. These are thought to arise from a disturbance in a self-produced action-monitoring system, something dealt with in detail in chapters 18–21. Jeannerod and colleagues (chapter 18) and subsequently Blakemore and Frith (chapter 19) unfold their notion of action (and thought) recognition in others and oneself. They present empirical data and conclude that some symptoms in schizophrenia are a result of an alteration in the 'Who is the source of the action?' system. Fu and McGuire (chapter 20) apply these ideas particularly to the auditory and speech system and describe a functional imaging aproach to investigate them. These models cannot explain schizophrenic ego-disturbances (*Ichstörungen*) claim Walter and Spitzer (chapter 21), who present a modified theory based on right hemispheric and dopamine-system dysfunction.

In chapter 22, the final chapter, Kircher and David attempt to integrate most of what has been proposed in this volume by presenting a model of consciousness and self-consciousness. Psychopathological and empirical findings are related to this model.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (DSM IV), 4th edn. Washington, DC: American Psychiatric Association.

Block, N., Flanagan, O. & Güzeldere, G. (eds.) (1997). *The Nature of Consciousness.* Cambridge, MA: Bradford Book/MIT Press.

Crow, T.J. (1985). The two syndrome concept: origins and current status. *Schizophrenia Bulletin*, **11**, 471–86.

Cutting, J. (1985). *The Psychology of Schizophrenia.* Edinburgh: Churchill Livingstone.

Gallagher, S. & Shear, J. (eds.) (1999). *Models of the Self.* Thorverton, UK: Imprint Academic.

Gelder, M., Lopez-Ibor, J.J. & Andreasen, N. (2000). *Oxford Textbook of Psychiatry.* Oxford: Oxford University Press.

McKenna, P. (ed.) (1994). *Schizophrenia and Related Disorders.* Oxford: Oxford University Press.

Metzinger, T. (ed.) (1995). *Conscious Experience.* Paderborn: Schöningh-Verlag.

Rosenthal, D. (ed.) (1991). *The Nature of Mind.* Oxford: Oxford University Press.

Zahavi, D. (ed.) (2000). *Exploring the Self.* Amsterdam: John Benjamin.

Part I

Conceptual background

1

The self and psychiatry: a conceptual history

German E. Berrios[1] and Ivana S. Marková[2]

[1] Department of Psychiatry, University of Cambridge, UK
[2] Department of Psychiatry, University of Hull, UK

Abstract

The concept of self is a construct. It is not a 'natural kind' sited somewhere in the human brain. The western concept of self emphasizes individualism and autonomy but this view is cultural and no more scientific or truthful or advanced than the syncytial or collective view of self developed in other cultures and which revolves around family or clan rather than individual. Originally meant by St Augustine to be just a metaphorical or virtual space within which theological models of responsibility, guilt and sin could be played out, the self regained importance in the hands of Luther who started its reification as a private cave where god and man would regularly meet to sort out their differences. During the seventeenth century, the metaphors of the Reformation become secularized and built into liberalism and capitalism. The self survived by becoming a conceptual prop for bourgeois notions such as individual ownership, natural rights and democracy.

Wanting to reinforce the political status quo, nineteenth-century science transformed the political self into a psychological entity and proceeded to 'naturalize it' (i.e. render it into a natural kind). This additional reification engendered curious inferences. One was the belief that a 'self' really existed inside the European mind and brain. This self was characterized as driving, organized, executive and with a capacity for leadership and domination. Another curious inference was that, since the self was an ontological blob, it could be affected by pathological lesions and disease; and that this could be 'visualized' if only the right technique was available.

For a time, alienists took to the self with great gusto. For example, based on the tautological view that the self is truly impaired only in schizophrenia (*Ichstörungen*), much effort was invested in trying to find out whether this disease had to do with a disintegration of the 'boundaries' of the self; a reduction in its power and energy; its incapacity to discriminate between itself and the environment. This debate was harmless enough when played out in the territory of phenomenology or popular existentialist literature. However, during the last 30 years new techniques have encouraged researchers to reify the self further. Unfortunately, no new conceptualization has arisen and hence no interesting questions are being asked.

This is a pity for there is enough historical information available to see that the self is a linguistic trope, a yarn, a mode of talking about people and their reasons for doing things.

The self was never meant to be a solid object like a stone, a horse or a weed, nor even a concept to be considered as semantically tantamount to changes in blood flow or test scores. Of course, patients with disordered minds do sport hurting, afflicted and cursing selves but not as they may do carcinomas or broken legs. Their selves live in the same realm as do their virtues, vices, beliefs and aspirations, and that is where they should remain.

The self and psychiatry: a conceptual history

The term *self*[1] is currently used to refer to a putative *core*,[2] assumedly a defining feature of humanity and responsible, inter alia, for the experience of the so-called *sens intime*.[3] The appearance of the 'self' is part of the wider process whereby 'person-related' concepts (Laurent, 1993)[4] were constructed in western culture.[5] To this day, philosophical, psychological, theological and moral versions of the self vie for supremacy[6] and the debate continues on whether they have a divine, evolutionary or cultural origin[7] (Lévy-Bruhl, 1928; Marsella et al., 1985; Renaut, 1997; Gergen, 1998). Because of its unclear boundaries (Gallagher & Shear, 1999) and voluminous literature, writing on the history of the self is a hard task (Danziger, 1997b). After offering a summary of its history, this chapter will focus on the period during which the self was incorporated into psychiatry.

The beginnings

Plato (McCabe, 1994) and Aristotle[8] (Hartman, 1977) were amongst the first to discuss the need for a theory of *individuation* of objects and entities; indeed, it has been claimed that an inchoate 'form' of self was already present in Aristotle's notion of memory (Annas, 1992). However, the sense in which the self will be discussed in this chapter was only achieved during the post-Aristotelian period (Snell, 1953; Onians, 1954) and it is marked by the moment in which the philosophers of the first Stoa[9] redefined the Platonic term imagination (*phantasia*) as a collection of *individual* imaginations, *phantasiai*. The problem of how, then, did individuals 'recognize' multiple experience as theirs (Sandbach, 1994) was resolved by stating that the *hêgemonikon* (the highest component of the soul)[10] actually tagged up or personalized individual *phantasiai*: 'The *hêgemonikon* provided the Stoics with the concept of unitary self, actively engaged as whole in all moments of an animal's experience' (p. 107, Long, 1991). In Plotinus (1966), the *hêgemonikon* acquired a sense or feeling of 'privacy' and 'interiority'.[11] Concepts such as introspection (Reesor, 1989; Rappe, 2000), consciousness and awareness of function and content of function (O'Meara, 1995) started life during this same period. In his effort to consolidate the idea that man and God needed a private venue to meet,[12] St Augustine (1991) developed the idea of the self as a 'private inner space'.[13] It has

been suggested that such a concept was based on Greek notions of 'subjectivity' and 'sin' and on Plotinus' mechanism of self-reflection (Mondolfo, 1955).

The seventeenth century

Descartes (1596–1650) (1967)[14] identified the self with the *res cogitans* (thinking substance) and believed it to be the basis for the belief in the existence of the external world. The legitimacy of this foundationalist claim, the nature of his dualism and the force of 'I think, therefore I am' (*cogito ergo sum*) as a logical entailment have since been subject to scrutiny (Frondizi, 1952).[15] Whether out of conviction or convenience, eighteenth-century neuroscientists followed a naive 'dualist' interpretation of Cartesianism so that they could claim that knowledge gained on the *res extensa* (the brain) had no theological implications (in regards to the soul or *res cogitans*) (Bynum, 1976).[16] The same interpretation of the Cartesian self (as an absolute knower) was built by nineteenth-century alienists into their own concepts of mental symptom and disease.

Descartes restarted the seventeenth century debate on 'interiority', 'self' and 'self-identity' (Garber & Ayers, 1998; Schoenfeldt, 1999). John Locke (1632–1704) (1959), one of the participants in the debate, set the scene in his analysis of the *principium individuationis*:[17]

> to find wherein personal identity consists, we must consider what person stands for;- which, I think, is a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does only by that consciousness which is inseparable from thinking, and, as it seems to me, essential to it: it being impossible for any one to perceive without perceiving that he does perceive. When we see, hear, smell, taste, feel, meditate, or will anything, we know that we do so. Thus it is always as to our present sensations and perceptions: and by this every one is to himself that which he calls self (Locke, 1959, II, xxvii).

John Locke also proposed that the feeling of continuity of the self was based on a concatenation of memories[18]:

> Make these intervals of memory and forgetfulness to take their turns regularly by day and night, and you have two persons with the same immaterial spirit, as much as in the former instance two persons with the same body. So that self is not determined by identity or diversity of substance, which it cannot be sure of, but only by identity of consciousness (Locke, 1959, II, xxvii, 23).

However, by suggesting that it was mappable on to a region of the *body*, Locke encouraged the development of a 'psychological' self; indeed, like all other images resulting from perception, the self also revealed itself as a projection on to the camera obscura of the mind – which was the way John Locke was to allegorize its functioning.[19] Via the French version of the *Essay*,[20] Locke's ideas became