Cathexis Revisited

Corticolimbic Resonance and the Adaptive Control of Memory

DON M. TUCKER and PHAN LUU

Psychology Department, University of Oregon, Eugene, Oregon 97403, USA
Electrical Geodesics, Inc., Riverfront Research Park, Eugene, Oregon 97403, USA

Several issues that Freud (1895) addressed in his Project, a “psychology for neurologists,” remain fundamental problems for a neural model of mental processes. By theorizing how neurons could represent experience and organize behavior adaptively, Freud confronted the problem of establishing memories in the same neural mechanism that must be immutable in response to perceptual events. That is, according to Freud there is an inherent inconsistency between the changes that must occur in the functions of neurons in order for memory traces to be formed and the stability that must also be present in order for perceptual processing to be unprejudiced. By proposing a neural architecture for motivating cognition and behavior adaptively, he attempted to understand how needs and values come to shape the structure of memory and behavior. Perhaps most importantly, by framing his neurophysiological model to address broad psychological questions, Freud proposed a specific theoretical approach for bringing the evolutionary insights of nineteenth-century biology to bear on practical questions of human behavior.

In this paper we revisit Freud’s neurophysiological speculations in light of current thinking in cognitive neuroscience. For important aspects of Freud’s approach, such as the emphasis on adaptive and homeostatic mechanisms in regulating cognition, modern neuropsychological theory could still be informed by Freud’s work. With the concept of cathexis, for example, Freud integrated the mechanisms of memory with those of motivation in a way that is quite relevant to understanding the current evidence on corticolimbic interaction in memory and emotion. The idea that emotional excitement may be mediated directly by mechanisms of neuronal excitement remains a direct and compelling hypothesis in light of the modern evidence on emotion, kindling, and neural plasticity in corticolimbic networks.

For other aspects of Freud’s theorizing, such as his separation of perceptual networks from memory networks, we think the modern evidence shows that Freud’s proposals were wrong. Memory and perception are very difficult to separate, and in fact they appear to be achieved within common networks in the cortical pathways linking sensory areas with limbic structures. Even here, however, it was an important insight to recognize the inherent incompatibility between perceptual stability and mnemonic plasticity. This incompatibility—now known as “catastrophic interference” or “the stability/plasticity dilemma”—has proven to be a key problem in computational approaches to neural models in the last half of the century.

In this paper we entertain the hypothesis that the stability–plasticity dilemma may not be restricted to neural network simulations. It may reflect a fundamental fact of

This work was supported by NIMH Research Grants MH42128 and MH42669 to the University of Oregon and Small Business Innovation Research Grants MH50409 and MH51069 to Electrical Geodesics, Inc.
human experience. We argue that Freud's attempt to compartmentalize the mind, and thereby to avoid the dilemma, can be shown to be unworkable in light of the modern evidence on the neuroanatomy of memory systems. Because perception, memory, and action are organized within common neural networks, the stability-plasticity dilemma is an implicit reality of each significant life experience.

THE NEURAL MODEL FOR PSYCHOANALYSIS

Freud eventually attempted to destroy his “psychology for neurologists” manuscript before it could be published (as the Project for a Scientific Psychology). He apparently felt the neurologizing was inadequate for what had become the psychological theory of psychoanalysis (Pribram & Gill, 1976). However, any student of psychoanalysis can see the foundations for some of the most basic psychoanalytic concepts in the constructs of neuronal activity formulated in the Project. For example, the notions of cathexis and catharsis were proposed to describe how neurons become “charged” with energy and “discharged” of it in the course of reacting to the flow of activity from the perceptual organs and homeostatic core. Beginning with an electrical metaphor for neuronal activity (charge), Freud generalized the construct to describe how motivational excitement or energy could become attached with specific mental contents, such as those observed in hysteria and obsessions. Although Freud's theory of the motivational control of mental processes soon became formulated in exclusively psychological terms, it began as a neurophysiological model, and it seems clear the psychological model of psychoanalysis retained important qualities (excitement, charge, and discharge) of the original neuroelectric metaphor. As we will see below, Freud's metaphoric description of cathexis and catharsis may still prove useful for explaining the neurophysiology of memory and its motivational control.

Another example of the theoretical primordia in the Project is Freud's exploration of the subdivisions of the mind that may explain psychological experience. The Project shows Freud experimenting with ideas about domains of the mind that operate according to different rules. These constructs were clearly influential on his later formulations of mental topology, although major transformations were required before the phi, psi, and omega domains took on the character of the id, ego, and superego. The tendency toward rapid discharge of activity (catharsis) is shared by the phi domain and the id, whereas both the psi domain and the ego are capable of containing affective energy for more extended development (cathexis). However, in the Project it is the psi domain that receives direct motivational controls from the endocrine system, whereas this becomes a property of the id rather than the ego in psychoanalytic theory. Indeed, it is within the project that Freud first introduced the ego. He conceptualized the ego as follows:

For these two processes indicate that an organization has been formed in psi whose presence interferes with passages (quantity) which on the first occasion occurred in a particular way (i.e., accompanied by satisfaction or pain). This organization is called the “ego.” (Freud, 1895/1953, pp. 322–333)

This conceptualization of the ego as memory traces has profound implications for the conceptualization of the self in modern neuroscience that will be discussed in the following sections. Parallels also may be drawn between the supervisory functions of the omega domain and the superego, although the notion of parental introjects resulted in a quite different character for the superego from the mental domains found in the Project.

From these parallels we can see that, although there were to be major theoretical de-
velopments in psychoanalysis from the original formulations, Freud made several key assumptions about the mind in the *Project* that were to shape his later work. Remarkably, several of these assumptions were derived from his effort to model the mechanisms of memory and motivation in neurophysiological terms. Thus, the distinction between primary and secondary processes, perhaps the central element of the psychoanalytic conception of thought, reflects assumptions about cathexis and catharsis that were formulated in the *Project's* neuronal model. Primary process cognition operates toward immediate catharsis of motivational energy. Secondary process cognition cathects motivational charge, brought on by the requirement to deal with "the exigencies of life." It thereby achieves both the capacity for memory and the capacity for more "volitional" delayed gratification, and thus goal-directed behavior. Although these distinctions were refined on the basis of Freud's extensive reflections on psychological experience, they were initially drawn to cope with the assumptions of the neuronal model of the *Project*. Freud believed that neurons whose activation is responsive to sensory data (the phi domain) must not have the capacity for memory (and thus would operate on primary process rules). Neurons whose activation encodes experience in memory (the psi domain), through facilitation of conductance pathways, achieve the capacity for stability of representation and thus the coherent identity of the ego.

On the centennial of the *Project*, it may be appropriate to question how the fundamental assumptions about primary and secondary processes can be evaluated in terms of today's cognitive neuroscience. As shown below, we find that Freud's close integration of motivation and memory, and even his electrical metaphor of cathexis and catharsis, may be useful theoretical tools for understanding the adaptive control of corticolimbic interaction. On the other hand, we take issue with the assumption that the mind may be separated into domains that maintain, or do not maintain, memories. From observing the mental processes of computers, cognitive neuroscientists have gained insights that Freud would have appreciated. In applying these insights to the human brain and mind, certain implications prove to be both surprising and relevant to basic questions of psychoanalytic theory.

**CATHEXIS: THE CHARGED IDEA**

Within Freud's neurophysiological theory, the neuron's contact barrier, known today as the synaptic cleft, forms the foundation of memory. Through cathexis, a neuron gains charge, and, through catharsis, releases it, but because the contact barriers act as resistors to conductance, Freud thought each catharsis altered the conductance capacity of the barriers such that memory is formed. Integral to this neuron model, and to Freud's later psychology, was the assumption that the charge reflects emotional (endogenous) as well as exogenous neural energy. When an object of experience becomes cathected, it is represented in the mind as emotionally charged, as significant to the self. When Freud described catharsis in the psychotherapy of his patients, he described a discharge of emotional energy in relation to personally significant memories. In this metaphor of cathexis and catharsis, motivation and memory are fused. The representation of an experience in memory occurs because the experience is emotionally charged.

Modern cognitive models of memory rely on concepts of "activation" of information elements, and even "spreading activation" across the elements as a function of their association (McKoon & Ratcliff, 1980). But there is typically little concern with motivating the activation of cognitive elements in these models, as if the cognitive system operates according to Brownian motion (spontaneously and in a vacuum). Emphasis on the motivational control of memory is remarkably absent in the modern cognitive
neuroscience approach. Through clinical observations on patients with amnesia (Squire, 1986) and through experimental lesion studies of visual memory in monkeys (Ungerleider & Mishkin, 1982), the essential role of limbic structures in memory has been determined. But with few exceptions (Kornhuber, 1973; Nadel, 1991, 1992), the extensive literature on limbic contributions to memory ignores, or at least does not make explicit, the obvious implication of the fact that the “primitive” limbic structures are required for memory: Memory has evolved as an adaptive operation. In practical terms, Kornhuber recognized that the brain is not an unlimited storage device. This limited capacity is now formally recognized in biologically inspired computational models of memory (McClelland et al., 1995; Treves & Rolls, 1994). Kornhuber argued that the limbic architecture allows for events to be adaptively evaluated before memory storage.

From modern neuroanatomy it is apparent that the entire neocortex of humans continues to be regulated by the paralimbic (archicortex and paleocortex) regions from which it evolved (Sanides, 1970; Pandya et al., 1988). Information is represented in memory as a direct function of its motivational significance to the organism. For a more ecological cognitive theory, it may be useful to revisit Freud’s formulation of cathexis and catharsis as fusing motivational control with the mnemonic representation. Certainly there are intrinsic qualities of the information, such as its coherence or imaginability, that influence memory. But in a naturalistic context the reason that the limbic structures control memory consolidation may be that it is the adaptive significance of the information—the extent to which it resonates with core concepts of needs and values—that determines whether that information will be given existence in a limited memory store of the hippocampus, which supports extended consolidation required for integration into long-term memory.

AN ELECTRICAL METAPHOR FOR THE NEUROPHYSIOLOGICAL MECHANISMS

Although uninformed critics have described psychoanalysis as a “hydraulic” model, in which emotional pressure may not be contained, students of the Project have recognized that if there is an implicit metaphor of mechanism in Freud’s theorizing, it is an electrical one (Pribram & Gill, 1976). In drawing from the models of neuronal function of his day, Freud considered their properties as electrical, and theorized about their operations as involving the storage and management of quantity of energy.

In a laboratory science, it is often assumed that an explanation is unsatisfactory if it is metaphoric rather than literal. At the metaphoric level, Freud’s formulation of cathexis as a kind of electrical process, regulating memory as a function of emotional charge, is quite relevant to today’s theoretical problems of understanding the limbic adaptive control of memory. More than that, the model of neuroelectrical charge and discharge may be literally descriptive of important aspects of corticolimbic neurophysiology.

Lesions of limbic structures and paralimbic cortex impair the formation of new memories, but they do not destroy long-standing memories (Squire, 1992). The implication is that memories are eventually represented in neocortical networks, but the consolidation process to achieve this requires some involvement of limbic structures. Because limbic lesions produce a degree of retrograde amnesia, the corticolimbic interaction appears to operate over days, weeks, and maybe even years to consolidate experience within long-term memory.

As described below, the architecture of the mammalian cortex reflects the evolution of neocortex from the archicortex (dorsally) with its hippocampal base and the paleocortex (ventrally) with its base in olfactory cortex and close links to the amygdala
In each sensory pathway, there are multiple levels of representation between primary sensory (neocortex) and paralimbic networks, with reentrant connections linking each level. Whereas the anatomical evidence shows that these reentrant connections must be functional to allow memory consolidation to operate across the network hierarchy, the functional question is what happens in these networks to alter the neural activity, and thus activity-dependent synaptic modifications, that achieves the representations. Studies of the electrical activity of the cortex have provided intriguing suggestions.

Within the hippocampus, studies of long-term potentiation (LTP) have shown that increasing the activity level of a neuron (through antidromic stimulation) causes an enduring strengthening or potentiation of afferent synapses that are active at that time (Squire, 1987). Because the hippocampus is particularly sensitive to LTP, this mechanism has been considered to be a model for associative memory generally. Both the amygdala and the hippocampus are excitatory focal points for another model of corticolimbic activity, kindling. Kindling occurs when repeated electrical stimulation of a site of cortex results in increasing discharges with each successive stimulation. It is different from LTP in several respects and the most salient ones are that it is produced by interstimulus intervals that range from 20 minutes to 24 hours and once established may be permanent. Regardless of where the cortex is stimulated, kindling tends to focalize in the limbic areas (Doane & Livingston, 1986) and, moreover, it is generally accepted that the sites around the amygdala are the most sensitive to kindling. Once a site has been kindled it tends to recruit other sites in limbic cortices such that over time the secondary sites will show afterdischarges as great in amplitude as those seen in the primary site (Racine & McIntyre, 1986). This spread of afterdischarge to secondary sites indicates a long-term potentiation in activated pathways. The electrical reactivity of limbic cortex has been considered as an explanation for why seizures tend to focalize in the medial temporal lobe (Doane & Livingston, 1986; Bear, 1979).

Do these phenomena of electrical reactivity provide clues to the normal physiology of corticolimbic interaction in motive-memory consolidation? Clearly, both kindling and seizures are pathological processes. However, kindling has been shown to be classically conditioned (Janowsky et al., 1980), suggesting that it may be controlled by normal learning mechanisms. As a general metaphor for emotional reactivity, a kindling model has been used to describe the increased magnitude of depressive affect with each successive episode of depression (Post, 1986). Harkness (1996) has hypothesized that a more literal kindling interpretation may be appropriate for understanding the neuropsychological response to traumatic stress. A kindling of limbic reactivity may explain how trauma victims become sensitive to even minor psychological stresses.

The phenomena of dissociation and other forms of memory disorganization in trauma victims suggests that strong emotional activation may produce pathological as well as adaptive influences on memory. Remarkably, a high proportion of women psychiatric patients with a history of sexual abuse have been found to suffer from seizure disorders (Shearer et al., 1990). Epileptiform EEGs and other signs of seizure disorder have also been observed in children who have suffered psychological abuse, even when the incidence of physical abuse (and head injury) are controlled statistically (Teicher et al., 1993). The temporary improvement of depression by electroconvulsive therapy could suggest that the depressive condition involves some state of corticolimbic tension that is discharged by the seizure process. In the classical phenomenon of “forced normalization,” epileptics with psychiatric symptoms may show a worsening of these symptoms when given anticonvulsant drugs, and a remission of psychiatric symptoms when seizures are allowed to return (Flor-Henry, 1983).

The implication from this line of evidence may be that both emotional states and memory mechanisms are mediated by the neurophysiology of corticolimbic interaction.
It would seem unlikely that these phenomena share the same neural networks accidentally. The neurophysiological reactivity of limbic tissue—through engaging widespread neocortical networks—appears to mediate both emotional reactivity and the normal processes of memory organization. With extreme emotional stress, there may not only be emotional sensitization and memory disorganization, but frank pathology of corticolimbic function.

In cases of temporal lobe epilepsy of unknown or varied etiology, the changes in personality function may suggest exaggerated limbic motivational control over cognition. Bear and Fedio (1977) contrasted the psychological problems of patients with right and left temporal lobe epilepsy. The patients with a right temporal focus were emotionally labile, as if the right hemisphere’s affective processes were poorly regulated. Perhaps more interesting were the patients with the left temporal lobe focus, who showed a preoccupation with intellectual concerns, such as religious or philosophical issues. Although these patients showed exaggerated verbal, intellectual ideation suggestive of exaggerated left hemisphere cognition, this ideation was strongly (and pathologically) motivated, as if the disordered limbic physiology of the left hemisphere provided an exaggerated motivational control for left hemisphere cognition. Bear and Fedio (1977) suggested that the epileptic process gradually results in a “functional hyperconnection” of limbic with cortical networks that produces a distortion in the motivational processes of the individual’s personality (Fedio & Martin, 1983).

In the daily organization of experience, the limbic networks seem to resonate to the motivational (i.e., personal) significance of each event. In doing so, they engage the consolidation of that event in proportion to its significance. Because the adaptive control is integral to the representational process, the phenomenon of “memory” could be redefined as “motive-memory.” The significance of each event is integral to the representation itself. In Freud’s terms, an event becomes organized in memory to the extent that it is affectively “cathected.”

PERMEABLE NEURONS AND THE PROTECTION OF PERCEPTION

Freud proposes that perceptual mechanisms may be separated, and thereby protected, from the impermeable psi (memory) neurons. Because the psi neurons were permeable, the passage of energy through them produced no changes. This permeable property left the perceptual mechanisms unprejudiced by experience. As mentioned previously, the neurophysiological domains of the Project were the first partitions of the mind in psychoanalysis, and they foreshadow important features of the later mental topography. In the later formulation of primary and secondary processes, the motivational influences are reinterpreted to operate on the permeable domain (primary processes), whereas they had been seen as operating on the impermeable domain (psi neurons) in the Project. But Freud maintained the distinction between a domain dominated by discharge or catharsis (primary process cognition) and one able to maintain a motive or cathexis in support of goal-directed behavior (secondary process cognition). By assuming that these domains of the mind operate according to fundamentally different principles, Freud prepared psychoanalysis, and the twentieth-century popular concept of the mind that it so strongly influenced, to consider the possibility that the mind is not unitary, but may comprise potentially discordant mental domains.

There are major functional divisions of the human brain in modern theoretical neuroscience that are relevant to the domains of the mind that Freud considered in psychoanalysis. The large frontal cortex of the human brain has long been recognized to be interdependent with limbic cortex (Pribram & MacLean, 1953; Nauta, 1971), suggesting a close relation between motivational processes and the higher executive func-
tions. Indeed, it is the prefrontal lobe’s extensive connections with limbic structures that lead Nauta (1964) and Pribram (1960) to view it as an extension of the limbic system. Nauta suggested that the prefrontal cortex may be the isocortical representative of limbic functions in much the same sense as the motor isocortex is the superstructure of those lower areas involved in motor functions.

The increasing recognition of the emotional sophistication of the nonverbal right hemisphere has led to speculations that interhemispheric dynamics may mediate what Freud understood as intrapsychic dynamics (Galin, 1974; Tucker, 1981). More recently, a reconsideration of functional domains of the mind has been stimulated by the recognition that the neocortex evolved from two sources: archicortical (dorsal spatial pathway) and paleocortical (ventral object pathway) (Liotti & Tucker, 1995; Tucker et al., 1995). These cortical networks appear to be separable functional divisions by reason of their patterns of intrinsic connectivity. This anatomical insight raises the question of whether there is a character of motivational control of memory in the ventral cortical networks that differs fundamentally from that in the dorsal cortical networks. Nadel (1992) has argued for two systems of memory: one that encodes the location information and the other that encodes object information. We have speculated on how these two memory systems may reflect the motivational bias of each trend (Luu & Tucker, 1998).

Although each of these ways of parcellating functional divisions of the human brain may shed light on aspects of motivation and personality that are relevant to psychoanalytic theory, none of these divisions achieves the separation of perception (permeable neurons) from memory (impermeable neurons) that Freud attempted in the Project. As we will see below, any viable functional parcellation of the cortex must include both limbic and neocortical networks, and at that point the interdigitation of perceptual operations and memory operations becomes unavoidable. However, by recognizing the difficulty of encoding memory and mirroring perception in the same neurons, Freud achieved a fundamental insight into neuronal models that was to be rediscovered several times during the later twentieth century.

**DISTRIBUTED REPRESENTATIONS AND THE STABILITY–PLASTICITY DILEMMA**

In theorizing about the mechanism of memory, Freud started with the reasonable assumption that memory is based on neuronal activation. In fact, explaining mental processes in terms of a physical quantity (i.e., the activation of neurons and its passage through the neurons) seemed for Freud to be the fundamental requirement for a scientific psychology. In the modern paradigm, although studies of single-unit activity remain important, it has proven more useful to model memories as distributed, represented in the pattern of the strengths of connections among a large network of neurons.

The insight that memories could be represented in distributed networks, which in retrospect was at least implicit in Freud’s formulation, was achieved mid-century (McCulloch & Pitts, 1943), and it proved an integral insight to the development of electronic cybernetics (Heims, 1991). With their numeric simulations of “nerve nets,” McCulloch and Pitts showed that networks of simple neuron elements could be combined to achieve more complex patterns of information storage than could be achieved by 1:1 mapping of information bits onto neuronal activations. Hebb’s general formulation (1949) of learning rules for assemblies of neurons provided a model for how representations could be formed by the patterns of connectivity.

Within a few years, the behavior of cell assemblies and nerve nets was sufficiently understood that a fundamental problem was recognized. If new information is pre-
sented to an existing nerve net model, the ability to store the new information will be constrained by its similarity to the existing memories (Milner, 1957). Furthermore, making adjustments to cope with the new information will inherently degrade the existing memory. All events are represented by the same connection strengths, so incorporating new information necessarily alters existing connections. For connectionist models that must store all events with the same distributed connection architecture, this problem, known as “catastrophic interference” or the “stability-plasticity dilemma,” appears to be a serious limitation.

The fundamental nature of this dilemma was recognized perhaps most clearly by Grossberg, who has devoted substantial theoretical effort toward possible solutions (Grossberg, 1984, 1980). Interestingly, Grossberg's primary solution is quite similar to Freud's: to separate the processing system into a domain of protected memory that is separate from the domain of perceptual input. A control process gates access to the memory domain as a function of the novelty and adaptive significance of the input information (Grossberg, 1984, 1980).

Graceful Degradation Through Interleaving

Avoiding catastrophic interference has been addressed recently in a neuroanatomical model of consolidation proposed by McClelland, McNaughton, and O'Reilly (1995). They point out that hippocampal lesions produce both a deficit in memory for new events (anterograde amnesia) and a temporally graded deficit in remembering events from the past (retrograde amnesia). The implication is that the hippocampus is required to consolidate information into a store that is then hippocampally independent (i.e., the neocortex). The anterograde amnesia seems to occur because the new learning is not possible without the hippocampus, and the graded retrograde amnesia is due to the disruption of the consolidation that was ongoing at the time of the lesion.

Thus McClelland et al. propose that the mammalian nervous system comprises two memory systems. The first is the neocortical processing system, in which the processing of information adaptively alters the synaptic connections. In traditional connectionist terms, the patterns represented within the alteration of these connections is knowledge/memory. The synaptic changes that occur within this network are developed over repeated exposures. In connectionist models, learning within a distributed network may be facilitated by gradual training. However, if the information is to be learned quickly or training is focused (i.e., without the interleaving of other information), the performance of this network abruptly degrades. This is the problem of catastrophic interference. Therefore a second system, of which the hippocampus must be a critical component, achieves a short-term representation of an event. It then slowly “teaches” the neocortical system, via its back-projections, by gradually reinstating the neocortical patterns over time. The result may be a pattern of connection weights in the neocortex that can accommodate the new information with a minimal disruption of the existing representations.

This division of memory into a temporary and a long-term store has profound implications for the self and sensory processing. Freud believed that it was crucial to protect the perceptual mechanisms from being prejudiced by experience, but it seems to be the case that experience inevitably alters the perceptual mechanisms of the neocortex. Moreover, as each new experience is incorporated into neocortical networks, the perceptual mechanism is also altered. In other words, if the ego is memory-based, then the evidence suggests that it is intimately involved with all aspects of perception; the self is inextricably bound up with every new encounter and the prejudice of prior experience cannot be avoided.
CATHECTED RESONANCE AT THE PARALIMBIC CORE

By formulating a connectionist model in terms of corticolimbic networks and the process of interleaving, McClelland et al. have brought their theory of memory into contact with both neurophysiological and anatomical evidence. Through a mechanism such as interleaving, the connection patterns of the network may be rearranged to accommodate new information with a minimal disruption of the existing pattern (memory). Although this mechanism allows for the management of the stability–plasticity dilemma, it does not avoid it. Each new event requires that the existing representation be altered to accommodate it. Given the trade-off inherent in the dilemma, a mechanism is required to evaluate the significance of the new event, and to regulate the degree of interleaving or consolidation as a function of this significance. We propose that Freud's construct of cathexis may be an accurate description of how the limbic regions motivate the consolidation process.

Motivational and Electrical Excitement

As reviewed above, the amygdala, hippocampus, and paralimbic cortex are highly reactive to electrical stimulation. The phenomenon of kindling shows the tendency of limbic regions to maintain and exaggerate electrical activity originating in the neocortex. Such a tendency would be required if the hippocampus and other limbic areas serve to regulate and maintain extended consolidation. Through connections with the hypothalamus, striatum, and other subcortical structures, the limbic areas are well-positioned to integrate homeostatic processes and motivational biases to direct process of memory organization.

In psychological terms, the personality changes interpreted by Bear and Fedio (1977) to reflect corticolimbic hyperconnection may show how an exaggerated sense of personal significance may constrain and degrade both left and right hemisphere contributions to cognition. In the normal brain, the paralimbic contribution to the experience of personal, motivational significance may serve as a guide to the cognitive ruminations (conscious or otherwise) that serves as the gatekeeper for memory. In psychopathology, ideas of reference occur as an exaggerated sense of personal significance constrains perception and cognition.

The Structure of Memory

If we understand that human memory is instantiated in the connectivity of corticolimbic networks, then we may reason directly from the architecture of these networks to interpret the functional structure of memory. Although the gross anatomy of the mammalian cortex is familiar to any biology student, the connectional structure has only been clarified in recent years by advances in anatomical labeling methods. As with many questions in biology, a strictly descriptive analysis of cortical anatomy leads to a confusing and apparently arbitrary assortment of representational areas and their patterns of interconnectivity, whereas an evolutionary analysis provides fairly simple organizational principles for understanding the origins, and thus implicit evolutionary plan, of cortical architecture.

Salamanders evidence a neotenic or “dedifferentiated” morphology that may resemble that of early vertebrate forms (Gould, 1977). Based on his study of the tiger salamander, Herrick (1948) proposed that cortical morphogenesis has been shaped by the segregation and increasing localization of sensory input to the undifferentiated
primordium of the cerebral cortex, the pallium. Herrick noted that in the salamander this functionally directed morphogenesis seems to have resulted in the pattern of connections between the hypothalamus and the medial wall of the pallium, on the one hand, and between the sensory thalamus and the lateral wall of the pallium, on the other, suggesting that the primordial memory operations of the cortex integrated both homeostatic controls and the sensory model of the environment.

Also considering the differential innervation of cortical areas by subcortical structures, Nauta (1964) noticed that the prefrontal cortex organizes its efferent connections through dual routes. One route leads from the mediodorsal cortices through the cingulum bundle and terminates mainly in the subiculum and entorhinal area, which in turn send massive connections to the hippocampus. The second route involves the orbitofrontal connections with the temporal pole, insula, and amygdala through the uncinate faciculus. This suggested to Nauta that there may be a corresponding dualism of prefrontal-limbic connections.

Sanides (1970) proposed a model of cortical evolution that is consistent with Nauta's observations, and with Herrick's formulation of cortical morphogenesis. Sanides argued that the differentiation of cortices in each cerebral hemisphere proceeds in a ring-like manner from two moieties or divisions: the paleocortex (olfactory) and archicortex (hippocampal). Each new cortical area seems to have differentiated out of heterogeneous (undifferentiated) cortex in such a way that the new area is centered, like an island, within the older cortex. With each wave of cortical differentiation, sensory input became increasingly segregated from the rest of the cortex. In addition to the segregation of the newly evolved cortical area, each wave of cortical differentiation resulted in additional lamination and additional cells in layers III and IV. Within each lobe of the cerebral hemisphere, the sensory, motor, and "association" cortices can be traced back to these two primordial moieties.

In the prefrontal lobe, the paleocortex gives rise to the orbitofrontal and ventrolateral cortices, whereas the archicortex gives rise to the medial and dorsolateral cortices (Pandya et al., 1988; Pandya & Yeterian, 1990). The principal sulcus of the prefrontal lobe in monkeys and the inferior frontal sulcus in humans demarcate the boundary on the dorsolateral surface of the two trends. The intraparietal sulcus of the monkey marks the boundary between the ventral and dorsal trends in the parietal lobe; the superior parietal lobe is Brodmann's area 5 and the inferior parietal lobe is Brodmann's area 7. However, area OPT and caudal area PG of the inferior parietal lobe belong to the archicortical trend (Pandya et al., 1988). In humans, instead of the intraparietal sulcus, parts of the angular gyrus appear to represent the transition between the ventral and dorsal trends (Eidelberg & Galaburda, 1984). Unlike the primate brain, areas 5 and 7 of the human brain belong to the superior parietal lobe and areas 39 and 40 belong to the inferior parietal lobe.

Echoing the work of Nauta, Pandya and associates (1988) noted that the pattern of cortical projections also follow the sequence of cortical evolution, such that the most recently evolved areas are densely connected with those areas that are immediately adjacent to them in the evolutionary sequence, and they show sparse connections with areas of distant evolutionary relation. Moreover, projections from newer areas to older areas originate from supragranular layers (above layer IV) and terminate in layer IV, whereas projections from older cortices to more recent cortices originate from the infragranular layers (below layer IV) and terminate in layer I.

This anatomical analysis may suggest that the evolution of neocortex has involved increasing sensory and motor segregation, resulting in greater articulation and precision of the organism's sensorimotor interface with the environment (Tucker, 1992). In contrast, at the core of the brain's connectional architecture, are the paralimbic networks, which appear to be strongly regulated by adaptive influences from bodily and
homeostatic centers (Derryberry & Tucker, 1991). Information is passed from the processing layers of primary sensory areas (supragranular layers) to the receptive layers of older cortex (layer IV). The integrative core of limbic and paralimbic cortices may be in a position to integrate sensory information with internal states and thereby provide the mechanisms by which sensory information is adaptively evaluated for memory storage (Kornhuber, 1973). Moreover, through its connections back to the sensory systems, the integrative core is in a position to adaptively regulate the physiological activity of neocortical networks. This regulation may influence the ontogenetic plasticity and thus differentiation of the cortical networks as well as the ongoing activities of corticolimbic consolidation (Tucker, 1992; Luu & Tucker, 1998).

The implication of this anatomical evidence is that paralimbic cortices serve as the focal points of neocortical networks, a suitable architecture for regulating consolidation adaptively. Moreover, the pattern of connectivity of the corticolimbic architecture has clear implications for the organization of mental representations. The most dense interconnectivity, within each of the archicortical and paleocortical trends as well as between the two trends, is found within paralimbic regions. In contrast, in the neocortical regions comprising sensory and motor cortices, the networks are highly restricted to local connectivity, and thus isolated from global integration with other brain systems. Certainly the intermediate zones traditionally described as “association” cortex, such as the prefrontal cortex or inferior parietal lobule, must have important integrative functions. In fact, the global connectivity of the paralimbic cortex may have important disadvantages, because each representation would be heavily embedded in the global organismic and environmental context.

Yet if form (connectivity) defines function (memory), the most general organization of cognition must occur in paralimbic networks. The neocortex of each sensory region is limited to that modality; the “association” cortex of temporal and parietal cortex is limited to various stages of perceptual integration, and even with its extensive connectivity, the frontal lobe is occupied with the task of organizing action. At the core of the cerebral hemispheres, in the paralimbic cortices, not only are there extensive reciprocal projections with all neocortical regions, but the dense local connectivity creates a highly generalized, and thus abstract, level of mental representation.

A representation is abstract to the extent that it is not stimulus-bound (or, in the case of the motor system, response-bound) and that it is able to integrate influences from many specific instances. Given their extensive connectivity, at the crossroads of all neocortical regions, the paralimbic representations may achieve the most abstract level of cognition within the cortical networks. In fact, clinical observations of aphasia have suggested that when lesions produce deficits in semantic processing of language, they invariably encroach on limbic regions (Bown, 1987).

Recursive Arbitration of Self and Context

Given the close integration of visceral, homeostatic, and postural influences within the paralimbic regions, we can surmise that the abstractions formed there include a holistic integration of the internal state of the organism. These abstractions may be described as syncretic (Tucker, 1981; Werner, 1957) in that multiple elements are fused rather than organized in any systematic fashion. Derryberry and Tucker (1991) have suggested that the extensive integration of autonomic and kinesthetic processes in paralimbic cortices implies that these areas maintain an ongoing concept of the bodily self. This emphasis on a representation of the self may be important to understanding the nature of the concepts formed in paralimbic networks. Reasoning from the neurophysiological evidence, these paralimbic representations appear to be highly “charged”
or reactive to relevant stimulation. Reasoning from the memory evidence, this charge is translated into a consolidation process, in which the excitability of paralimbic regions results in an ongoing recruitment of extensive regions of the neocortex.

Within the corticolimbic architecture, perception seems to occur as a kind of arbitration between the differentiated sensory representations of the articulated neocortical (primary sensory) networks and the syncretic mnemonic representations of the primitive paralimbic networks. The intermediate (“association”) cortex must carry out this arbitration, through representations that are intermediate between the generic mnemonic concept and the articulated sensory map.

In psychological terms, Shepard (1984) has described perception as an arbitration between expectancy and sensation, as “hallucination constrained by the sensory data.” From his psychological analysis, Shepard proposed that the sensory data are perceived to the extent that they “resonate” with the expectancies developed from experience. Although Shepard does not emphasize the motivational control of these expectancies, motivation has been central to previous feed-forward models of perception. The “new look” movement of the 1950s described perception as an operation of the personality (Bruner, 1957; Kragh & Smith, 1970). Similarly, Gibson (1979) defined perception as an operation of matching organismic requirements (needs) to functional opportunities in the environment (affordances).

Corticolimbic interaction must integrate the generic, historical, and motivationally constrained representations of the paralimbic regions with the specific, immediate, sensorily constrained representations of sensory neocortex (Tucker, 1992). The nature of the connectivity between cortical and limbic networks suggests certain functional characteristics. In the visual system, the neocortical networks send primary outputs to layer IV of the adjacent (toward limbic) networks (Pandya & Yeterian, 1985). Because in sensory cortex layer IV is the target of the thalamic (sensory) data, the architecture of the multiple networks (representational areas) between neocortex and limbic cortex appears to reflect a cascade of projections in which a cortical network is interposed between the primordial limbic cortex and the thalamic projection. The “back-projections” (or feed-forward) from limbic toward neocortical networks proceed from deep layers toward the superficial (I and II) layers, apparently influencing the local or recurrent processing of the target areas.

In terms of the anatomical accounts described above, the evolution of neocortex could be seen as a progression of increasingly articulated models of the sensory data, beginning with limbic cortex, each nested within the previous one, and each replacing its progenitor as the target for thalamic input. The resultant architecture retains the paralimbic core as the holistic integration of the bodily evaluation and general organismic context. Within each nested, increasingly articulated network, the sensory interface with the environment became modeled with progressively greater precision, and with greater distance from the motivational constraints inherent to primitive general cortex. The evolution of motor neocortex appears to have followed a similar but inverse progression, with each nested network gaining control over the motor output circuits (striato-thalamic and, later, the direct pyramidal projections) while operating under progressively refined constraints from limbic (deep to superficial) projections.

For both sensory and motor processes, corticolimbic interaction must involve a kind of recursive processing (Tucker, 1992). The same information is repeated in multiple representational levels, supported by point-to-point reentrant connections, and it becomes transformed within each level by the recurrent intra-level connections. Although cortical processing is often seen to be a one-directional process (sensory to limbic), the phenomenon of consolidation suggests that it is ongoing and bidirectional, with extensive reentrant communication up and down the hierarchy of linked networks (Treves & Rolls, 1994). Corticolimbic interaction may be recursive in that it is self-
transformational, becoming altered in each reentrant iteration by the constraints operating at that representational level.

At the neocortical level, the constraints are imposed by the sensory qualities and motor afforances of the environment. The neocortex could be seen as a progressive modeling of the interface with the environmental context (Tucker, 1992). At the paralimbic level, the constraints are those of the internal bodily representation of the self, with the particular emotional excitability to regulate the consolidation of the memory/motive. This consolidation process can thus be seen as an ongoing arbitration between the bodily self and the sensorimotor interface with the environmental context (Tucker, 1992).

THE STABILITY–PLASTICITY HYPOTHESIS

The process of corticolimbic arbitration may also represent a negotiation of the stability–plasticity dilemma. Because memory is organized within the same networks as sensory and motor processes, there are no regions of the cortex that can be separated as protected regions of memory, and vice versa, there may be very few sensory processing areas (such as the ocular dominance columns) that are protected from the changes produced by a lifetime of experience. There does seem to be differential plasticity across the multiple levels, with the early myelination of sensory and motor neocortex resulting in less plasticity of these regions after the initial experience-expectant specification (Greenough & Black, 1992), and with extended plasticity maintained in paralimbic regions. Yet the amnesia evidence suggests that the consolidation process engages widespread neocortical as well as limbic networks, so that the cortex as a whole is susceptible to the interference created by encoding a new perceptual event. An interleaving process maintained by the hippocampus (McClelland et al., 1995) may allow a more graceful reorganization of the network, but the stability–plasticity dilemma remains. The result of each consolidation process is a wholly reorganized network. In psychological terms, there is only one representation, one concept, in the mind. To accommodate each new experience, the entire corpus of personal memories must be adjusted to create a new concept that accommodates both remote and recent history, as well as the implicit expectancy for the future.

If so, Freud’s proposed separation of memory and perception may not be found in the mammalian brain. Even for the cortical networks that may represent functional domains of the mind, such as frontal versus posterior systems, left versus right hemispheres, or archicortical versus paleocortical moieties, the fundamental integration of limbic with neocortical representations must occur within each corticolimbic network hierarchy. Because of this, the stability of existing representations in the network may be challenged by the consolidation of each new experience.

Until there is evidence that the human brain has solved, and not just managed, the stability–plasticity dilemma, we should consider the null hypothesis: There is no solution, this dilemma is inherent to human experience, and we have failed to recognize it. In fact, several of Freud’s psychological observations may be congruent with the hypothesis that the stability–plasticity dilemma is real.

Infantile Amnesia and the Wisdom of Age

Freud thought it remarkable that older children and adults cannot remember the events of their lives before the ages of 4 or 5. However, if we consider the plasticity of the young brain, with its overabundant synapse formation and vigorous pruning of
connections in response to functional interactions with the environment (Innocenti, 1983), it becomes clear that the brain's connectional architecture is progressively reorganized during the first years of life. The result may be that stability—at least in terms of a memorable continuity of experience—is sacrificed for this extensive plasticity. There is “memory” to the extent that the early foundations shape the connectional base of the network, but the very young experiences are inaccessible to the older mind.

As the child matures, there is greater continuity of the self, as experience becomes organized within a memory system that is coherent, and self-addressable, over time. In a connectionist model, a “fresh,” untrained system is highly unstable, and it is only through experience derived from training with varied instances that the system stabilizes. For the young child each experience is transformational, and the self (the memory system) is correspondingly labile. For a 10-year-old, a summer is a major epoch of subjective time, and at the end of that epoch the child may emerge with a substantially new self and a substantially new world view. Plasticity is extensive, and stability is therefore readily abandoned. For a 50-year-old, on the other hand, the summer comes and goes in a flash, hardly denting the accumulated connectional residuals of the old self. As if it were metered by the incremental changes of the self, subjective time compresses with each year of experience.

Similarly, in scientific work, we find that new theories are understood only by the graduate students, whose intellectual identities are then wholly transformed by their enthusiasm for the new insights. In contrast, the senior professors are burdened with such connectional inertia that when they encounter new ideas there is no apparent effect, other than an occasional vague irritation.

**Motivated Interleaving and the Corrective Emotional Experience**

Although there may be a linear developmental progression toward stability, mediated simply by the accumulation of connectional inertia, neural controls on the balance between stability and plasticity would be critical to the evolution of an effective memory system, as recognized by Grossberg (1984) and by McClelland et al. (1995). Even the brief consideration of corticolimbic neurophysiology described above suggests that the neural control of the motive-memory system is clearly more complex than the simple model of cathexis and catharsis of neurons in Freud’s *Project*. Yet what is missing from current studies of corticolimbic interaction is an appreciation of the essential role of motivational processes in memory, an appreciation that Freud clearly developed from his clinical observations.

Freud’s interest in catharsis was stimulated by his clinical experience with patients who improved psychologically when they were able to express the emotional distress associated with a traumatic memory. The implication seemed to be that the catharsis relieved a pathological cathexis, much like lancing a boil. However, more experience with psychotherapy taught Freud that catharsis alone may not result in effective psychological change. A constructive, cognitive integration seemed required as well, leading to the method of cognitive interpretation in psychoanalysis.

Yet the critical role of emotional arousal in the reorganization of memories, and the self, continued to be a key issue in the development of psychotherapy as a discipline. Even therapists who do not stress cathartic methods recognized that personal change (substantive reorganization of memory) does not occur with cognitive insight alone, unless there is a “corrective emotional experience.” (Rogers, 1951).

We may take the lesson from psychotherapy that effective corticolimbic recursion requires motivation. The mechanism may be a simple one of interleaving. An emotionally significant event captures attention (the ruminating of working memory, conscious
or unconscious). For example, the perception of an attractive object is a subjectively vivid event, and the memory of this perception returns to the imagination at idle moments. Similarly, even a minor social insult seems to gain a purchase in memory, to resurface and recruit indignation long after the episode would normally be forgotten. By gaining access to recursive processing, motivationally significant information gains access to the mechanism of self-organization.

Motivational control of plasticity must begin with the activity-dependent specification of the infant's brain. Whatever controls activity of the neuraxis controls the synaptic traffic, and thus connectional viability. The brain has evolved to control activity adaptively. Particularly for the young brain, emotional trauma may lead to a degree of catastrophic interference that threatens the integrity of the self. As plasticity declines, and the self becomes more fully differentiated, the self becomes less permeable to disruption by any given emotional episode. But self-organization and memory, even in the relatively rigid adult brain, are still likely to be gated by emotional arousal.

**De Facto Defense Mechanisms**

Freud observed that some memories appear to be “repressed” or kept from awareness because they are unacceptable to the ego. The recent epidemic in the United States of recovered memories of childhood traumas has caused renewed public interest in whether repression actually occurs. In recognition of the incidence of documented sexual abuse of girls in today’s families, it becomes difficult to accept Freud’s assertion that the traumatic memories of his patients reflected oedipal fantasies rather than fragmentary veridical memories. On the other hand, the fact that so many apparently normal people become convinced of past traumas, including alien abductions and satanic rituals, must show that the line between human memory and human fantasy is a thin one.

In this light, it is interesting to consider the stability–plasticity dilemma from the point of view of the self, that is, the existing memory system prior to the alteration of synaptic patterns required to encode a new, possibly highly cathected, experience. It may be that catastrophic interference is completely automatic, such that the self has no forewarning until the catastrophe is complete. In this case, the motivational energy to stimulate consolidation may be a primitive, uncontrolled process, such that following an emotional trauma the consolidation is unchecked until the degradation of the network is complete.

But it is possible to imagine that the ongoing process of consolidation entails some degree of self-monitoring, particularly if it is sustained over weeks and months. If the stability–plasticity hypothesis is accurate, an experience that is incompatible with the self demands a reorganization of the self. A difficult negotiation of the stability–plasticity dilemma requires sacrificing either the self or the memory of the experience. Repression therefore may not require a highly organized internal conspiracy, but may occur as a *de facto* result of the inability of integrating a difficult experience without creating a substantial disruption of the self. The possibility may then exist for incomplete consolidation, such that fragmentary representations of experiences are encoded in the network but not fully integrated within the self.

Extreme emotional trauma may not be necessary to create a difficult negotiation. In the social psychology literature on cognitive dissonance, subjects who are convinced to act in ways that are incompatible with their attitudes (self-representations) have been found to change their attitudes, often in surprising ways (Festinger & Carlsmith, 1959). The encoding of the new actions may force a reorganization of the representation of the self. According to the stability–plasticity hypothesis, the attitude change would be
required not by the actions themselves, but by the commitment of the actions to memory. In psychopathology, the phenomena of dissociations, fugue states, and other disorders of memory are closely associated with disorganization of the self, such as in multiple personality disorder (American Psychiatric Association, 1994). If stability-plasticity is a practical dilemma for the mind, memory consolidation and continuous self-organization may be inseparable outcomes of the same neurophysiological process.

Another phenomenon known as perceptual defense (Bruner, 1957), can also be accounted for in a de facto manner when we realize that memory is also stored in perceptual/motor networks. For Bruner perceptual defense is a mechanism that wards off events that are threatening, an idea that is very much in the Freudian tradition. However, unlike Freud and very similar to the position that we take, Bruner believed that perceptual defense is just a result of the way perception operates. He believed that perceptual defense can best be interpreted as perceptual interference. The interference is a direct result of existing categorizations based upon experience. These categories produce states of perceptual readiness that either capture poorly fitting events or that may create excessive background noise. In the framework of the present paper, because memory (the self) is represented in the perceptual network, information not consistent with the existing structure cannot be processed (Luu et al., 1996). The resulting effect is, in a sense, a protection of the existing ego; inconsistent information cannot be inserted into the consolidation process.

A Re-Interpretation of Dreams

Freud soon moved on from his neurophysiological theory to take up the psychological theory of the mind explored in the Interpretation of Dreams (Freud, 1953). He became convinced that the mental operations of dreams, such as condensation and displacement, provide insights into the workings of the unconscious mind. In terms of function, Freud theorized that the symbolic work of dreams may be protective, disguising the potentially threatening content in order to avoid waking the dreamer.

There may be a more practical function of dreams. If stability-plasticity is unavoidable, and if consolidation involves an interleaving of new patterns with existing connections, then the entire memory system must be adjusted to achieve a new concept that incorporates recent experience. To the extent that new synaptic connection weights must convolve recent experience with remote memories, the entire contents of past memory, at least those that are to be retained, would need to be activated during the consolidation process.

As the cortex was elaborated in mammalian evolution, its new memory capacities may have raised the stability-plasticity dilemma, creating the need to interleave each episode of learning with the cumulative memory system. REM (rapid eye movement) sleep appeared in evolution at the same time as the corticolimbic structures of the mammalian brain. It may be that dreaming evolved to support an extended consolidation operation on a daily basis. Substantial evidence supports a functional role for dreams in animal and human memory organization. Recent work with the rat shows that hippocampal cells with overlapping spatial fields are highly correlated during the learning of a spatial task. The correlated activity of these same cells during slow-wave sleep prior to and after the training show a significant increase after training (Wilson & McNaughton, 1994). In humans, interruption of REM sleep disrupts memory consolidation for implicit (procedural) memory (Karni et al., 1994). The motivational process (cathexis) that must be involved in memory consolidation during wakefulness may also be important for consolidation during sleep. There is evidence of a close relation between REM sleep and intracranial self-stimulation (ICSS) (Steiner & Ellman,
1972), an animal model of reward. REM deprivation lowers the ICSS threshold and ICSS during REM sleep deprivation reduces REM rebound.

Freud noted that dreams early in the night reflect “day residue” or the events of that day, whereas dreams later in the night’s sleep include content from childhood experiences. Considering the process of memory consolidation in terms of the negotiation between stability and plasticity, it may be that dreams allow the day’s experiences to be incorporated within the memory system. Isolated from enactment by the pontine motor inhibition, dreams provide a fully energized consolidation, supported by strong affective and autonomic activity. A key question for cognitive neuroscience is why the REM dream process is mediated by some brainstem neuromodulator systems (ACh) and not others (NE, 5-HT). In psychological terms, a related question is why certain aspects of the memory system are exercised with the bizarre mentation of REM, whereas other aspects appear to be reorganized through the more methodical and sequential non-REM mentation. In both cases, the dream process is not remembered, although we must assume that its effects are.

The progression from day residue toward more remote memories through successive REM cycles suggests that the dream interleaving process is an iterative one, reorganizing successive layers of the historical self. In a sense, the self seems to be destroyed during each night’s sleep and reborn each morning, following a convolution of significant recent experiences and the activated residuals of the past.

REFERENCES


POST, R. M. (1986). Does limbic system dysfunction play a role in affective illness? In *The limbic


