# Emergence and the Cognitive Neuroscience Approach to Psychiatry*

Henrik Walter

Abteilung Psychiatrie III, Universitätsklinikum Ulm, Leimgrubenweg 12, D-89075 Ulm, Germany

Z. Naturforsch. **53c**, 723–737 (1998); received April 22, 1998

Emergence, Prefrontal Cortex, Neuropsychiatry, Decision Making, Volition Agency

In this paper, I introduce the philosophical notion of strong emergence and argue that it is almost exclusively applied to properties related to (conscious) subjective experience. Contrary to the still common attitude of refraining anxiously from these topics, I argue that we have a promising scientific approach to them: the cognitive neurosciences. I list a spectrum of interesting, potentially strongly emergent properties already investigated, and discuss the example of volition in more detail. Psychiatric disorders, like antisocial personality disorder, schizophrenia, and obsessive compulsive disorder, open new ways for understanding aspects of volition such as willed action, decision making, and agency. I conclude that the notion of strong emergence is only a preliminary label which, however, might be understood as a challenge for empirical scientists to explain and understand phenomena related to subjectivity and consciousness.

## Introduction: The Concept of Emergence

"Emergence" is a term that has become popular in sciences dealing with mental phenomena or complex systems (Beckerman *et al.,* 1992; Stephan 1998b; this issue, pp. 639–656). In a dictionary sense "to emerge" means "to come into view, to become known or apparent, to rise from obscure or inferior conditions". This is pretty close to the use of the term in ordinary language and scientific practice. Order "emerges" out of chaos, patterns out of noise, new types of organisms "emerged" in evolution and mental properties "emerge" out of neuronal acitivity. However, recently emergence has turned into a technical term. I will introduce here two distinctions: diachronic/synchronic and weak/strong emergence. If something new or unexpected comes into being for the first time, then we speak of *diachronic* emergence. I will not deal with diachronic emergence in this paper. *Synchronic* emergence, instead, refers to properties of complex systems composed of simple parts arranged hierarchially. Its weak variant is characterized by the following features: (i) Emergent properties are *system properties*, i.e. properties exhibited only by the whole system but not by its elements (weight for example is not a system property). (ii) The properties of the system are causally determined by the properties of its components, their arrangement and their interactions (synchronic or *microdetermination*). (iii) Emergent phenomena are only caused by physical properties, not by supernatural powers, forces like élan vital or a mental substance (*physical monism*).

Without doubt such weakly emergent properties exist, e.g. the fluidity of water or the eigenfrequency of a resonance circuit. Initial and boundary conditions given they can be explained by or deduced from the laws of nature. In contrast, *strong* emergence is defined as weak emergence plus the thesis that (iv) the system properties are – in principle – *explanatorily unreducible* to the properties of the system's components, their arrangement and their interactions *although* the former are synchronically determined by the latter. To call a property emergent in the strong sense is not only to claim that those properties are *not yet* explained. Rather, it is to claim that there *never* will be explanatory reduction, regardless of

* This communication is a contribution to the workshop on "Natural Organisms, Artificial Organisms, and Their Brains" at the Zentrum für interdisziplinäre Forschung (ZiF) in Bielefeld (Germany) on March 8–12, 1998.

Reprint requests to Dr. Dr. H. Walter.
Fax: (0731) 5026751.
E-mail: henrik.walter@medizin.uni-ulm.de

0939–5075/98/0700–0723 $ 06.00 

how much we will know, even if we were to know *everything* about the physical world.

## Qualia et alia

Do any emergent properties in the strong sense exist? This is highly controversial. Some think that they do (Stieve, 1997, Chalmers, 1996), some think that they do not (Vollmer, 1992, Churchland, 1995). But which properties *might* be strongly emergent in the first place? Explanatory irreducibility is almost exclusively claimed for subjective, mental properties directly or indirectly related to consciousness: qualia (the "what it is like" of a experience, e.g. the redness-quality experienced when seeing something red), emotions and feelings, meaning and intentionality, the unity of self, self-consciousness, or free will.[1] Philosophers usually discuss principal arguments about why such properties might not be explanatorily reducible. One example is the argument of the nonalgorithmicity of meaning, another, the Tibetian-mill-argument for the nonreducibility of qualitative experiences (for a critical discussion cf. Buschlinger, Vollmer and Walter, 1998). However, all these properties may also be the subjects of empirical science. After a long period of prohibition, scientists have started again to explore subjective states and consciousness. The last decade has seen a tremendous increase of interest in these topics especially in cognitive psychology and neuroscience. Conferences are held, new journals are published, and renowned scientists set out to draw theories of consciousness. Rather than discussing these theories, I will present some concrete examples of conscious and/or subjective phenomena that are candidates for strong emergence. Furthermore I make a methodological claim: The only way not to become involved in discussions originating from rather doubtful thought experiments (for a critic see Wilkes, 1993, Buschlinger, 1993) is to attempt to reduce the properties, i.e. to be a methodological reductionist. Why? – If you succed, then you have falsified the "in principalists". If you do not succed, then you might at least better understand *why* those properties cannot be reduced. Or, even better, you might generate new hypotheses about how to do so. So, even advocats of strong emergence should be methodological reductionists for heuristic reasons.

However, I do not claim to be able to present satisfying explanations for any of the mental properties mentioned above. Rather, I want to show the follwing: Firstly, there are a lot of interesting, real and relevant problems for empirical science related to the notion of subjectivity and consciousness. Secondly, there are ways to approach them. Thirdly, if we do not decide to ignore them we have no choice: We have to confront rather philosophical questions when investigating these phenomena (Walter, 1996a).

In table 1, a selection of phenomena related to subjectivity and conscious experience is shown. All of them are currently investigated by the cognitive neurosciences. (Many of them are described in more detail in Ramachandran and Hirstein (1997), Joseph (1996), Feinberg and Farah (1997), Spitzer and Maher (1998)). These studies have already yielded a huge amount of surprising answers and new questions for philosophy of mind as well as neuroscience itself (cf. Metzinger, 1995). However, for the purpose of this paper, I restrict myself to the problem of volition.

## Cognitive Neuroscience

Cognitive neuroscience is a new "emerging" discipline (in the diachronic sense) which aims to understand and explain cognitive phenomena by investigating their neural bases (Gazzaniga, 1995). There are three major ways in which cognitive neuroscience ultimatively may contribute to the explanation of allegedly strongly emergent mental properties: lesions studies, neuroimaging experiments and neurocomputational modelling.[2] I will explicate all of them shortly. *Lesion studies*: Patients with lesions in certain brain areas show certain functional deficits, e.g. aphasia. Neuropsychol-

[1] The only exception are properties in quantum physics. However, because they may be strongly emergent contemporary physicists sometimes claim that they are closely related to the phenomenon of consciousness, e.g. Penrose (1994).

[2] Obviously there are other scientific approaches like clinical studies, psychological experiments or animal modells. But for the question of explanatory reducibility the above mentioned approaches are the most important.

ogists conclude from this that the lesioned areas are necessary for this function in intact brains (cf. Damasio and Damasio, 1989, Farah and Feinberg 1997). However, this approach is limited because of several reasons: The lesions might comprise only one of several critical structures that are part of a widely distributed network subserving this function. Furthermore, it might be the case that the lesioned structure normally suppresses some neuronal processes which are released after the lesion and result in disturbed function. Finally, as there is neuroplasticity in the adult brain, the lesioned brain reorganizes itself at least partly and thus may not be a good model for intact brains. In *neuroimaging* techniques are used that measure brain acitivity directly (EEG: electroencephalography; MEG: magentencephalography) or indirectly (SPECT: single photon emission tomography; PET: positron emission tomography; fMRI; functional magnetic resonance imaging). Whereas EEG has been used since long, the other techniques are quite new. They allow not only to measure brain activity but also to map it onto the anatomical structure, so that localization of the activity becomes possible. Measures of electromagnetic activity have a good time resolution (in the range of msecs), whereas measures of regional cerebral blood flow (rCBF; mainly PET and fMRI) have a better spatial resolution (in the range of mms). The most widely used technique in cognitive neuroscience has been PET (Posner and Raichle, 1994; Frackowiak *et al.,* 1997), although in the last years the fMRI technique is used increasingly. It eventually might replace PET as the gold standard because of its non-invasiveness. The technical limitations and implicit assumptions of neuroimaging techniques bear problems of their own when applied to problems in philosophy of mind which I have discussed elsewhere (Walter, 1997a). Nonetheless, neuroimaging is a rich source for neurophilosophy. In *neurocomputational modelling* (Churchland and Sejnowski, 1992) mental tasks can be simulated by a computer using neural networks (cf. Stein, 1998). Although it is clear that simulating phenomena implies that the modell omits something (cf. Buschlinger, Vollmer and Walter, 1998) this approach has a clear advantage over other approaches: A simulation model allows to change parameters and to look what happens after certain changes. Thus, mental phenomena can be varied systematically, predictions can be made and new results can be obtained which are not explicitly built into the model. This experimental manipulation of the mind is not possible in the same way with lesions or neuroimaging studies. To date, there are nearly no neurocomputational models of volition (but cf. Changeux and Dehaene, 1996). In what follows I consequently refer primarily to lesion and neuroimaging studies of normal subjects and psychiatric patients.

## The Problem of Volition

"Volition" is the act or power of making one's own choices or decisions. "The will" has traditionally been regarded as a mental faculty. At least three varieties of will can be distinguished (cf. Kane, 1996): *desiderative* will (what I wish), *rational* will (what I choose), and *striving* will (what I try). In the early days of experimental psychology, about one hundred years ago, will was studied experimentally, e.g. by Rieger, Külpe, Ach, and Wundt (cf. Heckhausen and Gollwitzer, 1987). One reason for the interest in will was related to the practice of hypnosis that became popular at that time. Especially the phenomenon of posthypnotic tasks[3] raised the question in how far people are in control of what they are doing. Although some psychologists have continued to work in this field, volition and will have been almost totally ignored in academic psychology for a long time. Probably, this was due to the fact that will has been studied mostly by relying on subjective reports, and also by its "dangerous" relatedness to the concept of free will. Recently, however, scientists have regained interest in volition (e.g. Libet *et al.,* 1983; Heckhausen *et al.,* 1987; Frith *et al.,* 1991; von Cranach and Foppa, 1996; Spence *et al.,* 1997; Hyder *et al.,* 1997; Walter, 1997b; Prinz, 1998). Recent work in the neurobiology of volition can be discussed from a dedicated neurophilosophical point of view (Walter, 1998). However, here I will take a more neuroscientific stand on these results.

[3] Posthypnotic task: People in the hypnotic state are advised to perform a certain act after hypnosis, e.g. to open the window. When doing so, people often give reasons unrelated to hypnosis like: "It was too hot" or "I just felt like doing so". They do no remember that they had been instructed.

## Conscious Will: Half a Second Behind?

One of the most famous examples for the cognitive neuroscience approach to volition is the study of the so called readiness potential (RP) (Libet, 1993). The RP is a negative electrical potential occuring about one second before a self-paced movement is executed. This surprising long delay has led to a number of experiments which triggered a wild discussion on the role of consciousness for willed action. The question arose: How is it possible that the "neuronal adequaecy" for a willed action is established *before* we become aware (are conscious) of the intention to move? This problem was raised by an experiment done by Libet *et al.* (1983). People were instructed to perform simple movements whenever they wanted to do so. The readiness potential was recorded. Simultaneously, the subjects were looking at a clock that had no arms but a fast revolving point which completed one cyle within 256 msec so that the 12 numbers at each "5 s" position corresponded to approximately 21 msec of real time. Thus the subjects could remember the clock position in relation to subjective mental events. The result was interesting: The onset of premovement neural activity (RP) started nearly half a second before people felt the "urge to move". With respect to similar experiments utilizing sensory stimuli, Libet concluded that consciousness is always half a second behind: The neural processes would already be initiated before we become aware (conscious) of what is going on. This seems to be a challenge to what we call free will.

Unsurprisingly, Libets discussion received heavy fire from empirical scientists as well as from philosophers. It is impossible to recapitulate this intricate discussion without being superficial. However, the three most important arguments are the following: (i) One major drawback is the experimental setting itself. The subjects did not react *immediately* to their "becoming conscious" of the urge to move. Rather, they had to remember it and reported it only some seconds later verbally. As memory processes may easily be distorted, as is known from several studies, this is a questionable timing method. (ii) There is an implicit assumption that there exists a point in time (or at least a very short time span) at which consciousness "enters the stage". Especially Dennett (1994) has criticized this "Cartesian theater". He argued that becoming conscious is not a precisely timed event but rather a – we might say "emerging" – result of parallel and intertwinned events of information processing extended in time. (iii) Another critique seems the most crucial to me: Based on their own experiments with the readiness potential, Keller and Heckhausen (1990) have argued convincingly that no process of conscious volition in a strict sense is involved here. The "intention to move" had been formed already at the outset of the experiment when subjects internalized the instructions. The "urge to move" should not be interpreted as the beginning of a conscious volitional process but rather as the felt crossing of a threshold, elicited e.g. by sensory reafferences (for a related line of argument cf. Walter, 1998).

Nonetheless, the Libet experiments have triggered a lot of experimental work and theoretical discussions about the nature of will and consciousness, and have led to progress. For example the readiness potential can not only be recorded when people do *execute* movements, but also when they only *imagine* them ("motor imagery", cf. Jeannerod, 1997). Thus, in some way, it becomes possible to study "pure thought".

With the help of pattern recognition by neural networks, even online thought reading has become reality. In order to see specific premovement neural activity it is necessary to average at least a few dozens of recordings because of the low signal-to-noise ratio. Neural networks, however, are able to recognize specific premovement activity ("desynchronization") in a single trial if properly trained. Experiments were carried out where subjects were instructed to perform or imagine one of three movements. Simulataneously, a computer tried to recognize premovement activity online. Feedback was given to the subjects about which movement the computer recognized. Although the computer was not perfect, the number of hits were significant above chance (Kalcher *et al.*, 1994). In a similar experiment, the neurosurgeon Walter Grey directly recorded from the cortex by means of implanted electrodes over the motor cortex (cf. Dennett, 1994). Grey asked the patients to look at pictures of a caroussel projector and told them that they could move to the next picture by pressing a button. What the patients did not know: the caroussel was not moved by the button press, but by the reinforced signal of their motorcortex which actually started before the movement was executed. Subjects were quite irritated because they had the impression that the projector anticipated their decisions!

As should be obvious by now, many interesting questions arise from these paradigms of mind-brain reading. They are not only relevant for potential therapeutic purposes (controlling artificial limbs) but also touch – nolens, volens – philosophical questions about what it is "to be in control", i.e. the problem of free will. Now they can be approached empirically.

## Imaging Willed Action

Recent work in the field of volition takes advantage of the relatively new techniques of PET and fMRI. Both techniques measure cerebral blood flow which is an indicator of neural activity. In PET, resting cerebral blood flow can be measured quantitatively. In activation studies, a control condition is compared with an activation condition so that only specific task related blood flow shows up. Starting with the now classical paper by Frith *et al.* (1991) some studies have been carried out investigating "willed action". Typical experiments use reaction tasks as the control condition (e.g. moving a touched finger, repeating words silently) and "willed action" as the activation condition (e.g. *choosing* a finger to move, *freely generating* new words beginning with the same letter as the stimulus). The stimuli in the two tasks are similar. In the former only a stimulus driven reaction is required whereas in the latter *internally generated* action and thinking is involved what operationally is defined as "willed action" (Frith *et al.,* 1991; Hyder *et al.,* 1997; Phelps *et al.,* 1997). Calculating the difference in brain activation between the two tasks shows that a large part of the dorsolateral prefrontal cortex and – to a lesser extent – also the anterior cingulate cortex are activated. Classical lesion neuropsychology has already postulated the prefrontal cortex as the site of executive control, especially when performance is required that deviates from routine action (Shallice, 1988). The anterior cingulate has only recently gained firm interest for action and volition. Posner and colleagues have attributed the anterior cingulate a critical role for attention (Posner and Raichle, 1994). Paus and colleagues (1993) have shown in a PET study that novel actions are correlated with activity in the anterior cingulate cortex with the precise location of the activity being dependent on the modality (manual, oculomotor, and speech). In general, the anterior cingulate is considered an important cortical region mediating "drive" and the "selection for action" (for a good review cf. Devinsky *et al.,* 1995). Its location between prefrontal, limbic and premotor systems makes it well suited to perform these functions. As the experiments discussed below show, emotional factors mediated partly by the anterior cingulate play an important role in volitional processes that are involved in decision-making.

## Feeling Reasons

To define "willed action" as "interally generated, not stimulus driven action" ist quite crude. As the example of obsessive compulsive disorders shows not every internally generated action is willed. In more sophisticated approaches to volition decision tasks are studied that are related to what I have called "rational will". (Choosing between alternatives is also the paradigmatic situation discussed in the philosophy of free will.) I will discuss two experimental paradigms of decision making. Both aim at simulating real life decisions such that decisions are made under uncertainty or by use of incomplete information. Both of them show that emotions play a crucial role in practical reasoning. In this context, a hitherto neglected structure of the prefrontal cortex, namely its ventromedial part, plays a prominent role. Broadly spoken, the ventromedial prefrontal cortex comprises all the parts of prefrontal cortex which are not dorsolateral, motor or premotor cortex. It also includes the anterior part of the cingulate cortex. In the literature, the term "orbitofrontal cortex" is also used quite often, which refers to the prefrontal cortex above the eye cavity. Although overlapping, the extensions of both expressions are not identical.[4]

[4] The orbitofrontal cortex comprises the part of the prefrontal cortex above the orbita (Brodman Areas 12, 11, 13, 25 and 47). The term ventro, medial also in cludes more medial parts especially part of the anterior congulate (Brodman Area 24).

One test sensitive for ventromedial dysfunction is the so called "gambling task" (Bechara *et al.,* 1994; 1997; 1998). In this task, subjects choose cards from four decks and receive monetary reward and punishments. In the long run, two decks are winning decks and the two others losing decks. The task is constructed such that when starting the subjects do not have any idea which decks are the good and which are the bad ones. They have to find it out by trial and error. How did normal subjects perform? After turning about 20 cards, they started to prefer the good ones. The intermittent debriefing showed that at that point the subjects were not aware of their strategy. However, psychophysiological measures showed that something had happened. After some punishments had been encountered the electrodermal skin conductance response (SCR, an indicator of emotional involvement) started to increase before the decision for one deck. With continuing to select cards, subjects then started to report gut feelings that something is wrong with certain decks. Finally (after about 80 cards), most of the subjects (80%) understood which decks were good for what reason and performed accordingly, i.e. preferred the good ones. The conclusion is that emotions influence decisions under conditions of uncertainty (as it is almost always the case in real life), even before subjects are aware of their feelings or their explicit strategies. But the story is even more interesting.

Originally, the gambling task was constructed to measure deficits of patients with brain lesions in the ventromedial prefrontal cortex. Patients almost never have lesions confined only to that region. But sometimes they have, usually as a result of an operation of an olfactory meningeoma, or of a rupture of an aneurysm in the anterior communicating artery. These patients show a striking clinical picture. Although they do not suffer from the executive deficits as patients with dorsolateral prefrontal damage do, and although their intelligence is preserved, their lives change dramatically after the lesion. They perform badly in social life, are unable to reason appropriately in circumstances characterized by uncertainty, make real life decisions in ways conducive to the maintenance and betterment of themselves and of their family, and are no longer capable of succeeding as independent human beings. Additionally, their emotional life is disturbed in subtle ways (Damasio, 1994; Rolls *et al.,* 1994; Hornak *et al.,* 1996).

So, how did patients with "acquired sociopathy" (the expression used in the literature), perform on the gambling task? First of all, they did not change their decision strategy during the task but continued to prefer the bad decks. This behavior was associated with a lack of anticipatory SCRs, although the patients were able to show SCRs in other emotional situations. Additionally, the patients did not report feelings that some decks were strange. Most interestingly, however, patients *still* chose the wrong strategy even *after* they had reached the stage where they knew what the good and what the bad decks were (about 50% of the patients)! Somehow, the disturbed emotional feedback prevents that people act according to their rational insights. With the help of a delayed response task and by analysing the exact location of the lesions it is possible to dissociate the contributions of working memory and decision-making per se in this experimental task (Bechara *et al.,* 1998).

Another study involving decision making with emotional feedback was undertaken by Elliott *et al.* (1997). This experiment, a variant of the tower of Hanoi task, has a two by three factorial design: two experimental tasks (planning, guessing) and three feedback conditions (positive, negative, none). In the planning task, healthy subjects were presented with two arrays of coloured balls and asked to work out the minimum number of moves needed to transform one array to the other within a certain time limit, according to rules explained prior to scanning. In the guessing task, subjects were presented with two identical arrays of coloured balls. Without solving any task (no planning), they had to press one of six buttons and were told that three of them were randomly assigned as correct. Both conditions were carried out with positive, negative, or no feedback. The feedback was not related to the correctness of the answers but followed a fixed schedule. When subjects were debriefed after scanning, they all claimed to have realized that the feedback was irrelevant to performance, approximately after they absolved half way through the total number of scans. But even then (in the second half), subjects performed objectively worse in the negative feedback condition compared to positive or neutral feedback. This strongly suggests that feedback carried affective salience, even when subjects *were aware* of its non-relevance! The cerebral activity during the task was measured with PET. The main result (for the discussion of volition) is that there was an activation of the ventromedial prefrontal cortex associated with the presence of feedback in the guessing task. This is further evidence for the involvement of the ventromedial cortex as a critical structure mediating emotional feedback influencing decisions under uncertainty ("real life conditions").

Notice that for both experiments it is necessary to take into account the subjective and conscious experience during task performance. The subjects had to perform the task consciously: They did not only react to stimuli (as in the control condition), but had to think about them in a certain way, and had to be aware of certain features of the task. Most importantly, for the interpretation of the results, it is necessary to consider what the subjects felt and thought. In conclusion: We cannot ignore the subjective or the conscious component in these experiments. Rather, we have to consider them as central for the investigation.

## The Body and the Self

In the gambling experiment the emotional salience of decisions plays an important role for practical decision making. But how are emotions "measured"? And what *are* emotions and subjective feelings? Emergentist often claim that we never will be able to explain them. But there is progress in understanding the neural machinery for decision making and the emotions involved (Damasio *et al.,* 1996). According to Damasio (1994), an emotion is the combination of a mental evaluative process with dispositional responses to that process. For the evaluation the body plays a central role. Whenever a person is thinking or imagining something, this causes changes of her body state. The body state is continuously represented in the brain. The feeling of an emotion is the experience of body state changes in juxtaposition to mental images. Damasio calls the experienced body states "somatic markers". Why? – Because the reaction of the body, or rather its central representation, "marks" imagined or experienced events as good or bad, positive or negative and thus serves as a bias for action. However, the somatic markers are very crude biases: They do not indicate anything specific about a certain situation. Their evaluative nature is analogous to the evaluative nature of pain: Experiencing pain only means that something is going wrong, that there is some possible harm and that something has to be changed.

How does all that relate to volition and decision-making? If a person has to make a decision the following happens: The person imagines possible outcomes or scenarios. These imaginations are generated by means of the prefrontal cortex. The prefrontal activity influences (via the ventromedial cortex) certain brain areas (especially the amygdala and the hypothalamus) which then induce changes in body states. Those changes are represented online in the central nervous system. (There is also the hypothesis that an "as-if route" exists which is able to influence the central body representation directly.) If those induced changes depend on the individual past experience of the person, Damasio calls them "secondary emotions". The feeling of a secondary emotion is the juxtaposition of a certain mental image and the associated body states. Which body states are associated with a certain situation depends on the personal history of the individual. For example, if a person has gone often through painfull or unpleasant situations which are similar to the imagined scenario, the latter will induce body states which are felt as negative evaluative somatic markers. They do serve as a guideline for actions depending on the imagination of possible outcomes. This theory explains why patients with lesions in the ventromedial cortex are handicapped in their real life decisions. They miss the usual emotional bias so that their actions are guided mainly by short term gains, and do not take into account prior experiences in a serious way. In philosophical terms, we might say that these patients have a defect in practical reason.

Although this explanation is not the only game in town (for an alternative explanation cf. Rolls, 1996), and although it does not explain everything, e.g. why body states are felt at all (the qualia problem), we can state some progress. Instead of just claiming explanatory irreducibility the attempt for explanation leads to new questions. Only because we have started to formulate empirical informed hypotheses about allegedly irreducible, i.e.

strongly emergent phenomena, we may proceed further. For example, neuroimaging studies could be done with subjects performing the gambling task or similar tasks. What we expect is not only activity in the ventromedial prefrontal cortex, but also in the network comprising the central body representation. The relevant representation of the body is postulated to reside in the secondary and tertiary sensorimotor areas (like SII or the insula) and some other parts, especially the right parietal cortex. According to Damasio (1994), the body representation is not only the neural correlate for somatic markers but it also serves as the neural basis for a representation of the self.

Similar ideas are (again) becoming popular in philosophical thinking (Bermúdez *et al.,* 1995). "Agency", the sense of being the cause and initiator of one's own action, is an important ingredient of what is called free will. If we conceptualize volition as the self generation of alternatives and choosing between them, we can now begin to construct a picture of the neuronal processes underlying volition. With a neurobiologically informed theory of the self in mind, we now might study phenomena which involve disturbances of the self. There also are neural network models at hand that model the generation of alternatives and their selection (Changeux and Dehaene, 1996).

With the help of these and other results it is possible to approach the problem of free will. Even if there is no free will in a strong philosophical interpretation we might understand what it means to be autonomous from a naturalistic point of view. In order to construct a theory of natural autonomy which comprises the most important aspects of free will (being able to do otherwise, intelligibility and agency) in a realistic interpretation, we need to get into contact with such empirical approaches (Walter, 1998). In the domain of neuropsychiatry there are a lot of phenomena related to disturbances of volition and the self. They may be used to elaborate the above mentioned hypotheses. From this rich reservoir, I will discuss three examples.

## Antisocial Personality Disorder

The clinicial picture of ventromedially lesioned patients is called "acquired sociopathy" because it resembles very much the picture of what used to be called "sociopathy" or "psychopathy" and today is called "antisocial personality disorder". It seems evident to study these patients in order to find out if there is a dysfunction of the cortical-subcortical loops mentioned above. Surprisingly, this has not been done yet. Why? – On the one hand, it is not easy to find patients with pure antisocial personality disorder without any concomitant psychiatric disorders. On the other hand, these patients are not easily accessible. One large group is found in prison. (Most of the neuropsychological studies with psychopaths have been performed with imprisoned patients.) The other group is even more difficult to study, and much less is known about them: the socially succesfull sociopaths. Many psychologist think that persons with antisocial personality disorder who are intelligent and had a good education are not imprisoned, but rather become soldiers, explorers, or politicians.

However, there are at least some studies which show evidence for ventromedial dysfunction. Lapierre *et al.* (1994) have studied 30 psychopaths and 30 non-psychopaths in prison with six neuropsychological test. Three of them are thought to be specific for orbiofrontal or ventromedial dysfunction (a go/no go task, the Porteus Maze Test, and the modular smell identification test), one for dorsolateral dysfunction (Wisconsin card sorting test) and two for posterorolandic dysfunction (mental rotation task, similarity subset of the Ottawa-Wechsler intelligence scale). The patients with an antisocial personality disorder were significantly impaired only on the orbitofrontal-ventromedial tests. There are also some preliminary neuroimaging studies which point to a possible ventromedial dysfunction in murderers and patients with personality disorder (Raine *et al.,* 1994; Goyer *et al.,* 1994). From earlier studies it is known that the SCR is diminished in sociopaths (Hare, 1971) which is in good accordance with the somatic marker theory, although not yet conclusive. To further elucidate the theory of decision making, the same real-life-simulating decision tasks used in studies of acquired sociopathy should be adminstered to subjects with antisocial personality disorder, preferably with neuroimaging methods, as suggested elsewhere (Walter, 1996b).

## Schizophrenia, Volition and Agency

Another interesting disease for the study of volition disorders is schizophrenia. Certain forms of this disorder are characterized by negative symptoms like apathy, loss of interest, reduced drive and reduced daily activity which can be described as hypovolition. Liddle (1987) distinguished three syndromes of schizophrenia which are characterized by the dominance of certain symptom clusters: The psychomotor poverty syndrome (reduced drive and acitivity), the disorganization syndrome (incoherent speech and inappropriate affect), and the reality distortion syndrome (delusions and hallucinations). He studied resting cerebral blood flow with PET in all three groups in order to find out if there are symptom specific patterns of cerebral activity (Liddle *et al.,* 1992; Liddle, 1994).

For our purpose the results in the first two groups are of interest. In the psychomotor poverty group reduced activity in the dorsolateral prefrontal cortex (DLPFC) was found. The finding of "hypofrontality" has been confirmed in a lot of other studies, although there are also some controdictory result (for a review cf. Andreasen, 1997). However, hypofrontality seems to be rather symptom specific than disease specific as a very similar symptom (psychomotor retardation) in a different disorder (major depression) shows the same pattern of reduced activity in the DLPFC (Bench *et al.,* 1993). In the disorganization group of schizophrenic patients, another pattern showed up: hypoactivity of the ventromedial parts and hyperactivity in the right anterior cingulate. This finding can be interpreted as a decoupling of prefrontal-basal-ganglia circuits resulting in a decoupling of action generation and planning on one side (prefrontal cortex) and selection for action (anterior cingulate) on the other side. This might possibly lead to an indiscriminate stringing together (disorganization) of thoughts and actions. This is exactly what is found in this group of patients. So, "hypovolition" and "disorganization" seem to be associated with disturbed activities in dorsolateral prefrontal and cingulate cortex.

However, most of these studies have been performed with patients being in a resting state. Rather than scanning patients during rest, it is possible to have them perform a task during the scan. But this leads to another problem of interpretation. Activation studies in patients are compromised by the fact that controls and patients perform differently on the task. The question arises if dorsolateral hypoactivation is nothing more than the neural correlate of poor task performance which may be caused by a variety of pathomechanisms. So, studies are warranted which also control for task performance. Friston *et al.* (1994) have studied controls and patients with a paced verbal fluency task in which performance of normals was adjusted to that of controls in order to be comparable. Controls showed findings that were expected: Increased activation in DLPFC on the left, in the anterior cingulate and in the thalamus. Additionally, there was decreased activation in the posterior cingulate and in superior temporal cortex. The results for schizophrenic patients were rather surprising. They also showed increased activation in the DLPFC, which in 6 patients with especially poor performance was actually *larger* than for normals. The authors argue that schizophrenics have to activate their DLPFC *more* than normals in order to achieve the same level of performance. The only consistent difference between both groups was that schizophrenic patients showed no decreased activation in superior temporal cortex.

In another study with 12 unedicated patients the authors could replicate this finding and additionally found a failure of activation of the anterior cingulate in a verbal fluency task (Dolan *et al.,* 1995; Flechter *et al.,* 1996). The authors then did manipulate dopaminergic transmission (which is known to be distrubed in schizophrenia) by adminstering the dopamine agonist apomorphine. They found a significant normalization of anterior cingulate activation and an attenuation of the abnormal superior temporal activation, expressed in the non-drug state, in the schizophrenic subjects. The authors conclude that schizophrenics show a disturbed fronto-temporal interaction which can partly be normalized by manipulation of dopaminergic transmission. This is in good accordance with the known structural temporal abnormalities found in schizophrenia.

In a study by Spence *et al.* (1997), symptoms were controlled while patients were performing a motor decision task. This study is of interest here because it relates to the hypothesis of the body representation as the basis for the neural self as spelled out above. Spence *et al.* studied schizophrenic patients showing passivity phenomena, also called delusions of alien control (cf. Table I). Delusions of alien control can be described phenomenologically as a loss of agency: Patients have the feeling that they are controlled by other people or alien forces. For example, the patients in the study said things like: "I feel like an automaton guided by a female spirit who has entered me during the scan session", "The voices have some kind of control over my movements", "Other people make me think certain thoughts" or "I could feel God guiding me". Schizophrenic patients with this symptom were studied twice performing paced joystick movements on two occasions 4–6 weeks apart. During the first scan, passivity symptoms were maximal, while by the second scan these symptoms had significantly improved in five of seven patients. Two control groups were scanned: Normal subjects as well as deluded schizophrenic patients without passivity phenomena. Apart from the activity associated with the motor act per se, there was specific acitivity associated with the symptom of passivity: hyperactivity of the anterior cingulate as well as hyperactivity in the right parietal cortex. The specifity of this activity was demonstrated by the fact that it remitted in the control scan in those subjects in whom passivity decreased over time.

These results are in good accordance with neurologic findings of alienation (cf. Table I): The alien hand sign (Goldberg and Bloom, 1990, Gasquoine, 1993) results from extensive lesion of the medial frontal cortex of the hemisphere contralateral to the affected hand and/or the corpus callosum (which connects both hemispheres). Anosognosia (Feinberg and Farah, 1997), sometimes referred to as a delusionary belief in health, is almost exclusively found after lesions of the right hemisphere. The localisation of these lesions as well as the results of the study of Spence *et al.* (1997) support the hypothesis of a bodily based neuromatrix of the self mediating the feeling of agency, e.g. the conviction of having *self* generated actions and thoughts.

## Compulsive-obsessive disorder as a hypervolitional disorder

If we simplify the results concerning certain types of schizophrenia we might say that they represent a hypovolitional disorder. Are there other psychiatric diseases showing disorders of volition? Yes, there are. Obsessive-compulsive disorders (OCD) are characterized by the repetitive experience of unwanted and recurrent, perseverative thoughts (obsessions) and/or a compulsion to repetitively perform certain acts. Patients often report that they have thoughts or do things against their own will. Although this disorder traditionally has been regarded as a neurotic condition, accumulating evidence suggests an association with dysfunction in specific brain regions. The first studies of PET in OCD patients showed hyperactivation of the right head of the caudate nucleus (a subcortical motor structure) and the orbitofrontal cortex. More recent studies show symptom specific activation in those brain areas as well as in the anterior cingulate and several more brain regions (Baxter *et al.,* 1992; McGuire *et al.,* 1994; Breiter *et al.,* 1996). Baxter *et al.* (1990) have proposed a dysregulation in frontal-basal-ganglia loops as the pathomechanism of OCD. They hypothesized that a reduced filter function of the head of the caudate would allow "orbital 'worry' inputs" to direct the output of the caudate, resulting in a reduced inhibition of the thalamus. This has two consequences: On the one hand orbitofrontal activity is enhanced, which creates a potentially self-reinforcing loop which is difficult to break. On the other hand, orbital worry inputs may come to have an indue, rigid influence on thalamic outputs to other brain region mediating OCD symptoms that are not localized to the orbitofrontal cortex.

These hypotheses get support from a recent fMRI study (Breiter *et al.,* 1996). When symptoms were provoked with aversive stimuli during the scan sessions an enhanced activity could be observed, amongst other regions, in the medial orbitofrontal cortex, the lateral inferior prefrontal cortex, and the anterior cingulate. In further accordance with an orbitofrontal dysfunction, Abbruzzese *et al.* (1997) could show that OCD-patients, compared to normal controls and schizophrenics, are selectively impaired on an orbitofrontal test (OAT: object alternation test), while performing normally on a dorsolateral test (Wisconsin card sorting test).

With another neuroimaging study with OCD-patients it could be shown for the first time directly that both succesfull psychotherapy and

Table i: Psychological, neurological and psychiatric phenomena already studied by the cognitive neurosciences which refer to subjectivity and/or consciousness and thus may be used to investigate 'strongly emergent' properties.

| Phenomenon | Description (and a recent reference) |
|---|---|
| *Blindsight* | People with an infarction in the occipital lobe do deny seeing anything, but demonstrably actually do use visual information in their "blind" visual field, e.g. when they are forced to guess a number presented to their "blind" field (Stoerig and Cowey, 1997). |
| *Charles-Bonnett-Syndrom* | People with damage to the visual pathway causing blindness in either a large portion or in the entire visual field experience vivid visual hallucinations in their "blind" field (Fernandez *et al.,* 1997). |
| *Neglect* | The failure to report, respond, or orient to stimuli that are presented contralateral to a brain lesion which is not in the primary cortex. For example, people do eat only the left side of their meat on a dish but claim they have finished (Farah and Feinberg, 1997). |
| *Anosognosia* | Unawareness of neurologic defects or illness. People with right hemisphere lesions often deny their left hemiplegia. Usually after large right hemispheric lesion on the right (Farah and Feinberg, 1997) → Anton's syndrome. |
| *Anton's syndrome* | Blind people, e.g. after bilateral occipital infarction, deny that they are blind (Farah and Feinberg, 1997). |
| *Prosapognosia* | The inability to recognize people's face. Usually caused by bilateral lesions in the inferior temporal lobes (Farah, 1990). |
| *Synaesthesia* | Sensations evoked by stimuli in one modality produce vivid "qualia" normally associated with another modality; patients "see" sounds or "hear" colours (Paulesu *et al.,* 1995). |
| *Phantom limbs* | Amputees feel their missing link, mostly in the form of pain, and might, by tricky experimental manipulation, even experience moving it (Ramachandran, 1994). |
| *Misidentification syndromes* | Patients incorrectly identifiy or reduplicate persons, places, objects or events. Associated with psychotic states or right hemisphere damage with bifrontal or diffuse cortical disturbance (Moselhy and Oyebode, 1997) → Capgras syndrome |
| *Capgras syndrome* | The delusional belief that a person or persons, generally close to the patient, have been replaced by "doubles" of imposters. Often seen in psychotic states, but 1/3 of all cases are seen in conjunction with traumatic brain lesions (Hirstein and Ramchandran, 1997). |
| *Akinetic mutism* | Loss of will after lesion of the anterior cingulate. People do and say nothing, although they are not aphasic or paralyzed (Damasio and van Hoesen, 1983). |
| *Alien hand sign* | One hand, usally the left, acts "autonomous" and is no longer experienced as ones own. For example, the hand will grasp something without its owner wishing to do so and the owner might have to detach it with his healthy hand. Associated with lesions in the medial anterior part of the brain (Goldberg and Bloom, 1990). |
| *Doppelgänger* | The distinct feeling that some invisible being is close by. Often reported by mountaineers, explorers, single-hand sailors, and castaways but also after brain lesions in both hemispheres (Brugger *et al.,* 1997). |
| *Temporal lobe seizures* | Epileptic seizures originating in the temporal lobes, especially the left one, show many features we asscociate usually with awareness and consciousness (Joseph, 1996). |
| *Imagery* | Actions may not be performed but only imagined. It is possible to study the correlated brain activity and thus to study "pure thought" (Faw, 1997). |
| *Sleep dream* | It is possible to study the cerebral correlates of sleep dream as a special form of consciousness with neuroimaging techniques (Braun *et al.,* 1998). |
| *Lucid dreaming* | Dream states in which people are aware of the fact that they are dreaming. Accordingly, they dare to do unusal things, like jumping from a bridge, going through walls, etc. Lucid dreaming can be trained (Gackenbach and LaBerge, 1988). |
| *Hallucinations* | Vividly hearing or seeing things which are not there. Mostly seen in schizophrenia (Andreasen, 1997, Spitzer and Maher, 1998). |
| *Delusions* | A subjectivly sure conviction, uncorrectable by empirical evidence, that something is the case what is objectively wrong, e.g. to be persecuted. Typically seen in schizophrenia, but sometimes also associated with brain lesions (Andreasen, 1997, Spitzer and Maher, 1998) → passivity phenomena. , misidentification syndromes, anosognosia |
| *Passivity phenomena* | Being convinced that someone or something is controlling ones actions or thoughts. In Anglo-american psychiatry refered to as "delusions of alien control", in German psychiatry as "Ich-störungen" (I-disorders). The border between "I" and world becomes permeable; psychic processes are not longer experienced as belonging to the "I" but as made from the outside (Spence *et al.,* 1997). |
| *Obsessions and compulsions* | Acting or thinking something against ones will without being able to stop doing so. For example, people cannot help washing repeatedly their hands or thinking obscene thoughts. Probably a neurobiological disorder (Breiter *et al.,* 1996). |
| *Conversion* | Pseudoneurologic symptoms (disorder of voluntary movements or sensory functions) that are due to an organic lesion but also not simulated (not consciously pretended). The disorder is thought to be the result of psychic conflicts (Mace, 1992). |
| *Dissociative identity disorder* | The existence of two or more distinguishable identities or personal states, which repeatadly take control over the behavior of a person combined with an extraordinary inability to remember important personal information. Formerly known as *multiple personality disorder.* Sometimes doubted to exist (Miller and Triggiano, 1992). |

pharmacotherapy actually alter brain activity (Baxter *et al.*, 1992). The only *common* finding after both therapies was a reduction in the pretreatment hyperactivation of the right head of the caudate nucleus. In contrast the alteration of activity in the anterior cingulate gyrus did depend on the *type* of therapy: Whereas drug responders showed no correlation between the extent of clinical improvement and metabolism in this area behavior therapy responders tended to show an increase of activity in the anterior cingulate. Schwartz *et al.* (1992) speculate that this may be explained by the fact that an element of will is necessary for succesfull behavior therapy because the patients need to exert an effort to control behavioral responses.

In the light of the other studies on pathological volition discussed above a hypothesis may be formulated: Obsessive compulsive disorder can be interpreted as a "hypervolitional syndrome". Due to a pathological positive reinforcement, the orbitofrontal-basal ganglia loop gets decoupled from the representation of the neural self (the representation of body changes in the right hemisphere), so that a feeling of alienation results ("*I* do not want to do this!"). In this respect OCD is a counterpart to the psychomotor poverty syndrom in schizophrenia.

Schizophrenia as well as OCD are good examples for the fact that neuroimaging studies are only intelligibly interpretable in the light of hypotheses and theories. The neural proceses determining mental functions are much more complex than simple hypo- or hyperactivation hypotheses do suggest. For many complex mental functions, we will not just find one brain area lighting up more strongly here or less intense there. Rather, we must interpret observed activities as signs of dysregulations of interconnected and interacting cortical areas resulting in specific clinical pictures.

## Conclusion

The most promising candidates of strong emergent properties are to be found in the realm of mental phenomena. To call a mental property strongly emergent means to claim that it can not be explanatorily reduced to the properties of the underlying neural systems. I have argued that this should be understood as a challenge for scientist to try to explain subjective phenomena associated with consciousness. I have presented a list of phenomena studied in the cognitive neurosciences which are interesting in that respect. Furthermore, I have discussed empirical investigations concerning willed action, decision making and agency in more detail. It goes without saying that we do not yet have explanatorily reduced those phenomena in a satisfying way. However, the discussion might give more credit to the point of view that we should be methodological reductionists *even if* those mental phenomena finally should be emergent in a strong sense: It is the only way to make scientific progress.


Abbruzzese M., Ferri S. and Scarone S. (1997), The selective breakdown of frontal functions in patients with obsessive-compulsive disorder and in patients with schizophrenia: a double dissociation experimental finding. Neuropsychologia **35**, 907–912.

Andreasen N. C. (1997), Linking mind and brain in the study of mental illnesses: a project for a scientific psychopathology **275**, 1286–1593.

Apkarain A. V. (1995), Functional imaging of pain: new insights regarding the role of the cerebral cortex in human brain perception. Sem. Neurosci. **7**, 279–293.

Baxter L. R, Schwartz J. M., Guze B. H., Bergman K. and Szuba M. P. (1990), Neuroimaging in obsessive compulsive disorder: seeking the mediating neuroanatomy. In: Obsessive-Compulsive Disorders: Theory and Managment (M. A. Jenike, L. Baer and W. E: Minichiello, eds.). Mosby Yearbook, St. Louis, pp. 167–188.

Baxter L. R., Schwartz J. M., Bergman K. S. *et al.* (1992), Caudate glucose metabolic rate changes with both drug and behavior therapy for obsessive compulsive disorder. Arch. Gen. Psychiatry **49**, 681–89.

Bechara A., Damasio A. R., Damasio H. and Anderson S. W. (1994), Insensitivity to future consequences following damage to the prefrontal cortex. Cognition **50**, 7–15.

Bechara A., Damasio H., Tranel D and Damasio A. R. (1997), Deciding advantageously before knowing the advantageous strategy. Science **275**, 1293–1294.

Bechara A., Damasio H., Tranel D. and Anderson S. W. (1998), Dissociation of working memory from decision making within the human prefrontal cortex. J. Neurosci. **18**, 428–437.

Beckermann A., Flohr H. and Kim J., eds. (1992), Emergence or Reduction? Springer, Berlin, New York.

Bermúdez J. L., Marcel A. and Eilan N. (1995), The Body and the Self. MIT Press, Cambridge, MA.
Braun A. R., Balkin T. J., Wesensten N. J., Gwadry F., Carson R. E., Varga M., Baldwin P., Belenky G. and Hersowitch P. (1998), Dissociated pattern of activity in visual cortices and their projections during human rapid eye movement sleep. Science **279**, 91–95.
Breiter H. C., Rauch S. I., Kwong K. K., Baker J. R., Weisskoff R. M., Kennedy D. N., Kendrick A. D., Davis T. L., Jiang A., Cohen M. S., Stern C. E., Belliveau J. W., Gaer L., O'Sullivan R. L., Savage C. R., Jenike M. A. and Rosen B. R. (1996), Functional magnetic resonance imaging of symptom provocation in obsessive compulsive disorder. Arch. Gen. Psychiatry **53**, 595–606.
Brugger P., Regard M. and Landis T. (1996), Unilaterally felt "presences": the neuropsychiatry of one's invisible *Doppelgänger*. Neuropsychiat. Neuropsychol. Behav. Neurol. **9**, 114–122.
Buschlinger W. (1993), Denkkapriolen. Könighausen und Neumann, Würzburg.
Buschlinger W., Vollmer G. and Walter H. (1998), The mental life of artifacts. Explications, Questions, Arguments. Z. Naturforsch. **53c**, 455–479.
Chalmers D. J. (1996), The Conscious Mind. Oxford University Press, Oxford.
Changeux J.-P. and Dehaene S. (1996), Neuronal models of cognitive functions associated with the prefrontal cortex. In: Neurobiology of Decision-Making (A. R. Damasio, H. Damasio, H. and Y. Christen, eds.). Springer, New York, pp. 125–144.
Churchland P. M. (1995), The Engine of Reason, the Seat of the Soul. MIT Press, Cambridge, MA.
Churchland P. S. and Sejnowski T. J. (1992), The Computational Brain. MIT Press, Cambridge, MA.
Damasio A. R. (1994), Descartes' Error. Emotion, Reason and the Human Brain. Putnam' Son, New York.
Damasio A. R. and Damasio H. (1989), Lesion Analysis in Neuropsychology. New York.
Damasio A. R. and van Hoesen G. W. (1983), Focal lesions of the limbic frontal lobe. In: Neuropsychology of Human Emotion (K. M. Heilman, P. Satz, eds.). Guilford Press, New York, pp. 85–110.
Damasio A. R., Damasio, H. and Christen Y, eds. (1996), Neurobiology of Decision-Making. Springer, New York.
Damasio H. (1996), Human neuroanatomy of decision-making. In: Neurobiology of Decision-Making (A. R. Damasio, H. Damasio, Y. Christen, eds.). Springer, New York, Berlin, pp. 1-12.
David A. S. and Cutting J. C., eds. (1994), The Neuropsychology of Schizophrenia. Erlbaum, Hillsdale, NJ.
Dennett D. C. (1994), Consciousness Explained. Little, Brown & Company, Boston, Toronto, London.
Devinsky O., Putnam F., Grafman J., Bromfield E. and Theodore W. H. (1995), Contributions of anterior cingulate cortex to behavior. Brain **118**, 279–306.
Dolan R. J., Flechter P., Frith C. D., Friston K. J., Frackowiak R. S. J. and Grasby P. J. (1995), Dopaminergic modulation of an impaired cognitive activation in the anterior cingulate cortex in schizophrenia. Nature **378**, 180–182.
Elliott R., Frith C:D. and Dolan R. J. (1997), Differential neural response to positive and negative feedback in planning and guessing tasks. Neuropsychologia **35**, 1395–1404.
Farah M. (1990), Visual Agnosia. MIT Press, Cambridge, MA.
Faw B. (1997), Outlining a model of mental imaging abilities. Neurosci. Biobehav. Rev. **21**, 283–288.
Feinberg T. E. and Farah M. J. (1997), Behavioral Neurology and Neuropsychology. McGraw-Hill, New York.
Fernandez A., Lichtshein G. and Vieweg W. V. R. (1997), The Charles Bonnet syndrome: a review. J. Nerv. Ment. Dis. **185**, 195–200.
Flechter P. C., Frith C. D., Grasby P. M., Friston K. J. and Dolan R. J. (1996), Local and distributed effects of apomorphine on fronto-temporal function in acute unmedicated schizophrenia. J. Neurosci. **16**, 7055–7062.
Frackowiak R. S. J., Friston K. J., Frith C. D., Dolan R. J. and Mazziotta J. C. (1997), Human brain functioning. Academic Press, San Diego, London.
Frith C. D., Friston K. J., Liddle, PIF. and Frackowiak, R. S. J. (1991), Willed action and prefrontal cortex in man: a study with PET. Proc. Roy. Soc. B, **244**, 241–246.
Gackenbach J. and LaBerge S. (1998), Consious Mind, Sleeping Brain. Plenum, New York, London.
Gasquoine P. G. (1993), Alien hand sign. J. Clin. Exp. Neuropsych. **15**, 653–667.
Gazzaniga M., ed. (1995), The Cognitive Neurosciences. MIT Press, Cambridge, MA.
Goldberg G., Bloom K. K. (1990), Alien hand sign: localization, lateralization and recovery. Am. J. Phys. Med. Rehab. **69**, 228–338.
Gollwitzer P. M. and Bargh J. A., eds. (1996), The Psychology of Action. Guilford Press, New York.
Goyer P. F., Andreason P. J., Sempl W. E., Clayton A. H., King A. C., Compton-Toth B. A., Schulz S. C. and Cohen R. M. (1994), Positron-emission tomography and personality disorders. Neuropsychopharm. **10**, 21–28.
Hare R. D. and Quinn, M. J. (1971), Psychopathy and autonomic conditioning, J. Abnorm. Psychol. **77**, 223–235.
Heckhausen H., Gollwitzer P. M., Weinert F. E., eds. (1987), Jenseits des Rubikon. Der Wille in den Humanwissenschaften. Springer, Berlin, New York.
Hirstein W. and Ramachandran V. S. (1997), Capgras syndrome: a novel probe for understanding the neural representation of the identity and familarity of persons. Proc. R. Soc. Lond. B **264**, 437–444.
Hornak J., Rolls E. T. and Wade (1996), Face and voice expression identification in patients with emotional and behavioral changes following frontal lobe damage. Neuropsychologia **34**, 247–261.
Hyder R., Phelps E. A., Wiggins C. J., Labar K. S., Blamire A. M., Shulman R. G. (1997), "Willed action": a

functional MRI study of the human prefrontal cortex during a sensorimotor task. Proc. Nat. Acad. Sci. USA **94**, 6989–6994.
Jeannerod M. (1997), The Cognitive Neuroscience of Action. Blackwell, Oxford.
Joseph R. (1996), Neuropsychiatry, Neuropsychology, and Clinical Neuroscience, 2nd ed. Williams & Wilkins, Baltimore.
Kalcher J., Flotzinger D., Gölly S., Neuper CH. Pfurtscheller G. (1994), Graz Brain-Computer Interface (BCI) II. In: Computers for Handicapped Persons (W. L. Zagler, G. Busby and R. R. Wagner, eds.). Springer, Berlin, New York, pp. 170–176.
Kane R. (1996), The Significance of Free Will. Oxford University Press, Oxford, New York.
Keller I. and Heckhausen H. (1990), Readiness potentials preceding spontaneous motor acts: voluntary versus involuntary control. Electroenc. Clin. Neurophys. **76**, 351–361.
Lapierre D., Braun C. M. J. and Hodgins S. (1995), Ventral frontal deficits in psychopathy: neuropsychological test findings. Neuropsychologia **33**, 139–151.
Libet B. (1993), The neural time factor in conscious and unconsious events. In: Experimental and Theoretical Studies of Consciousness (Ciba Foundation Symposium). Wiley, Chicester, pp. 123–146.
Libet B., Gleason C. A., Wright E. W. and Pearl, D. K. (1983), Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). Brain **106**, 623–642.
Liddle P. F. (1987), The symptoms of chronic schizophrenia: a reexamination of the positive-negative dichotomy. Br. J. Psychiatry **151**, 145–151.
Liddle P. F. (1994), Volition and schizophrenia. In: The Neuropsychology of Schizophrenia. (A. S. David and J. C. Cutting, eds.). Erlbaum, Hillsdale, N.J, pp. 39–49.
Liddle P. F., Friston K. J., Frith C. D., Hirsch S. R., Jones T. and Frackowiak, R. S. J. (1992), Patterns of cerebral blood flow in schizophrenia. Br. J. Psychiat. **160**, 19–186.
Mace C. J. (1992), Hysterical conversion. II: A critique. Brit. J. Psychiat. **161**, 378–389.
McGuire P. K., Bench C. J., Frith C. D., Marks I. M., Frackowiak R. S. J. and R. J. Dolan (1994), Functional anatomy of obsessive-compulsive phenomena. Br. J. Psychiat. **164**, 459–468.
Metzinger T., ed. (1995), Conscious Experience. Schöningh/Imprint Academic, Exeter, UK.
Miller S. D. and Triggiano P. J. (1992), The psychophysiological investigation of multiple personality disorder: review and update. Am. J. Clin. Hypnos. **35**, 47–61.
Moselhy Z. H. and Oyebode F. (1997), Delusional misidentification syndromes: a review of the anglophone literature. Neur. Psychiat. Br. Res. **5**, 21–26.
Northoff G., ed. (1997), Neuropsychiatrie und Neurophilosophie. Schöningh, Paderborn.
Paulesu E., Harrsion J, Baron-Cohen S., Watson J. D. G., Goldstein L. H., Heather J., Frackowiak R. S. J. and Frith C. D. (1995), The physiology of coloured hearing. A PET activation study of colour-word synaesthesia. Brain **118**, 661–676.
Paus T, Petrides M., Evans A. and Meyer E. (1993), Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: a positron emission tomography study. J. Neurophysiol. **70**, 453–469.
Penrose R. (1994), Shadows of the mind. Oxford University Press, New York.
Phelps E. A., Hyder F., Blamire A. M. and Shulman R. G. (1997), FMRI of the prefrontal cortex during overt verbal fluency. Neuroreport **8**, 561–565.
Posner M. I. and Raichle M. E. (1994), Images of the mind. Freeman and Company, New York.
Prinz W. (1998), Die Reaktion als Willenshandlung. Psychol. Rundschau **49(1)**, 10–20.
Raine A., Buchsbaum M. S., Stanley J., Lottenberg S., Abel L. and Stoddard J. (1994), Selective reductions in prefrontal glucose metabolism in murderers. Biol. Psychiatry **36**, 365–373.
Ramachandran V. S. (1994), Phantom limb, neglect syndromes, repressed memories and Freudian psychology. Int. Rev. Neurobiol. **37**, 291–333.
Ramachandran V. S., Hirstein W. (1997), Three laws of qualia: what neurology tells us about the biological functions of consciousness. J. Consciousn. Stud. **4**, 429–457.
Rolls E. T. (1996), The orbitofrontal cortex. Phil. Trans. R. Soc. Lond. B. **351**, 1433–1444.
Rolls E. T., Hornak J., Wade D. and McGrath J. (1994), Emotion related learning in patients with social and emotional changes associated with frontal lobe damage. J. Neur. Neurosurg. Psychiat., **57**, 1518–1524.
Schwartz J. M., Martin K. M. and Baxter L. R. (1992), Neuroimaging and cognitive-biobehavioral self-treatment for obsessive compulsive disorder: practical and philosophical considerations. In: Obsessive-Compulsive Disorders (I. Hand, W. K. Goodman and U. Evers, eds.). Springer, Berlin, Heidelberg, pp. 82–101.
Shallice T. (1988), From Neuropsychology to Mental Structure. Cambridge University Press, Cambridge, MA.
Spence S. A., Brooks D. J., Hirsch S. R., Liddle P. F., Meehan J, and Grasby P. M. (1997), A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). Brain **120**, 1997–2011.
Spitzer M. and Maher B. (1998), Experimental Psychopathology. Cambridge University Press, Cambridge, MA (in press).
Stein D. (1998), Neural Networks and Psychopathology. Cambridge University Press, Cambridge, UK (in press).
Stephan A. (1998a), Varietes of emergence in artificial and natural systems. Z. Naturforsch. **53c**, 639–656.
Stephan A. (1998b), Emergenz. Historisch-systematische Studien zu einem zentralen Begriff der Metaphysik und Wissenschaftstheorie. Karlsruhe: Habilitationschrift.
Stieve H. (1997), Limits of natural science: brain research and computers. Z. Naturforsch. **50c**, 317–336.
Stoerig P. and Cowey A. (1997), Blindsight in Man and Monkey. Brain **120**, 535–559.

Vollmer G. (1992), Das Ganze und seine Teile. Holismus, Emergenz, Erklärung und Reduktion. In: Wissenschaftstheorien in der Medizin (W. Deppert, W. Kliemt, H. Lohff and J. Schaeffer, eds.). de Gruyter, Berlin, New York, pp. 183–223.
von Cranach M and Foppa K. (1996), Freiheit des Entscheidens und Handelns. Ein Problem der nomologischen Psychologie. Asanger, Heidelberg.
Walter H. (1996a), Minimale Neurophilosophie. In: Cognitio humana (Ch. Hubig, H. Poser, eds.). Berlin, pp. 1515–1522.
Walter H. (1996b), Die Neurobiologie menschlicher Autonomie. Nervenarzt **67** (Suppl. 1), 19.
Walter H. (1997a), Neuroimaging und philosophy of mind. In: Neuropsychiatrie und Neurophilosophie (G. Northoff, ed.). Schöningh, Paderborn, pp. 193–222.
Walter H. (1997b), Authentische Entscheidungen und emotive Neurowissenschaft, Philosophia naturalis **34**, 147–174.
Walter H. (1998), Neurophilosophie der Willensfreiheit. Von libertarischen Illusionen zum Konzept natürlicher Autonomie. Schöningh, Paderborn.
Wilke K. V:. (1993), Real People. Personal Identity without Thought Experiments. Clarendon Press, Oxford.