Principles of Neuronal Integration and Defense Mechanisms: Neuropsychoanalytic Hypothesis

Georg Northoff (Magdeburg) & Heinz Boeker (Zurich)

Current progress in the cognitive and affective neurosciences affects the development of psychoanalytic theory and practice. However, despite the emerging dialogue between neuroscience and psychoanalysis, the neuronal processes underlying defense mechanisms remain unclear. We consider defense mechanisms as complex emotional-cognitive constellations. As such they require specific functional mechanisms to integrate neuronal activity across several brain regions—that is, neuronal integration. We develop hypotheses about the relationship between principles of neuronal integration and certain defense mechanisms. Neuropsychological and imaging findings in neuropsychiatric disorders are discussed in relation to neuronal integration and specific defense mechanisms. We conclude that the investigation of neuronal integration contributes to revealing the psychophysiological mechanisms underlying defense mechanisms.

Those endorsing the integration of psychoanalysis and neuroscience usually refer to Freud's "Project for a Scientific Psychology" (Freud 1950 [1895]), in which he sought a unitary conception of mind and brain (McCrone, 2004). However, the dialogue between psychoanalysis and neuroscience has only recently emerged (Beutel, Stern, et al., 2003; Brockman, 2001; Gabbard, 1992a, 1992b; Kandel, 1998, 1999; Northoff, 2000; Northoff, Böker, et al., 2003; Shore, El Khashlan, et al., 2003; Shore, Sumner, et al., 2003; Solms, 1997; Solms & Lechevalier, 2002; Westen, 1999, 2002; Westen & Gabbard, 2002a, 2002b). Kandel (1999) recently articulated the hope "that biology might reinvigorate the psychoanalytic exploration of the mind." He emphasized the need for combined research approaches in psychopathology (implicit and explicit memory, development, etc.) and psychotherapy (Anderson, 2004; Anderson, Hackett, et al., 2004; Beutel, Stern, et al., 2003; Pugh, 2002). This led some authors to investigate the neuronal mechanisms of psychodynamic processes. For example, emphasizing the developmental features of the right orbitofrontal cortex, Shore (Shore, El Khashlan, et al., 2003; Shore, Sumner, et al., 2003) proposes neurobiological mechanisms of unconscious processes like projective identification. Based on psychoanalytic treatment of patients with orbitofrontal cortical lesions, Solms (1997; Solms, Turnbull, et al., 1998) makes inferences about the localization of early, immature, and sensorimotor regression mechanisms. Furthermore, Northoff, Böker, et al. (2003) were able to link sensorimotor regression, as observed in catatonic patients, to a complex neural network, including the orbitofrontal, medial prefrontal, and premotor cortex (see later for details). The orbitofrontal cortex also plays a role in emotional-cognitive interaction, which Westen and Gabbard (2002a, 2002b) consider to be crucial in conflict and compromise. However, despite these studies and the crucial role of defense mechanisms in psychoanalytic theory and practice, their underlying neuronal mechanisms remain yet to be explored.

Defense mechanisms and conflict are two concepts that have remained at the core of psychodynamic approaches to understanding and treating clinical psychopathology. The term "defense mechanisms" was coined over a hundred years ago to describe psychological mechanisms for coping with intrapsychic conflicts (Freud, 1894). From a psychoanalytic...
perspective, defense mechanisms mediate between an individual’s wishes, needs, and affects on the one hand, and both internalized object relations and external reality on the other. Through specific constellations of affective and cognitive function, defense mechanisms help to resolve conflicts whether triggered by internal stressors or external stressors. Since defense mechanisms can be considered complex emotional-cognitive constellations, the recent progress in affective and cognitive neuroscience (Damasio, 1999; Gazzaniga, 2004; Panksepp, 1998a, 1998b, 2003) introduces the possibility of mapping out the neuronal processes of defenses.

Skeptics (Edelson, 1984; Jones, 2003; Schneider & Northoff, 2006) doubt that there can be common conceptual grounds for linking the hypothesis about unconscious aspects of the mind (as advanced by psychoanalysis) such as defense mechanisms and knowledge about the brain (as framed by neuroscience). They claim that the complexity and richness of defense mechanisms will be lost in empirical neuroscientific investigation. For example, defense mechanisms are neither purely cognitive nor emotional. Instead, complex emotional-cognitive interaction dominates, which remains elusive to empirical investigation. Moreover, they argue that it remains impossible to associate the complex subjective experience to neuronal activity in specific brains regions; they thus argue against neuronal localization of defense mechanisms (Schneider & Northoff, 2006).

In order to yield some answer to the first claim, we consider defense mechanisms to be complex modes of emotional-cognitive interaction. We argue that they are constituted by interaction between emotional and cognitive function reflecting specific emotional-cognitive constellations. Different defense mechanisms might then correspond to different forms of emotional-cognitive interaction.

To account for the second claim of the skeptics, we focus on neuronal integration rather than neuronal localizationism. We hypothesize that different forms of neuronal integration might account for different ways of emotional-cognitive interaction and could therefore correspond to different defense mechanisms. In what follows, we develop a hypothesis of the relationship between neuronal integration, emotional-cognitive integration, and defense mechanisms. We describe what is meant by the term neuronal integration. We then describe different principles of neuronal integration and relate them to specific defense mechanisms, especially those in depression, catatonia, and phobia, developing hypotheses with predictions and suggestions for further empirical testing.

Neuronal integration

Neuronal integration describes the coordination and adjustment of neuronal activity across multiple brain regions. The interaction between distant and remote brain areas is considered necessary for a complex function to occur, such as emotion or cognition (Friston, 2003; Price & Friston, 2002). Neuronal integration focusing on the interaction between two or more brain regions must be distinguished from neuronal segregation (Friston, 2003; Price & Friston, 2002). Here, a particular cognitive or emotional function or processing capacity is ascribed to neural activity in a single area that is both necessary and sufficient; one can subsequently speak of neuronal specialization and localization. We assume that defense mechanisms, as complex emotional-cognitive interactions, cannot be localized in specialized or segregated brain regions. Instead, we assume that defense mechanisms require interaction between different brain regions and thus neuronal integration.

For neuronal integration to be possible, distant and remote brain regions have to be linked together, which is possible for by connectivity. Connectivity describes the relation between neural activity in different brain areas. For anatomical connectivity we use the term “connections” in order to clearly distinguish it from functional connectivity. In addition, Friston and Price (2001) distinguish between functional and effective connectivity. Functional connectivity describes the “correlation between remote neurophysiological events,” which might be due to either direct interaction between the events or other factors mediating both events. A correlation can indicate either a direct influence of one brain area on another or their indirect linkage via other factors. In the first case, the correlation is due to the interaction itself, whereas in the second, the correlation might be due to other, rather indirect factors such as, for example, stimuli based on common inputs. In contrast, effective connectivity describes the direct interaction between brain areas; it “refers explicitly to the (direct) influence that one neural system exerts over another, either at a synaptic or population level” (Friston & Price, 2001). Here, effective connectivity is considered on the population level because this corresponds best to the level of different brain regions investigated in this paper. For example, the prefrontal cortex might modulate its effective connectivity with subcortical regions, thereby influencing specific functions like, for example, interoceptive processing.

Based upon connectivity, neural activity between distant and remote brain regions has to be adjusted, coordinated, and harmonized. Coordination and ad-
justment of neural activity might not be arbitrary but be guided by certain principles of neuronal integration (Northoff, Heinzel, et al., 2004). These principles describe functional mechanisms according to which the neural activity between remote and distant brain regions is organized and coordinated. We want to focus here on four particular principles of neuronal integration—reciprocal modulation, modulation by functional unity, top-down modulation, and modulation by reversal—as revealed in recent imaging studies on emotional-cognitive interaction. We hypothesize that each of the four principles of neuronal integration can be associated with a specific defense mechanism; this allows for developing predictions and suggestions for future empirical testing.

Top-down modulation and somatization

We introduce the principle of top-down modulation that integrates cortical and subcortical regions. Our hypothesis is that top-down modulation is related to the defense mechanism of somatization.

Top-down modulation

Top-down modulation might be described as modulation of hierarchically lower regions by those that are higher in the hierarchy. Often top-down modulation concerns modulation of neural activity in subcortical regions by cortical regions. For example, premotor/motor cortical regions might modulate neural activity in subcortical basal ganglia like the caudate and striatum (Masterman & Cummings, 1997; Northoff, 2002a, 2002b). Yet another example is top-down modulation of primary visual cortex by prefrontal cortical regions, which has been shown to be essential in visual processing (Lamme, 2004). Top-down modulation might be related to the concepts of “re-entrant circuitry” (Tononi & Edelman, 2000) and feedback modulation (Lamme, 2001). These concepts allow for circuiting of information and readjustment of neural activity in one area according to another, rather distant area. This provides the possibility of adjusting, filtering, and tuning neural activity in the lower area according to the one in the higher area. For example, top-down modulation allows for attentional modulation of visual input, which makes selective visual perception possible (Lamme, 2001).

We here want to focus on the medial prefrontal cortex. Neural activity in both the medial prefrontal cortex and amygdala has been shown to be involved in emotional processing (Murphy, Nimmo-Smith, et al., 2003; Phan, Wager, et al., 2002). The functional relationship between these is supposed to be characterized by top-down modulation of the amygdala by the medial prefrontal cortex (Davidson, 2002; Pessoa, McKenna, et al., 2002; Pessoa & Ungerleider, 2004; Shin, Wright, et al., 2005). Medial prefrontal cortical regions seem also to exert top-down control of neural activity in the insula (Nagai, Critchley, et al., 2004), which is densely and reciprocally connected with subcortical medial regions like the hypothalamus, the periaqueductal gray (PAG), the substantia nigra, and various brainstem nuclei such as the raphe nuclei and the locus coeruleus (Panksepp, 1998a, 1998b).

Both the amygdala and the subcortical medial regions are involved in regulating internal bodily functions, whereas medial prefrontal cortical regions have been associated with emotional processing (Murphy, Nimmo-Smith, et al., 2003; Northoff & Berman, 2004; Phan, Wager, et al., 2002). The three regions—medial prefrontal cortex, amygdala, and subcortical medial regions—show dense and reciprocal connections (Ongur & Price, 2000; Panksepp, 1998a, 1998b). One might, therefore, assume modulation between all of them. This might not only include top-down modulation, but also the reverse kind of modulation, bottom-up modulation, as illustrated in Figure 1. In the case of bottom-up modulation, a hierarchically lower area modulates activity in an area that is higher in the hierarchy. For example, subcortical midline regions might modulate neural activity in medial prefrontal cortex via the insula, thus concerning the same regions as top-down modulation. Accordingly, bottom-up and top-down modulation might co-occur across the same regions (see Figure 1).

Functionally, this co-occurrence of bottom-up and top-down modulation might allow for reciprocal adjustment between emotional and internal bodily processing. Internal bodily processing concerns only stimuli from the own-body, so-called internal self-related stimuli. These include, for example, stimuli from autonomic-vegetative or other humoral functions. Emotional processing, however, concerns both internal self-related bodily stimuli and external self-related stimuli from the environment. For example, emotional processing might be induced by specific events within the environment that, in turn, might induce internal bodily stimuli. Since, however, studies about the functional relationship either between the three regions or between both kinds of processing have not yet been reported, our assumptions must be considered preliminary and speculative.

Psychologically, the co-occurrence of top-down and bottom-up modulation might correspond to the
processing of emotional stimuli:  
Top-down modulation  

Processing of bodily stimuli:  
Bottom-up modulation  

Figure 1. Schematic illustration of bottom-up and top-down modulation.

co-occurrence between emotional and bodily awareness. We are aware of the emotions associated with certain events in the environment. This co-occurs with awareness of one’s own body, which usually remains in the background. Such co-occurrence might account for our predominantly outward focus, directing our attention towards other persons and events in our environment, whereas the inward focus, directing our attention towards our own body, is not as central and predominant and seems to remain in the background.

Somatization

The functional balance between bottom-up and top-down modulation may be altered during regressive psychic functioning. Regression to an early stage might be characterized by the predominance of bodily reactions rather than emotional or cognitive reactions. The patients become hyperaware of any subtle changes in their own body while, at the same time, they become less aware of their environment (see later for further details). Such specific form of mental functioning connected with a strong focus on bodily complaints might therefore be called somatization, which is the term adopted in the following discussions. The defense mechanisms of somatization can be observed paradigmatically in patients with depression, who often subjectively experience bodily changes, especially autonomic-vegetative symptoms. We hypothesize that somatization in depression might be related to an abnormal balance between bottom-up and top-down modulation with respect to emotional and internal bodily processing.

Functionally, somatization might indicate predominance of internal bodily processing compared to emotional processing. Processing of signals from internal bodily control centers predominates over emotional processing, originating either internally or externally. Thus internal bodily processing takes over, whereas emotional processing, unlike in the healthy case, shifts into the background. The balance between internal bodily and emotional processing is subsequently readjusted and shifted to a novel level. We therefore predict that depressed patients with strong somatization react considerably more to internal bodily stimuli than to internal and especially external emotional stimuli. Depressed patients might, indeed, also show abnormally strong autonomic-vegetative reactivity; though this is still a matter of debate, some studies show abnormal vegetative reactivity (such as heart-rate variability) in depression (Bar, Greiner, et al., 2004; Guinjoan, Bernabo, et al., 1995). Finally, we suggest that depressed patients might be less responsive to externally induced emotions arising, for example, in social interactions, as was postulated by the interactional theory of depression, which describes depression as a self-perpetuating interpersonal system. In particular, this system is driven by a growing discrepancy between the verbal content and the affective quality of others’ responses (Coyne, 1976a, 1976b, 1985; Coyne, Kessler, et al., 1987; Lewinsohn, Hoberman, et al., 1985).

Psychologically, somatization is reflected in a person’s heightened awareness of his or her own body and internal bodily functions. Depressed patients shift their focus of attention from their own and others’ emotions to their own bodily functions. Instead of experiencing
their own emotions, they experience their own bodily functions. Instead of observing others’ emotions, they observe their own body. We therefore predict that depressed patients with strong somatization show abnormally strong bodily attention and decreased emotional attention. This might be accompanied by deficits in theory-of-mind tasks where observation of social interactions is required; depressed patients might show deficits in these tasks because of their shift from an outward focus to an inward focus (Inoue, Tonooka, et al., 2004; Kerr, Dunbar, et al., 2003) and attention (Murphy, Sahakian, et al., 1999; Paradiso, Lamberty, et al., 1997; Sheppard, 2004). What remains to be shown, however, is that the balance between emotional and bodily attention is shifted towards the latter in depressed subjects.

Physiologically, the altered functional balance between bottom-up and top-down modulation might correspond to altered neural activity in the medial prefrontal cortex, amygdala, and subcortical medial regions. This is in accordance with the current empirical findings (Elliott, Rubinsztein, et al., 2002; Liotti, Mayberg, et al., 2002; Mayberg, 2003). However, the involvement of subcortical medial regions in depression has not yet been clearly demonstrated in human functional imaging, possibly due to methodological limitations with respect to these regions. Furthermore, the mechanisms of top-down modulation itself and its change in depression has not been demonstrated. To further specify these findings, we predict that depressed patients will show abnormal reactivity to bodily stimuli compared to emotional stimuli. Moreover, future studies should compare depressed patients with strong somatization/somatic symptoms with patients without such symptoms. Finally, we predict abnormal connectivity between the amygdala, medial prefrontal cortex, and subcortical medial regions in depressed patients; we predict stronger bottom-up connectivity from subcortical to cortical regions and weaker top-down connectivity from cortical to subcortical regions. Recent studies investigated connectivity in depression and reported anomalies (Seminowicz et al., 2004); however, the focus was on limbic-cortical rather than on subcortical-cortical connectivity.

**Reciprocal modulation and introjection**

We now introduce the principle of reciprocal modulation, which modulates neural activity across different cortical regions. Our hypothesis is that reciprocal modulation is related to the defense mechanism of introjection.

Reciprocal modulation

Recent studies (Goel & Dolan, 2003a, 2003b; Northoff & Bermpohl, 2004; Northoff, Heinzel, et al., 2004) demonstrate a pattern of opposite-signal changes in the medial and lateral prefrontal cortex during emotional-cognitive interaction. These results are compatible with the assumption of functional mechanisms of reciprocal modulation and reciprocal attenuation during emotional-cognitive interaction. Reciprocal modulation can be defined by signal changes in opposite directions (i.e., signal increases and decreases) in different regions. For example, emotional-picture viewing is known to lead to signal increases in the medial prefrontal cortical regions and concurrent signal decreases in the lateral prefrontal cortex (Murphy, Nimmo-Smith, et al., 2003; Northoff, Heinzel, et al., 2004; Phan, Wager et al., 2002). In contrast, cognitive tasks like judgment or evaluation induce the reverse pattern with signal increases in the lateral prefrontal cortex and signal decreases in the medial prefrontal cortex. This is compatible with the functional mechanism of reciprocal modulation (see Figure 2; see also Northoff, Heinzel, et al., 2004). Interestingly, analogous patterns of reciprocal modulation have been observed in other cortical regions, including the medial and lateral orbitofrontal cortex (Northoff & Bermpohl, 2004; Northoff, Heinzel

![Figure 2](attachment:image.png)

**Figure 2.** Schematic illustration of reciprocal modulation (A) and reciprocal attenuation (B).
et al., 2004; Northoff, Richter et al., 2000; Northoff, Witzel et al., 2002; O’Doherty, Kringelbach et al., 2001; O’Doherty, Rolls et al., 2001), right and left motor cortex (Allison, Meador et al., 2000), striate and extrastriate visual cortex (Kleinschmidt, Buchel et al., 1998), subgenual anterior cingulate and right prefrontal cortex (Liotti, Mayberg et al., 2002), sub/pre- and supragenual anterior cingulate (Bush, Luu et al., 2000), as well as visual and auditory cortex (Laurienti, Burdette et al., 2002; Laurienti, Field et al., 2002).

Emotional-cognitive interaction is then associated with the functional mechanism of reciprocal attenuation: inclusion of an emotional component into a cognitive task resulting in, for example, emotional judgment leads to smaller signal decreases in medial prefrontal cortical regions and, at the same time, smaller signal increases in lateral prefrontal cortical regions; this has been called attenuation (Northoff, Heinzel et al., 2004). Since attenuation concerned both medial and lateral prefrontal cortical regions in opposite directions (i.e., smaller signal decreases/increases, respectively), one can speak of reciprocal attenuation.

**Introjection**

Reciprocal modulation and attenuation might be altered in the defense mechanism of introjection. Introjection is not always a defense. Considering the definition of introjection offered by Kernberg (1976), introjection is the reproduction and fixation of an interaction with the environment by means of an organized cluster of memory traces, implying at least three components: (1) the image of an object, (2) the image of the self in interaction with that object, and (3) the affective coloring of both the object-image and the self-image.

From this perspective, introjection is a developmental process on a continuum with identification, which may or may not be pathological. Introjection may be looked upon as a pathological mechanism of defense, if it is used as part of a regressive process to overcome the unbearable loss of the object and the differentiation between subject and object. Regarding the structural level of the developmental process, introjection thus may be differentiated from identification (cf. Mentzos, 1982). As a component of the depressive-melancholic mode of compensation and defense (cf. Mentzos, 1982, pp. 250), introjection is connected with a distorted perception of inner and outer reality. Dependent on the ambivalent object, the introjection in depressed subjects is connected with different vicious circles of growing social isolation on the one hand, and the development of negative thoughts (cf. Beck, 1974) on the other. Finally, this specific cognitive-emotional processing in depression results in the development of an inward focus. Thus, the subject–object relationship is no longer directed outward but inward: the experience of the outward world is transformed into an experience of the inward self.

Introjects may be characterized as intropsychic representations that organize and direct behavior as the subject “checks” impulses against imagined responses of the introjected object. In the case of depression, conflicts with significant objects are often internalized, and the resulting aggression is often redirected against the depressed person’s own self. We hypothesize that introduction in depression could be related to abnormal reciprocal modulation during emotional-cognitive interaction (Adroer, 1998; Berman & McCann, 1995; Deci, Eghrari et al., 1994; Malancharuvil, 2004).

Functionally and psychologically, emotional-cognitive processing might be altered in depressed patients with strong introjection. Depressed patients are no longer able to evaluate their own emotional and bodily experience appropriately; the judgments of their own states are “subjectively” distorted and decoupled from “objective” reality. Subjective distortion is manifest in the extreme negativity of their judgments concerning either their own emotions and their own body or emotions in other persons and events in their environment. This extreme negativity corresponds to what psychologically has been described as the “negative bias” (Elliott, Rubinsztein et al., 2002; Gotlib, Krasnoperova et al., 2004). A recent study on faces (Gotlib, Krasnoperova et al., 2004) has demonstrated that the “negative or attentional bias” might be related to interpersonal dysfunction in these patients. However, further studies are necessary to show how and why the “negative bias” apparently induces decoupling of “subjective” judgments from “objective reality.”

We predict that abnormal reciprocal modulation and attenuation during emotional-cognitive interaction might be crucial in generating introjection in depression. Due to abnormal reciprocal modulation, the depressed subject might no longer be able to cognize and thus to introspect and evaluate his or her own experience appropriately. The depressed subject might no longer be able to connect and link his or her emotional experience with his or her own cognition, both becoming decoupled from each other. If, however, emotion and cognition are decoupled from each other, reciprocal adjustment remains impossible. The depressed subject’s emotional experience has no access to cognition any more, making any appropriate judgment impossible. Similarly, the perceived emotions of others have no access to cognitions either, making any appropriate
emotional judgment about them impossible. Others' emotion thus become directed inward rather than outward. Emotional conflicts with other persons are internalized and directed against one's own person rather than against the other. Introspection as characterized by strong inward focus and weak outward focus results. Neuropsychologically, this might be tested in tasks distinguishing own and others' emotional stimuli as well as in tasks employing social situations like theory of mind. We predict that depressed patients with strong introjection remain unable to appropriately distinguish between own and others' stimuli. This would be compatible with empirical findings in depressed patients showing deficits in theory-of-mind tasks (see above).

Physiologically, altered reciprocal modulation and attenuation in depressed subjects with strong introjection might be reflected in altered neural activity in medial and lateral prefrontal cortex. Depressed patients show hyperactivity in the medial prefrontal cortex and hypoactivity in the lateral prefrontal cortex during emotional stimulation (Elliott, Rubinsztein, et al., 2002; Liotti, Mayberg, et al., 2002; Mayberg, 2003). This corresponds indeed to abnormal reciprocal modulation between medial and lateral prefrontal cortex. What remains to be shown, however, is that this abnormal neural activity in medial and lateral prefrontal cortex is related to emotional and cognitive dysfunction. Furthermore, abnormal reciprocal attenuation during emotional-cognitive interaction has not yet been demonstrated in depressed patients.

Finally, we want to raise a methodological question regarding an appropriate control group. Ideally, this would be depressed patients without introjection; whether, however, this is possible at all remains questionable since introjection is one of the core defense mechanisms observed in depression.

Modulation by reversal and displacement

We introduce the principle of modulation by reversal, which modulates neural activity within one region across different stimuli. Our hypothesis is that modulation by reversal is related to the defense mechanism of displacement.

Modulation by reversal

Several studies have demonstrated reversal of signal changes in the opposite direction within the same region. Signal changes within the orbital and medial prefrontal cortex (OMPFC), were, for example, reversed (from signal increases to signal decreases) by either preceding (e.g., expectancy) or simultaneous (e.g., distraction or increased focus) attentional manipulation of emotional stimulation. To account for attentional modulation in the paradigm, emotional pictures were preceded either by an expectancy period or by a flickering stimulus (Keightley, Winocur, et al., 2003; Ploghaus, Becerra, et al., 2003; Ramman & Owen, 2004; Simpson, Drevets, et al., 2001; Simpson, Snyder, et al., 2001). A recent study (Northoff, Richter, et al., 2005) observed reversal of signal increases to signal decreases in both OMPFC and posterior cingulate cortex when emotional judgment was preceded by an expectancy period. Similar changes in signal direction were also observed in cognitive tasks during attentional modulation. For example, noun generation, object knowledge, and impersonal/personal word-judgment tasks (Ferstl & von Cramon, 2002; Kelley, Macrae, et al., 2002; Mitchell, Colledge, et al., 2002; Simpson, Drevets, et al., 2001; Simpson, Snyder, et al., 2001) induced similar signal decreases in OMPFC. In other words, neural activity in response to an emotional stimulus is dependent on whether the stimulus was expected or not; expected stimuli results signal increases and unexpected stimuli results in decreases.

Analogous signal changes in OMPFC have also been observed in reward studies in both humans and monkeys. Expectancy of reward delivery induces signal increases (in humans) and neuronal excitation (in monkeys) in the OMPFC during the expectancy period itself. If, however, the reward delivery is delayed, omitted, or devalued, signal decreases (in humans) and neuronal inhibition (in monkeys) can be observed in OMPFC (Anderson & Sobel, 2003; Braver & Brown, 2003; Gottfried, O'Doherty, et al., 2003; McClure, Berns, et al., 2003; Kringelbach & Rolls, 2004; Montague & Berns, 2002; Rolls, 2000; O'Doherty, Dayan, et al., 2003; Rolls, Tovee, et al., 1999; Schultz, 2000; Schultz, Tremblay, et al., 2000; Tobler, Dickinson, et al., 2003; Tremblay & Schultz, 1999, 2000a, 2000b). Moreover, modulation by reversal has also been observed in OMPFC during a switch from abstract reward to punishment. The OMPFC showed true signal increases during reward that were reversed into signal decreases during punishment (Critchley, Mathias, et al., 2001; O'Doherty, Deichmann, et al., 2002; O'Doherty, Dayan, et al., 2003; O'Doherty, Kringelbach, et al., 2001; O'Doherty, Rolls, et al., 2001). Taken together, these findings fit well with the functional mechanism of modulation by reversal: a modulating factor (i.e., expectancy) reverses the type of neural activity within a specific region (i.e., OMPFC).
during a particular task (i.e., emotional judgment). The functional mechanism of modulation by reversal is illustrated schematically in Figure 3.

Physiologically, the exact nature of signal decreases in fMRI, as distinguished from signal increases reflecting neuronal excitation, has not yet been elucidated (Gusnard & Raichle, 2001; Logothetis, Pauls, et al., 2001). However, a recent study strongly suggests that signal decreases are largely a result of active neuronal inhibition (Shmuel, Yacoub, et al., 2002). If signal decreases do indeed reflect neuronal inhibition and signal increases neuronal excitation, then attentional manipulation reverses the type of predominant neural activity in a specific region during an emotional or cognitive task. The transformation of signal increases into signal decreases would then correspond to reversal of neuronal excitation into neuronal inhibition, the latter predominating over the former.

Functionally, such reversal in signal (or neuronal activity) type is supposed to reflect a new linkage between an externally induced stimulus and an internally generated behavioral response—that is, stimulus–response associations (and consequently response–reward associations) (Nobre, Coull, et al., 1999).

Psychologically, the neuronal mechanisms of modulation by reversal in OMPFC might account for the suppressed perception of either reward stimuli or emotional stimuli. If reward is modulated by either delay or punishing stimuli, signal changes in OMPFC are apparently reversed. Since delayed or punishing stimuli change the level of attention, these results suggest that signal decreases in OMPFC might be associated with attentional modulation. This indicates that signal changes during a specific emotional or cognitive task might also depend on the respective psychological context, as reflected, at least partially, in preceding or simultaneous attention.

Displacement

Attentional modulation might be altered in the defense mechanism of displacement. Displacement is characterized by detachment of emotional reactions from their original contents, which are replaced by other contents. The defense mechanism of displacement can be observed paradigmatically in phobia where the anxiety is, for example, associated with little animals, social situations, or rooms rather than with its original contents—that is, conflicts.

One of the main characteristics of displacement is that the association of emotional reactions with the substituted contents is fixed and rigid. Modulation by attention remains impossible because the attentional focus can no longer be shifted. We hypothesize that this lack of attentional modulation in displacement might be related to abnormal modulation by reversal. Note that modulation by reversal is supposed to account only for one particular characteristic of displacement, the fixed and rigid emotion–stimulus association. In contrast, we do not discuss the mechanisms underlying the generation of displacement; these might possibly be different from modulation by reversal (cf. Rudden, Busch, et al., 2003). It can be assumed that lack of attentional modulation accounts for the rigidity of the phobia, but not necessarily its origin, in which case the phenomenon has to do with resistance rather than displacement. Accordingly, displacement has two components: first, the switch from the original content to a substitute content, and then a resistance to incorporating new emotional responses.

Functionally, new linkage of internally generated emotional reaction to other externally induced stimuli or contents seems to be prevented in displacement. This makes new stimulus–response or stimulus–reward associations impossible. The subject characterized by
strong displacement remains incapable of associating his or her emotional reactions to other contents—the subject has no choice other than reacting with phobia when exposed to the respective stimuli. We therefore predict that displacement in phobic patients is accompanied by deficits in stimulus–response linkage and, more specifically, stimulus–reward association. This is supported by one study showing reduced reward dependence in patients with social phobia (Kim & Hoover, 1996). We suggest that this deficit in stimulus–reward association might at least partially correspond to the fixed and rigid association of emotional reactions to particular stimuli.

Psychologically, the phobic subject is characterized by strong and abnormal attention to that particular content/stimulus in relation to a particular emotional reaction. We suggest deficits in preceding attention, simultaneous attention, and selective and shifting attention. There is heightened preceding attention—that is, expectancy that reaches abnormally high levels. Thoughts merely indicating the particular stimulus without its actual appearance can already be sufficient to elicit emotional reaction. Simultaneous attention might also be abnormally increased in these patients who remain unable to shift their emotional attention. We therefore predict that displacement in phobic patients is characterized by attentional alterations concerning preceding attention—that is, expectancy, simultaneous attention, and selective and shifting attention. This is supported by recent studies showing increased selective attention to the respective stimuli in phobic patients (Bogels & Mansell, 2004; Pfliugshaupt, Mosimann, et al., 2005; Rinck & Becker, 2005).

Physiologically, as described above, the OMPFC seems to play a crucial role in attentional modulation and stimulus–response/reward association. Therefore, we suggest that cognitive induction of signal decreases in OMPFC might be altered in phobic patients. This leads us to predict that, during attentional modulation, phobic patients remain unable to reverse neural activity in OMPFC from signal increases to signal decreases. Preceding expectancy might no longer induce signal reversal in OMPFC from signal increases to decreases. Instead, there might be raised signal increases in OMPFC which could remain immune to attentional modulation. This is indeed supported by recent imaging studies with phobic patients showing stronger signal increases in medial cortical regions such as the OMPFC and the anterior cingulate cortex as well as in closely connected regions such as the amygdala and the insula during exposure to the respective stimuli (Dilger, Straube, et al., 2003; Lorberbaum, Kose, et al., 2004; Pissiota, Frans et al., 2003; Straube, Kolassa, et al., 2004; Straube, Mentzel, et al., 2004; Veltman, Tuinbreijer, et al., 2004). However, abnormal attentional modulation of neural activity in these regions by the different forms of attention remains to be shown.

**Modulation by functional unity and sensorimotor regression**

We introduce the principle of functional unity that integrates neural activity from different regions with respect to a particular function. Our hypothesis is that functional unity is related to the defense mechanism of sensorimotor regression.

**Modulation by functional unity**

Another example of a possible functional mechanism of emotional–cognitive interaction is the constitution of functional unities. Functional unity can be described as the coordination of neural activity over a limited time period by means of which different regions are linked together with respect to a particular function. Such transient functional unities might be identified based on the psychophysiological characteristics or the functional connectivity of the respective regions (Friston, 1998, 2003; Friston, Fletcher, et al., 1998; Friston, Harrison, et al., 2003; Friston, Josephs, et al., 1998; Friston & Penny, 2003; Friston & Price, 2003). The medial regions in the brain’s cortex, the so-called cortical midline structures (CMS), can be considered a functional unity (Northoff & Bermpohl, 2004), which is supported by different lines of evidence. First, one can often observe co-involvement and coactivation of different midline regions. For example, the mechanisms described above of modulation by reversal can be observed not only in the OMPFC but also in the posterior cingulate (Northoff, Richter, et al., 2005). Other studies on emotions and cognitions show similar co-involvement of anterior and posterior midline regions (Iacoboni et al., 2004, Northoff & Bermpohl, 2004; Northoff, Heinzel, et al., 2004). Second, unlike other more lateral cortical regions and subcortical regions, the CMS show a continuous high level of neural activity during resting conditions (e.g., fixation on a cross) (Gusnard, Akbudak, et al., 2001; Gusnard & Raichle, 2001; Mazoyer, Zago, et al., 2001; Raichle, 2001; Raichle, MacLeod, et al., 2001). Third, regions in the CMS are characterized by close anatomical connections and tight functional connectivity. For example, Greicius, Krasnow, et al. (2003) investigated the functional connectivity among CMS regions in
both the resting and the activation state. They observed increased functional connectivity between anterior and posterior CMS regions in the resting state, whereas it decreased during active cognitive tasks.

Taken together, these findings are compatible with the functional mechanisms of modulation by functional unity (see Figure 4). The data described provide compelling evidence for the existence of CMS as a functional unity that seems to be particularly active and cohesive in the resting state (Greicius, Krasnow, et al., 2003; Wicker, Keysers, et al., 2003; Wicker, Perrett, et al., 2003; Wicker, Ruby, et al., 2003).

Sensorimotor regression

Modulation by functional unity might be altered in the defense mechanism of sensorimotor regression. With sensorimotor regression, conflicts are no longer solved by cognitive or emotional means but, rather, by bodily means and thus by sensorimotor function. The defense mechanisms of sensorimotor regression can be observed paradigmatically in patients with catatonia. We hypothesize that sensorimotor regression in catatonia could be related to abnormal modulation by functional unity across CMS.

Catatonia is a psychomotor syndrome showing a unique constellation of affective, behavioral, and motor symptoms (Bush, Fink, et al., 1996a, 1996b; Fink, 1993; Northoff, 2002a, 2002b). Most strikingly, acute catatonic patients are totally immobilized, posturing in bizarre positions, and becoming totally mute.

Psychodynamically, bizarre motor behavior has been interpreted as sensorimotor regression reflecting an immature and emotionally guided defense mechanism against the uncontrollable overflow of anxieties—that is, “immobilization by anxieties” (Northoff, 2002a, 2002b; Perkins, 1982). At the same time, more mature and cognitively guided defense mechanisms such as internalization and externalization seem to break down in patients’ handling of their emotional conflicts (Böker & Lempa, 1996; Johnson, 1984). According to psychodynamic theory, unlike schizophrenic patients, catatonic patients are no longer able to develop paranoid ideas with respect to their environment and thus to externalize their emotional conflicts. Unlike affective patients, catatonic patients remain unable to attribute the emotional conflicts to themselves in, for example, ideas of guilt and sin and thus to internalize them (Johnson, 1984). These cognitive mechanisms of internalization or externalization are apparently no longer available for catatonic patients, whose sensorimotor regression aims at compensating their overwhelming anxieties (Arieti, 1972; Böker & Lempa, 1996; Johnson, 1984). In self-evaluation, catatonic patients characterize their self by “low emotional arousal,” “low self-esteem,” and “lack of social contact” (according to the Landfield categories that were used to systematically categorize individual constructs of the patients generated by means of the Repertory Grid Technique: cf. Böker, 2004; Böker, Northoff, et al., 2000a, Böker, von Schmeling, et al., 2000b; Northoff, Böker, et al., 2003).

Imaging studies during emotional stimulation showed altered patterns of signal changes in medial and lateral orbitofrontal cortex (MOFC, LOFC, respectively) in catatonic patients compared to noncatatonic psychiatric and healthy controls. Specifically, we observed reduced signal changes in the MOFC and enhanced signal changes in the LOFC during negative stimulus presentation (for details see Northoff & Bermpohl, 2004; Northoff, Heinzel, et al., 2004). Correlation analysis of functional connectivity between OFC, OMPFC, and premotor and motor cortex in catatonic patients as compared to noncatatonic psychiatric and healthy controls revealed the following differences: Catatonic patients showed significantly lower scores for functional connectivity from the orbitofrontal cortex to the medial prefrontal and the premotor/motor cortex when compared to noncatatonic psychiatric controls and healthy subjects. Correlation analysis revealed the following specific findings in catatonic patients: (1) Emotional arousal and self-esteem significantly correlated with decreased signal changes in the medial orbitofrontal cortex. (2) The dimension of social contact correlated significantly with motor symptoms as well as with decreased signal changes in the orbitofrontal and medial prefrontal cortex and their lower connectivity to the premotor cortex.

Findings of abnormal connectivity from orbitofrontal cortex over medial prefrontal cortex to premotor/
motor cortex suggest abnormal modulation by functional unity in these patients. The functional unity across anterior cortical midline structures seems to be less coherent than in healthy subjects. This less coherent functional unity might, in turn, facilitate abnormal transformation of emotional symptoms into motor symptoms.

The anterior CMS are also densely connected to subcortical motor regions such as the basal ganglia (in cortico-striato-pallido-thalamic-cortical loops) that are involved in generating behavior. If these subcortical regions are abnormally modulated by CMS, abnormal behavior might be generated (Northoff, Witzel et al., 2002; Rolls, Tovee et al., 1999). This seems to be the case in catatonia. The low self-esteem and the abnormal social contact in catatonic patients, as both are apparently associated with disturbed function in CMS, are accompanied by abnormal behavior possibly resulting from abnormal cortico-cortical and cortico-subcortical modulation. This, in turn, might provide the basis for recruiting sensorimotor-regression mechanisms as a substitute for the failing cognitively oriented defense (see above). The recruitment of sensorimotor regression mechanisms might correspond to abnormal modulation of dense connectivity from orbitofrontal over medial prefrontal to premotor cortex. Abnormal connectivity might then block neuronal activation in premotor and motor cortex and, at the same time, alter neural activity in connected basal ganglia, ultimately resulting in catatonic motor symptoms with sensorimotor regression.

The relationship between an abnormal functional unity of anterior CMS and sensorimotor-regression mechanisms is further supported by studies in patients with conversion-disorder symptoms. Conversion disorder with hysterical paralysis might be regarded as a form of sensorimotor regression in relation to emotional–cognitive conflict that can no longer be solved by cognitive defense exclusively. Imaging studies in acute paralytic patients revealed deficits in various regions of the anterior CMS including the orbitofrontal and the premotor/motor cortex (Halligan, Athwal, et al., 2000; Marshall, Halligan, et al., 1997; Spence, Crimlisk, et al., 2000; Vuilleumier, Chicherio, et al., 2001). Why, however, is there a symptomatic difference in sensorimotor regression between hysterical and catatonic patients, the former showing conversion disorder and the latter sensorimotor regression? It should first be noted that hysterical patients can show a catatonic-like picture and, conversely, catatonic patients can appear strongly hysterical (Modestin & Bachmann, 1992a, 1992b; Northoff, 2002a, 2002b; Northoff, Eckert, et al., 1997). Such symptomatic overlap suggests that both catatonia and hysteria have in common the neuronal mechanisms underlying sensorimotor regression. They might share the abnormal functional unity of anterior CMS resulting in abnormal motor behavior. However, modulation by functional unity might concern not only overlapping regions but also different regions; this in turn might explain the symptomatic differences between catatonia and hysterical conversion. Future studies specifically targeting single catatonic or hysterical symptoms might reveal those neuronal mechanisms specifically associated with hysterical conversion disorder as distinguished from those related to catatonic sensorimotor regression.

Conclusion

We have investigated the relationship between neuronal integration and defense mechanisms. Defenses are typically thought of as being psychically motivated and as being instantiated against something. We focused on compensation mechanisms of defense that are usually connected with specific forms of failure of psychic functioning (especially severe depression, catatonia). Under these conditions, psychotherapists may be confronted with depressive symptoms that initially have no direct link to the therapeutic process. Therefore, the examples in our paper are not merely illustrations of organic deficits or the manifestation of actual impairment. Our intention is to describe the neural physiological correlates of regressive processes on the basis of a neuropsychoanalytic model for the disturbances of emotional self-referential processes. From a descriptive, phenomenological perspective, these dynamic processes may result in a specific symptomatology (cognitive disturbances, autonomic vegetative symptoms, and, last but not least, disturbances of movement initiation). Psychically motivated defenses (against anxiety, shame, feelings of indebtedness) can often be observed before the automatization of the somato-psychic-psychosomatic vicious circle takes place or after the remission of the actual depressive symptomatology and the overcoming of the somatically rooted automatization patterns—for instance, during the course of psychotherapeutic treatment of depressed patients.

Particular principles of neuronal integration are hypothesized to be related to specific defense mechanisms. For example, displacement may be related to an inability to reverse signal changes by attention—that is, to abnormal modulation by reversal. Somatization may be accounted for at least partially by abnormal functional balance between top-down and bottom-
up modulation across medial cortical and subcortical regions. Introspection is hypothesized to be related to abnormal reciprocal modulation and attenuation of neural activity in medial and lateral prefrontal cortex during emotional-cognitive interaction. Finally, we suggest that sensorimotor regression corresponds to abnormal modulation of functional unity across medial prefrontal cortical regions. We conclude that, though our hypotheses must be considered preliminary, they at least provide a starting point for future empirical testing of psychophysiological mechanisms underlying defense mechanisms.

REFERENCES

Fink, M. (1993). Catatonia and psychotic (delusional) depres-


Liotti, M., Mayberg, H. S. et al. (2002). Unmasking disease-


Panksepp, J. (1998b). The periconscious substrates of con-


