From the
Department of Clinical Neuroscience
Cognitive Neurophysiology Research Group
Karolinska Institutet
171 76 Stockholm Sweden

LEARNING AND MEMORY IN THE HUMAN BRAIN

Karl Magnus Petersson

Stockholm 2005

ISBN 91-7140-304-3

'We turn now to psychology! Recall the notation and terminology of chapter 5. There we defined a total computable function $\sigma(e, x, t)$ that codes the state of the computation $P_e(x)$ after t steps; $\sigma(e, x, t)$ contains information about the contents of the registers and the number of the next instruction to be obeyed at stage t. It is clear, then, that complete details of the first t steps of the computation $P_e(x)$ are encoded by the number

$$\sigma^*(e, x, t) = \prod_{i \leq t} p_{i+1}^{\sigma(e, x, i)}.$$

Let us call the number $\sigma^*(e, x, t)$ the code of the computation $P_e(x)$ to t steps. Clearly σ^* is computable.

Suppose now that we are given a total computable function ψ and a program P. By the *Ψ-analysis* of the computation P(x) we mean the code of the computation P(x) to $\psi(x)$ steps. We call a program P *Ψ-introspective at x* if P(x) converges and gives as output its own Ψ-analysis; we call P *totally Ψ-introspective* if it is Ψ-introspective at all x.

Theorem

There is a program P that is totally Ψ-introspective.

Proof. Simply apply corollary 1.4 to the computable function $f(e, x) = \sigma^*(e, x, \psi(x))$, obtaining a number n such that

$$\phi_n(x) = f(n, x) = \text{the } \Psi\text{-analysis of } P_n(x). \qquad \text{'}$$

Cutland, N. J. (1980). *Computability: An Introduction to Recursive Function Theory*. pp. 204-205. Cambridge, UK: Cambridge University Press.

TABLE OF CONTENTS

0. PREFACE

In writing this thesis, 'Learning and Memory in the Human Brain', I make no claims of originality. This should be apparent from the list of references. Most, if not all, ideas, insights, and concepts are already well-known. Perhaps not always in neuroscience or cognitive neuroscience, but in other important related fields like biology, psychology, linguistics, cognitive science, computational and computer science, physics and mathematics. I apologize for any unintended misrepresentation of concepts and lack of understanding of the ideas of others. I have sometimes chosen to stay relatively close to the original sources in an attempt to avoid this. Finally, being but a small stepping stone in the development of insight into human cognition and the workings of the human brain, I would like to suggest that the contemporary understanding of the very many complex issues involved in this enterprise is only in its beginnings. Although tremendous progress have been made due to the collective efforts in the field, we should not be surprised, but rather expect, that the present day ideas and insights will be radically transformed over time. I suspect that only the most general concepts and models will stand the tooth of time and this is primarily due to their lack of specific empirical content. Other concepts and ideas are also likely to survive, but for different reasons; they will survive on a terminological level with radically different content as a consequence of creative reinterpretation. We only need a brief look at the scientific development in physics, chemistry, and biology over the last 500 years to understand that cognitive neuroscience has a brave new future. Therefore, no contemporary model or interpretation of empirical data in cognitive neuroscience should be taken too seriously, including my own.

'Learning and Memory in the Human Brain' is based on two lines of empirical quest. The first attempts to investigate learning and memory in normal healthy young adult brains, while the second investigates the effects of literacy on the adult human brain. In the second line of experimental investigation, we take the view that the educational system is an institutionalized cultural process. Given that the educational system is an important source for structured cultural transmission, the study of illiterate subjects and their matched literate controls represents one opportunity to investigate the interaction between neurobiological and cultural factors on the outcome of cognitive development and learning (Petersson & Reis, 2005, in press;

Reis, Guerreiro, & Petersson, 2003). These two lines of empirical inquest are based on cognitive-behavioral laboratory experiments in combination with functional neuroimaging methods.

The thesis encompasses seven chapters, a reference list, and the eight papers on which the thesis is based. The first five chapters provide background material and in chapter 6 we discuss the experimental studies that form the basis of the thesis. In the first chapter, we provide a brief review of the brain, its structure and physiology, as well as cognition from the point of view of information processing in physical systems, including an outline of information processing as conceived of within the classical framework of cognitive science. We show how this perspective can be understood in terms of information processing in a certain class of dynamical systems (Church-Turing computable) and we indicate how this view of cognition can be generalized to general dynamical systems. In the second chapter, we integrate this dynamical view of cognition with learning and development. Here, cognition and learning as well as development are viewed as coupled (i.e., interacting) dynamical systems. Innately dependant constraints is conceptualized in terms of genetically dependent initial conditions as well as constraints on the form of the system dynamics, the space of cognitive states, as well as the space of learning/development parameters. In chapter 3 we describe the methodological background for the experimental studies that are discussed at some length in chapter 6. In chapter 4, we review the cognitive neuroscience of human memory systems and chapter 5 provides a review of experimental work on literate and illiterate subjects, in particular work on our study population of Olhão in the southern Portugal.

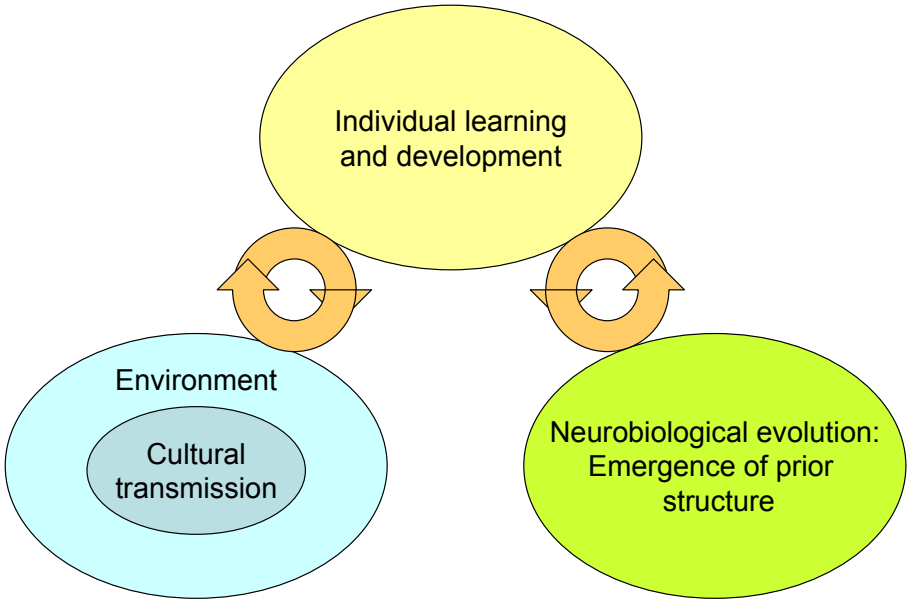
The first experimental study discussed in chapter 6 outlines several approaches to the study of learning related effects in the human brain with hemodynamically based functional neuroimaging methods (Petersson, Elfgren, & Ingvar, 1999b). Two of these approaches are applied in the second and third study, where we take the view that learning can be viewed as processes by which the brain functionally restructures its processing pathways or its representations of information. In the context of the second and third study, several previous lines of research have suggested that repeated reactivations of the neocortical representations of declarative memories strengthen the neocortical interconnections so that the neocortical memory network eventually can

support retrieval independently of the medial temporal lobe (MTL). In these studies (Petersson, Elfgren, & Ingvar, 1997; Petersson, Elfgren, & Ingvar, 1999a) it was assumed that practice and consequent reactivation of the relevant neocortical regions would strengthen the network interconnections in such a way that the neocortex could support memory retrieval less dependent on the interaction with the MTL. An additional perspective on these studies is provided by the concepts of controlled and automatic processing, where controlled processing is relatively more dependent on attentional and working memory processes related to the anterior cingulate and fronto-parietal networks. The natural prediction then, as retrieval in some sense become more automatic with practice, is that retrieval should be less dependent on these brain networks. The experimental results reported are broadly consistent with these suggestions. These investigations of learning related modulation of functional retrieval networks were further explored in two different experimental paradigms in the fourth study (Petersson, Sandblom, Gisselgård, & Ingvar, 2001). This allowed us to investigate material specific effects on learning related modulation of retrieval as well as to investigate the effects of performance. In the fifth study (Petersson, Reis, Castro-Caldas, & Ingvar, 1999) a group of healthy older illiterate women was investigated on an auditory word-pair association cued-recall paradigm. We report that effective declarative encoding correlated positively with the level of activation observed in the MTL as well as the inferior prefrontal region. In study 6, 7, and 8, illiterate subjects and their matched literate controls were investigated during simple auditory-verbal language tasks. In study 6, literate (4 years of schooling) and illiterate participants were compared on immediate verbal repetition of words and pseudowords (Castro-Caldas, Peterson, Reis, Stone-Elander, & Ingvar, 1998). The experimental results provided the first indication that learning to read and write during childhood influences the functional organization of the adult human brain. The follow-up study (Petersson, Reis, Askelöf, Castro-Caldas, & Ingvar, 2000) suggested that the parallel interactive processing characteristics of the underlying language-processing network differ between literate and illiterate subjects during immediate verbal repetition. Finally, in the 8th study, the activation levels of the right and left inferior parietal regions were investigated in two independent groups of illiterate subjects and their matched literate controls (Petersson, Reis, Castro-Caldas, & Ingvar, submitted).

Overall, the results suggested that literate subjects are relatively more left lateralized compared to illiterate subjects. Based on these results, we suggested that acquiring reading and writing skills at the appropriate age shapes not only the local morphology of the corpus callosum (Thompson et al., 2000; Zaidel & Iacoboni, 2003) but also the degree of functional specialization as well as the pattern of interaction between the interconnected regions of the inferior parietal cortex.

Karl Magnus Petersson
2004-07-17

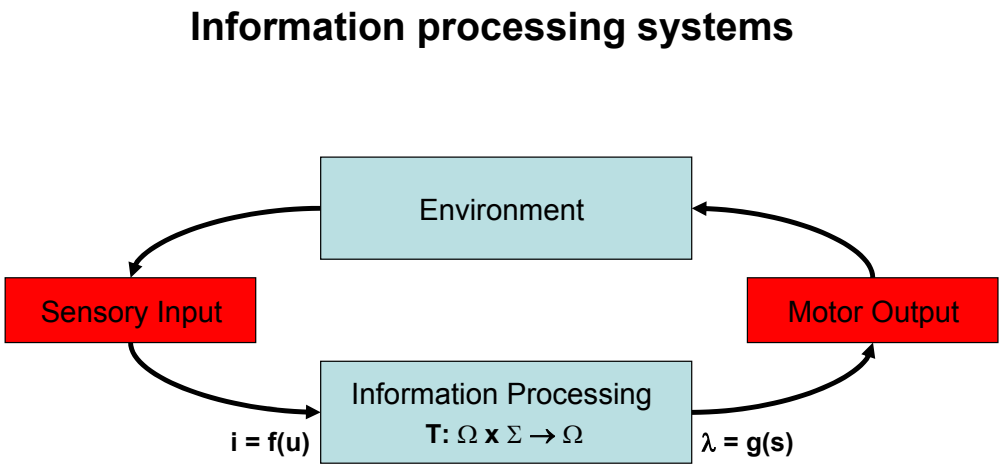
1. GENERAL REFLECTIONS ON COGNITIVE BRAIN FUNCTIONS



[Figure 1.1] An adaptive cognitive system situated between its evolutionary history and current environment. Neurobiological systems represent evolved biological systems and in order to fully understand the significance of their different features it seems reasonable to take not only their individual histories (ontogenesis) into account but also the evolution of the whole system (phylogenesis). For example, the capacity of an embodied cognitive system to learn and develop provides a necessary basis for the possibility of cultural and evolutionary interaction.

We begin by reviewing some structural and functional facts about neural systems that are relevant from a cognitive neuroscience point of view. We will also briefly outline the classical cognitive perspective on psychological explanation; that is, the standard framework of Church-Turing computability for information processing systems. In the next chapter we will sketch a generalized non-standard computability framework based on a

dynamical systems perspective on cognition. This latter framework incorporates the classical perspective as a special case and encompasses the class of neural networks as a natural model for cognition. We will also try to integrate these perspectives with some contemporary ideas on the functional architecture of the human brain, learning and adaptation at different characteristic time-scales, and more broadly the interaction, via individual learning, between factors determined by neurobiological evolution as well as the environment of the human cognitive system, including social and cultural transmission (Figure 1.1). This and the following chapter are expanded versions of Petersson (2004), Petersson (2004, in press), and Petersson, Grenholm, and Forkstam (in preparation).



[Figure 1.2] Information processing systems. Cognition is equated with internal information processing. Here the cognitive system is portrayed as interfacing with the external environment. However, it should be noted that the processing (sub-)system equally well can be viewed as interfacing with other sub-systems; i.e., the processing system is an internal sub-component that receives input from and transmit output to other sub-systems. In the figure, the space of internal states, s, is represented by Ω (i.e., $s \in \Omega$). The processing

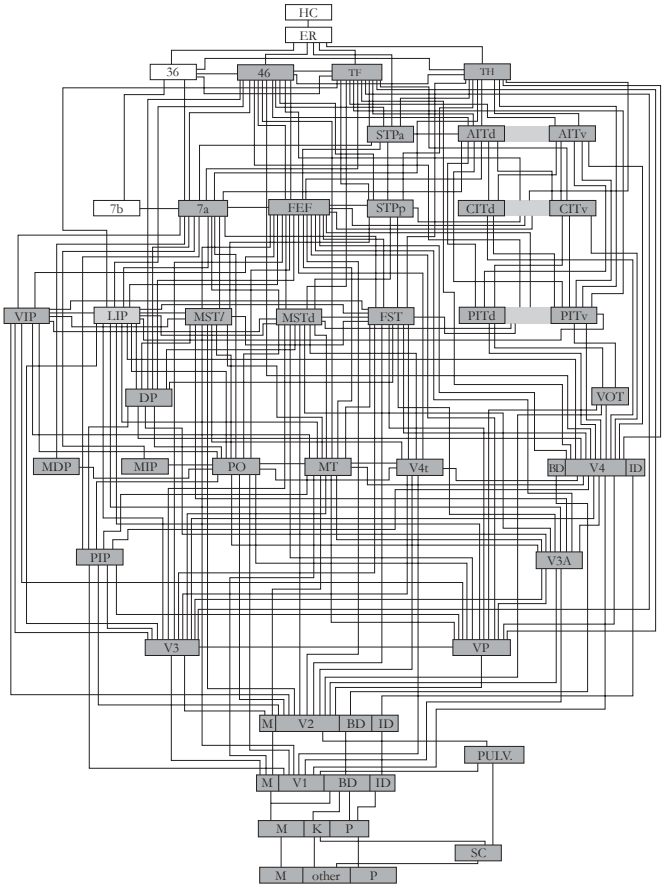
of information is governed by dynamical principles, T, which for simplicity here is represented as a cognitive transition function $T:\Omega x\Sigma \rightarrow \Omega$: Given an internal state $s \in \Omega$ and input u, here transformed according to $i = f(u) \in \Sigma$, T specifies (deterministically or indeterministically) a new internal state $T(s, i) \in \Omega$ and output is generated according to an output transformation $\lambda = g(s)$.

In general we will consider a physical system as an information processing device (Figure 1.2; i.e., a computational system in a general sense), when a subclass of its physical states ($s \in \Omega$; cf. Figure 1.2) can be viewed as representational (or rather, cognitive, in the sense of Jackendoff, 2002 pp. 19-23) and transitions ($T:\Omega x\Sigma \rightarrow \Omega.$; cf. Figure 1.2) between these can be conceptualized as a process operating on these cognitive structures (i.e., in some sense implementing well-defined operations on the representational structures). More generally, information processing, that is, the state transitions, can be conceptualized as trajectories in a state space (cf., discussion below). We shall use the terms 'representational' and 'cognitive' interchangeably. It is important to note from the outset that when we are using 'representational', this is not meant to implicitly entail an idea or conceptualization of meaning in terms of a 'referential' or 'representational semantics'. Rather, 'representational' or 'cognitive' is referring to the functional role of a physical state with respect to the relevant processing machinery, and thus does not have an independent status separate from the information processing device as such. In other words, meaning is inherent or created by the processing system as a whole, though various degrees of internal isolation in terms of natural sub-systems is conceivable. Thus, the 'internal semantics' of the system is at best only in complex and indirect ways related to the exterior of the system (via the sensorimotor interfaces and the corresponding processing sub-systems) and there may be important aspects which only has an internal significance. We also note that since the brain can only represent 'numbers' in terms of membrane potentials, inter-spike-intervals, or any appropriate set of dynamical variables, it is clear that the human brain does not represent cognitive structures in a simple transparent manner. However, it is well-known that the class of so-called symbolic processing models can be captured within the Church-Turing framework of computability, which is equivalent to the class of partially recursive functions

(Cutland, 1980; Davis, Sigal, & Weyuker, 1994; Rogers, 2002). Hence it is possible to simulate all finitely specified symbolic models as processes on numbers. Furthermore, it has recently becom known that these models can be emulated in dynamical systems, including generic first-order recurrent networks (for a reviews see, Siegelmann & Fishman, 1998; Siegelmann, 1999) and low-dimensional smooth dynamical systems (Moore 1991). For example, the analog recurrent network architecture can be viewed as a finite set of analog registers (e.g., membrane potentials) that processes information interactively and concurrently (cf., section 1.4, 1.6.2, and 2.1.2).

1.1 A BRIEF OVERVIEW OF THE STRUCTURAL AND FUNCTIONAL COMPLEXITY OF THE BRAIN

We will in the following sub-section follow the general ideas as outlined by Koch and Laurent in their interesting and thought provoking "Complexity and the nervous system" (1999). The human brain - a cognitive system - of which presumably relevant aspects can be conceptualized in terms of information processing, is one of the more (if not the most) complex systems in the known universe. Macroscopically the human brain can be characterized as approximately 1.5 kg of grey and white matter; the grey matter is formatted into a convoluted surface of gyri and sulci that contains neurons as well as local and more long distance neuronal interconnectivity, while the white matter contains long distance cortico-cortical regional and cortico-subcortical interconnectivity, sensory input as well as motor output fiber tracts and inter-hemispheric tracts (Nieuwenhuys, Voogd, & van Huijzen, 1988). Besides the neocortex, grey matter is also localized to the medial temporal cortex (including the hippocampus), the basal ganglia, the cerebellar cortex and nuclei, as well as various other subcortical nuclei in the mesencephalon and brainstem (Nieuwenhuys et al., 1988). Microscopically, the brain is composed of about 10^{10} – 10^{12} neuronal processing units (i.e., the neurons), each supporting on average 10^3 – 10^4 axonal output connections and receiving, on average, the same number of dendritic and somatic input connections. The connectivity comprises in total some hundreds of trillions of interconnections and many thousand kilometers of cabling (Koch & Laurent, 1999; Shepherd, 1997).



[Figure 1.3] The structural organization of the human brain. Brain connectivity resembles a (weakly) hierarchically structured, recurrently connected network composed of different functionally specialized brain regions, which consists of several types of processing elements (neurons) and synaptic connections (Felleman & Van Essen, 1991; Shepherd, 1997). (Adapted from Felleman & van Essen, 1991; courtesy of Frauke Hellwig).

The functional complexity of the nervous system arises from the non-linear, non-stationary, and adaptive characteristics of the neuronal processing units (including synaptic parameters that can change across multiple time-scales of behavioral relevance), and the spatially non-homogeneous, parallel and interactive patterns of interconnectivity (Figure 1.3). These characteristics are one reason it is difficult to analyze and understand the nervous system as an information processing system (note that the terms 'non-linearity' and 'non-stationarity' are not well-defined properties but rather reflect the absence of 'linearity'

and 'stationarity' – fundamentally, this is also the reason why there is no general method of attack for the analysis of this type of systems (cf., McCauley, 1993a, p. 2)).

The structural organization of brain connectivity resembles that of a (weakly) hierarchically structured, recurrently connected network composed of different functionally specialized brain regions, which consists of several types of processing elements and synaptic connections (Felleman & Van Essen, 1991; Shepherd, 1997). Interestingly, the classification scheme used by Felleman and van Essen (1991) in their survey (Figure 1.3) gradually breaks down at the higher processing levels. This is consistent with the hypothesis that there is no focus for process control (cf. the CPU of the Von Neumann architecture, Tanenbaum, 1990). This point is also illustrated in maps of functional connectivity, which are apparently lacking a central processing focus (Stephan et al., 2000). Additional data from Goldman-Rakic and colleagues (e.g., 1988) support these suggestions, indicating that higher order, domain general structures like for example the prefrontal cortex, the cingulate cortex, and the medial temporal lobe depart from the connectivity patterns of lower order, domain specific regions. In addition, recent work in cognitive neuroscience (see e.g., Gazzaniga, 1999) indicate that organizational principles for cognitive brain functions depend on distributed connectivity patterns between functionally specialized brain regions as well as functional segregation of interacting processing streams (the dominant pattern of interconnectivity being recurrent). Now, the processing properties of a given brain region is clearly determined by its extrinsic and intrinsic connectivity pattern, its neuronal subtypes, their properties (including e.g. the distribution of receptor types and ion channels) as well as the local connectivity. However, given the surprisingly uniform basic outline of the neocortical architecture, the functional role of a given brain region might to a not yet well-understood degree be determined by its place in the neocortical macro-circuitry. Structural and functional evidence supporting this hypothesis were recently reviewed by Passingham and colleagues (Passingham, Stephan, & Kötter, 2002), and they suggest that each cytoarchitectonic area has a unique pattern of input and output connectivity and a corresponding pattern of task dependent functional connectivity. However, this rather static view is likely to be revised, given the possibility of dynamically (i.e., dependent on the processing context) established functional networks, issues to which we will return further on in this text (cf., section 1.4.1 and Figure 1.7).

One may ask why the brain is so heavily recurrently interconnected. This complexity is unnecessary for a system based on linear, sequential, hierarchical feedforward information transfer, but is essential for network processors that support interactive recurrent distributed processing. In consequence, it appears that parallel interconnected distributed anatomical networks, characterized by recurrent interconnectivity and functional integration across cortical networks are essential processing characteristics of the brain. For a network to have the capacity to realize a wide range of dynamic behaviors, functional feedback supported by recurrent anatomical connectivity is necessary. The specifics of the input and output connectivity, as well as the local architecture of a given brain region, are as we have already noted important determinants of the region's behavioral and cognitive significance; in other words, its functional specialization and its range of functional integration options in relation to other brain regions. The dynamics of the interfaces between the functionally specialized regions characterize, at least partly, the specifics of functional integration in a given processing context. Additional determinants of the functional architecture are the mechanisms that enable the processing systems to incorporate adaptive changes, allowing the system to learn as a functional consequence of information processing. Thus, the system is non-stationary and the class of realizable dynamical models consequently becomes richer and can be viewed as being parameterized by the adaptable parameters of the network (cf., sub-sections 1.5 and chapter 2).

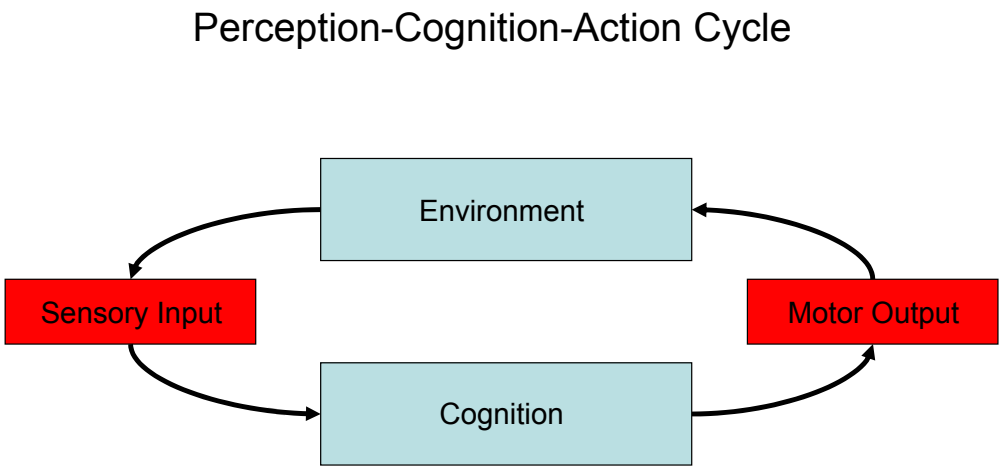
Since the network circuit hypothesis of McCulloch and Pitts (1943; see also Minsky, 1967) and the neuronal assembly hypothesis of Donald O. Hebb (1949), several approaches to addressing information processing in neural systems have suggested that information is represented as distributed activity in the brain and that information processing, subserving complex cognitive functions, emerge from the interactions between different functionally specialized regions or neuronal groups. These approaches include the perspectives of theoretical modeling (Amit, 1989; Arbib, 2003; Haykin, 1998; Hertz, Krogh, & Palmer, 1991; Trappenberg, 2002), cognitive psychology (Horgan & Tienson, 1996a; Macdonald & Macdonald, 1995; McClelland & Rumelhart, 1986), and cognitive neuroscience (Koch & Davis, 1994; Mesulam, 1998), as well as lesion approaches (Eichenbaum & Cohen, 2001; Squire, 1992; Zola-Morgan & Squire, 1993) and functional

neuroimaging based on electrophysiological (Varela, Lachaux, Rodriguez, & Martinerie, 2001) and hemodynamic methods (Friston, 1994; Horwitz, 1998; Horwitz, Tagamets, & McIntosh, 1999; McIntosh & Gonzalez-Lima, 1994). Fundamentally, all these approaches suggest that cognitive functions emerge from the global dynamics of interacting sub-networks. Moreover, despite the fact that at least some neurons and neural systems appear to perform at levels not too far off from what is physically possible, given the input and hardware characteristics (Rieke, Warland, van Steveninck, & Bialek, 1996), it appears that the basic computational units of the brain (i.e., neuron or its synapses) and their interconnections (Koch & Segev, 1998) are relatively slow and, perhaps, imprecise in relation to the real-time task demands on processing performance and this seems to be related to inherent processing limitations of neurons.

In conclusion, the neural system is likely to represent information in terms of neural assemblies and population codes (Arbib, 2003; Gerstain, Bedenbaugh, & Aertsen, 1989; Gerstner & Kistler, 2002; Trappenberg, 2002), and although some neurons appears to integrate inputs regardless of their temporal structure, substantial evidence exists that the relative timing of action potentials carries information, allowing for combinatorial spatiotemporal codes (cf. e.g., Arbib, 2003; Gerstner & Kistler, 2002; Koch & Davis, 1994; Koch & Laurent, 1999). Furthermore, it seems plausible that the brain processes information interactively in parallel and that rapid, fault tolerant, and robust processing properties emerges from these processing principles (cf. e.g., Amit, 1989; Arbib, 1995; Haykin, 1998; Hertz et al., 1991). In this context, it is interesting to note that, given the intricate complexity at multiple levels of structure as well as function, Koch and Laurent (1999) suggest that continued reductionism is not likely on its own to lead to a fundamental understanding of cognitive brain functions from a complex systems perspective. Instead, they argue, that the detailed investigation of the nervous system has to be complemented by investigations at several different system levels (cf., Amit, 1989, 1998; Arbib, 2003; Trappenberg, 2002). At present, higher cognitive functions of the nervous system are commonly characterized in terms of large-scale/macroscopic concepts that are relevant at a behavioral level. An important objective of cognitive and computational neuroscience is therefore to bridge between the properties that characterize neurons, or neuronal assemblies, and the processing units and processing principles that are subserved by neural

networks and are relevant to cognition. Clearly, the most crucially outstanding issue is related to questions about the neural code and how functional descriptions are to be translated into this code.

1.2 THE PERCEPTION-COGNITION-ACTION- AND THE ENCODING-STORAGE-RETRIEVAL CYCLE



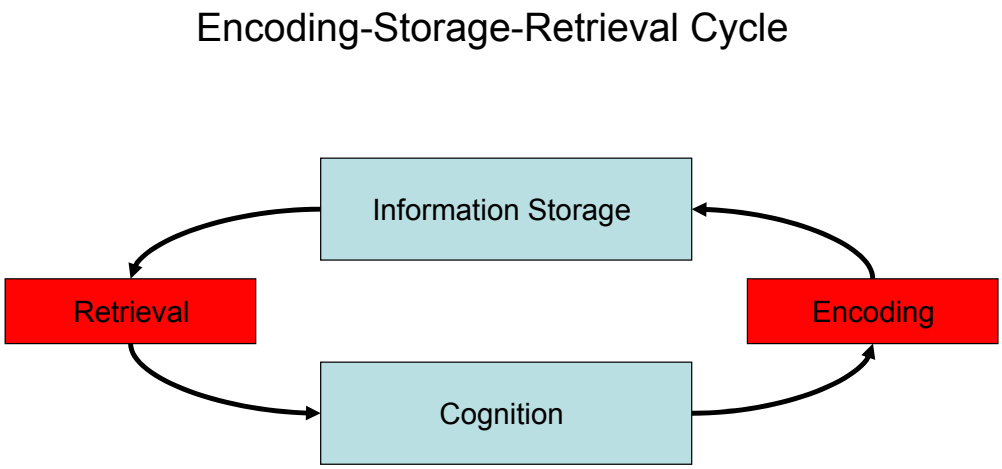
[Figure 1.4] The perception-cognition-action cycle. The perceptual systems allow the human brain to extract relevant patterns of information from, at times, a noisy, changing, and unpredictable environment, while the motor output apparatus allows it to temporally organize behaviorally relevant actions and act in a goal directed fashion in its environment (including e.g. the creation of artifacts, communicating with conspecifics, as well as to effect changes in the physical and socio-cultural environment). Here cognition is equated with internal information processing. Note the similarity with the conceptualization of an information processing system in Figure 1.2.

It is of heuristic value to ascribe the brain an overarching function and think of this in terms the perception-cognition-action cycle (Rosenbleuth, Wiener, & Bigelow, 1943; Wiener, 1948); for example, to protect the individual and its kin within a particular ecosystem and to increase the likelihood of propagating its genetic information. The individual brain interfaces its environment, through sensory input surfaces and motor output machinery, in what may be called the perception-cognition-action cycle: sensory input → perceptual processing → cognitive processing → temporal organization of motor output → action (Figure 1.4). The brain receives perceptual information through several sensory modalities and coordinates actions in the form of movements of the skeleto-muscular apparatus, glandular responses (regulated by the autonomic nervous system), as well as other soft (e.g., the larynx and tongue) appendages.

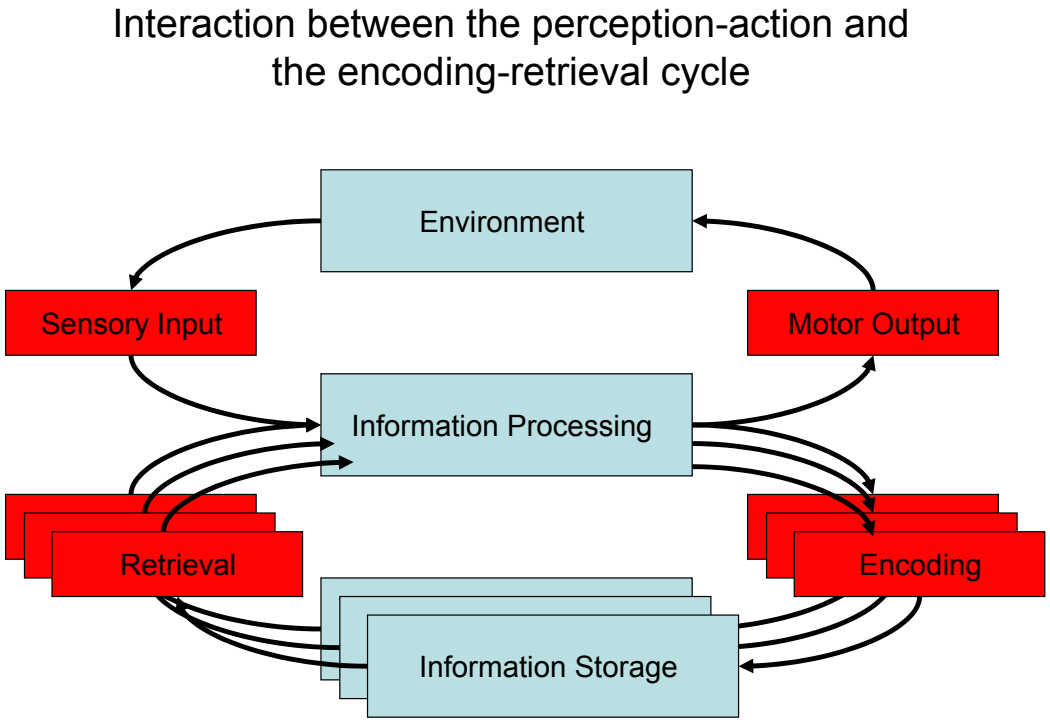
Beyond the previous remarks, brain complexity is also reflected in the structural composition of its processing units (neurons), including the composition of the dendritic tree and neuronal soma, its synaptic organization and passive as well as active membrane properties supported by voltage- and neurotransmitter-gated ion-channels, and its axonal arborization. These characteristics provide neurons with adaptable nonlinear dynamical properties (Koch, 1999; Shepherd, 1997). Chemical synapses show a number of different forms of plasticity with characteristic time-scales that span at least nine orders of magnitude, from milliseconds to weeks, providing a necessary substrate for learning and memory (Anderson, 2002; Koch, 1999; Koch & Laurent, 1999).

Generally, information is received through the input synapses of a neuron and flows from the dendritic tree, via the soma, to the axon hillock where an action potential may (or may not) be triggered, spreading along the axon and the final terminal arborization, where neurotransmitters are (stochastically) released into the synapse from the pre-synaptic membrane, which then diffuses across the synapse and activate post-synaptic receptors thus generating a post-synaptic potential; and the whole process starts anew in the downstream neuron. In all its roles, the nervous system invokes neuronal processing, store information, through memory formation and changes in its adaptable properties, generating models or representations relevant given its dynamic processing environment. From a cognitive neuroscience perspective the perception-cognition-action cycle thus needs to be

complemented by the encoding-storage-retrieval cycle (Figure 1.5). The perception-action cycle and the encoding-retrieval cycle interact through active processing of information subserved by various forms of short-term working memories. In addition, one has to imagine that there is not only one encoding-retrieval cycle but several, and likewise, that there are several parallel perception-action cycles. This gives rise to the idea of different memory systems as well as interacting cognitive modules (Figure 1.6).



[Figure 1.5] The encoding-retrieval cycle. Learning can be defined as the processes by which the brain functionally restructures its processing networks and/or its cognitive representations as a function of experience. The stored information (i.e., the memory trace) can then be viewed as the resulting changes in the processing system. The processing system is thus non-stationary, and from this perspective, learning in a neural network is the dynamic consequence of processing and network plasticity. In all its roles, the nervous system invokes neuronal processing, store information, through memory formation and changes in its adaptable properties, generating models or representations relevant given its dynamic processing environment.



[Figure 1.6] The interaction between perception and encoding-retrieval cycles. In order to incorporate the capacities for memory, learning, and adaptation explicitly, the perception-cognition-action cycle needs to be complemented with the encoding-retrieval cycle. These cycles interact through the on-going active information processing in for example working memory. Here learning and adaptation is conceptualized as a functional consequence of information processing.

1.3 MODULARITY

The neural system controls behavior with local and global consequences in terms of survival and reproductive success. We can attempt to understand important aspects of neural processing within an evolutionary framework considering that the human brain has an evolutionary history on the order of 1 billion years (Koch & Laurent, 1999).

Evolvability, the property of a genetic system to tolerate mutations and modify the genotype without seriously reducing its phenotypic fitness, must have provided, it seems, a selective advantage. Koch and Laurent (1999) suggest that the property of evolvability favored compartmentalization (or modularity), redundancy, weak and multiple (parallel) linkages between regulatory processes as well as component robustness (for a somewhat different perspective, see Fodor, 2000). The idea is that sufficient stability and tolerance for evolutionary modification is provided if several (many) of the constituent components and their coupling links are not crucial for survival but can serve as a substrate for evolutionary tinkering (i.e., search in fitness space). It therefore seems reasonable to assume that such indirect evolutionary pressures should lead to neural systems replete with specialized circuits, parallel pathways, and redundant mechanisms (Koch & Laurent, 1999). The effects of neurobiological evolution can thus be conceptualized as a mechanism for the incorporation of prior structure into the processing infrastructure and we will return to this important issue later on in chapter 2.

Rarely do cognitive models of brain functions detail explicit models for the processing infrastructure or the underlying neurophysiological events or processes that support them (see e.g., Charniak & McDermott, 1985; Fodor, 1983; Newell, 1990; Posner, 1989; Stillings et al., 1995). In contrast, (artificial) neural network approaches make assumptions regarding interactive parallel processing elements and base their ideas on models of various degrees of neurobiological plausibility (see e.g., Amit, 1989; Arbib, 2003; Churchland & Sejnowski, 1992; McClelland & Rumelhart, 1986; Trappenberg, 2002). Independent of whether cognition is best conceptualized in terms of the classical cognitive rule-based symbolic processing paradigm (Fodor & Pylyshyn, 1990) or in terms of parallel sub-symbolic processing at one level of abstraction or another (Shastri, 1995; Shastri & Ajjanagadde, 1993; Smolensky, 1988), it is clearly the case that cognitive functions are implemented in the network architecture of the brain and depend on the processing characteristics of such networks.

Before we proceed to briefly outline the classical cognitive paradigm, we note that it is important to realize the differences between brains and computers. The logical gates from which a computer is constructed are homogeneous and non-adaptive (though this of course does not rule out memory). Moreover, the connectivity density of gates is commonly

low compared that of the brain (cf. e.g., Savage, 1998; Tanenbaum, 1990, for concrete examples). In the central processing unit of any microprocessor, one gate is connected, on average, to on the order of 1 - 10 other gates, approximately a factor 1000 - 10000 less than inter-neuronal convergence and divergence. More importantly, neural systems wire themselves during ontogenetic development and this circuitry is modifiable by learning (also throughout adult life). While we often conceptualize brain function in terms of information processing, the character of the brain as a dynamical system differ significantly from present day computer architectures in the scale of structural and dynamic complexity. For example, a computer (e.g., a Von Neumann Machine) can viewed as incorporating a general purpose "homunculus" in the form of a central processing unit exerting finite state control over the process flow (Minsky, 1967; Savage, 1998), and while the processing in a computer is highly coordinated and synchronized globally (explicitly through a clock-frequency or implicitly through different versions of just-in-time processing), these features seems absent in neural systems as described above. The apparent absence of global process coordination represents an outstanding challenge for cognitive neuroscience to better understand. In addition, the classical cognitive science perspective is not easily translated into the processing characteristics of neural systems and some have taken this as evidence indicating that there may be a fundamental problem with the classical view (e.g., Charniak, 1993; Churchland & Sejnowski, 1992; Edelman, 1990; Rumelhart & McClelland, 1986), while for example Chomsky (2000b) has described this as a problem for neuroscience rather than cognitive science. However, as we will briefly review in chapter 2, recent advances in the understanding of non-classical information processing in dynamical systems allows us to begin to imagine how we might integrate the classical cognitive science framework within a more general dynamical systems framework which also includes the recurrent neural networks as a natural class from an analog information processing perspective.

1.4 CLASSIC COGNITIVE MODELS

The framework of classical cognitive science and artificial intelligence (cf. e.g., Charniak & McDermott, 1985; Fodor, 1983; Newell, 1990; Posner, 1989; Stillings et al., 1995) assumes that information is coded by structured representations ("data structures") and that

cognitive processing is accomplished by the execution of algorithmic operations ('‘rules'') on the basic representations ("symbols") making up the structured representations. This processing paradigm, sometimes called rule based symbolic processing (cf. e.g., Horgan & Tienson, 1996a; Wilson & Keil, 2001) suggests that cognitive phenomena can be modeled within the framework of Church-Turing computability. In other words, this perspective effectively takes the view that isomorphic models of cognition can be found within the framework of Church-Turing computability (cf. e.g., Cutland, 1980; Davis, Sigal, & Weyuker, 1994; Lewis & Papadimitriou, 1981; Rogers, 2002). From this perspective, a cognitive system consist of a state space of internal states (represented by Ω in Figure 1.2) and computations are instantiated as transitions (represented by $T:\Omega x\Sigma \rightarrow \Omega$; Figure 1.2) between states while optionally receiving input ($i = f(u) \in \Sigma$; Figure 1.2) and generating output ($\lambda = g(s)$; Figure 1.2) as determined by a cognitive transition function (deterministic computation) or transition relation (non-deterministic computation) and thereby generating trajectories in state space (Cutland, 1980; Davis et al., 1994; Lewis & Papadimitriou, 1981; Savage, 1998).

Here, we will formulate computation and the framework of Church-Turing computability from a dynamical systems perspective (cf. equation [1]). Consider the simpler case of a cognitive transition function. This is no restriction since non-deterministic transition relations only add descriptive convenience but no additional computational power. So, let Σ be the space of inputs i ($i \in \Sigma$), Ω the space of internal states s ($s \in \Omega$), and Λ the space of outputs λ ($\lambda \in \Lambda$). The possible cognitive transitions T are then determined or governed by a transition function $T:\Omega x\Sigma \rightarrow \Omega x\Lambda$ (i.e., $T:\Omega x\Sigma \rightarrow \Omega$ extended with $\lambda:\Omega x\Sigma \rightarrow \Lambda$ for convenience). In other words, suppose at processing step n, the system receives input i(n) when in state s(n), then the system changes state into s(n+1) and outputs λ(n+1) according to:

$$[s(n+1), \lambda(n+1)] = T[s(n), i(n)] \qquad [1]$$

In this way, the processing system traces a trajectory in state space, …, s(n), s(n+1), …, while reading the input stream …, i(n), i(n+1), …, and generating the output …, λ(n),

λ(n+1), … (cf., Figure 1.2). Equation [1] is a description of a time-discrete dynamical system. Within the framework of Church-Turing computability, it is assumed that Σ, Ω, and Λ are all finite. In this context, Equation [1] describes a forced (i.e., driven by the input …, i(n), i(n+1), …) time-discrete dynamical system, which generates trajectories or orbits (i.e., …, s(n), s(n+1), …) in a finite state space, the combined effect of which is a constructed sequence of actions …, λ(n), λ(n+1), … . Here, we have not explicitly described the memory organization of the computational system (cf., Table 1). In principle, this is crucial because the properties of the memory organization in terms of storage capacity (e.g., finite or infinite), and accessibility (e.g., stack- or random access) determine in important respects the computational power of the processing architecture (for details see e.g., Davis et al., 1994; Lewis & Papadimitriou, 1981; Savage, 1998).

Architecture	**Complexity**	**Memory organization**			
		States	***Registers***	***Stack***	***Accessibility***
FSA	Finite	Finite	-	-	-
PDA	Finite	Finite	-	Unlimited	Top of stack
LBA	Finite	Finite	Unlimited[1]	-	Random access
URA	Finite	Finite	Unlimited	-	Random access

Table 1. The Chomsky hierarchy and the memory organization of respective architecture. In the table, complexity refers to machine complexity. FSA = finite state architecture, PDA = (non-deterministic) push-down architecture, LBA = (non-deterministic) linearly bounded architecture, URA = unlimited register architecture (which is equivalent to the Turing architecture). [1] Linearly bound in the input size with a universal constant.

It is important to distinguish between the complexity of the computational mechanism of the architecture (machine complexity) and the complexity of its memory organization. We will briefly focus on just one aspect of the memory organization, its storage capacity; in particular, whether this is finite or infinite. This turns out to be crucial for the expressivity

of the system, one important aspect of which is the types of recursive structure that can or are expressed in the generated output. For all classical architectures, the transition function T can be realized in a finite-state architecture. For example, in the case of the universal Turing architecture, the transition function T in (1) can be implemented as a finite state machine (finite-state control, cf., Savage, 1998). Thus, within the classical framework, information processing is subserved by transitions between internal states, while in general receiving input, storing intermediate results of the computation in memory, and generating output. Thus, with respect to the mechanism subserving transitions between internal states there is no fundamental distinction in terms of machine complexity between the different computational architectures (Table 1, cf., Savage, 1998). However, as indicated by the strict inclusion in the Chomsky hierarchy (Table 1, cf., Davis et al., 1994), there are differences in expressivity. These differences are fundamentally related to the interaction between the generating mechanism and the available memory organization. The most important determinant of structural expressivity is the availability (or absence) of infinite storage capacity. Thus, it is the characteristics of the memory organization, which in a fundamental sense, allow the architecture to recursively (Cutland, 1980; Rogers, 2002) employ its processing capacities inherent in T, to realize functions of high complexity or achieve complex levels of expressivity (Petersson, 2004, in press). However, the Chomsky hierarchy is only one of the simplest examples of a complexity hierarchy and is of limited significance from an implementational view (Petersson, 2004, in press). Instead, much of the more recent work in complexity theory (e.g., Papadimitriou, 1994) focuses on more fine-grained complexity hierarchies related to realizability requirements and computational costs in terms of processing time and memory space requirements for effective general solutions to problem classes.

Language modeling in theoretical linguistics and psycholinguistics, among other cognitive domains (cf. e.g., Charniak & McDermott, 1985; Newell, 1990; Russel & Norvig, 1995), represents one example in which the classical framework clearly has served us well (cf. e.g., Partee, ter Meulen, & Wall, 1990; Sag, Wasow, & Bender, 2003). A fundamental hypothesis of generative grammar (Chomsky, 1957) is that it is possible to give an explicit recursive definition of natural language (or at least for syntax) and all commonly used

formal language models can be described within the classical framework (Partee et al., 1990; Wasow, 1989).

1.4.1 LEVELS OF DESCRIPTION

Cognitive models of information processing, formulated within the classical framework, can profitably be analyzed at (least) three levels of description (Marr, 1982): 1) the *functional/computational level*, which specifies in formal terms which function results from the processes of the system, that is, a formal theory for the function computed by the system (generally a partial recursive function, cf., Rogers (2002)); 2) the *procedural/algorithmic level*, which, given a formal theory, specifies the representations and procedures for processing these representations (i.e., Σ, Ω, Λ, and $T{:}\Omega x\Sigma \rightarrow \Omega x\Lambda$ above); 3) the *implementational/hardware level*, which, given an algorithmic description, specifies how the representations and procedures are implemented in a physical system.

A central idea of classical cognitive science, so-called functionalism, is that the fundamental architectural aspects of cognition are independent of any particular implementation, but can be captured in terms of an abstract functional organization by virtue of which the physical state transitions are systematically (homomorphically) related to. The mathematical description briefly outlined above is useful in order to characterize this functional organization and constitutes the design of cognition according to the classical view. However, an important constraint for models of cognition, which claims to model physically realizable systems, is that processing has to be feasible to implement in a physical device. This constraint has been elaborated in terms of tractable computability (Horgan & Tienson, 1996a, see also the preface of Charniak (1993) for some interesting reflections on this issue). Tractable computability requires that it is possible to implement an algorithmic description in a physical device (e.g., meeting the constraint of a finite memory organization as well as real-time constraints) and that this can be achieved within reasonable computational complexity, logical depth and machine complexity (cf., Savage, 1998). Here reasonable is often taken to mean that the implementation does not consume computational resources that scales exponentially in time and space with the problem size. In other words, only algorithmic descriptions of polynomial complexity (Hopcroft, Motwani, & Ullman, 2000; Papadimitriou, 1994) are feasible. However, for efficient

solutions in real systems, a general polynomial complexity constraint might not be sufficiently tight, but might require a low-order polynomial constraint, at least in time (e.g., given the sluggishness of neurons compared to silicon-hardware; while, constraints in memory space might be somewhat more lax, taking the view that each neuron corresponds to a memory register). In practice this means that the implementation must meet real-time and space constraints of the task the system is set to handle and these are determined by on-line processing time and other limitations of the physical device. It is of interest to note that it has been suggested that some aspects of cognition may be non-tractable, from the perspective of classical computational theory (e.g., Horgan & Tienson, 1996b; for an alternative viewpoint from physics in general, see McCauley, 1993b). The demands in terms of computational complexity, it is argued, seems to be to great in terms of time- and/or memory-space complexity to be tractably implementable, and perhaps even computably unsolvable (Fodor, 2000). There also appears to be problems of tractable computability in unconstrained models of natural language; for example, aspects of language performance related syntactic parsing and comprehension display complexity characteristics which might be problematic from a tractability point of view, unless the models are further constrained (Barton, Berwick, & Ristad, 1987; Wasow, 1989).

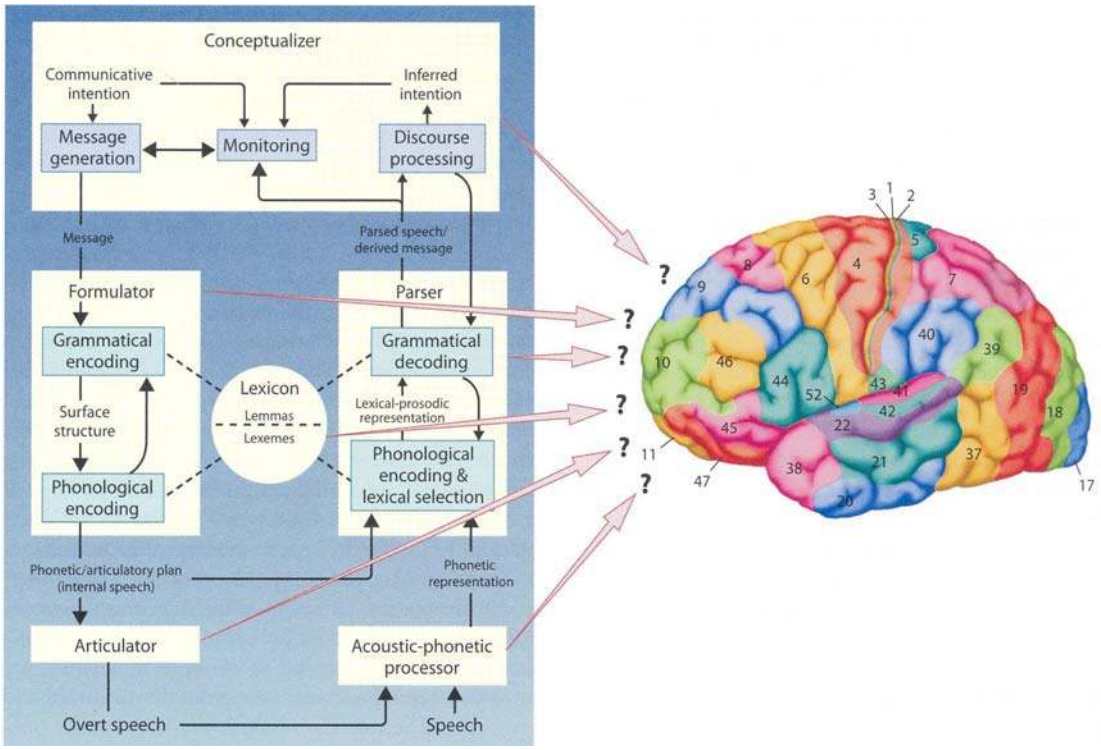
Classical cognitive science is also associated with the idea of modularity, that is, the cognitive architecture is conceived as being divided into well-defined sub-components, which interact communicatively. Classical cognitive modularity is closely associated with, but not necessarily dependent on, the idea of (in relevant respects) genetically determined and informational encapsulated structures. In other words, these modules are viewed as input-output devices, which are isolated from lateral or top-down influences between modules and are feeding a central domain-general processing module. This is essentially the classical high-level feed-forward perspective on cognition outlined by Jerry Fodor (1983) in the ‘Modularity of Mind’. In this way, cognition is commonly divided into functional domains, including for example, sensory-perceptual, different types of short-term working- and long-term memory, language, emotion, attention, planning, problem solving, and the temporal organization of behavior. Furthermore, these domains are commonly elaborated and divided into further sub-domains and cognitive components/processes. Some evidence for cognitive modularity has come from

neuropsychology. Neuropsychological lesion data have been interpreted as supporting a modular view of brain function, not only in functional terms, but also in structural-anatomical terms. In particular, data indicating double dissociations have been interpreted as support for the usefulness of this framework. In addition, data on developmental disorders have also been interpreted as support for the view that cognitive modules are to some extent innately/genetically specified. However, recent studies have indicated that it might be possible to understand these findings in a different way (e.g., Elman et al., 1996; Paterson, Brown, Gsödl, Johnson, & Karmiloff-Smith, 1999; Plaut, 1995; Young, Hilgetag, & Scannell, 2000). Furthermore, there is at present no accepted canonical way of deconstructing cognition into domains except perhaps at a very coarse level.

Why, then, is the classical perspective of modularity difficult to integrate with a neurophysiological perspective on brain function? From the preceding discussion it should be clear that the short answer is: we know far too little about cognition and its implementation in the brain in conjunction with a lack of understanding of the coding or representational as well as the processing principles of the nervous system (cf. e.g., Arbib, 2003; Gerstner & Kistler, 2002).

To recapture, the organization of the brain resembles a hierarchically structured, recurrently connected network, in which brain regions and neural assemblies interact in parallel and in an integrative fashion. Functional neuroimaging data are entirely consistent with this latter perspective (e.g., Ingvar & Petersson, 2000) and adds a complication to simple ideas of how functional properties are mapped onto anatomical structures. It is therefore unlikely that the structure-function mapping is direct and transparent and this of course has important consequences on the interpretability of data generated from functional neuroimaging as well as behavioral experiments (see chapters 3-5). In particular, a given brain region may serve different functions depending on the functional context in which it operates at any given moment of processing. More specifically, a given brain region may dynamically participate in several functional networks and it is the functional network which, at least partly, determines the functional role of the region. Furthermore, since information is believed to be represented as distributed activity and information processing is thought to emerge from the interactions between different specialized regions, these processing characteristics suggest that the structure-function relationship is complex

(Figure 1.7). However, from a neuroscience perspective on cognition it is natural to think about cognitive function in terms of interactive, parallel distributed, processing principles, and the great challenge is to understand how cognitive function can arise from network architectures such as the brain.



[Figure 1.7] Cognitive-functional and anatomical-structural modularity. The left part of the figure represents the psycholinguistic model of language processing by Levelt (1989). Generally, there is no accepted canonical way of deconstructing cognition into domains except perhaps at a very coarse level at present. To some extent this also holds for anatomical-structural segmentation of the brain. The organization of the brain resembles a hierarchically structured, recurrently connected network, in which brain regions and neural assemblies interact in parallel and in an integrative fashion. A given brain region may dynamically participate in several functional networks and it is the functional network which, at least partly, determines the functional role of the region. Furthermore, since information is believed to be represented as distributed activity and information processing

is thought to emerge from the interactions between different specialized regions, these processing characteristics suggest that the structure-function relationship is complex. Adapted from (Gazzaniga, Ivry, & Mangun, 1998). permissions@wwnorton.com

1.5 A DEVELOPMENTAL PERSPECTIVE ON COGNITION - THE CLASSICAL VIEW

For simplicity, let us assume some version of cognitive modularity, and let us focus on some particular module C, which is fundamental in the sense that all normal individuals develop cognitive capacities related to C. As a preliminary assumption, it is then reasonable to view aspects of the module C as a species-wide adaptation. At any point in time t one can imagine C being in a given state $m_C(t)$ [Note that here state refers to the model instantiated by C rather than an internal state in state space. This state (or model) is more akin to a point in the space of adaptive parameters, cf. below]. If we suppose that C incorporates an innately specified prior structure, we can capture this by the notion of a structured initial state of C, $m_C(t_0)$. If the system has adaptive characteristics we can conceptualize the development of the system as a trajectory in its accessible model space M = {m| m can be instantiated by C} driven by the interaction with the environment and in conjunction with innately specified developmental processes. Thus, as C develops, it traces out a trajectory in M determined by its adaptive (or developmental) dynamics L, and the input i(t) it receives, according to:

$$m_C(t+\Delta t) = L(m_C(t), i(t), \Delta t, t) \qquad [2]$$

where the explicit dependence on time, t, in L captures the idea of an innately specified developmental process (maturation) as well as the possible dependence on the previous developmental history. If C and L are such that it (approximately) converges on a final model $m_C[F]$, this will characterize the end-state of the developmental process reached after time Δt_F, that is, $m_C(t_0+ \Delta t_F) \approx m_C[F]$.

Within the classical cognitive framework of equation [1], m_C determines the transition function T in the following sense: T can be viewed as a function of m_C, that is, T

is parameterized by m_C according to $T = T[m_C]$. In other words, we take the view that the accessible model space M can be viewed as a functional form $T[\cdot]$ that represent a parameterized model space M (i.e., we have $T:M\times\Omega\times\Sigma \rightarrow \Omega\times\Lambda$ instead of $T:\Omega\times\Sigma \rightarrow \Omega\times\Lambda$, which from a developmental perspective is a static model). Equation [1] should thus be modified according to:

$$[s(n+1), \lambda(n+1)] = T[m_C][s(n), i(n)] \qquad [1']$$

where $m_C = m_C(n)$ is updated according to the adaptive dynamics:

$$m_C(n+1) = L(m_C(n), i(n), n) \qquad [2']$$

We thus see that development (as well as learning) of a cognitive system can be conceptualized in terms of a forced system of coupled (i.e., interacting) equations. This is in essential respects similar to Chomsky's well-known hypothesis concerning language acquisition (e.g., Chomsky, 1980; Chomsky, 1986) where the module C is taken to be the faculty of language, L the language acquisition device, and the model space M the set of natural languages, which determine the universal grammar. Different aspects of the universal grammar, including constraints and principles (Chomsky, 2000b), are captured by M, L, and the initial state $m_C(t_0)$. Language acquisition and prior knowledge of language can arguably be viewed as a species-wide adaptation. Chomsky and others have argued extensively that the inherent properties of M, L, and $m_C(t_0)$ are determined by innately specified (genetic) factors, genetically determined morphogenetic processes, in interaction with the physiochemical processes of cells. One might attempt to translate the theory of principles and parameters (Chomsky & Lasnik, 1995) into the present framework where the principles and constraints are related to aspects of M, L, and $m_C(t_0)$, and the parameters are related to the adaptive aspects of $m_C(t)$. From this point of view, natural language acquisition is the result of an interaction between two sources of information: 1) *innate prior structure*, which is likely to be both of a language specific nature as well as of a more general non-language specific type (e.g., this would include both characteristics of the initial state as well as the characteristics of an innately specified learning mechanisms); 2)

the *environment*, both the linguistic and the extra-linguistic, which can be viewed as an interactive boundary condition for the developing system. The type of learning that characterizes language acquisition seems to be implicit in nature and the knowledge acquired is to a large extent unconscious (Chomsky, 1986). One may thus suggest that a relevant learning paradigm for language acquisition, from the point of view of learning theory (Arbib, 2003; Haykin, 1998; Jain, Osherson, Royer, & Sharma, 1999; Vapnik, 1998), can essentially be captured by a mixture of innately constrained unsupervised/self-organizing (e.g., Arbib, 1995; Haykin, 1998) and perhaps modern reinforcement learning (e.g., Sutton & Barto, 1998). In conclusion, it thus appears that language development, as an example of cognitive development in general, is the result of the interaction between genetically determined factors and processes as well as the environment. However, it should be emphasized that the outline captured in equation [2] is not necessarily related to the classical cognitive framework per se (i.e., equations [1'] and [2']) but can be viewed as a more general recipe that can be applied also to non-classical frameworks, a perspective to which we will return to in chapter 2.

1.6 COGNITIVE NEUROSCIENCE

Fundamental objectives of cognitive neuroscience are to understand how different cognitive brain functions are implemented in the neural processing infrastructure and to understand the detailed relationship between the structure of the brain, cognitive function, and behavior. In order to achieve these objectives, it is necessary to adequately characterize brain structure at the relevant levels of description, to formulate a general framework for conceptualizing cognitive brain function, and to measure relevant neurophysiological events and processes as well as adequately characterize behavior.

1.6.1 WEAK REDUCTIONISM

Emergent complex high-level phenomena necessarily presuppose interaction between systems constituents. The attempt to understand complex systems in terms of their systems-level organization has recently received new interest in biology (cf., Csete & Doyle, 2002; Kitano, 2002). These approaches to systems biology have adopted the perspectives of control theory (Isidori, 1995; Sontag, 1998; Wiener, 1948) and attempts to understand

complex systems in terms of interacting modules or components and their interface properties from a reversed engineering perspective (Csete & Doyle, 2002). High-level phenomena like cognition supposedly represent emergent system properties that depend on low-level phenomena in some more or less systematic fashion. The complexity of the brain is realistically described in terms of non-linear, non-stationary, and adaptable processing elements interconnected in a highly parallel distributed and non-homogeneous network topology. This led Koch & Laurent (1999) to suggest that a fundamental understanding of the brain can probably not be achieved by continued reductionism and atomization, at least not at the present stage. One may question whether it is possible, or even meaningful, to attempt a complete reduction of one level of description to another, given the dependence on abstraction (in a technical sense) when going from a lower-level to a higher-level of description. When we attempt to bridge the gap between cognition and neurophysiology in a substantial sense, it may be the case that we can only hope for what Chomsky has termed "unification through accommodation" (2000b). Chomsky (2000b) provides a number of examples of what he has in mind. For example, the explanation of planetary motion in terms of contact mechanics was demonstrated by Newton to be unsolvable but was overcome by introducing immaterial forces (i.e., gravitation); the problem of reducing electromagnetism to mechanics was resolved by accepting fields as real physical entities, while the problem of reducing chemistry to physics was only overcome by introducing "even weirder hypotheses about the nature of the physical world" (i.e., quantum physics). Thus, he argues, in each of these cases "unification was achieved and the problem resolved not by reduction, but by quite different forms of accommodations" (Chomsky, 2000a). Another example, of what will be called a weakly reductive explanation, is represented by statistical physics (e.g., Huang, 2001; Mackey, 1992; Mandl, 1988; Reif, 1965). Statistical physics exemplifies one of the most well-understood methods to analyze the macroscopic properties of high-dimensional systems composed of weakly interacting microscopic constituents. Moreover, methods from statistical physics have been applied to models of brain functions, in particular learning and memory, as well as to information processing more generally (e.g., Amit, 1989; Arbib, 2003; Engel & Van den Broeck, 2001; Hertz et al., 1991; Leff & Rex, 1990; Nishimori, 2001). Now, take for example the extremely simple case of an ideal gas, an ensemble of weakly interacting particles. This system is described

by on the order of 10^{23} degrees of freedom (dimensions) at the microscopic level. However, the macroscopic behavior can to a good approximation be described by 3 degrees of freedom determined by the simple equation pV = nkT. Thus, a nominal extremely high-dimensional system effectively reduces its macroscopic behavior to a low-dimensional system. Thus we may hope, although this is altogether unclear at present, that in a similar fashion it will become feasible to relate the microscopic description of the processing infrastructure of the brain to a macroscopic cognitive-behavioral description. However, it should be kept in mind that low-dimensionality as such does not imply simple system behavior. On the contrary, low-dimensional non-linear systems can display behavior of any imaginable level of complexity, including deterministic systems (Beck & Schlögl, 1993; Lasota & Mackey, 1994a; McCauley, 1993a; Moore, 1991a, 1991b; Ott, 2002).

In order to capture cognition, it seems clear that linear interactions are not sufficient. Instead, non-linear types of interaction have to be at play for interesting phenomena to emerge. The interaction between neurons is characteristically weak; the influence of a single neuron on another is relative small, typically on the order of 1% of the firing threshold. This implies that cortical neurons rely on convergent and cooperative afferent input, some of which may be part of the 'spontaneous' background activity, to activate a single neuron. Thus, it is clear that the functional significance of a single neurons behavior to a large extent is determined by its processing context. Furthermore, synaptic transmission appears to be stochastic in nature. For example, the probability of synaptic release given an action potential (AP), P[release | AP], can be as low as P[release | AP] ~ 0.1 (Koch, 1999). In addition, P[release | AP] is non-stationary and adaptable - it depends on the stimulus history. In addition, the postsynaptic outcome in terms of the postsynaptic potential can also be variable (Koch, 1999).

Returning to the issue of bridging the gap between a microscopic (e.g., neurophysiology) and a macroscopic (e.g., cognition/behavior) description in a substantial sense, one has to remember that the principles and units of analysis for macroscopic phenomena are not necessarily the same as those for describing and analyzing the microscopic phenomena. This difference is roughly captured by intuition that there is a difference between our understanding of how various graphical user interfaces work and the principles for organizing the circuit logic of a computer and how the computer's

instruction set come to have a functional role in the logical circuitry (Savage, 1998; Tanenbaum, 1990). There is always an irreducible element inherent in the description of high-level phenomena (explicitly or implicitly), which has to do with how the microscopic variables are composed or aggregated to explain macroscopic phenomena. For example, a macro (e.g., in some assembler language, cf. e.g., Cutland, 1980; Davis et al., 1994) is implemented as a compositional structure. The macro is composed from the instruction set, and is in one sense dependent on the particular instruction set used. However, the macro is not determined by the instruction set, since a complete description of the macro requires a specification of the compositional structure. Moreover, the functional description of the macro is in important respects independent of the chosen instruction set, since it does not matter which instruction set and logical circuitry is chosen to implement the functional description. This corresponds to the idea of functionalism in classical cognitive science, which suggests that the fundamental architectural aspects of cognition are independent of any particular implementation. This also happens to be the fundamental reason why recursive function theory (Rogers, 2002), which does not refer to or depend on any particular implementation, precisely corresponds to any particular universal computational implementation (i.e., the Church-Turing thesis), including for example, the Turing architecture (Lewis & Papadimitriou, 1981), the unlimited register architecture (Cutland, 1980), semi-Thue production systems (Davis et al., 1994), and Post systems (Minsky, 1967). This has been succinctly stated as: 'Hardware and software are logically equivalent' (Tanenbaum, 1990, p. 11).

It is important to realize that a simple statement of system dependence on microscopic variables is insufficient for a reductive explanation. More importantly, the emergent macroscopic form is in essential ways dependent on the functional form of the interaction between microscopic constituents, and in the case of stochastic systems, also on statistical properties of ensembles of interacting constituents (e.g., averaging properties like ergodicity, mixing etc., see e.g., Billingsley, 1995; Lasota & Mackey, 1994a). It should be realized that these forms or properties of the interaction, in a narrow sense, represents an irreducible system property. By this we mean that the form of interaction between microscopic constituents is not explainable in terms of the constituents themselves in isolation. Instead, there is necessarily an added element in the specification of the system in

terms of the form of the interaction between constituent components as well as boundary conditions not determined by the system as such. This emphasizes the importance of understanding the interactive and integrative principles governing the processing system as well as the context in which the processing takes place.

Finally, and from a systems perspective, it should be noted that the microscopic details of the constituents may turn out to be less important for high-level explanations than the details of the interactions between constituents, opening up for the possibility to study reduced models in a meaningful manner (Amit, 1989, 1998; Gerstner & Kistler, 2002; Trappenberg, 2002). Concepts like structural stability and canonical models (Devaney, 1989; Hoppensteadt & Izhikevich, 1997) might also be of importance as a foundation for the study of reduced models.

1.6.2 A GENERAL FRAMEWORK FOR CONCEPTUALIZING COGNITIVE FUNCTION – INFORMATION PROCESSING IN DYNAMICAL SYSTEMS

In order to capture the essentials of the classical cognitive perspective and outline a broader framework, which we suggest might be more amenable for understanding cognition form a neurobiological perspective, we will first generalize Marr's three levels of analysis: the computational, the algorithmic, and the implementation level (Marr, 1982). We will follow Horgan and Tienson (1996a) quite closely and within this generalized framework, the three levels of analysis correspond to: 1) the *cognitive level* - a formal (mathematical) theory of structured representational states and the cognitive transition system, which specifies in formal terms which transition function results from information processing of the system; 2) the *dynamical system level* - given a formal cognitive theory, a state space is specified and processing is formulated in terms of dynamical systems; 3) the *implementational level* - given a dynamical system, this level specifies the physical hardware implementation of the dynamical system; for example, how the dynamical system is realized in the neural networks of the brain. Here we end by noting that a dynamical system instantiated in a neural network is determined by its local dynamics (i.e., local processing in computational units) and its network topology, and thus the temporal transition/evolution of global states of the network will in general be extremely complex, with the entire structure of the network affecting it.

As previously emphasized, a system can be understood as processing information, if the system in some well-defined sense can be said to represent information or functional properties by a set of internal states in its state space and that the system processes information or performs computations (in a general sense, i.e., not necessarily computable in the sense of Church-Turing) on these representations by its internal dynamics, being driven (i.e., forced) by input or spontaneous internal activity, and optionally generates output. Again, we emphasize that to 'represent information'' here does not presume a referential theory of representation or semantics (cf., section 1 and Jackendoff, 2002). Furthermore, it should be noted that, a priori, the concept of "processing" here entails little more than a description of the evolution of the dynamical system along trajectories in state space (cf. e.g., Haykin, 1998; Lasota & Mackey, 1994b; Ott, 1993). Intervening states along the state space trajectories, between representational states, do not necessarily have to be meaningful from an external point of view or in terms of cognitive processing, other than in terms of 'mediating states' enforced by the particular dynamics or implementation. These mediating states may thus have a purely implementational or system internal 'semantics'. The perspective taken here thus conceptualizes processing of information as implemented by the dynamics of a dynamical system.

This framework generalizes that of classical cognitive science, which represents a subset of dynamical systems (as already noted; cf., equation [1]). In attempting to relate cognitive function to the brain within the suggested framework, one has to characterize: 1) what information is represented and the coding principles by which this information is encoded; that is, a specification of the structure of representations; 2) the representational dynamics; that is, the processing principles of the system. For example, the nervous system can be considered as a physical dynamical system with causal and stochastic interactions that generate trajectories in an appropriate state space (e.g., representing membrane potential of each neuron along a particular dimension in a high-dimensional state space; i.e., viewing each neuron as an analog register). In other words, the temporal evolution of a brain system, determined by the neuronal dynamics, can be seen to correspond to trajectories in a state space, representing transitions between physical states. Within this framework, this is all that is meant by the idea of instantiating information processing in a physical system. It is interesting to note that this perspective has recently been revived,

under the umbrella of non-standard or analog computation, in theoretical computer science (for recent reviews see e.g., Siegelmann, 1999; Siegelmann & Fishman, 1998; Siegelmann & Sontag, 1994).

Here we want to emphasize that the recurrent architecture of the brain seems to be a prerequisite for functional integration and serves as a basis for the dynamical systems perspective. The flip side of functional integration is naturally functional specialization. Functional specialization, as presently understood, may depend on the functional processing context, that is, the functional network in which a given region is engaged during processing: a given brain region can subserve several functional roles depending on the network it is dynamically 'plugged into'. This suggests a weaker form functional modularity, or *dynamic functional modularity*, and this point of view entails the possibility of interacting dynamical sub-systems (e.g., sub-networks). Dynamic functional modularity, as well as the concept of networks of networks (functional specialization/integration), predicts that the brain should support an inhomogeneous dynamics, that is, the local characteristics of the dynamics in state space is not translational invariant. This seems necessary, if a structured parallel and interactive processing perspective shall be meaningful. Finally, the question regarding the most appropriate way to resolve the information processing system of the brain along structural-anatomical and cognitive-functional dimensions remains open.

2. INTERACTION OF ADAPTABLE SYSTEMS AT DIFFERENT TIME-SCALES

As neurobiological systems are evolved biological systems, in order to fully understand their different features it seems necessary to take not only their individual histories (ontogenesis) into account but also the evolution of the whole system (phylogenesis), which thus provide a functional context for the different features at any given point in time. Moreover, it is important to realize that evolved structures may or may not be, in some sense, close to optimal 'solutions' because it can be assumed that evolution generates 'good enough' rather than optimal 'solutions'.

Neural systems are examples of naturally evolved information processing systems which have evolved under tight energy, space, and real-time processing constraints. These constraints include the available energy flows and the available space (anatomical) for the neural system as an integral part of the organism; the processing time and space (memory) available to perform computation in order to appropriately solve a task on a behaviorally relevant time-scale (e.g., input signals need to be recognized and response patterns organized and executed on the relevant time-scale). With respect to perception, the characteristic time-scale of computation must match that of the external world. The same holds for output control, the characteristic time-scale of motor output must match the time it takes to organize a coordinated response that is behaviorally relevant. Moreover, when the outputs from different processing components need to be integrated, then the time-scales of the various processors involved must also match. There are also constraints set in terms of energy turnover, the physical and biochemical infrastructure, as well as spatial constraints. It seems safe to assume that these types of general constraints must have had an important influence on the brain from an implementational point of view. It is unknown to what extent it has had any influence on the functional organization at the cognitive/computational level but this seems likely at the dynamical system/algorithmic level. However, there are good reasons to believe that the nervous system is not fully specified at a phylogenetic (i.e., genetic) level – the existence of learning and adaptation speaks clearly on this issue – but it would also seem too restrictive, ineffective, or too costly to pre-specify every detail of the functional organization of the brain at a phylogenetic level. Instead, ontogenetic development and learning represent viable complements.

During brain development, from the fertilized egg to the adult brain, the normal individual acquires an amazing range of cognitive skills; this includes for example sensori-percepto-motor skills, natural language and communicative skills, procedural skills, as well as general/semantic and episodic knowledge, and so forth. Furthermore, the capacity for learning and development provide a necessary basis for the possibility of cultural and evolutionary interaction (cf., Figure 1.1). Although, it seems clear that evolved prior structure can in principle influence culture in a general sense through various cognitive constraints, it is unclear to what extent the reverse is the case, though still in principle possible (e.g., it has been suggested that the emergence of the dairy farming culture selected for adult lactose tolerance (Feldman & Cavalli-Sforza, 1989)). It should also be kept in mind that these processes, phylogenetic evolution, ontogenetic development, and individual learning, operate on different time-scales relevant for changes at the individual-, cultural- and species level.

In what ways do cognitive and neural processes interact during development, and what are the consequences of this interaction for theories of learning? Quartz and Sejnowski (1997) attempted to sketch a neural framework for addressing these issues. Development, learning, and cognitive skill acquisition implies that the neural infrastructure changes its processing characteristics as a result of these processes. Given the co-localization of memory (i.e., here in the general sense of adaptive changes) and processing in the brain, this entails a system with properties that are time-dependent ('non-stationary') and the effected changes in the processing characteristics are driven by and result from the outcome of an interaction between neurobiological maturation and experiential learning processes. Quartz and Sejnowski (1997) suggested that two themes emerged in their review of structural measures of representational complexity: (1) development comes with a progressive increase in the structural complexity, which underlies representational complexity, and (2) this increase in the structural complexity depends on interaction with a structured environment as a guide to the development. They also suggested that this favored a neo-Piagetian view, which they called neural constructivism, implicating that there is an active constructive interaction between the developing system and the environment in which it is embedded. Post-natal human cerebral development is a progressive process, which last at least until early adulthood. This suggests the possibility

of a complex interaction between environmentally derived information and prior genetic structure, which takes place during ontogenetic development, in constructing mental representations and processing networks (Thompson et al., 2001; Thompson et al., 2000). This form of environmentally guided neural circuit building is a form of learning, called constructive learning by Quartz and Sejnowski (1997). Quartz and Sejnowski (1997) suggest that the central problem confronting a cognitive system is to find an appropriate class of representations for specific problem domains and that this problem is resolved by constructive development/learning principles, which creates sufficiently efficient representations under the influence of the environment, in interaction with general constraints imposed by the neural architecture. This perspective emphasizes the interaction, via individual learning, between evolutionary determined neurobiological constraints and experiential factors, including cultural transmission. Recently, Li (2003) reviewed the re-emerging co-constructive conceptions of development and outlined a framework for cognitive and behavioral development across the life span. Li (2003) suggested that new insights might be gained from an integrated perspectives on cultural and experiential influences with behavioral genetics and cognitive neuroscience.

In this chapter, we will take human language as an illustration of the issues involved (cf., Hauser, Chomsky, & Fitch, 2002; Nowak, Komarova, & Niyogi, 2002). Human language is a major vehicle for cultural transmission. It has been suggested that natural language arises from three distinct but interacting adaptive systems: individual learning, cultural transmission, and biological evolution (Christiansen & Kirby, 2003). As will be outlined below, these systems can be viewed as adaptive as well as interacting. The adaptive character of evolution as well as individual learning ('adaptation of the individual's knowledge') is undisputed but this is less clear for cultural development. Moreover, Christiansen and Kirby (2003) argue that the knowledge of particular languages persists over time by being repeatedly used to generate language output and this output represents input to the language acquisition device of the individual learner. It is likely that aspects of natural languages have adapted to the constraints set by the language acquisition device. Constraints on language transmission are thus set by prior structures determined by the outcome of biological evolution. However, Christiansen and Kirby (2003) suggest that if there are features of language that must be acquired by all learners, and there are

constraints or selection pressures on the reliable and rapid acquisition of those features, then an individual who is born with such acquisition properties will have an advantage, exemplifying the so-called Baldwin effect of genetic assimilation, whereby acquired features can become innate. These suggestions are entirely consistent with Chomsky's perspective that the universal competence grammar is determined by the language acquisition device and the initial state of the individual (cf., discussion in section 1.5). Christiansen and Kirby (2003) suggest further that in order to understand the human language capacity and its evolution it is necessary to understand the workings of the human brain; the structure and use of language; how cultural change affect language and how language distinguishes itself from communication systems; the evolution of hominids; how language acquisition takes place during ontogenetic development; and how learning, culture and evolution interact. So, after Chomsky (1965; 1986) suggested that prior innate constraints are necessary in order to simplify the acquisition task to such a degree that language acquisition becomes possible for the learner, and thus in effect transferring the problem of learnability (Gold, 1967; Nowak et al., 2002; Pinker, 1991) to a problem of the evolutionary origins of language, we seem to made a full circle. However, the (un)learnability problem might not be as severe as is commonly thought (for brief reviews, see e.g., Petersson, 2004, in press; Petersson, Forkstam, & Ingvar, 2004). The (un)learnability paradox is often associated with a well-known result in formal learning theory (Gold, 1967; see also Jain et al., 1999), which states that no super-finite class of languages is in general learnable from positive examples alone without additional constraints on the learning paradigm. It has also been suggested that this is the case when statistical learning mechanisms (cf. e.g., Cherkassky & Mulier, 1998; Duda, Hart, & Stork, 2001; Vapnik, 1998) are employed (Nowak et al., 2002). However, already Gold (1967) noted that under suitable circumstances this (un)learnability paradox might be avoided. Recent results in formal learning theory confirm Gold's (1967) suggestion that, if the class of possible languages is restricted, then it is possible to learn infinite languages in infinite classes of formal languages from positive examples (Shinohara, 1994; Shinohara & Arimura, 2000). It should be noted that these prior constraints on the class of possible (or accessible) languages are of a general type and not 'language specific' per se (e.g., restrictions on the maximal number of rules employed by the languages in the class). As

noted by Scholz and Pullum (2002), there exists classes of formal languages rich enough to encompass the 'string-sets' of human languages and at the same time being identifiable from a finite sequence of positive examples. Furthermore, the acquisition task becomes potentially more tractable if there are additional structure in the input or if only 'probable approximate' identification is required (cf. e.g., Anthony & Bartlett (1999) for an outline of the probably approximately correct learning paradigm). One possibility is to generate expectations based on an internal model for prediction (Petersson, 2004, in press). Within an unsupervised learning framework, the internal model can be acquired through a learning process which is driven by the difference between input and internally generated predictions (i.e., self-organized based on general constraints). A simple example of this is the simple recurrent network (SRN) architecture (e.g., Elman, 1990). Recent results suggest that this may be a viable approach for network models of finite recursion (Christiansen & Chater, 2001).

2.1 THE NEUROBIOLOGY OF CHANGE – LEARNING AND ADAPTATION AT DIFFERENT CHARACTERISTIC TIME-SCALES

Evolutionary/phylogenetic processes are driven by genetic adaptation in terms of random variation and subsequent selection based on the phenotypic expression of the genotype. This process can be viewed as a mechanism for incorporating prior knowledge in terms of genetically determined specifications and constraints relevant to the developing brain – evolutionary learning at the phylogenetic level. In this view, the genotype can loosely be likened to a compressed file, an evolutionary memory, which in some sense is uncompressed, transformed, and compiled piecewise into executables that are executed during development.

In order to effectively solve real world learning problems it is typically necessary to incorporate relevant prior structure (i.e., prior knowledge) in the functional architecture. This is a general and well-accepted insight from work in formal learning theory (e.g., Jain et al., 1999) as well as in statistical learning theory (e.g., Cherkassky & Mulier, 1998; Geman, Bienenstock, & Doursat, 1992; Haykin, 1998; Vapnik, 1998). The extent to which prior information is invoked in an explanatory scheme has the effect of shifting the explanatory burden from ontogenetic development and learning to phylogenetic adaptation,

which of course requires its own explanation. On the other hand, prior innate structure can significantly reduce the complexity of the acquisition problem by constraining the available model space M (for a given cognitive component, cf., section 1.5), thus alleviating the extent of the search problem that the child is confronted with in order to converge on an appropriate final model $m_C[F]$. Chomsky's hypothesis of the existence of a universal competence grammar, determined by the language acquisition device and the initial state of the individual's language faculty (Chomsky, 1965, 1986) is an example of this (cf., section 1.5).

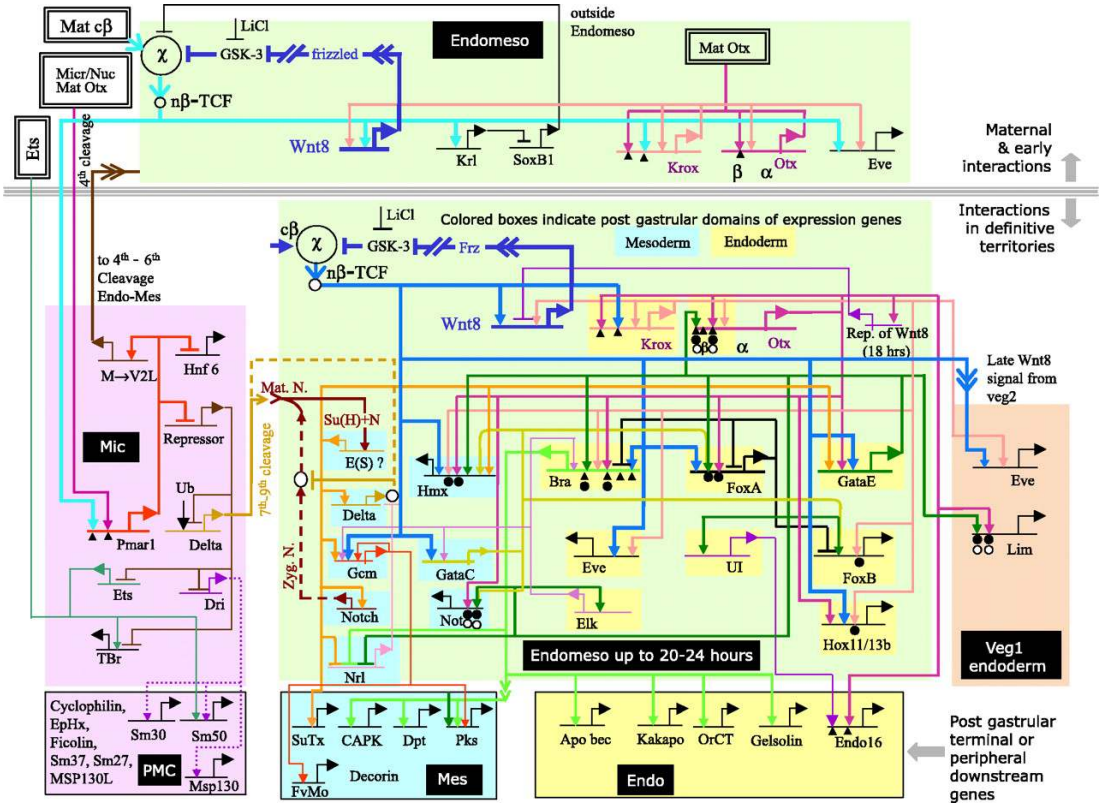
It was recently suggested that individual ontogeny is hierarchically organized within an open developmental system and that developmental phenomena need to be investigated by jointly considering interactions between endogenous and exogenous processes at various levels (Li, 2003). Developmental processes, operating on an ontogenetic time-scale, are commonly viewed as processes that depend on the interaction between genetic and environmental/experiential factors (Bota, Dong, & Swanson, 2003; Davidson et al., 2002; Elman et al., 1996; Johnson, 1997; Karmiloff-Smith, 1993, 1994). For example, it has been argued on empirical as well as theoretical grounds that prior genetic structure and its effects in terms of brain organizations impart reflect a socio-cultural influence on evolution (for a recent review see Li, 2003). For example, the suggestion that the biological evolution of brain encephalization was, in part, driven by the increase of social group size and the emergence of language as a more efficient means for handling complex social interactions (cf. e.g., Dunbar, 1998, 2003). On the other hand, brain encephalization might have been a factor in allowing for larger social group sizes in the first place. This precedence ambiguity provides an example of the fact that the relation between cognition and evolution might profitably be conceptualized in terms of co-evolution.

Recent advances in systems biology indicate that heritable developmental programs are regulated by sophisticated multi-component mechanisms with environmental interfaces (e.g., maternal molecules of regulatory significance: Figure 2.1). It is conceivable that there also exist purely genetically specified/driven maturational processes (cf., Davidson et al., 2002). However, the idea of purely experientially dependent developmental processes is probably not well-conceptualized, since all adaptive neural mechanisms have to be instantiated in genetically specified neural processes and hardware at a sufficiently level of

detail – in simplistic terms, the system does not learn the acquisition mechanism, but uses the acquisition mechanism to learn and develop. Rather, as a first order approximation, the relevant question appears to be to what extent the outcome of the developmental process is genetically constrained. However, this is a far too coarse way of posing the question, because what we want to understand is how the genetically specified information and experience enters into the developmental process. A stated objective for systems neuroscience, is to investigate the relationship between gene networks and brain networks as well as their role in the emergence of cognition and behavior (cf., Bota et al., 2003). It is not unlikely that genetic and experiential factors interact through out most of the life time of a neural system, although the degree of plasticity of the system appears to decrease over time. For example, it is conceivable that the over-expression of neurons and synapses as well as neuronal cell death and synaptic pruning (Hutterlocher, 1990), are related to activity-dependent competitive and selectionist principles driven in part by experience (Quartz & Sejnowski, 1997). With respect to the nervous system, one may suggest that there is a genetically guided basic outline of the brain architecture, which represents the development of an effective infrastructure for information processing, development, and learning – the unfolding of prior structure implicitly represented and regulated by the relevant parts of the genotype.

In order to see how we can conceive of learning and memory, which properties are needed for an information processing system to be able to learn and adapt in a non-stationary environment, we will start by outline a theoretical perspective on learning and memory as fundamental brain functions. Learning, here in the general sense of adaptation, can be defined as the processes by which the brain functionally restructures its processing networks and/or its representations of information as a function of experience. The stored information (i.e., the memory trace) can then be viewed as the resulting changes of the processing system. From this perspective, learning in a neural network is the dynamic consequence of information processing and network plasticity. Furthermore, fundamentally, it appears that processing and memory are co-localized at a microscopic level. Thus a neural processing system with the capacity to learn requires that aspects of its organization is capable to change in the light of experience, that is, some aspects of the neural infrastructure are adaptable. For simplicity, let us call these aspects adaptable

learning parameters. Thus, when the system interacts with its environment, the system undergoes some type of change as a result of information processing and these changes are captured by the adaptable parameters.



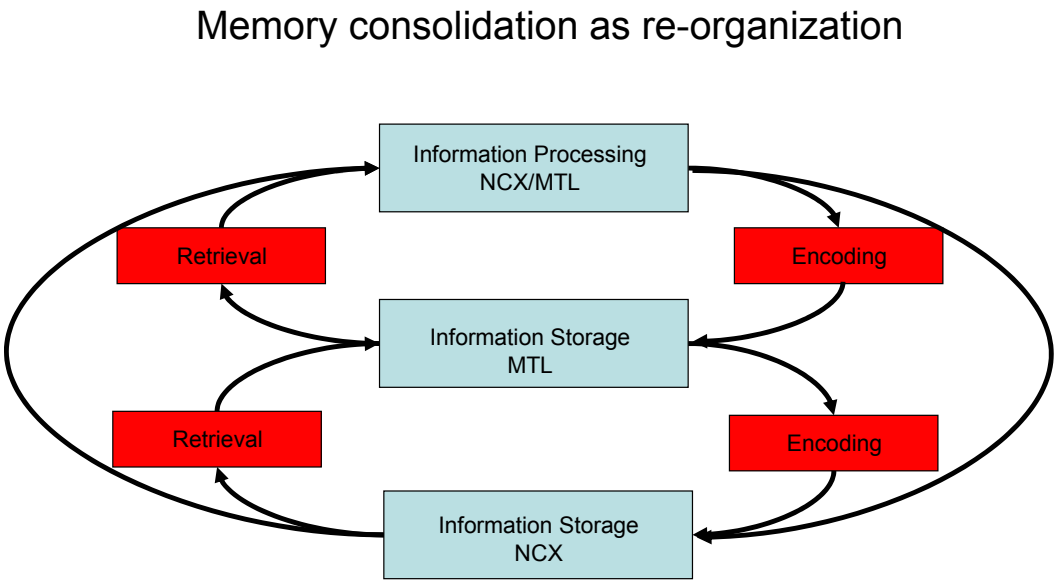
[Figure 2.1] Regulatory gene network. Recent advances in systems biology indicate that heritable developmental programs are regulated by sophisticated multi-component mechanisms. Emergent complex high-level phenomena necessarily presuppose interaction between systems constituents. Developmental processes, operating on an ontogenetic time-scale, depend on the interaction between genetic and environmental/experiential factors. Recent attempts to understand complex biological systems at a systems-level in terms of interacting components and their interface properties have adopted perspectives from control theory and reversed engineering (cf., Csete & Doyle, 2002; Kitano, 2002). The figure is an example of a regulatory gene network for endomesoderm specification (cf., Davidson et al., 2002). With permission (esandler@aaas.org).

Since different acquisition or learning problems are not equivalent in the sense of task requirements, it seems likely that different acquisition problems are supported by different adaptation/learning processes (Figure 1.6). Thus it has been suggested on both theoretical and empirical grounds that the brain is equipped with multiple memory systems (e.g., Eichenbaum & Cohen, 2001; Schacter & Tulving, 1994; Squire, Knowlton, & Musen, 1993; Stadler & Frensch, 1998). Furthermore, it has been suggested that learning related changes are prevalent in most (if not all) brain systems (Eichenbaum & Cohen, 2001). These memory systems serve different purposes and store different types of information with different spatio-temporal characteristics.

Memory can be decomposed into several processing stages, including on-line encoding, memory formation and storage, consolidation, re-organization and maintenance, as well as retrieval. Memory failure can be attributed to any of these stages, including encoding and memory formation failure, storage and maintenance failure as well as retrieval failure (i.e., failure to reactivate stored information based on the retrieval cues present). As an aside, we might ask what could possibly be the relevance of forgetting. Here we make a distinction between memory failure and forgetting. It may be the case that some aspects of forgetting are necessary side consequences of being able to create an effective 'knowledge base' of experience and general information. Forgetting processes thus might allow the system to restructure its 'knowledge base' in such a way that only relevant aspects of the information are kept, increasing the efficiency of the system in terms of content and retrievability. There is a creative element in this type of forgetting processes, in the sense that they support the emergence of generalized representation – akin to a semantization of episodic memory. Examples of neurophysiological mechanisms hypothesized to support learning and long-term memory are long-term potentiation and long-term depression (e.g., Bear & Kirkwood, 1993; Bliss & Collingridge, 1993). Other examples of short-term adaptable mechanisms are post-spike adaptation, habituation, and short-term potentiation (cf., Johnston & Wu, 1995; Koch, 1999). As noted in chapter 1, human learning and adaptive brain processes operate at many characteristic time-scales, spanning some seven or nine orders of magnitude. Newell (1990) has conceptualized this in terms of bands of cognition and there are serious attempts underway in applied cognitive

science to make use of large parts of this spectrum of time-scales in cognitive modeling (Anderson, 2002). Here we conclude that different storage systems operate at different time-scales and show different forgetting characteristics.

A closely related notion, and an independent rational for the existence of multiple memory systems, is related to the serial learning problem, also called the stability-plasticity dilemma (Grossberg, 1988; Haykin, 1998). The dilemma relates to the problem of up-dating the knowledge base in the light of novel information to be stored and integrated with previously acquired information.



[Figure 2.2] Memory consolidation as re-organization. David Marr (1971) was first to suggest that the medial temporal lobe (MTL), which is characterized by rapid and extensive plasticity, creates pointers of incoming information from the neocortex (NCX) for rapid storage in the hippocampus. These pointers are then thought to participate in the reorganization, integration, and consolidation of neocortical representations. In line with this suggestion, Squire and colleagues (2004) suggest that the neocortex interacts with the medial temporal lobe in order to establish (formation), store and maintain, as well as

retrieve long-term information. It is suggested that this form of memory, declarative memory, ultimately becomes independent of the medial temporal lobe through the process of consolidation. This hypothesis, then, represents one example of the idea of a processing system with multiple interacting memory systems operating at several different characteristic time-scales.

Here there is a trade-off between stability and plasticity: stability is necessary to ensure robust process reliability but its very nature precludes any possibility for adaptation or learning from experience; on the other hand, plasticity is necessary for the acquisition of new information as well as flexible adaptation to a changing environment; however, this also introduces instability that might cause reliability problems; in other words, the processing system becomes non-robust. For example, too much plasticity might allow that important previously acquired information gets overwritten or lost due to excessive changes imposed by the newly acquired information, so-called catastrophic interference. One suggested solution to the serial learning problem is to first store information in a memory structure suitable for rapid acquisition, and subsequently consolidate this in an integrated fashion in a different storage system, thus reducing the risk of catastrophic interference while familiar or less relevant information is allowed to be forgotten (e.g., McClelland, McNaughton, & O'Reilly, 1995; Petersson et al., 1999a; Robins, 1995, 1996a, 1996b). More generally, the serial learning problem can be effectively handled by a processing system with multiple interacting memory systems operating at several different characteristic time-scales (Figure 2.2). Given that the stability-plasticity trade-off is appropriately handled, it is clear that information processing systems with adaptive properties and learning capacities can act with greater flexibility in a non-stationary environment and adapt to present and near future conditions within a time frame that is relevant to the individual. Development and the acquisition of cognitive skills depend on prior knowledge (in the sense of Chomsky, 1986) as well as fine tuning of the appropriate network architecture. It remains a challenge to disentangle development from learning, and perhaps it is more fruitful to view these as two interacting processes determining the developmental trajectory as well as the outcome of development.

In neural network models, the processing of active representations and learning/adaptive processes are commonly modeled as two (several) sets of dynamical variables. Typically, one set with rapid (millisecond time-scale) and the other with slower dynamics (depending on the characteristic time-scale of the learning process modeled). These sets of dynamical variables represent, on the one hand, the active on-going processing of information (representational dynamics), and on the other, the learning dynamics. For example, in a simple network model, learning and adaptation can be modeled as synaptic dynamics, but more complex models are of course possible. Formally, the relation between the representational dynamics and the learning dynamics is captured by the notion of coupled (interactive) dynamical systems (cf., section 1.5 and Figure 1.6). Ultimately, it may perhaps become necessary to formulate such a description in terms of for example coupled systems of stochastic differential/difference equations, corresponding to representational and learning dynamics at several different time-scales (cf., section 2.1.2. Note that more general formulations in terms semi-groups/semi-dynamical systems are also possible, see e.g., Lasota & Mackey, 1994b). Simulations of biologically plausible network models with adaptive characteristics represent initial attempts in this direction (Amit, 1998; Arbib, 2003; Gerstner & Kistler, 2002; Trappenberg, 2002).

In contrast to simple information storage, it is also possible to view learning and adaptation, as a process of generalization or a process of sequential estimation in a stochastic environment (cf. e.g., Anthony & Bartlett, 1999; Arbib, 2003; Cherkassky & Mulier, 1998; Haykin, 1998; Hertz et al., 1991; Vapnik, 1998). Here, generalization can be viewed as an instance of the model selection problem (i.e., the problem of learning to generalize from a limited (finite) amount of noisy data), a fundamentally complex and difficult problem. We argue, as have many others, that unconstrained model space generalization is not possible. As noted in the introduction of this chapter, statistical and formal learning theory show that it is important to incorporate relevant prior knowledge in the functional architecture of the learning system in order to ensure an effective acquisition capacity. An equivalent way of stating this is that the accessible model space (cf., section 1.5) has to be constrained in order for appropriate generalization to occur. It is of course important that the functional architecture reflects appropriate prior information in order for the learning system to be efficient (Geman et al., 1992). This is closely related to the so-

called bias-variance dilemma (Geman et al., 1992; Haykin, 1998). In appendix A2.1 we provide a general formulation, give a detailed mathematical treatment, and suggest some relevant interpretations for cognitive neuroscience setting of this important result.

In the context of neural networks, prior information can be incorporated in different ways, including specifics related to structural aspects of the network (e.g., network topology and parameters related to the computational units), the structure of the learning mechanisms, and the structure of the initial condition of the system. This has the consequence of imposing constraints on the accessible model space. In the case of simple network models, this translates into constraints on the accessible manifold in the space of learning parameters (e.g., the space of synaptic parameters).

2.2 LEARNING PARADIGMS – DIFFERENT WAYS OF INTERACTING WITH THE ENVIRONMENT

Different memory systems may require different learning modes and in this sub-section we will briefly outline the most common conceptualizations of various learning paradigms; in other words, different ways of interacting with the environment. Learning by instruction, often called supervised learning, presupposes a rich source of external feedback - a teacher. An example of supervised learning is the error-based learning paradigm in which detailed directional information is provided and utilized (e.g., various gradient descent approaches like error back-propagation) by the learning system to improve performance (Arbib, 2003; Haykin, 1998). A weaker form of environmental interaction, which also is dependent on external feedback, is reinforcement learning. Reinforcement learning is a kind of trial-and-error based learning. The reinforcement paradigm thus implies that the system learns, through trial-and-error interaction with the environment, to gradually select appropriate actions by being provided external feedback in the form of reward signals. This can be conceptualized as a marked random search through the available model space and introduces an important trade-off. This trade-off has been called the exploration-exploitation dilemma: how should the learning system allocate its temporal resources, given a finite life-time; with respect to exploration: how much time should the system spend attempting improve its model of the environment with the objective of optimizing exploitation opportunities (i.e., performance); with respect to exploitation: how much time

should be devoted to utilizing what has already been acquired in order to achieve the primary objectives of the learning system in the first place (cf., Haykin, 1998; Sutton & Barto, 1998). Thus, reinforcement learning represents a type of guided learning where positive and/or negative feedback is provided based on the outcome of an action. However no detailed directional information is provided, as is the case in the supervised learning paradigm, in the sense that the learning system is not instructed how to change its internal workings but only an evaluation of whether a certain choice of action was in some sense appropriate or not (cf., Sutton & Barto, 1998). In contrast, the correct response to a given stimulus is provided by the teacher in the simplest versions of the supervised learning paradigm. Finally, learning and adaptation can take place without any external feedback, so-called unsupervised or self-organized learning (Arbib, 2003; Haykin, 1998). This basically implies that the outcome of the acquisition process is determined by the interaction between the input experienced and the prior structure as well as properties of the learning system. For example, a self-organized learning process may structure a neural network to represent the type of environmental structure it encounters and the adaptive process is sensitive to (e.g., correlation structures in the environment, if the system is sensitive and can adapt with respect to this type of structure, cf., Rieke et al., 1996). Another example of self-organized learning is based on internal monitoring, that is, monitoring of system performance based on internal measures of error, or more generally, on internal measures of consistency. These internal measures can be used to improve internal representations or processing pathways. For example, one part of the nervous system may monitor another part and provide adequate teaching feedback. It is important to realize that primary monitoring mechanisms and internal measures of consistency presupposes prior information (i.e., the system comes with these properties in place when embarking on the acquisition task; they are not acquired but a priori given), while secondary monitoring mechanisms can be acquired as for example in predictive model based processing. In predictive model based processing, the learning system constructs (estimates) an internal model of relevant aspects of the input. This acquired forward model can subsequently be used to generate model dependent predictions or expectations. These predictions can be compared with the actual outcome and entails the possibility of an unsupervised learning framework in which internally generated error information (=

difference[input, prediction]; i.e., the input represents the 'outcome') drives the learning process (cf. e.g., Duda et al., 2001; Haykin, 1998). Simple examples of this are predictive adaptive time series models and various predictive learning schemes for recurrent neural networks.

2.3 INTERACTIVE STOCHASTIC DYNAMICS – LEARNING AND ADAPTATION IN INFORMATION PROCESSING SYSTEMS

The simple dynamical systems framework outlined in section 1.6.2 of chapter 1, which generalized the classical framework described in section 1.4, can easily be adapted to incorporate the idea of learning and development, thus generalizing the classical view summarized in section 1.5. In this section we will provide a very general conceptualization of cognitive systems with capacity to learn and develop. In appendix A2.2 we provide an illustration of this framework by analyzing and in detail work through a simple concrete example a continuous-time analog dynamical system: the Bayesian confidence propagation network (Sandberg, Lansner, Petersson, & Ekeberg, 2002). As noted in chapter 1, several non-standard models of information processing have recently been outlined (for recent reviews see e.g., Siegelmann, 1999; Siegelmann & Fishman, 1998). For example, one way to generalize the Church-Turing framework of computability is to employ analog instead of discrete representations (Siegelmann, 1999). Another is to use parameterized models in combination with adaptive dynamics. As an aside, note that the universal Turing machine U can be viewed as parameterized by the 'program number' p in a von Neumann type architecture and thus incorporating all realizable Turing machines R (cf., Davis et al., 1994). More specifically, suppose the Turing machine R corresponds to the program number p, then the outcome of R computing on the input i is given by, $R(i) = U(p,i)$.

Building on results which show that it is possible to embed Turing machines in discrete-time analog dynamical systems on 2-dimensional compact manifolds (Moore, 1991a, 1991b), Siegelmann and Sontag (1994) have shown that it is possible to implement Turing machines in discrete-time recurrent networks with rational synaptic weights. This generalizes the ground breaking work of McCulloch and Pitts (1943; see also, Minsky, 1967), who showed that the class of networks of thresholding units is equivalent to the class of finite-state machines. Now, it is well known that Turing machines can be

implemented as a finite-state machine coupled to two stack memories. Siegelmann and Sontag (1994) took advantage of the fact that stack memories as well as Turing tapes, as indicated by for example Moore (1991a; 1991b), can be simulated in rational arithmetic with piecewise affine transformations. Thus it was possible to show that discrete-time recurrent networks have computational processing power that depends on, among other things, the type of numbers utilized as synaptic weights: it turns out to be the case that natural-, rational-, and real numbers corresponds precisely to networks that are computationally equivalent to the finite-state, Turing, and super-Turing models of processing. Moreover, a large class of discrete-time dynamical systems do not have greater processing capacities than the discrete-time analog recurrent network architecture (Siegelmann, 1999). However, the dependence on infinite precision processing implies that these capacities generally are sensitive to system noise. Importantly, there appears to be several brain internal noise sources (e.g., Gerstner & Kistler, 2002; Koch, 1999; Rieke et al., 1996). Now, it seems clear that any reasonable analog model of a brain system will have a state-space in the form of a compact manifold (i.e., closed and bounded, cf., Dudley, 2002). Here the mathematical property of compactness represents the natural generalization of finiteness in the classical framework (cf., section 1.5). Moreover, finite precision computations or realistic noise levels would have the effect of coarse graining the state-space, thus effectively discretizing the state-space into a finite number of 'voxels' of roughly equivalent states. This follows from the compactness property. It thus appears that even if we model a brain system as an analog dynamical system, this would behave (approximately) as a finite state analogue (Petersson, 2004, in press). Under the additional assumption of finitely available processing time, the same conclusions follow in the case of continuous-time evolution of state variables if finite temporal precision is assumed. Similar results hold under the assumption of finitely available processing time, and the same reasoning applies, even if one introduces continuous temporal evolution of state variables and finite temporal precision or realistic temporal noise is assumed.

Returning to the conceptualization of cognitive learning systems, which generally can be framed as the interaction between a state space dynamics and a learning dynamics (cf., section 1.5). Here an information processing system C with adaptive properties is specified as an ordered triplet, C = <functional architecture, representational dynamics,

learning dynamics>; the *functional architecture* is a specification of the structural organization of the systems; for example, an architectural outline related to weak functional modularity subserving integrative interactive processing and incorporating different sorts of constraints representing prior structure (cf., the discussion above, section 1.5, and appendix A2.1). The *representational dynamics* includes a specification of a state space, Ω, of state variables, s, carrying/representing information ($s \in \Omega$; e.g., membrane potentials), and dynamical principles, T (i.e., $T:\Omega x M x \Sigma \rightarrow \Omega$), governing the active processing information in state space; the active representational dynamics is commonly conceived of as taking place on a rapid (short) characteristic time-scale.

Similarly, the *learning dynamics* includes a specification of learning (adaptive) variables/parameters, m (e.g., synaptic parameters), for information storage (memory formation) and dynamical principles, L (i.e., a 'learning algorithm'; e.g., co-occurrence or covariance based Hebbian learning) governing the temporal evolution of the learning variables in the model space M ($m \in M$). The temporal evolution of the adaptive parameters depends on the active processing of information and the learning dynamics is commonly conceived of as taking place on a slower (longer) characteristic time-scale than that of the representational dynamics. In order to be more explicit, this can for example be formulated within the framework of stochastic differential/difference equation (e.g., Øksendal, 2000), here with additive noise $\xi(t)$ and $\eta(t)$:

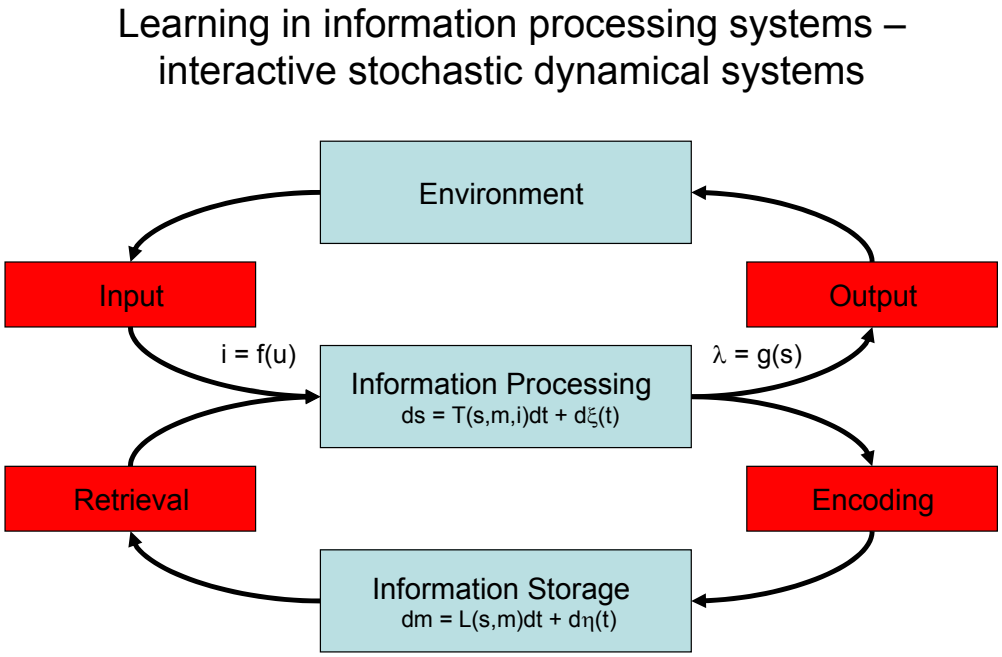
$$ds = T(s,m,i)dt + d\xi(t) \qquad [3]$$

$$dm = L(s,m)dt + d\eta(t) \qquad [4]$$

where i is the input representation the system receives (i.e., $i = f(u)$) and the output λ is a function of s (i.e., $\lambda = g(s)$), see Figure 2.3. As in the classic cognitive framework (cf., equation [1], section 1.4, and equations [1'] and [2'] of section 1.5), these equations determines trajectories in state space $s = s(t)$; the temporal evolution of s as the system receives input $i = f(u(t))$ and generates output trajectories $\lambda = g(s(t))$. In addition, the system traces a trajectory $m = m(t)$ in the model space; in the present case, the space of learning parameters. Thus information processing and learning can be formulated as a system of

coupled equation as illustrated by [3] and [4], and it is clear that learning represents a dynamical consequence of information processing and system plasticity (Petersson et al., 1997). This outline can easily be elaborated to include classes of adaptive parameters operating at different characteristic time-scales (Figure 2.4) as well as parameters describing developmental processes. In short, developmental systems can also be modeled as a coupled dynamical system representing processing as well as learning and development. It is also clear that this view represents a generalization of the classical view on learning and development (cf., section 1.5).

General dynamical system theory (level 2 in the sense of section 1.6.2) is in some sense (obviously) too rich as a framework for formulating explicit models of cognitive brain function. For example, it turns out that for any given state space one can find a universal dynamical system whose traces (a kind of non-linear projection) will generate any possible trajectory in the state space (for the continuous case cf. e.g., Lasota & Mackey, 1994a). Thus, what is needed is a specification of cognitively relevant constraints and processing principles (level 1 constraints in the sense of section 1.6.2) as well as constraints and processing principles relevant for the neurobiological networks subserving information processing in the brain (level 3 constraints in the sense of section 1.6.2).



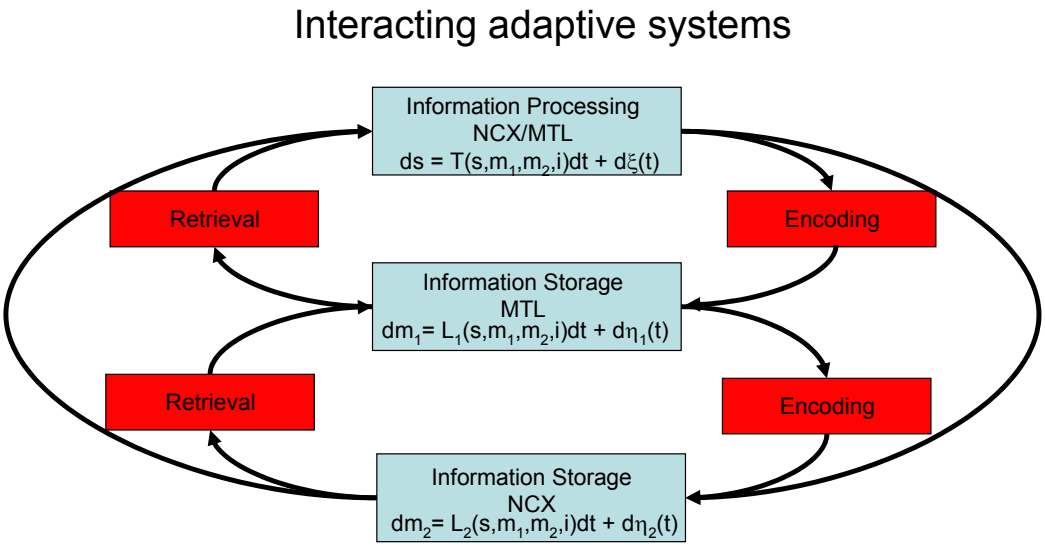
$$ds = T(s,m,i)dt + d\xi(t) \qquad [3]$$

$$dm = L(s,m)dt + d\eta(t) \qquad [4]$$

[Figure 2.3] Learning and adaptation in information processing systems. A cognitive processing system C with adaptive properties is specified as an ordered triplet, C = <functional architecture, representational dynamics, learning dynamics>. The representational dynamics corresponds to equation [3], while the learning dynamics corresponds to equation [4]. These equations represent a system of coupled stochastic differential/difference equation, which allows information processing to interact with the learning dynamics. For example, equation [3] and [4] can be related to the interaction between the perception-cognition-action and encoding-storage-retrieval cycle (Figure 1.6), where [3] is related to the active processing of information in short-term working memory and [4] is related to the encoding-retrieval cycle.

Any real progress on this front would represent a significant generalization of Chomsky's concept of knowledge and competence (Chomsky, 1965, 1986, 2000b). An important set of constraint comes from the requirements of tractable processing, that is, our models of

cognition has to be physically implementable in brain tissue and perform within given limits of real-time processing and short-term and long-term memory capacities, assuming we in important respects are dealing with a finite system.



$$ds = T(s, m_1, m_2, i)dt + d\xi(t)$$

$$dm_1 = L(s, m_1, m_2, i)dt + d\eta_1(t)$$

$$dm_2 = L(s, m_1, m_2, i)dt + d\eta_2(t)$$

[Figure 2.4] Interacting adaptive systems. The functional architecture of the brain is specified by its structural organization, here exemplified by the neocortex (NCX) and the medial temporal lobe (MTL). The representational dynamics subserves active on-line information processing. Two different learning systems are represented by two different sets of adaptive variables, m_1 and m_2, for information storage (memory formation). For example, the two systems could represent short-term memory and long-term memory. Alternatively, in line with ideas related to memory consolidation as re-organization (Figure 2.2) the neocortex interacts with the medial temporal lobe in order to establish and retrieve declarative information. It has been suggested that this form of memory ultimately becomes independent of the medial temporal lobe through the process of consolidation (Squire, 1992). These examples represent two examples of the idea of a processing system with

multiple interacting memory systems, which operate at different characteristic time-scales (Petersson, 2004).

On the final note, the existence of universal dynamical systems, which can emulate any state space dynamics, suggests another possibility. In general, these universal systems are infinite dimensional. Now, Vapnik's support vector machine approach takes advantage of the fact mapping data non-linearly into a high-dimensional space typically has the consequence of making the data linearly separable and thus easier to learn (Vapnik, 1998). If the neural infrastructure can support dynamics of very high-dimensionality this might provide a clue to why human brains are able to learn and acquire such a rich spectrum of cognitive skills using what appears to be a surprisingly stereotypic network architecture at a microscopic level.

APPENDIX

A2.1 NOISE, ESTIMATION, AND APPROXIMATION ERRORS – SUGGESTED IMPLICATION FOR ADAPTABLE COGNITIVE SYSTEMS

For a comprehensive background to the mathematical concepts, tools, and their properties that will be used in this appendix, consult for example the work of Billingsley (1995) or Dudley (2002). The objective of this appendix is to derive the bias-variance trade-off for a very broad class of adaptable systems in a more general setting than is commonly done. We also suggest how this can be translated into the context of cognitive neuroscience, indicating the importance of prior structure in order to ensure effective learnability for an adaptable cognitive system faced with a complex acquisition tasks entailing generalization based on model selection. Here the prior structure is inherent to the adaptive system's accessible model space. From a neurobiological perspective, prior knowledge and information can be associated with the idea of an innately determined prior knowledge.

In order to get started, we first derive the generalized regression model. So, let (Ω, F, P) be a probability space and $Y:\Omega \rightarrow R^N \in L^1(\Omega, F, P)$ an integrable real random vector. Let $X:\Omega \rightarrow \chi \in M(\Omega, F, \chi, A)$ be a measurable random variable on some general measurable space (χ, A). We will be using the conditional expectation operator $E[Y|G]$, meaning the conditional expectation of Y with respect to the σ-algebra $G \subseteq F$ (i.e., the Radon-Nikodym derivative of $\nu(A) = \int_A Y(\omega)dP(\omega) = \int_A YdP, \forall A \in G$, with respect to the probability measure $P: G \rightarrow [0,1]$). In all cases, we will condition on the σ-algebra $\sigma(X)$ generated by a random variable X, and we indicate this by $E[Y|X]$, which exist since $Y \in L^1(\Omega, F, P)$. In fact, $E[Y|X]$ is a function of X (i.e., $\exists$ measurable $\beta:\chi \rightarrow R^N$ such that $E[Y|X] = \beta(X)$).

Define ε according to $\varepsilon = Y - E[Y|X]$. It follows from the linearity and tower properties of the conditional expectation operator that $E[\varepsilon|X] = E[Y - E[Y|X]|X] = E[Y|X] - E[E[Y|X]|X] = E[Y|X] - E[Y|X] = 0$, and in particular $E[\varepsilon] = E[E[\varepsilon|X]] = 0$. Let $\langle\cdot,\cdot\rangle:R^N \times R^N \rightarrow R$ denote the inner-product on R^N, then for any measurable $g: \chi \rightarrow R^N$ such that $\langle \varepsilon, g(X)\rangle \in L^1(\Omega, \mathcal{F}, P)$ we also have, by linearity and the tower property:

$$E[\langle \varepsilon, g(X)\rangle|X] = \langle g(X), E[\varepsilon|X]\rangle = 0 \text{ and } E[\langle \varepsilon, g(X)\rangle] = E[E[\langle \varepsilon, g(X)\rangle|X]] = 0. \quad [1]$$

According to above, $E[Y|X] = \beta(X)$ and thus we have arrived at the general regression model $Y = E[Y|X] + (Y - E[Y|X]) = \beta(X) + \varepsilon$, where $E[\varepsilon|X] = 0$. Conversely, let us assume a regression model for Y in relation to X, that is, $Y = f(X) + \varepsilon$, where $E[\varepsilon|X] = 0$, then it follows that $E[Y|X] = E[f(X)+\varepsilon|X] = f(X) + E[\varepsilon|X] = f(X)$; thus $f(X) = E[Y|X]$. Hence, given Y and X, the regression model for Y given X is uniquely determined by E[Y|X]. In the case $Y \in L^2(\Omega, F, P)$, turns out to be the orthogonal projection of Y on to the function space $\{Z \in L^2(\Omega, F, P) \mid \exists \text{ measurable } \phi:\chi \rightarrow R^N \text{ such that } Z = \phi(X)\}$.

Having derived and characterized the generalized regression model, we move on to derive the bias-variance trade-off. Suppose we have a model space defined by the function space $F:\chi x W \rightarrow R^N$, where W is the space of adaptive parameters $w \in W$ such that $\forall\, w \in W$, $F(\cdot, w):\chi \rightarrow R^N \in M(\chi, A)$ is measurable. Now, any learning process, which attempts to solve the generalization problem based on model selection, can be viewed as searching for or attempting to estimate a model $f: \chi \rightarrow R^N$ in the accessible model space $M = \{F(\cdot, w): \chi \rightarrow R^N \mid w \in W\}$. Suppose this estimation procedure is based on a finite measurable acquisition sample $T = \{T_1:\Omega \rightarrow T_1, \ldots, T_n:\Omega \rightarrow T_n\}$, where (T_k, S_k), $k = 1,\ldots,n$, are yet other measurable spaces. For example, $T_k = (X_k, Y_k)$ in the case of a supervised, or $T_k = X_k$ in the case of a self-organized (unsupervised) learning paradigm. Now, the learning process L (whatever its details) can be viewed as a measurable mapping from $T_1 x \ldots x T_n$ to W, which induces a mapping $W:\Omega \rightarrow W$, where the stochastic properties derives from the sample $\{T_1, \ldots, T_n\}$ and possibly other random sources, for example an additive noise source $\eta:\Omega \rightarrow W$ as in $W = L(T_1, \ldots, T_n) + \eta$, and addition is defined on W as is the case if for example $W = R^M$. W induces a probability distribution $\mu_W = PW^{-1}$ on the measurable space (W, V), that is, $\forall\, A \in V : \mu_W(A) = P(W^{-1}(A))$, where $W^{-1}(A) = \{\omega \in \Omega \mid W(\omega) \in A\}$. Assume in addition that (X, ε) is independent of W and that all relevant random variables/vectors belongs to $L^2(\Omega, F, P)$. Let $f(X) = E[Y|X]$ and consider the squared L^2-norm of $Y - F(X, W)$; that is, the averaged squared error of $F(X, W)$ as a model for Y, $E[||Y - F(X, W)||^2]$:

$$E[||Y - F(X, W)||^2] = E[||Y - f(X) + f(X) - F(X, W)||^2] =$$
$$= E[||Y - f(X)||^2 + ||f(X) - F(X, W)||^2 + 2\langle Y - f(X), f(X) - F(X, W)\rangle] =$$
$$= E[||Y - f(X)||^2] + E[||f(X) - F(X, W)||^2] + 2E[\langle Y - f(X), f(X) - F(X, W)\rangle] =$$

$= //$ where the last term = 0, see below $// = E[||\varepsilon||^2] + E[||f(X) - F(X, W)||^2]$. [2]

To show that $E[\langle Y - f(X), f(X) - F(X, W)\rangle] = 0$, let $g(X,W) = f(X) - F(X, W)$, and remember that $\varepsilon = Y - f(X)$, then:

$E[\langle Y - f(X), f(X) - F(X, W)\rangle] = E[\langle \varepsilon, g(X,W)\rangle] = \int_\Omega \langle \varepsilon, g(X, W)\rangle dP =$

$= \int_{Rx\ldots xRx\chi xW} \langle e, g(x, w)\rangle dP(\varepsilon,X,W)^{-1} = //$ (X, ε) and W are independent $// =$

$= \int_{Rx\ldots xRx\chi xW} \langle e, g(x, w)\rangle dP(\varepsilon,X)^{-1} dPW^{-1} = //$Fubini's theorem$// =$

$= \int_W \{\int_{Rx\ldots xRx\chi} \langle e, g(x, w)\rangle dP(\varepsilon,X)^{-1}\} dPW^{-1} = \int_W E[\langle \varepsilon, g(X, w)\rangle] dPW^{-1} =$

$= //$ according to [1], $E[\langle \varepsilon, g(X, w)\rangle] = 0, \forall\ w \in W // = 0.$

Thus, $E[||Y - F(X, W)||^2] = E[||\varepsilon||^2] + E[||f(X) - F(X, W)||^2]$. Furthermore,

$E[||f(X) - F(X, W)||^2] = \int_\Omega ||f(X) - F(X, W)||^2 dP = \int_{\chi xW} ||f(x) - F(x, w)||^2 dP(X,W)^{-1} =$

$= //$ X and W independent, Fubini's theorem $// =$

$= \int_\chi \{\int_W ||f(x) - F(x, w)||^2 dPW^{-1}\} dPX^{-1} =$

$= // \mu_X = PX^{-1}, \mu_W = PW^{-1}$, and define $E[F(x, W)] = \int_W F(x, w))d\mu_W // =$

$= \int_\chi \{\int_W ||f(x) - E[F(x, W)] + E[F(x, W)] - F(x, w)||^2 d\mu_W\} d\mu_X =$

$= \int_\chi \{\int_W ||f(x) - E[F(x, W)]||^2 d\mu_W\} d\mu_X + \int_\chi \{\int_W ||E[F(x, W)] - F(x, w)||^2 d\mu_W\} d\mu_X$

$+ 2 \int_\chi \{\int_W \langle f(x) - E[F(x, W)], E[F(x, W)] - F(x, w)\rangle d\mu_W\} d\mu_X =$

$= //\ \mu_W$ is a probability distribution on W $// =$

$= \int_\chi ||f(x) - E[F(x, W)]||^2 d\mu_X + \int_\chi \{\int_W ||E[F(x, W)] - F(x, w)||^2 d\mu_W\} d\mu_X$

$+ 2 \int_\chi \{\int_W \langle f(x) - E[F(x, W)], E[F(x, W)] - F(x, w)\rangle d\mu_W\} d\mu_X.$ [3]

Now, define the *bias* and *variance* terms B(x) and V(x) according to:

$B(x) = f(x) - E[F(x, W)],$

$V(x) = E[||F(x, W)] - E[F(x, W)]||^2] = \int_W ||F(x, w) - E[F(x, W)]||^2 d\mu_W.$

Then B(x) can be interpreted as the local approximation error (i.e., the approximation error at x) in F(·, w) averaged over the model space M of f(x) . V(x) is the estimation error or variance induced by the acquisition sample and the learning process. It follows from [3] that,

$E[||f(X) - F(X, W)||^2] = \int_\chi ||B(x)||^2 d\mu_X + \int_\chi V(x) d\mu_X +$

$+ 2\int_\chi \{\int_W \langle f(x) - E[F(x, W)], E[F(x, W)] - F(x, w)\rangle d\mu_W\} d\mu_X.$

The last term reduces to 0 according to:

$\int_\chi \{\int_W \langle f(x) - E[F(x, W)], E[F(x, W)] - F(x, w)\rangle d\mu_W\} d\mu_X =$

$= \int_{\chi} \langle f(x) - E[F(x, W)], E[F(x, W)] - \int_{W} F(x, w) d\mu_W \rangle d\mu_X =$

$= \int_{\chi} \langle f(x) - E[F(x, W)], E[F(x, W)] - E[F(x, W)] \rangle d\mu_X = 0.$

Thus, $E[||f(X) - F(X, W)||^2]$ is given by:

$$E[||f(X) - F(X, W)||^2] = \int_{\chi} ||B(x)||^2 d\mu_X + \int_{\chi} V(x) d\mu_X = E[||B(X)||^2] + E[V(X)]$$

and it follows from [2] and [3] that,

$E[||Y - F(X, W)||^2] = E[||\varepsilon||^2] + E[||f(X) - F(X, W)||^2] =$

$= E[||\varepsilon||^2] + E[||B(X)||^2] + E[V(X)].$

Hence, there are three contributions to the L^2-norm of $Y - F(X, W)$:

1/ The *regression variance* or *noise* $E[||\varepsilon||^2] = \int_{\Omega} ||\varepsilon(\omega)||^2 dP(\omega)$, inherent in the regression model

$Y = f(X) + \varepsilon$, that is, inherent environmental noise.

2/ The *average approximation error* or *bias* $E[||B(X)||^2] = \int_{\Omega} ||B(X(\omega))||^2 dP(\omega) =$

$= \int_{\chi} ||f(x) - E[F(x, W)]||^2 d\mu_X = \int_{\chi} ||f(x) - \int_{W} F(x, w)) d\mu_W||^2 d\mu_X$, due to an inherently biased model space $M = \{F(\cdot, w) \mid w \in W\}$.

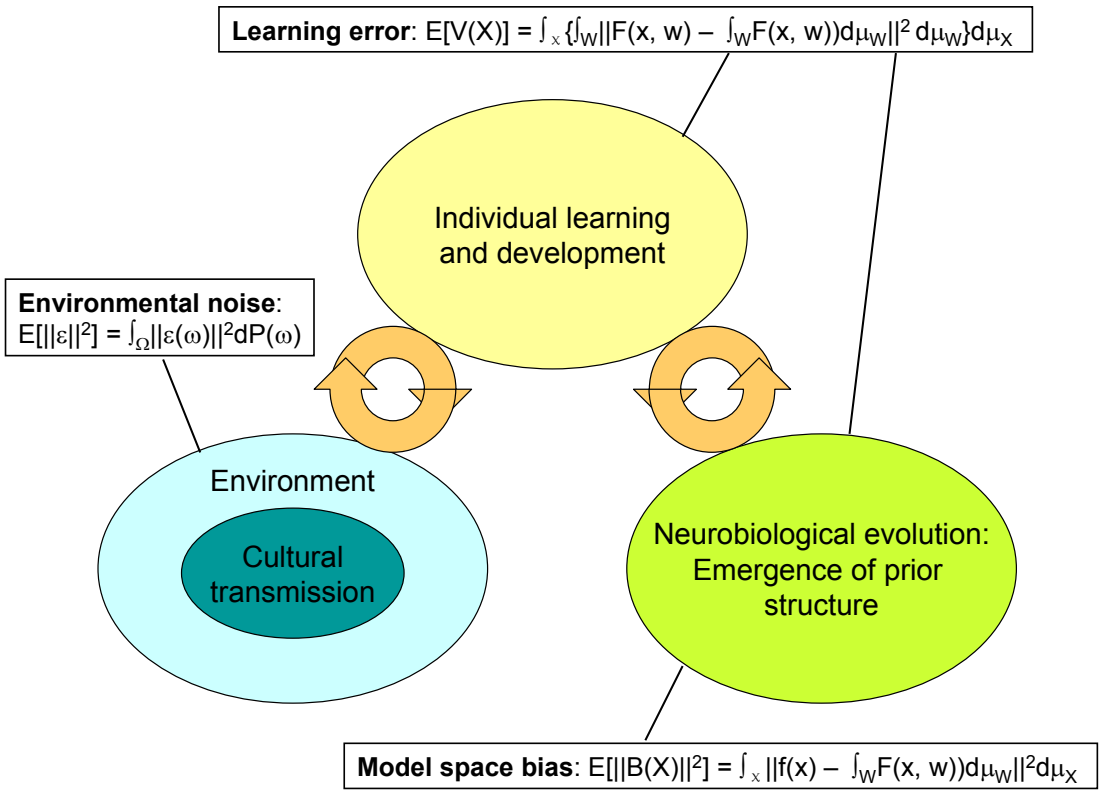
3/ The *average estimation error* or *variance* $E[V(X)] = \int_{\Omega} V(X(\omega)) dP(\omega) =$

$= \int_{\chi} \{\int_{W} ||F(x, w) - E[F(x, W)]||^2 d\mu_W\} d\mu_X =$

$= \int_{\chi} \{\int_{W} ||F(x, w) - \int_{W} F(x, w)) d\mu_W||^2 d\mu_W\} d\mu_X,$

which is induced by the acquisition sample and the learning process.

We conclude that there are three fundamental sources contributing to the lack of efficiency of an adaptive system in acquiring a proper generalization capacity: environmental noise, model space bias, and learning (estimation related) error (Figure A1). To achieve high acquisition efficiency it is necessary that the contribution from each of these sources is small. In order to reduce the model space bias term $E[||B(X)||^2]$, it is necessary to increase the expressive capacity of the model space, that is, to increase the set of accessible models $M = \{F(\cdot, w) \mid w \in W\}$. One way to achieve this is to increase the dimensionality of M, which implies that the number of adaptable parameters has to be increased. Given a fixed acquisition set $T = \{T_1: \Omega \rightarrow T_1, \ldots, T_n: \Omega \rightarrow T_n\}$, this typically implies that the variance term $E[V(X)]$ increases (cf. e.g., Haykin, 1998; Vapnik, 1998). A possible way to circumvent this is to increase the size of the acquisition set T and the time complexity of the learning problem in order to keep the overall error $E[||f(X) - F(X, W)||^2]$ under control.



[Figure A1] The Bias-variance trade-off. The overall performance of a learning system depends on three factors: (1) the inherent noise ε in the environment transmitted through the input data to the learning system; (2) the inherent bias B of the accessible model space (i.e., the average approximation error inherent to the learning system); and (3) the variance V induced by the acquisition sample and randomness inherent to the learning process as such (i.e., the average estimation error). In general, one might suggest that the proper prior model space bias is determined, at least partly, by innately specified factors, while the learning error is dependent on individual learning and development as well as innate factors specifying the acquisition mechanism.

Thus, generally increased acquisition efficiency by reducing the model space bias comes at the price of increased acquisition complexity. However, an alternative strategy to reduce the overall error is to incorporate relevant prior structure in the acquisition mechanism itself or into the structure of the accessible model space M, a so-called bias-reducing strategy.

The latter option ensures that there exist accessible models $F(\cdot, w):\chi \rightarrow R^N$ in M from the start that are guaranteed to approximate the acquisition problem $f:\chi \rightarrow R^N$ well and that these models are accessible to the learning process, given the properties that can be expected of a 'typical' acquisition set T.

In general, a 'proper' bias of the accessible model space has to be 'designed' specifically for each learning problem. From a neurobiological perspective, prior knowledge can be interpreted as an innately determined structure. Thus, for time and space restricted learning problems (e.g., a limited finite acquisition set T), the bias-variance trade-off strongly indicates the necessity of innately determined structure in order for a learning system, operating under complexity constraints, to acquire real complex skills or knowledge (cf., Gold, 1967; Jain et al., 1999; Nowak et al., 2002; Vapnik, 1998). Whether this prior structure is domain specific or not is in principle a different issue. However, given the specificity requirements of built in prior knowledge or bias for specific acquisition tasks, it will come as no surprise if parts of the prior knowledge turns out to be domain specific. This would seem to be the simplest 'solution' from an evolutionary perspective. The most prominent example of this line of thought is reflected in Chomsky's (1986) suggestion for the natural language domain that not only is prior innate constraints necessary but these prior constraints represents a linguistically specific competence in the form of a specifically structured initial state of the faculty of language and a specific language acquisition device. In summary, in order to succeed effectively on complex learning tasks, it seems necessary for a learning system to incorporate prior structure/knowledge in its accessible model space and in its learning mechanism. Given the complexity of many acquisition tasks confronting the human brain, we conclude that this is also the case for the brain (for further discussion of these issues see Petersson, 2004, in press; Petersson et al., 2004 and the references therein).

A2.2 THE BAYESIAN CONFIDENCE PROPAGATION NETWORK

In this appendix we will outline and work through a concrete example of the interactive dynamical systems framework for adaptive systems outlined in chapter 2. To recapitulate, information processing systems with adaptive properties were specified as ordered triplets,

C = ⟨functional architecture, representational dynamics, learning dynamics⟩ and we arrived at equations [3] and [4], here re-stated for convenience:

$$ds = T(s, m, i)dt + d\xi(t) \quad [1]$$

$$dm = L(s, m)dt + d\eta(t) \quad [2]$$

where $i = f(u)$ is the input and the output λ is a function of s (i.e., $\lambda = g(s)$). This framework can be viewed as a formalization of the interaction between the perception-action and the encoding-retrieval cycles (Figure 1.6 and 2.4). Here we will illustrate how the abstract formulation in [1] and [2] can be mapped on to a simple concrete example, the so-called Bayesian confidence propagation (BCP) network (Sandberg et al., 2002). The BCP network is an example of a continuous-time analog recurrent network.

In general, it is essential for a capacity limited real world learning system to give priority to the retention of relevant information that is appropriate to its operational objectives. In a non-stationary environment, the time of acquisition is one indicator of relevance. Thus a real-time on-line learning system with capacity limits needs to gradually forget old information in order to avoid catastrophic interference. This can be achieved by allowing new information to overwrite old. Memory systems with this property are called palimpsest memories. If the environment is non-stationary, it is generally important to give priority to more recently acquired information (note that this may take place at several different time-scales).

Auto-associative artificial neural networks (ANNs), for example McCulloch-Pitts associative memories and Hopfield networks, have been proposed as models for biological associative memory (cf., Arbib, 2003; Hopfield, 1982; McCulloch & Pitts, 1943; Minsky, 1967; Trappenberg, 2002). These represent one way of formalizing Donald Hebb's original ideas of synaptic plasticity and emerging cell assemblies (Hebb, 1949). Simulations have indicated that networks of cortical pyramidal and basket cells can operate as attractor networks (e.g., Fransen & Lansner, 1998). However, the standard correlation based learning rule for attractor ANN suffer from catastrophic interference, that is, all memories are lost as the system reaches a critical memory limit and becomes overloaded. Nadal et al. (1986) proposed the marginalist-learning paradigm as a way to handle the situation. The basic idea is to control the acquisition intensity and tune it to the level of crosstalk-noise (i.e., the correlation between memories). This has the consequence that the most recently

acquired information is more stable compared to older information; new patterns are stored on top of older, which gradually become overwritten and finally inaccessible. Another learning procedure with smooth forgetting characteristics is learning within bounds (Hopfield, 1982). This reduces the storage capacity compared to the standard Hopfield Hebbian-type learning rule in order to achieve long term memory stability.

A learning rule for attractor networks derived from Bayes' theorem (cf. e.g., Billingsley, 1995; Duda et al., 2001) was developed by Lansner and Ekeberg (1989), which represents a Hebbian-type learning process that reinforces connections between simultaneously active units and weakens or makes connections inhibitory between anti-correlated units. This learning process is based on a probabilistic view of learning and retrieval, with input and output unit activities representing confidence of feature detection and posterior probabilities of outcomes, respectively. The synaptic strengths are based on the probabilities of the units firing together, estimated by counting occurrences in the acquisition data. This procedure yields symmetric learning weights and thus allows for fixed-point attractor dynamics. It also generates a balance between excitation and inhibition, avoiding the need for external means of threshold regulation. We have described a modification of the Bayesian learning rule in order to achieve a real-time on-line learning system with palimpsest memory properties (for mathematical details, properties, and simulation results cf., Sandberg, Lansner, & Petersson, 2001; Sandberg, Lansner, Petersson, & Ekeberg, 2000; Sandberg et al., 2002). This incremental learning process is based on moving averages and the forgetting rate is controlled by a time constant. The BCP neural network with the incremental version of the Bayesian learning rule shows palimpsest memory properties and avoids catastrophic forgetting. It has a capacity dependent on the learning time constant and exhibits decreasing convergence speed for increasingly older information.

In the context of BCP networks, the functional architecture represents a specification of the types of neurons, making up the network, as well as their processing properties. The network consists of N neurons $i \in \{1, \ldots, N\}$. A neuron i first transforms the input with an affine transformation according to:

$$u_i = \Sigma_j \omega_{ij} s_j + \beta_i$$

where input s_j is weighted according to the synaptic parameters ω_{ij}, and β_i represent the

bias. The transfer function of the neuron i, is given by a truncated exponential: $\varphi(u_i) = \exp[u_i]$, when $u_i \leq 0$,and = 1, when $u_i > 0$. Thus,

$$s_j = \varphi(u_i) = \varphi(\Sigma_j \omega_{ij} s_j + \beta_i).$$

The structural organization of the network is determined by its connectivity matrix $[c_{ij}]_{NxN}$, which is also reflected in the weight matrix $[\omega_{ij}]_{NxN}$. The connectivity matrix, where $c_{ij} = 1$, if there is a connection from neuron j to neuron i, and = 0, otherwise, determines which computational nodes interact. In other words, the connectivity matrix determines the possible patterns of computational interaction or information flow in the network. In the present case, no self-interaction is allowed so $c_{ii} = 0$.

The representational dynamics includes a specification of the neuronal state variables (s, or alternatively, the 'membrane potentials' u) and dynamical principles governing information processing, T. The state space of the BCP network is N-dimensional and the dynamical variables of the state-space dynamics $s_i = s_i(t)$ represent the mean firing rate over some appropriate time-scale. The N-dimensional representational dynamics T can be broken down into its component form $T = [T_1, \ldots, T_i, \ldots, T_N]$ and is given by:

$$T_i = T_i(s, \omega, \beta) = \varphi(\beta_i + \Sigma_j \omega_{ij} s_j) - s_i.$$

The learning parameters ω and β are functions of an underlying set of adaptive parameters a and b, respectively. The components of $\omega = \omega(a, b)$ and $\beta = \beta(b)$ are given by: $\omega_{ij}(a_{ij}, b_i, b_j) = \log[a_{ij}/b_i b_j]$, and $\beta_i(b_i) = \log[b_i]$; in other words, $T = T(s, a, b) = T(s, \omega(a, b), \beta(b))$. Note how (a, b), or alternatively (ω, β), corresponds to m in equations [1] and [2]. Similarly, the learning dynamics includes a specification of learning (adaptive) variables/parameters, m, which here corresponds to $\omega = \omega(a, b)$ and $\beta = \beta(b)$, where a and b are the adaptive parameters, as well as dynamical principles determining the learning process, L. The temporal evolution of the adaptive parameters depends on the active processing of information carried by s. The learning dynamics L = L(s, a, b) of the BCP network operates in a N^2-dimensional space of learning parameters (i.e., model space); and broken down into component form $L = [L_1, L_{12}, \ldots, L_i, \ldots, L_{ij}, \ldots, L_{NN-1}]$, L is given by:

$$L_i = L_i(s, a, b) = [(1 - \lambda_0)s_i + \lambda_0] - b_i$$

and

$$L_{ij} = L_{ij}(s, a, b) = [(1 - (\lambda_0)^2)s_i + (\lambda_0)^2] - a_{ij}$$

where $0 < \lambda_0 << 1$ is a small positive constant which is necessary to introduce for technical reasons in order to avoid divergence problems of the logarithm in the neighborhood of 0. A processing interpretation of this constant is possible, in which λ_0 represent the averaged background of low-level network activity, in the absence of any input to the network (Sandberg et al., 2002). Thus the learning dynamics represent a form of learning within lower bounds. For a detailed outline of these ideas, the heuristic mathematical derivations of the BCP network and its learning process, see Sandberg et al. (2002).

In the final analysis of the BCP network, this yields a deterministic interactive dynamical system in which the representational and the adaptive dynamical variables interact according to:

$$\tau_C \cdot ds/dt = T(s, \omega(a, b), \beta(b))$$

$$\tau_L \cdot d[a, b]/dt = L(s, a, b)$$

where the τ_C is the 'membrane constant' of the processing units while τ_L determines the relevant time-scale for learning and forgetting. We define the learning rate as $\alpha_L = 1/\tau_L$ and set $\alpha_C = 1/\tau_C$. If we also include additive noise $\xi(t)$ and $\eta(t)$, generalizing somewhat the framework outlined in Sandberg et al. (2002), we end up with a stochastic dynamical system of the form:

$$\tau_C \cdot ds/dt = T(s, \omega(a, b), \beta(b)) + \sigma(t, \varepsilon)d\xi(t, \varepsilon) \quad [3]$$

$$\tau_L \cdot d[a, b]/dt = L(s, a, b) + \upsilon(t, \varepsilon)d\eta(t, \varepsilon) \quad [4]$$

where $\xi: RxE \rightarrow R^N$ (i.e., $\xi = \xi(t, \varepsilon)$, $t \in R$, $\varepsilon \in E$) is a normalized N-dimensional Ito process with zero mean and unit variance-covariance matrix (i.e., $E[\xi(t)] = 0$ and $Var[\xi_i(t)] = 1$), and a stochastic variance $\sigma = \sigma(t, \varepsilon)$, on a probability space (E, F, P). Similarly, $\eta: Rx E \rightarrow (R^N)^2$ is a normalized N^2-dimensional Ito process, while $\upsilon = \upsilon(t, \varepsilon)$ is a stochastic variance. In general we will leave the argument $\varepsilon \in E$ implicit in the following. Note that the form of [3] and [4] corresponds exactly to that of [1] and [2]. In detail, [3] and [4] thus takes the following form, for $i, j \in \{1, \ldots, N\}$, $i \neq j$:

$$ds_i/dt = \alpha_C \cdot T_i(s, \omega, \beta) = \alpha_C[\varphi(\beta_i + \Sigma_j \omega_{ij} s_j) - s_i] + \alpha_C \sigma_i(t) d\xi_i/dt \quad [5]$$

$$da_{ij}/dt = \alpha_L \cdot L_{ij}(s, a, b) = \alpha_L\{[(1 - (\lambda_0)^2)s_i s_j + (\lambda_0)^2] - a_{ij}\} + \alpha_L \upsilon_{ij}(t) d\eta_{ij}/dt \quad [6]$$

$$db_i/dt = \alpha_L \cdot L_i(s, a, b) = \alpha_L\{[(1 - \lambda_0)s_i + \lambda_0] - b_i\} + \alpha_L \upsilon_i(t) d\eta_i/dt. \quad [7]$$

If we temporarily departure from the on-line continuous perspective on learning and instead

take a batch perspective (i.e., keeping ω and β fixed in [5] while adapting a and b in [6] and [7] for a given time interval and subsequently updating ω and β at the end of this interval), it turns out that it is possible to explicitly integrate the system described by [6] and [7]. In order to do this we note that both systems of equations [6] and [7] have the form:

$$dF = \alpha\{(1 - c)\cdot s(t) + c] - F(t)\}dt + \sigma(t)d\xi(t).$$

Now, let $\theta(t) = (1 - c)\cdot s(t) + c$, then $dF/dt = \alpha\theta(t) - \alpha F(t) + \sigma(t)d\xi(t)$. Here, we can generalize the situation slightly and allow a time-varying $\alpha = \alpha(t)$. Thus, we have the general situation:

$$dF = \alpha(t)\theta(t) - \alpha(t)F(t)dt + \sigma(t)d\xi. \qquad [8]$$

In order to integrate [8] we introducing an integrating factor $G(t) = \exp[g(t)]$, where function $g = g(t)$ is defined according to: $g(t) = \int_{[0,\ t]} \alpha(\rho)d\rho \Rightarrow dg/dt = \alpha(t)$. Now, multiplying F(t) with the integrating factor G(t) and then taking the temporal derivative we arrive at:

$$d[GF] = (dG/dt)F(t)dt + G(t)dF = \qquad [9]$$

$$= //\ dG/dt = d\{\exp[g(t)]\}/dt = \exp[g(t)]\cdot dg/dt = \alpha(t)\exp[g(t)]\cdot = \alpha(t)G(t)\ //\ =$$

$$= \alpha(t)G(t)F(t)dt + G(t)dF = \alpha(t)G(t)F(t) + G(t)[\alpha(t)\theta(t) - \alpha(t)F(t)]dt + \sigma(t)d\xi =$$

$$= \alpha(t)\theta(t)G(t) + \sigma(t)d\xi/dt. \qquad [10]$$

Thus, by integrating [9] and [10], we arrive at:

$$G(t)F(t) = C + \int_{[0,\ t]} (d[GF]/d\rho)d\rho + \int_{[0,\ t]} \sigma(\rho)d\xi(\rho) = C + \int_{[0,\ t]} \alpha(\rho)\theta(\rho)G(\rho)d\rho$$

$$+ \int_{[0,\ t]} \sigma(\rho)d\xi(\rho), \text{ and } G(t) = \exp[g(t)] \text{ implies that:}$$

$$F(t) = C\cdot\exp[-g(t)] + \exp[-g(t)]\cdot\{\int_{[0,\ t]} \alpha(\rho)\theta(\rho)\exp[g(\rho)]d\rho + \int_{[0,\ t]} \sigma(\rho)d\xi(\rho)\} =$$

$$= C\cdot\exp[-g(t)] + \int_{[0,\ t]} \alpha(\rho)\theta(\rho)\exp[g(\rho) - g(t)]d\rho + \int_{[0,\ t]} \exp[-g(t)]\sigma(\rho)d\xi(\rho).$$

Now, if we assume that $\xi = \xi(t, \varepsilon)$ is a normalized Brownian motion and that the variance is non-random, that is, $\sigma(t, \varepsilon) = \sigma(t)$, then the noise term can be integrated by parts according to:

$$\int_{[0,\ t]} \exp[-g(t)]\sigma(\rho)d\xi(\rho) = \exp[-g(t)]\cdot\{\sigma(t)\xi(t, \varepsilon) - \int_{[0,\ t]} \xi(\rho, \varepsilon)d\sigma(\rho)\},$$

(for details see e.g. Øksendal (2000), theorem 4.1.5).

Furthermore, identifying $C = F(0)$ and defining an integration kernel $\chi(t, \rho)$ according to $\chi(t, \rho) = \alpha(\rho)\exp[g(\rho) - g(t)]$, if $-\infty < \rho \leq t$, and $= 0$, if $\rho > t$, we arrive at an explicit expression for F(t):

$$F(t, \varepsilon) = F(0)\cdot\exp[-g(t)] + \int_{[0, t]} \alpha(\rho)\theta(\rho)\exp[g(\rho) - g(t)]d\rho$$
$$+ \exp[-g(t)]\cdot\int_{[0, t]} \sigma(\rho)d\xi(\rho, \varepsilon) = F(0)\exp[-g(t)] + \int_R \theta(\rho)\chi(t, \rho)d\rho$$
$$+ \exp[-g(t)]\cdot\{\sigma(t)\xi(t, \varepsilon) - \int_{[0, t]} \xi(\rho, \varepsilon)d\sigma(\rho)\} \qquad [11]$$

The expression for $F(t, \varepsilon)$ in [11] includes a deterministic part $D(t) = F(0)\exp[-g(t)] + \int_R \theta(\rho)\chi(t, \rho)d\rho$ as well as a stochastic (non-deterministic) part $S(t, \varepsilon) = \exp[-g((t))]\{\sigma(t)\xi(t, \varepsilon) - \int_{[0, t]} \xi(\rho, \varepsilon)d\sigma(\rho)\}$. In short, within the batch-learning framework, $F(t, \varepsilon)= D(t) + S(t, \varepsilon)$, and this expression can be used to arrive at explicit expressions for a_{ij} and b_i in the simple case of a constant $\alpha(t) = \alpha_L$: $g(t) = \int_{[0, t]} \alpha(\rho)d\rho = \int_{[0, t]} \alpha_L d\rho = \alpha_L\cdot t$, and thus $\exp[g(\rho) - g(t)] = \exp[\alpha_L(\rho - t)] = \exp[-\alpha_L(t - \rho)]$. It follows that $\chi(t, \rho) = \alpha(\rho)\exp[g(\rho) - g(t)] = \alpha_L\exp[-\alpha_L(t - \rho)]$ becomes a time-invariant convolution kernel. In other words, $\chi(t, \rho)$ acts like a linear time-invariant filter of exponential decay. Now, given the network activity generated by the input from the environment, $\theta_{ij}(t) = [(1 - (\lambda_0)^2)s_i s_j + (\lambda_0)^2]$ and $\theta_i(t) = \alpha_L[(1 - \lambda_0)s_i + \lambda_0]$, respectively:

$$a_{ij}(t) = a_{ij}(0) + \alpha_L\cdot\int_{[0, t]} [(1 - (\lambda_0)^2)s_i(\rho)s_j(\rho)+ (\lambda_0)^2]\exp[-\alpha_L(t - \rho)]d\rho$$
$$+ \alpha_L\exp[-\alpha_L\cdot t]\cdot\{\sigma_{ij}(t)\eta_{ij}(t, \varepsilon) - \int_{[0, t]} \eta_{ij}(\rho, \varepsilon)d\sigma_{ij}(\rho), \text{ and}$$
$$b_i(t) = b_i(0) + \alpha_L\cdot\int_{[0, t]} [(1 - \lambda_0)s_i(\rho) + (\lambda_0)^2]\exp[-\alpha_L(t - \rho)]d\rho$$
$$+ \alpha_L\exp[-\alpha_L\cdot t])\{\sigma_i(t)\eta_i(t, \varepsilon) - \int_{[0, t]} \eta_i(\rho, \varepsilon)d\sigma_i(\rho).$$

Sandberg et al. (2002) also introduce a modular architecture within the BCP network framework in terms of so-called hyper-columns. This amounts to clustering the neurons in hyper-columns and imposing a weak prior structure in terms of a disjoint representation of abstract features and normalization of activity within hyper-columns according to: $S_{ik} = \varphi(u_{ik})/\Sigma_j \varphi(u_{jk})$; and this enters into equation [5].

Summing up, we have seen how the BCP network framework can be viewed as a particular example of the general framework specified by the equations [1] and [2]. The BCP network memory shows the palimpsest memory property and the time for the memory decay scales roughly as the learning time constant τ_L. The memory capacity increases linearly with τ_L up to a limit where it becomes equal to the standard counter BCP network. This means that the introduction of palimpsest properties has not reduced the maximal capacity as such. By setting the size of the network and the learning time constant the memory capacity can be regulated from a fast learning short-term working memory to a

slowly learning long-term memory (cf., Sandberg et al., 2001; Sandberg et al., 2000, 2002). Biological associative synaptic plasticity is generally assumed to be Hebbian-type in nature. This is also the case for the Bayesian-Hebbian-type BCP learning process outlined above. It exhibits a graded behavior with multiple synapse activations as well as a more step-wise behavior for single-synapse activation similar to experimental observations in LTP (Petersen, Malenka, Nicoll, & Hopfield, 1998). The BCP learning process displays both LTP- and LTD-like phenomena (cf. e.g., Artola & Singer, 1993; Bear & Kirkwood, 1993). Wahlgren and Lansner (2001) have shown that the Bayesian–Hebbian learning process, with some modifications, can provide a phenomenological model for synaptic long-term plasticity. Sandberg et al. (2001) indicate that when the learning rate is modulated by a relevance signal, the BCP network exhibit selective enhancement of the retrieval probability of the relevant information. This represent an example of a time-varying learning rate $\alpha_L = \alpha_L(t)$. Having a time-varying learning rate, that changes over time with the relevance of the information being processed, opens up for the possibility to control learning rate by various relevance or 'print-now' signals. This can be used to make the memory selective.

An alternative perspective on time-varying learning rates can also be taken. This can be related to our previous discussion emphasizing that different learning tasks are not equivalent and that the brain is equipped with multiple memory systems, storing different types of information of different spatio-temporal characteristics. As previously suggested, one the possible benefit of forgetting is that forgetting processes allows the system to restructure, integrate, and re-organize its knowledge base in such a way that only the relevant aspects of the information are preserved, thus increasing the efficiency of the system it terms of information content and retrievability (Figure 2.2 and 2.4). Hence, it is likely that the different storage systems operate on different time-scales and show different forgetting characteristics. In the BCP network, selecting a learning time constant sets a scale of temporal detail that the network is most sensitive to. The learning system will average out the events that occur at faster time-scales and adapt to slower changes. A rapidly adapting network would learn and remember actively represented information (short-term working memory), while a more slowly forgetting network might learn from single presentations, via working memory representations (episodic memory), and at an

even slower learning and forgetting rates, a memory system would average individual presentation events into a prototypic semantic memory. The BCP network equiped with several sets of adaptive parameters, operating at different characteristic time scales, can thus instantiate several forms of memory systems in the same network (cf., Figure 1.6 and 2.4).

3. METHODOLOGICAL BACKGROUND

To probe the mysteries of the human brain, cognitive neuroscience combine the experimental strategies of cognitive and experimental psychology with techniques that allows for detailed investigations of the brain activity that supports cognition. Functional neuroimaging methods provide experimental access to the living human brain and have developed rapidly during the last two decades. Hemodynamically based functional neuroimaging methods, such as positron emission tomography (PET) and functional magnetic resonance imaging (FMRI), are extensively used to investigate how neuronal processing correlates with changes in behavior or cognitive processing (Frackowiak, Friston, Frith, Dolan, & Mazziotta, 1997; Raichle, 1994; Toga, Mazziotta, & Frackowiak, 2000). A framework of well described theories and empirically validated methods are available (Frackowiak et al., 2004). The functional neuroimaging methods used differ in assumptions as well as the approximations employed. These need to be kept in mind for optimal use; of central importance is how well the empirical data fulfill these assumptions and approximations as well as the robustness of the methods used in analyzing the data. This notion emphasizes the importance of empirical validation, investigation of robustness, and the explicit characterization of the inherent limitations of a given method. nevertheless, the standard functional neuroimaging methods provide a useful means to investigate how networks of brain regions interact due to the whole brain coverage and the fact that primary data from different brain regions can be sampled on an approximately equal basis, at least with PET and to a lesser extent with FMRI (due to susceptibility artifacts and signal drop-out etc.). In addition, when the underlying assumptions and limitations are taken into account, the various standard approaches used generally serve their purposes well (Petersson, Nichols, Poline, & Holmes, 1999a, 1999b).

This chapter is largely base on the review by Petersson et al. (1999a; 1999b) in which the limitations of various functional neuroimaging methods were discussed. it is clear that these limitations are important to take into account when analyzing and interpreting functional neuroimaging data (Petersson, 1998). It should also be noted that the methods of the field are still developing rapidly (cf. e.g., Frackowiak et al., 2004; Friston, Glaser et al., 2002; Friston, Harrison, & Penny, 2003; Friston & Penny, 2003; Friston,

Penny et al., 2002; Genovese, Lazar, & Nichols, 2002; Ledberg, Fransson, Larsson, & Petersson, 2001).

3.1 THE COUPLING BETWEEN NEURAL ACTIVITY AND REGIONAL CEREBRAL BLOOD FLOW BLOOD

The neurophysiological basis of functional neuroimaging is the relatively tight and roughly linear coupling between the regional cerebral blood flow, the metabolic activity, and the neural activity as measured with electrophysiology (Gusnard & Raichle, 2001; Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Rees, Friston, & Koch, 2000; Scannell & Young, 1999; Siesjö, 1978). At rest the human brain consumes approximately 20% of the oxygen and metabolic supply needed by the body, although the brain accounts for only approximately 2% of the body mass. The oxygen is used in the oxidative metabolism of glucose to supply the brain with energy in the long-term (Raichle, 1998). Brief increases in neural activity of a given brain region implies that the energy and oxygen requirements in the given region increases and is accompanied by an increase in blood flow as well as glucose consumption that exceed the increase in oxygen consumption (Fox, Raichle, Mintun, & C., 1996). This means that the relationship between oxygen consumption and blood flow is not proportional. In a region of transient activity, the increase in glucose is partly broken down anaerobically by glycolysis despite of overcompensation in blood flow supply. As a result there is a lowered extraction fraction of oxygen that results in increased oxygen content in the blood nearby (Raichle, 2001). The robust empirical relationship between changes in brain activity and blood flow has been known for over a century (Raichle, 2001). The work of Logothetis and colleagues (Logothetis et al., 2001; Logothetis, Guggenberger, Pelea, & Pauls, 1999) indicates that a spatially restricted increase in the FMRI signal directly reflects an increase in neural activity. They recorded in parallel and correlated action potentials as well as local field potentials with the BOLD signal. Both these measures correlated with the observed BOLD signal, the local field potentials somewhat better than the action potentials. Local field potentials arise from the input as well as integrative processes within neurons. These findings are consistent with autoradiographic measurements of glucose consumption by different brain regions in rats

(Raichle, 2001). In his commentary, Raichle (2001) notes that the signal-to-noise ratios for neural signals recorded directly from the brain are much greater than the accompanying FMRI signal. This implies, for example, the absence of an FMRI signal does not necessarily imply that no information processing is going on in a particular brain region.

The energy turn-over in the brain is necessary as to maintain the ionic concentrations and membrane potentials at the appropriate levels. The dominant use of energy stems from maintenance of these membrane potentials. Neuronal signaling evokes ionic fluxes across membranes that need to be restored and most such fluxes are supported directly or indirectly by the Na/K-ATPase and other ionic pumps (Siesjö, 1978). Raichle (2001) suggests that the glutamate cycle act as a local driver for metabolism based on the abundance of glutamate as a transmittor. Thus the generation of ATP will be supplied anaerobically and lactate will be produced upon increases in neuronal work. This means that PET measures of regional cerebral blood flow will indicate proportionally higher increases in signal than the actual local increase in oxygen consumption. In FMRI, the sensitivity of the measured signal is based on the related increase in oxygenated blood locally. As noted, the BOLD signal correlates well with local field potentials and to a large extent these are generated in the postsynaptic dendritic component where large ionic fluxes that need restoration are generated from the neuronal input. There remains an unresolved question regarding the relationship between the type of brain activity and the signals measured with FMRI and PET. Are they related to excitation, inhibition, or both? The signal is a composite net-activity and it is difficult to imagine any regional activity that is not a mixture both excitatory and inhibitory components. However, given that inhibitory signals give rise to hyperpolarization and less ionic leakage post-synaptically, that the recorded signal might be more closely related to local excitatory activity (Shinohara, Dollinger, Brown, Rapoport, & Sokoloff, 1979).

3.2 PET ACQUISITION PROCEDURES

The functional neuroimaging data presented in chapter 6 were acquired with PET. In functional PET studies, a radioactive isotope (in the present case, [15-O]-butanol (Berridge, Cassidy, & Terris, 1990) or [15-O]-H2O (Fox & Mintun, 1989)) with a half-life of approximately 2 min is injected into the venous blood stream. Between 10 to 15 bolus

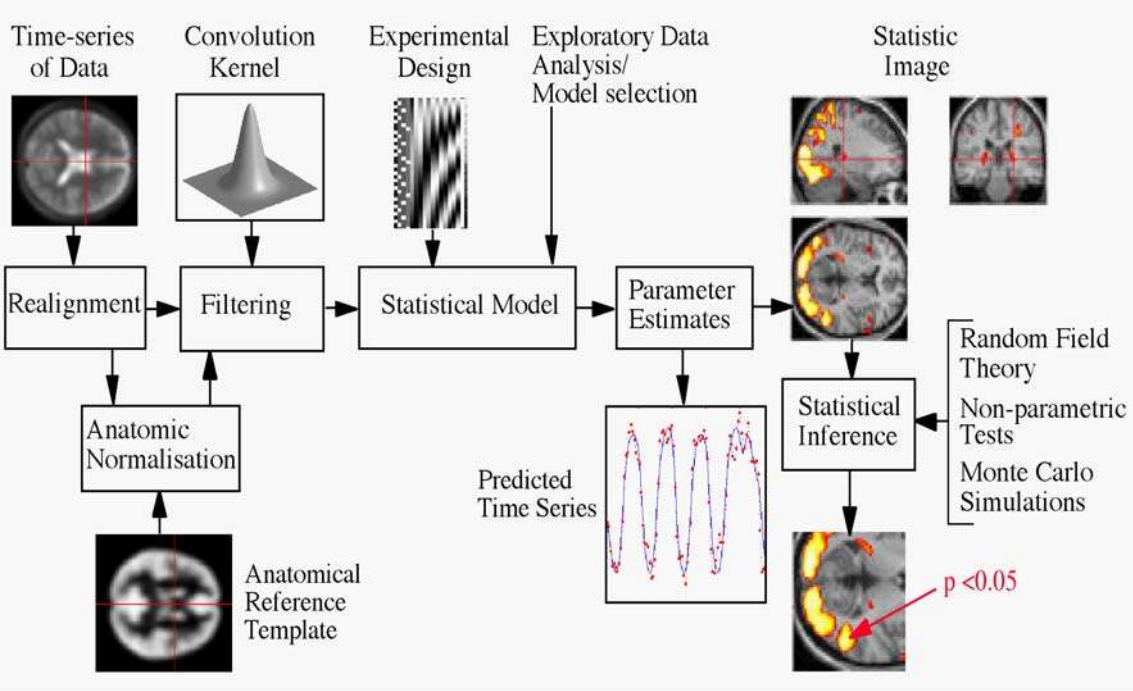
injections of 400-500 MBq (10-15 mCi) [15O]-butanol or [15O]-H2O were injected. In the basic PET scanner set-up used, the primary PET data acquisition started automatically when a predetermined activity threshold was passed upon bolus arrival to the brain. PET data was then acquired for the subsequent 40 s. In order for the injected radioactivity to return to the background level at least 5 x the half-life of [15-O] was allowed to pass (i.e., approximately 10 min between the PET scans). The tracer-molecules follow the blood flow and emit positrons that interact with electrons in their vicinity. Each positron-electron pair decays into two photons which then trace out a line-of-response in opposite direction. The photon radiation intensity along the different lines-of-response, determined by the field-of-view (i.e., the number and position of the detector rings of the PET scanner) depends on the tissue distribution of the tracer-molecules, which in turn is determined by the regional cerebral blood flow. The detector logic ensures that an appropriate estimate of the tissue distribution is acquired during the primary data acquisition. Various back-projection techniques can then be applied to the raw data to recover a 3D estimate of the tracer distribution and thereby an estimate of the regional cerebral blood flow. Because of the coupling between regional cerebral blood flow and neural activity, we arrive at an estimate of the underlying neural activity. In our case, the raw PET data (i.e., the sinograms) were reconstructed using the back-projection algorithms supplied by the CTI manufacturer for the 3D ECAT EXACT HR PET scanner (Wienhard et al., 1994) with a standard Hanning-filter setting (Hanning 5s) thus yielding the primary 3D PET data. Attenuation correction was routinely performed based on a transmission scan for each participant. Head movements were restrained by an individually fitted plastic helmet.

3.3 IMAGE PROCESSING AND STATISTICAL ANALYSIS

The primary PET data were then subjected to several steps of image- and statistical analysis which we will outline in the subsequent sections. Most functional neuroimaging studies are analyzed as group studies. Group investigations are conducted primarily because we are primarily interested in commonalities over participants, at this stage. In other words, we are interested in the effects that generalize at least to the group of subjects investigated. Another reason for conducting group investigations is to increase the signal-to-noise ratio

and thereby the statistical power of a given experiment. With respect to the particular studies outlined in chapter 6, we analyzed the PET data through out with the Karolinska Computerized Brain Atlas of Greitz (Greitz, Bohm, Holte, & Eriksson, 1991) as well as the SPM software (www.fil.ion.ucl.ac.uk).

During the last two decades a body of well described theories and empirically validated methods have been developed, providing a framework for investigating functional neuroimaging data and making scientific inferences based on statistical analysis. Statistical models make explicit as well as implicit assumptions about data. What is of importance in this context are not the assumptions or approximations per se but how well these are fulfilled by empirical data and the robustness of the methods used, when these assumptions or approximations are not fully met (Petersson et al., 1999a, 1999b).



[Figure 3.1] General outline of functional neuroimaging data analysis. The primary functional neuroimaging data are commonly pre-processed, that is, realigned, anatomically normalised, and spatially and temporally low-pass filtered; a statistical model for the data is

created; model parameters are subsequently estimated and a test statistic is chosen in order to conduct for statistical inference, taking into account the multiple non-independent comparisons.

Most functional neuroimaging methods are based on 3D voxel data, an approach pioneered by Fox and colleagues (Fox & Mintun, 1989; Fox, Mintun, Reiman, & Raichle, 1988) and Friston and colleagues (Friston et al., 1990; Friston, Frith, Liddle, & Frackowiak, 1991). The primary functional neuroimaging data are commonly pre-processed (e.g., realigned, anatomically normalised, and low-pass filtered), a statistical model and a test statistic is chosen, and model parameters estimated for statistical inference taking into account multiple non-independent comparisons and possible temporal autocorrelation (see Figure 3.1). In the following sections we will outline in some detail the different processing steps involved in analyzing functional neuroimaging data.

3.3.1 IMAGE PREPROCESSING

3.3.1.1 REALIGNMENT AND ANATOMICAL NORMALIZATION

In a functional neuroimaging study it is commonly the case that several measurements, a time-series of volumes (or 3d images), are acquired of a given participant in the different experimental conditions. For example, in our PET studies, 10 – 15 PET scans were acquired for each subject. The participants were asked to lie in a more or less comfortable position and as still as possible in the PET scanner during the experiment, in order to reduce head movement during the experiment. The head of the participant was fixated to the scanner with an individually fitted plastic helmet designed to minimize movement of the head in the scanner. Nevertheless, small head movements still occur (on the order of 1 – 3 mm). In order to compensate for this movement, the reconstructed PET images are automatically realigned (cf., Ashburner & Friston, 1997). The brain in each volume of the time-series of a given individual will thus occupy the same position in image space.

The brains of different individuals are anatomically different. A necessary requirement for group studies in functional neuroimaging is to represent data in a standardized anatomical space. This requires a method to transform, or warp, individual data into the standardized space, so-called anatomical normalization. Anatomical normalization aims to adjust for anatomical differences in order to allow data to be averaged across subjects. Anatomical normalization transforms the image time-series of the individual participant into a standardized anatomical space and in our studies we have commonly used the stereotactic space as defined by the SPM template (www.fil.ion.ucl.ac.uk), an approximate Talairach space (Talairach & Tournoux, 1988), and sometimes in combination with the Computerized Brain Atlas of Greitz (Greitz et al., 1991).

3.3.1.2 FUNCTIONAL-ANATOMICAL VARIABILITY AND SPATIAL FILTERING

There have been several attempts to assess the residual functional-anatomical variability after realignment and anatomical normalization in more or less low-pass filtered data. These attempts have often used the variability in location of the local maximum statistic (peak location). Several studies estimate inter-subject standard deviations of the peak coordinates to be on the order of 5-10 mm (Fox & Pardo, 1991; Hasnain, Fox, & Woldorff, 1998; Hunton et al., 1996; Ramsey et al., 1996). When, for example, PET data from different laboratories are compared, this variability increases (Poline, Vandenberghe, Holmes, Friston, & Frackowiak, 1996; Senda et al., 1998) and this indicates that activation foci that are less than 10 mm apart cannot always be reliably distinguished (Grabowski et al., 1996). Contrary to some claims in the literature, the intra-individual variability can also be significant, even for robust primary motor activations (Hunton et al., 1996).

The inter-individual residual variability in functional anatomy generally exhibits spatial structure and is dependent on the algorithm used for normalization. Simulation studies have indicated that a reduction of registration (realignment) error and a minimization of the residual anatomic variability can significantly improve the signal detection sensitivity (Worsley, Marrett, Neelin, & Evans, 1996b). In the presence of

residual functional-anatomical variability the effect of inter-subject averaging amounts to a spatial filtering or smoothing effect. Thus, if the spatial-scale of the filter matches the inherent scale of functional-anatomical variability in the population, no (or little) spatial information is lost. In general, using a voxel-based approach, it is important to reduce the impact of misregistration and inter-individual residual functional-anatomical variability. A common strategy is to low-pass filter, or smooth, the data either at reconstruction or with a suitably chosen convolution kernel (e.g., an isotropic 3D Gaussian kernel). Spatial filtering, which in effect is a local weighted averaging procedure, also increase the local equivalence of the voxel data across measurements and individuals and thus the validity of voxel-based statistical models.

Filtering data spatially may or may not increase the signal to noise ratio. This depends on the relation between the size and shape of the signal and the convolution kernel used. This relation between the signal size/shape and the characteristics of convolution kernel can understood in the light of the matched filter theorem (Rosenfeld & Kak, 1982). This theorem states that a signal in a background of white noise is detected with optimal sensitivity if a convolution kernel, which matches the size and shape of the signal, is used. It should be noted that the situation is slightly more complicated when the noise component is coloured (i.e., spatially autocorrelated). This is typically the case with hemodynamic functional neuroimaging data. Now, the result of the matched filter theorem serves as a good approximation, if the spatial extent of the signal is large compared to the extent of the autocorrelation. However if this is not the case, the choice of an optimal filter is more complicated. In this case the autocorrelation has to be taken into account when choosing the filter. This can be illustrated in the situation of stationary data (i.e., spatial homogeneity, or other words, translational invariant second moment characteristics of the probability distribution), which implies that the data can be whitened with a filter W. The matched filter theorem can then be applied to the whitened data generating a matched filter S. This is equivalent to directly apply the convolution of W and S, $W*S$, as a filter. We conclude then that the matched filter theorem suggests that the detection sensitivity is biased towards signals of similar spatial characteristics as the smoothing kernel.

In closing this section, which has mainly focused on functional-anatomical variability and signal detection in relation to image smoothing, it should be noted that the

arguments are general and also apply to signals in the temporal domain. In addition to outlining some complications related to these issues, we want to recapitulate that the objective of spatial filtering is mainly related to minimizing individual differences in residual functional-anatomy. It is also the case that spatial filtering typically increases the signal-to-noise ration, since the power of the residual spatial noise is usually dominant in the higher spatial frequencies. Moreover, convolving the data with a Gaussian kernel conditions the data to conform more closely to a Gaussian random field model (cf., section 3.3.3.3). One of the mechanisms behind this is the fact that filtering amounts to a weighted averaging and it follows heuristically from the central limit theorem of probability theory that random variables averaged in this way converge towards a Gaussian in distribution. In addition, it has recently been shown that as the projection counts in the PET reconstruction process (filtered back-projection) approach infinity the reconstructed images will become multivariate Gaussian distributed (Maitra, 1997). In our studies, we have generally filtered the data with a 3D isotropic Gaussian kernel of 14 mm full-width-at-half-maximum (FWHM).

3.3.2 STATISTICAL MODELING AND ESTIMATION

3.3.2.1 BASELINE FLUCTUATIONS AND GLOBAL NORMALIZATION

Functional neuroimaging experiments usually test hypotheses regarding regionally specific changes in neuronal activity. In the case of PET, these changes in neuronal activity are indirectly reflected in the associated changes in regional cerebral blood flow (rCBF) or regional cerebral counts (rCC), and in the case of BOLD FMRI, by changes in regional susceptibility. In the following, regional activity rA will represent rCBF, rCC, or regional FMRI BOLD signal, depending on the context.

For practical and other reasons, the imaging systems are commonly used in a non-quantitative mode. Therefore the focus is on relative regional changes which are then related to a baseline. This can be problematic since for example variability in global factors often induces baseline fluctuations. Different measures of global activity have been used to account for some of the baseline variability. An example of a simple estimator of global activity (gA) that has been used is the intra-cerebral average of regional activity rA. Global

activity defined in this way varies between subjects and over time. The variability depends on several variables (Frackowiak et al., 1997; McColl, Holmes, & Ford, 1994), for example, physiological (e.g., changes in pCO_2 levels and circulatory system changes), factors relating to the measurement procedures (e.g., differences in injected radioactive dose) and the imaging system (e.g., between-run variability in FMRI gain). Global changes are therefore difficult to interpret without quantification.

When there is a lack of absolute quantification and the experimentally induced regional changes are assessed relative to a baseline, changes in this baseline are often considered a nuisance effect. Since baseline fluctuations may be large, potentially hiding the effects of interest, it is necessary to account for or remove this variability in some appropriate manner. The notion of baseline variability as a nuisance effect implicitly assumes that the scan-to-scan baseline fluctuations are sufficiently independent of the experimental manipulations. In order to properly account for baseline variability there are two issues that need to be addressed: First, how to measure or estimate the baseline fluctuations, and second, how these measurements are used to explicitly model or remove the variability in baseline activity. Measurements of global effects, and consequently global normalization, are predicated on the assumption that the variability in global effects adequately represent the baseline fluctuations and that the experimentally induced regional changes are superimposed on this according to some model. Several approaches to account for global changes have been proposed and compared. For example, proportional scaling (Fox & Raichle, 1984; Kanno, Hatazawa, Shimosegawa, Ishii, & Fujita, 1996), log-linear regression models (Herholz, Kessler, Slansky, Mielke, & Heiss, 1993), histogram/rank equalization (Arndt, Cizadlo, O'Leary, Gold, & Andreasen, 1996), Z-score transformation of data (McIntosh et al., 1996), or modeled as a nuisance covariate in the general linear model (Friston et al., 1990). Both the ANCOVA (Friston, 1995; Friston et al., 1990; Ramsay et al., 1993) and the proportional scaling (Kanno et al., 1996) approaches have been empirically validated for PET data. In our studies we have consistently used the proportional scaling approach to global normalization.

The relation between rCBF and gCBF is most likely non-linear. However, over sufficiently constrained ranges this relation is well approximated by a linear model. For normal subjects and small ranges of gCBF, the incorporation of the gCBF as a covariate in

a linear model affords a reasonably good model of the relationship between rCBF and gCBF (Frackowiak et al., 1997). The additive ANCOVA model was proposed under the assumption that changes in gCBF and the experimentally induced changes in rCBF are well approximated as independent variables. However, it has been pointed out that the results of this approach may be problematic to interpret if changes in gCBF are correlated with the experimentally induced changes in rCBF (Ramsay et al., 1993; see also, Aguirre et al., 1998b and Andersson, 1997) or if the gCBF estimation is biased. This is also the case for proportional scaling.

The variability in gCC is often larger than in gCBF. Even if gCBF is relatively constant, subject differences in head fraction and variability in the introduced radioactive dose causes variability in gCC. In the case of count data, rCC is proportional to gCC when rCBF is constant. Therefore, if it can be expected that the variability in for example head fraction or introduced radioactive dose is dominant, proportional scaling is a reasonable choice. The empirical comparisons between various approaches, which have so far been performed, have yielded little differences between the various approaches to global normalization, both for PET data (Arndt et al., 1996; Frackowiak et al., 1997; Holmes, 1994; McIntosh et al., 1996) and FMRI data (Aguirre, Zarahn, & D'Esposito, 1998b). In other words, most functional neuroimaging studies on normal subjects yield similar results using either approach and thus robust results are roughly independent of the global normalization method chosen. However, when the global signal is significantly confounded with the experimental paradigm, it may be preferable in some situations to omit global normalization entirely and examine non-normalized changes (Aguirre, Zarahn, & D'Esposito, 1998a). An alternative strategy is to use more robust measures of gA; that is, attempting to estimate gA independent of the task induced changes in rA, in order to more accurately estimate the baseline fluctuations. One possibility is to examining brain regions known to be relatively unaffected by the experimental paradigm. An iterative solution to the latter suggestion has been proposed that successively eliminates voxels that indicate experimental effect from the set used to compute gA (Andersson, 1997). Moreover, the problem of estimating baseline fluctuations should be less complicated if closely matched activation and reference conditions are investigated. However, with increasing activation

differences between conditions, this may become a significant problem, emphasizing the need for carefully designed experiments that include active reference conditions.

3.3.2.2 THE CHOICE OF REFERENCE STATE

One principle that has emerged in cognitive neuroscience and functional neuroimaging is that of functional specialization or functional segregation. The idea of functional specialization rests on the hypothesis that different brain regions are specialized and implement different computations or operations on cognitive representations. This principle is reflected in the general linear modeling approach described in the section 3.3.2.3.

Functional neuroimaging data are typically analyzed in terms of a specific linear model, parameters are estimated, and subsequently various null-hypotheses tested. Under the assumption that the activation and reference conditions differ in some relevant specific aspect of cognitive processing, the locations of statistically significant differences in signal between conditions presumably define brain regions that are related to this difference. This approach is crucially dependent on an adequate choice of conditions to compare. Given an activation condition, the functional map will in general vary with the choice of reference condition. For example, in a simple subtraction, only parts of the underlying functional network may be observed, since common components activated to a similar degree will not be observed. Furthermore, results obtained with the subtraction approach can only be interpreted as relative differences since, at present, a canonical reference state or baseline condition seems difficult to define. This introduces a complication in the interpretation of functional maps, an ambiguity that is fundamental to the activation approach. In principle, a relative increase in rCBF in condition A compared to condition B might in relation to a third condition C represent an activation in condition A, a deactivation in condition B or a combination of both. The same holds for relative decreases in rCBF. It follows that, the formulation and specification of the reference condition is an important and difficult issue in functional neuroimaging. In general, the appropriate choice of the reference condition(s) is an issue that must be addressed at the design stage of a particular functional neuroimaging experiment. However, given an activation condition of interest, it is still an open question what is to be considered an appropriate reference condition(s). This depends crucially on the objectives of a given experiment, that is, the questions that the

experimental data are supposed to address. The central issue here is whether the chosen reference condition is well defined, in some suitable sense, to fulfill the objectives of a given functional neuroimaging experiment. One possible way forward is to use several reference conditions. This strategy allows for multiple perspectives on a given activation condition. For example, it is possible to use both closely matched control conditions and so-called low-level control conditions. A closely matched control condition differs ideally only in a single aspect from the condition of interest and can be used to test for specific effects. Instead a low-level control condition, for example rest with eyes closed or visual fixation, can be used to detect many, most, or all of the brain areas involved in a given task. Thus the simultaneous use of both closely matched and low-level control conditions can provide complementary information.

A further complication in the interpretation of results from functional neuroimaging studies based on the subtraction of functional images is whether a given brain function is well approximated as a linear additive decomposition into functional components, or, rather under which circumstances or comparisons they may be. In general, as indicated in chapter 1 and 2, brain functions cannot be expected to be linearly composed of a set of components. Rather, one can expect that complex brain functions are the result of non-linear interactions between components, that is, the regional brain activity associated with complex behavior may not be a sum of postulated constituents. If the regional brain activity associated with complex behavior is not the sum of apparent constituents, then the interpretation of results from the subtraction approach (in particular the compound hierarchical subtraction approach) is difficult and may depend strongly on the choice of experimental component tasks. This is particularly problematic if there is no canonical way of decomposing an overall task into components. This suggests that it may be necessary to develop new approaches that explicitly address the fact that brain functions emerge from non-linear interactions between components. Initial steps in this direction have been taken, as illustrated by various network approaches like structural equations modeling and dynamical causal modeling (Friston et al., 2003; McIntosh & Gonzalez-Lima, 1994, cf., section 3.4). In this context it should also be noted that it is only possible to detect quantitative differences with the activation approach. Qualitative differences in information processing that are not accompanied by quantitative changes will not be detected. However, the

complementary perspective represented by the network approach can in principle detect qualitative differences in the pattern of interactions between brain regions without any changes in mean activity (cf., sub-section 3.4). For further reflections on these issues see Ingvar and Petersson (2000).

3.3.2.3 THE GENERAL LINEAR MODEL

The general linear model (GLM) is a framework that encompasses all basic univariate models, including the ANOVA/ANCOVA and the multiple regression models. In the GLM framework n observations from a single image voxel are represented as column vector of length n, Y; the p effects and predictor variables are represented as p column vectors also of length n, forming an n x p matrix X called the design matrix. The fixed regression parameters are represented as a column vector β of length p; the residual random error is written as the column vector ε of length n. With the assumption of mean zero, independent and identically distributed error of magnitude σ^2, the concise representation of the GLM is:

$$E(Y) = X\beta \text{ and } Var(Y) = \sigma^2 I,$$

where I is the n × n identity matrix. Note that we have made no specific distributional assumptions; the usual normality assumption is only needed for statistical inference. Using only the general assumptions above, according to the Gauss-Markov theorem (Bickel & Docksum, 1977; Bilodeau & Brenner, 1999; Brockwell & Davis, 1991), the linear unbiased estimates of β and σ^2 that are best in terms of minimizing the squared estimation error are given by:

$$\beta = (X'X)^{-1} X'Y \text{ and } s^2 = 1/(n-p)(Y-X\beta)'(Y-X\beta),$$

where b and s^2 are the estimate of the unknown β and σ^2, respectively. The form of b can be found from algebraic manipulation of Y=Xβ. Note that Y-Xβ is the residuals, so that the form of s^2 is just the mean squared residuals (the n-p reflecting the dimensionality of the residuals that are left after fitting p independent effects). Tests of linear combinations of the parameters can be made under the normality assumption, which gives:

$$C\beta \sim N(C\beta, C(X'X)^{-1}C'),$$

where C is a row vector of length p, often called a contrast (cf., Frackowiak et al., 2004).

3.3.3 HYPOTHESIS TESTING AND STATISTICAL INFERENCE

We have briefly outlined the ways in which effects of interest, confounds, and nuisance variables can be modeled and estimated. The parameters are always assessed relative to their uncertainty in a statistical hypothesis-testing framework. Informally, we wish to know if the magnitude of the parameter (or contrast of parameters) is substantial with respect to its uncertainty (i.e., its standard deviation). Hypothesis testing proceeds as follows: The null hypothesis is assessed with a test statistic, a function of the data that is sensitive to departures from the null hypothesis and reflects the effects of interest; the observed statistic is compared to its distribution under the null hypothesis, yielding a P-value. A small P-value is interpreted as indicating that there is little support for the null hypothesis, though its interpretation is more subtle. The P-value is the probability of observing a statistic value as large or larger, under an identical replication of the experiment, and under the assumption that the null hypothesis is true. Hence, the P-value is a statement about the data under the null hypothesis, not the null hypothesis itself.

In the decision theoretic framework of hypothesis testing, a pre-specified level of significance is used to accept or reject the null hypothesis (Bickel & Docksum, 1977). In alternative frameworks, the smallness of the P-value is viewed as a measure of the strength of the empirical evidence against the null hypothesis (Edgington, 1995). This perspective views the size of P-value as representing a smooth transition from empirical evidence supporting the alternative hypothesis to empirical evidence in favor of the null hypothesis. Now, if one rejects the veracity of the null hypothesis whenever the P-value is below a critical value α then a valid test will control the false positive rate at α. The false negative rate β is closely related to the statistical power, $1-\beta$. The statistical power is the probability of rejecting the null hypothesis when it is false. While it would seem natural to focus attention on the power of the test, the power is a function of the unknown alternative, and the best that can be done is to use test statistics that maximizes power over all alternatives (relative to all other tests of the same class).

The regression approach in functional neuroimaging fits univariate models at every voxel (the number of voxels is typically on the order of 10^5), and effects of interest are tested in each individual model by generating and assessing a statistic image. Usually an image regression approach is used, which implies that the same univariate model is fitted at

each voxel. The common test procedures in functional neuroimaging conform to the standard structure of hypothesis testing. If a particular, pre-specified voxel is of interest, then standard univariate theory can be applied. Otherwise the statistic image is searched for, for example, voxels of significant magnitude using the local maximum statistic, or, given an intensity threshold, significant clusters using the supra-threshold cluster size statistic.

3.3.3.1 GENERAL ISSUES RELATED TO STATISTICAL INFERENCE

The statistical analysis of functional neuroimaging data typically implies that many hypotheses are tested on the same data set. Central to the multiple (e.g., voxel-by-voxel) hypothesis testing is an adequate handling of the multiple comparisons problem; that is, it is necessary to appropriately control the false positive rate. Ideally, the statistical inference procedure should handle the multiple comparisons problem effectively, avoiding any unnecessary loss of sensitivity and statistical power. Given the null-hypothesis H_0 and a test statistic T(X) of the data X, the test is said to be liberal, conservative, or exact, if for any given level α and rejection region $R(\alpha)$, the probability that T(X) belongs to the rejection region $R(\alpha)$, $P[T(X) \in R(\alpha))|H_0]$, is greater than, less than, or equal to α, respectively. Appropriate control of the false positive rate requires an exact or conservative test. In general, the more conservative the test is the lower the sensitivity of the test.

In order to handle the multiple comparisons problem (Hochberg & Tamhane, 1987) appropriately, the rejection criteria has to be chosen so that the probability of rejecting one or more of the null hypotheses when the rejected null hypotheses are actually true, is sufficiently small. Let the search volume $\Omega = \{v_1,...,v_K\}$ consist of K voxels $v_1,...,v_K$, and let $H_1,...,H_K$ be the null hypotheses for each voxel. The omnibus null hypothesis H_Ω is the (logical) conjunction of $H_1,...,H_K$, that is $H_\Omega = H_1 \cap ... \cap H_K$. To test $H_1,...,H_K$ we use a family of tests, $T_1,...,T_K$. For all $j \in \{1...K\}$ let E_j be the event that the test T_j incorrectly rejects H_j, that is $E_j = [T_j \in R(\alpha_j)]$, where $R(\alpha_j)$ is the corresponding rejection region at the level α_j. Suppose the test is exact or conservative, that is $P[E_j|H_\Omega] \leq \alpha_j$.

In the context of the family $T_1,...,T_K$ of tests, the family-wise error (FWE) rate is defined as the probability of falsely rejecting any of the null hypotheses $H_1,...,H_K$. Given

the level α, weak control over FWE requires that the probability of the rejecting the omnibus null hypothesis H_Ω, the union event $E_\Omega = E_1 \cup ... \cup E_K$, is at most α, $P[E_\Omega|H_\Omega] \leq \alpha$. Evidence against the omnibus hypothesis H_Ω indicates the presence of some activation somewhere. This implies that the test has no localizing power, meaning that the false positive rate is not controlled for individual voxels. Tests that have only weak control over FWE are called omnibus tests, and are useful to detect whether there is any experimentally induced effect at all, regardless of location. If, on the other hand, there is interest in not only detecting an experimentally induced signal but also reliably locating the effect, a test procedure with strong control over FWE is required. Strong control over FWE requires that FWE be controlled not just under H_Ω, but also under any subset of hypotheses. Specifically, for any subset of voxels $\omega \subseteq \Omega$ and corresponding omnibus hypothesis H_ω, $P[E_\omega|H_\omega] \leq \alpha$.That is, all possible subsets of hypotheses are tested with weak control over FWE. This ensures that the test is valid at every voxel, and that the validity of the test in any given region is not affected by the truth of the null hypothesis elsewhere. Thus, a test procedure with strong control over FWE has localizing power.

3.3.3.2 SPATIAL AUTOCORRELATION AND MULTIPLE NON-INDEPENDENT COMPARISONS

One way to achieve strong FWE control is to adjust the level of significance with which the different hypotheses $H_1,...,H_K$ are tested. The single step Bonferroni correction is an illustrative example of such a strategy. Suppose that $H_1,...,H_k$ are tested at an equal level, say b, that is, $P[E_1|H_\Omega] \leq b,..., P[E_K|H_\Omega] \leq b$. If all voxels have the same marginal distribution, then testing them at equal level amounts to thresholding the statistic image, giving a single threshold test. In general, $P[E_\Omega|H_\Omega] = P[E_1 \cup ... \cup E_K|H_\Omega] \leq P[E_1|H_\Omega] +...+ P[E_K|H_\Omega] = K \times b$. If b is chosen so that $K \times b = \alpha$, that is, $b = \alpha/K$, it follows that $P[E_\Omega|H_\Omega] \leq \alpha$. This so-called Bonferroni correction will be conservative when the individual tests are correlated, since then $P[E_\Omega|H_\Omega]$ will be substantially smaller than $P[E_1|H_\Omega] +...+ P[E_K|H_\Omega]$. For a large number of correlated tests, the Bonferroni correction results in a conservative procedure and an unnecessary loss of statistical power.

Functional neuroimaging data are often characterized by spatial autocorrelation, meaning that closely spaced voxels are correlated, due to the point spread function of the imaging system, physiological factors, as well as image smoothing. Given a non-trivial spatial autocorrelation in the statistic image this implies that multiple comparisons are non-independent and a simple Bonferroni correction would be unnecessarily conservative. Instead, an effective solution of the multiple non-independent comparisons problem is central to the voxel-by-voxel approach. There are several approaches to handle this problem. Broadly speaking, these divide into parametric, non-parametric, and Monte-Carlo simulation approaches (Petersson, 1998; Petersson et al., 1999b). The parametric approaches used in functional neuroimaging are usually based on some type of random field theory (Adler, 1981, 1998; Friston et al., 1995; Worsley et al., 1996) generating distributional approximations.

3.3.3.3 RANDOM FIELD THEORY

In our studies we have generally used the GLM framework for modeling and estimation, while we have based our hypothesis testing and statistical inference on parametric approaches founded in smooth random field theory. Random field theory has proved versatile in testing a number of test statistics (e.g., local maximum, cluster size statistic, or the number of regions with size greater than a given size). The smooth random field theory approach has been extensively validated on simulated data and empirical studies using real null data have indicated that this approach gives accurate results (e.g., Aguirre, Zarahn, & D'Esposito, 1997; Zarahn, Aguirre, & D'Esposito, 1997). In addition, investigations of the robustness and characterization of inherent limitations of the random field theory approach with respect to the various assumptions and parameters have been carried out extensively; including, for example, with respect to degrees of freedom (Worsley, Evans, Marrett, & Neelin, 1992), smoothness estimation (Poline, Worsley, Holmes, Frackowiak, & Friston, 1995), and variance heterogeneity (Worsley, 1996). In addition, non-parametric methods have been used as benchmarks for cross-validation of the random field theory approach and these investigations have also shown that the approach provides accurate results (e.g., Ledberg et al., 2001). Essentially, the random field theory approach allows for spatial correlation between voxels in the statistic image when correcting for multiple comparisons,

thereby improving on the Bonferroni correction and thus preserving statistical power. The approach has been developed to accommodate several statistical fields, such as Z, t, χ^2 and F fields, where all non-Gaussian random fields are derived from Gaussian random fields.

In the original work of Worsley et al. (1992), it was assumed that the excursion sets (roughly the potentially significant regions) did not intersect the boundary of the search volume, limiting the results to infinite search volumes. This is a reasonable approximation for finite search volumes provided the search volume is large relative to the surface area and the smoothness of the field. In addition, it was assumed that the covariance structure of the random field was stationary. These constraints have subsequently been relaxed and a unified approach described; the random field is transformed to an isotropic random field and the volume, surface area, and diameter are estimated in the space of resolution elements (i.e., resel space, Worsley et al., 1996) and the sationarity assumption has given way for random fields with local non-stationarities (Worsley, Andermann, Koulis, MacDonald, & Evans, Abstract presented at HBM99).

Another general assumption in the application of smooth random field theory to discrete statistic images is that the statistic image can be considered a well sampled version of the smooth random field, or conversely, that the smooth random field is a good approximation of the statistic image. In general, the frequency spectrum of smooth stochastic process is not bounded. However, in experimental data, the observable spatial frequencies are limited (i.e., only the spatial frequencies below half the frequency of the sampling process are observable by the Shannon-Nyqvist sampling theorem). The sampling issue becomes particularly important in the context of smoothness estimation. Smoothness estimation amounts to the estimation of a parameter related to the spatial auto-correlation. It should be noted that the smoothness estimation in random field theory relates to the spatial autocorrelation of the statistic image (which is described by the smoothness parameter(s)), and this is different from image smoothing or spatial filtering applied to the data during pre-processing. The estimation of the smoothness parameter should be independent of experimentally induced effects and thus smoothness estimation is generally made on the residual images. Furthermore, it is important to note that the smoothness estimate itself is the realization of a random variable (Poline et al., 1995). Poline et al. (1995) gives an approximate expression for the variance of this estimator. When estimated on a single

image, the variability of the resulting corrected p-value is found to be moderate (i.e., stdev(p)/E[p] is of the order of 20%).

The multivariate Gaussian assumption is fundamental to the results derived by Adler (1981) and Worsley et al. (1996 a,b). This assumption is difficult to check for functional neuroimaging data. However, with sufficient image smoothing and a sufficient number of effective degrees of freedom, then the multivariate central limit theorem (cf. e.g., Billingsley, 1995) lends support to this assumption. As suggested by the central limit theorem and the fact that the PET reconstruction process (filtered back-projection) implies the summations of a large number of Poisson distributed count data, the regional activity observed in reconstructed PET images can be expected to be (approximately) Gaussian distributed. It has recently been shown that as the projection counts approach infinity the reconstructed images will become multivariate Gaussian distributed (Maitra, 1997). For further discussion of assumptions, approximations, and limitations in functional neuroimaging, see Petersson et al. (1999a; 1999b).

3.4 FUNCTIONAL CONNECTIVITY AND NETWORK ANALYSIS

The statistical models described so far are used to investigate the relationship between the experimental paradigm and the changes induced in brain activity. These approaches study changes in regional activity and how these changes co-vary with specific external experimental manipulations or variables. As outlined in chapter 1 and 2, it has been suggested that higher cognitive functions are the result of network interactions between different brain regions. This suggests that the understanding of different cognitive brain functions can benefit from analyzing the interactions between different brain regions. Based on the idea that brain regions, which constitute components of a functional network, will have activities that are correlated one approach is to investigate the covariance pattern observed in functional neuroimaging data, so-called functional connectivity. Functional and effective connectivity was originally introduced in the context of electrophysiology (Aertsen, Gerstein, Habib, & Palm, 1989; Aertsen & Preissl, 1991) and these concepts were transferred to hemodynamically based functional neuroimaging approaches, typically with a slightly modified connotation. Functional connectivity was defined as the observed

correlations over time between different brain areas, independent of the sources of these correlations, and effective connectivity explicitly referred to the influence that one neural system exerts over another (Friston, 1994). In following sub-section, we will briefly discuss some issues related to covariance sources as well as outline a network approach to the analysis of effective connectivity based on structural equations modeling that we have used in studying a simple network model of immediate verbal repetition (Petersson et al., 2000).

Two different ways to estimate the covariances between brain regions in a given cognitive state have been described: over time within subject (Buechel & Friston, 1997) and over subjects (Horwitz, McIntosh, Haxby, & Grady, 1995). The basic hypothesis is that the intrinsic variability in the neural response of a cognitive state will emulate the relevant functional interactions and that these interactions will be reflected in the covariance structure. Several sources of interregional covariances have been proposed (Horwitz, Soncrant, & Haxby, 1992) and the actual sources of the observed covariances are largely unknown (Petersson et al., 1999a). If the covariances are estimated over subjects, it is necessary to assume that the subjects implement a sufficiently similar functional organization. The observed covariance structure can thus be viewed as reflecting an average common functional organization. However, it is conceivable that the functional organization can vary substantially between subjects; that is, the covariance structure of a subject may or may not be related to a common underlying functional organization. In PET studies the number of intra-subject observations is limited, so, in order to increase sensitivity, data is typically pooled over subjects. However, with FMRI, it is possible to study functional and effective connectivity in single subjects (Buechel & Friston, 1997).

3.5 STRUCTURAL EQUATIONS MODELING

It has been suggested that a comprehensive investigation using various network analysis approaches and large-scale neural modeling hold great potential and may add significantly to our understanding of human cognition (Horwitz, 1998; Horwitz et al., 1999). To characterize effective connectivity in functional neuroimaging data a network approach based on structural equations modeling (Bollen, 1989; Hayduk, 1987) was proposed by McIntosh and Gonzalez-Lima (1994). Structural equations modeling (SEM) provides an

opportunity to investigate functional-anatomical network models subserving different cognitive functions in terms of regions involved and their interactions. To characterize a functional network, a specific functional-anatomical model is used in conjunction with SEM to model the observed covariance structure between the regions included in the model. The functional-anatomical model is specified by selecting the network components and specifying the network topology (i.e., the connections between the components) based on theoretical and empirical considerations. Different constraints on the connections can also be specified. The interregional covariances are computed and the connection (or path) coefficients are estimated within condition. Differences between conditions or groups can then be evaluated using a stacked models approach (Bollen, 1989).

Structural equations modeling commonly use a linear system of equations to describe the interrelation between regions in the functional-anatomical model with the connection coefficients as free parameters. The connection coefficients are fitted in an optimization process. This procedure attempts to recreate the observed covariance structure between regions as closely as possible by finding optimal values of the path coefficients. There are several optimization algorithms available to estimate the connection strengths. Typically, the optimization process uses estimated starting values in combination with an iterative maximum likelihood procedure. For example, the standard implementation in LISREL (Boomsma, 1985; Jöreskog & Sörbom, 1996) a two-stage least squares approach in combination with the Davidon-Fletcher-Power algorithm and line search (other alternatives are available , cf., Jöreskog & Sörbom, 1996). With reasonably well-fitting models, the initial estimates are often close enough to the final maximum likelihood estimate for the optimization algorithms to quickly converge to this estimate. It should be noted that when the estimates depend non-linearly on the model parameters there is no guarantee that the global optimum will be reached with deterministic gradient descent algorithms or non-exhaustive search procedures. Alternatively, a simulated annealing approach to optimization can be used (Geman & Geman, 1984; Kirkpatrick, Gelatt, & Vecchi, 1983) even though practical annealing schedules can only guarantee good sub-optimal solutions.

The results of SEM analyses are potentially difficult to interpret for several reasons. There is no guarantee that the connections modeled actually reflect direct effective

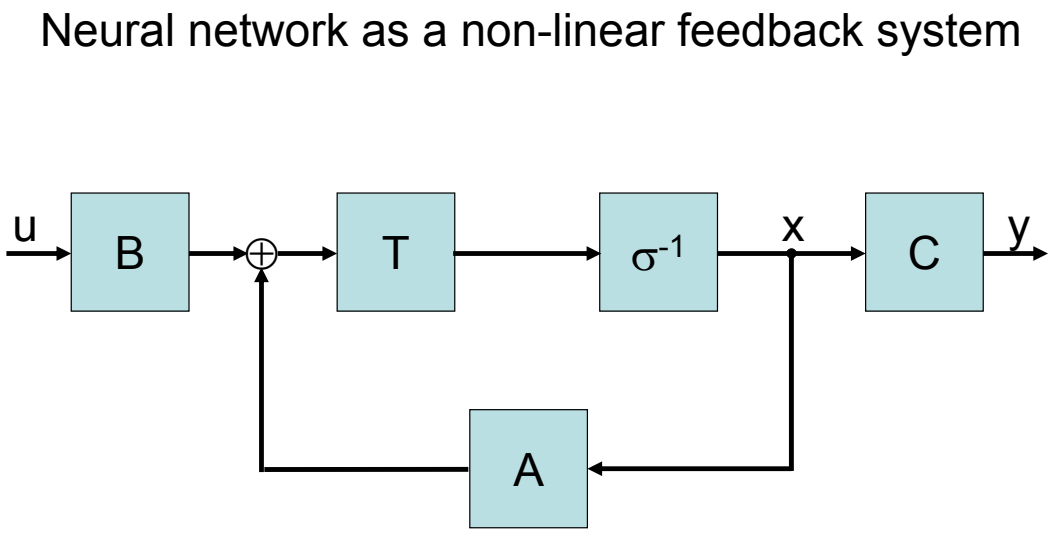
connections - it is possible that these are mediated through regions or connections not included in the model. Similarly, observed changes in the weights between conditions or groups may reflect common input from regions not modeled. Furthermore, unless reasonable goodness-of fit can be achieved with a given model in all conditions or groups investigated, the results of a stacked models comparison can also be difficult to interpret. For example, using an under-parameterized model to test differences in a stacked approach may yield results due to an ill-fitting model in one of the conditions or groups. The effect of using under-parameterized models (i.e., omission of network components, connections, or feedback loops) has been investigated in moderately complex models (McIntosh & Gonzalez-Lima, 1994). This simulation study indicates that the results from analysing moderately reduced models are fairly stable and that modification indices (Jöreskog & Sörbom, 1996) can provide indications of such omissions.

4. MEMORY

In this chapter, which is more empirically oriented compared to chapters 1 and 2, we will review some aspects of human memory from a cognitive neuroscience perspective. In chapter 2 we defined learning, in the general sense of adaptation, as the processes by which the brain functionally restructures its processing networks and/or its representations of information as a function of experience. We suggested that the memory trace (i.e., the stored information) can be seen as the resulting changes in the processing system. On this view, learning is a dynamic consequence of information processing and network plasticity. From this perspective, and in contrast to simple information storage, learning and adaptation can be viewed as a process of generalization. We also described memory as a process, decomposed into several processing stages, including on-line encoding (i.e., representation of the information to be stored), memory formation and storage, consolidation, re-organization and maintenance, as well as retrieval. We also noted that different acquisition problems require different learning processes instantiated in various memory systems in order to ensure effective solutions to learning problems and we argued for the idea of processing systems with multiple interacting memory systems, operating at several different characteristic time-scales (Figure 1.6). As noted in chapter 1, human learning and adaptive brain processes operate at many characteristic time-scales, spanning some seven to nine orders of magnitude and we concluded that different storage systems operate at different time-scales and show different forgetting characteristics. In addition, we outlined an independent rational for the existence of multiple memory systems, which is related to the serial learning problem (also called the stability-plasticity dilemma; cf., chapter 2 and Figure 2.2). This dilemma relates to the problem of up-dating the knowledge base in the light of novel information to be stored and integrated with previously acquired information. There is a trade-off between stability and plasticity: stability is necessary to ensure robust process reliability, while plasticity is necessary for the acquisition of new information; too much stability precludes sufficient plasticity, and conversely, too much plasticity threatens processing stability.

In our formulation of information processing systems with memory (adaptive) properties (cf., section 2.1.2) we conceptualized learning as the interaction between two (several) sets of dynamical variables, representational and adaptive, respectively. However,

there is another, more subtle form of memory, which can be instantiated in the state-space by the representational dynamics alone. This is a form of process memory which does not depend on learning instantiated in adaptive parameters. Fundamentally this form of memory is related to the fact that the current state on the state-space trajectory can be seen as representing aspects of the systems processing history. For example, if the state-space trajectory that the system is currently following crossed another possible trajectory at an earlier time point and no other trajectory-crossings take place, then the current state can be viewed as perfectly representing the segment between the trajectory-crossing and the current state. Thus, one sees that this form of process memory depends on the topology of possible state-space trajectories.



[Figure 4.1] Neural networks as non-linear feedback systems. A generic first order (artificial) discrete-time neural network is a non-linear forced dynamical system typically defined by the transfer functions of the component processors $T = [T_1, \ldots, T_N]$ operating on the input $u = u(t)$ according to: $\sigma[x_j(t)] = T_j(\Sigma a_{jk}x_k(t) + \Sigma b_{jk}u_k(t) + d_j)$, where σ is the time-shift operator defined by $\sigma[x(t)] = x(t+1)$. If we now define the matrices $A = [a_{jk}]$, $B = [b_{jk}]$,

and D = [d_j] and let T act componentwise on Ax(t) + Bu(t) + D, while absorbing D in A by extending x(t) with an additional component $x_{N+1}(t)$= 1, setting $a_{jN+1} = d_j$, and $a_{N+1N+1} = 1$, we arrive at the following equivalent expression x(t+1) = σ[x(t)] = T(Ax(t) + Bu(t)). In the figure, the matrix C selects the output units of the system. Thus, a first order discrete-time neural network can be seen to be a special case of a controlled discrete-time nonlinear dynamical system with feedback (cf. e.g., Isidori, 1995; Sontag, 1998).

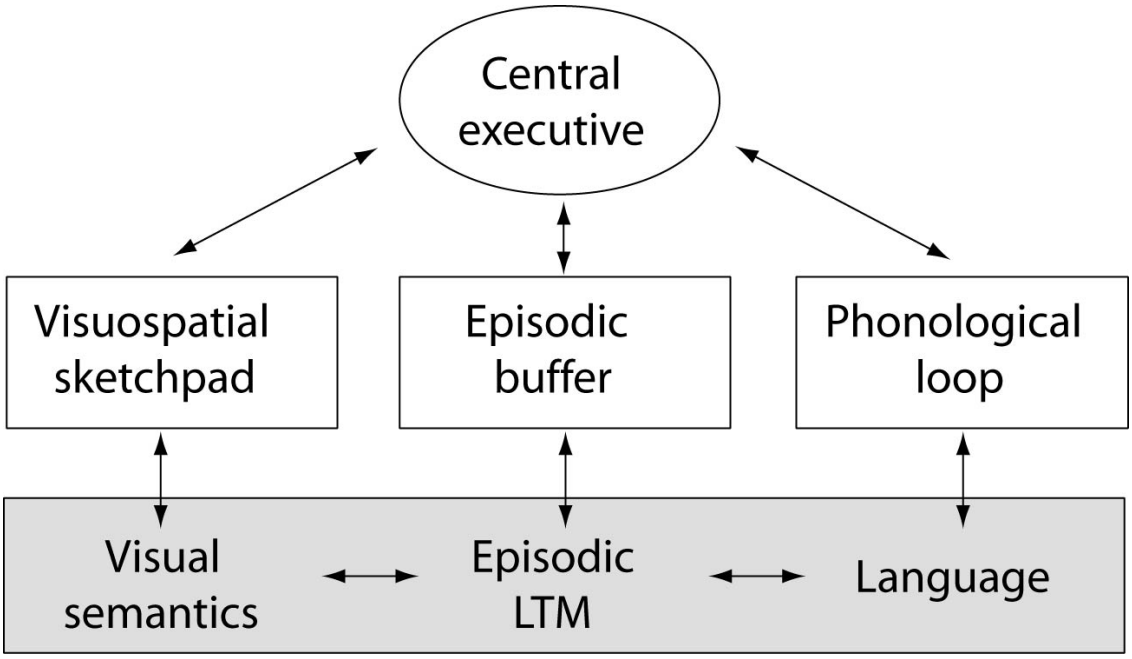
Incidentally, this is the form of memory that finite-state architectures can instantiate and is commonly used as a strategy to implement short-term memory properties in such processing systems (i.e., via state-space coding, cf. e.g., Hopcroft et al., 2000). It follows that the processing system's response to a given input depends on the current internal state; this is immediately obvious from both the classical cognitive formulation (T: Ω x Σ → Ω; cf. section 1.4 and Figure 1.2) as well as the general dynamical systems formulation (which is also described by T: Ω x Σ → Ω; cf. section 1.6), but with the added twist that the current internal state also represents the process history. A slightly different way of implementing process memory is exemplified by non-adaptable linear systems (Oppenheim, Willsky, Hamid, & Hamid Nawab, 1996). In these systems, process memory is implemented via feedback loops. In this context, it is of interest to note that the standard recurrent network architecture (Haykin, 1998) can be viewed as a controlled non-linear system with feedback loops (cf. e.g., Isidori, 1995; Sontag, 1998) and is thus capable of this form of memory (Figure 4.1).

Now, returning to more earthly matters, several memory researchers have argued on both theoretical and empirical grounds that the brain is equipped with multiple memory systems (e.g., Eichenbaum & Cohen, 2001; Schacter & Tulving, 1994; Squire et al., 1993; Stadler & Frensch, 1998). These memory systems serve different purposes and are therefore thought to store different types of information. Endel Tulving (1995) suggested that cognitive memory research, which has produced a tremendous amount of empirical data can be meaningfully 'ordered' with the help of two general concepts, memory systems and memory processes, both of which we have already outlined in some detail in chapter 2. Tulving (1995) proposed a simple model, the SPI model, for memory organization. The SPI

model states that cognitive memory systems are related to one another in terms of the principle memory processes encoding, storage, and retrieval. Tulving (1995) hypothesized that the relations of the various cognitive memory systems (he lists five: procedural/non-declarative, perceptual representation systems, semantic, short-term working, and episodic memory) are related in a process specific manner: information is serially (S) encoded into the systems in a contingent manner; the various memory traces are stored in parallel (P), while information can be retrieved independently (I) from each memory system. Despite its underspecified nature, the SPI model makes some empirical claims. The most important is that the relations between different memory systems are process specific. The serial character of encoding is consistent with the principle of co-localization of memory (in the general sense of adaptive changes) and information processing in the brain. Taking the more general view that information is encoded simultaneously in several interacting memory systems (cf., Figure 1.6), it is an empirical question if these interactions are best conceptualized as serial in character. Clearly there are dependencies between different processing systems, and in this sense the encoding of information in one system might be contingent on the processing of information in some other memory system (i.e., the output of one system is the input to another; e.g., episodic encoding might be dependent on semantic retrieval, while both processes are dependent on information being encoded, or represented, in the perceptual representation systems). Moreover, if the different cognitive memory systems correspond to different sets of physical systems (or sets of physical dynamical variables), then trivially information is stored in parallel and the fact that the same act of encoding can induce multiple effects in different parts of the brain is a natural consequence in a system with multiple and interacting memory systems. The strongest empirical consequence of the SPI model relates to the claim that information from each system (and subsystem) can be retrieved without the necessary implications for retrieval of corresponding information in other systems, which in this sense can be viewed as independent. However, the SPI model does not speak on the possibility that for example retrieval of information in one system might imply encoding and storage of information in another (e.g., retrieval of episodic or semantic information might induce encoding and storage in short-term working memory).

4.1 MULTIPLE MEMORY SYSTEMS

Human memory is commonly divided into memory systems which operate on different characteristic time-scales. An example of this is the coarse division of memory into short-term and long-term memory. One influential model of short-term memory was formulated by Baddeley & Hitch (1974). The Baddeley-Hitch model is not a simple model of short-term encoding, storage, and retrieval of information but includes components which are thought to support several higher cognitive functions, including reasoning and language (Baddeley, 1986). The Baddeley-Hitch model is therefore a model of working memory. In the original Baddeley-Hitch model, working memory consists of a central executive with two support systems, the phonological loop, for short-term encoding and storage of verbal information, and the visuo-spatial sketch pad, for short-term encoding and storage of visuo-spatial information. Recently Baddeley (2000) added another component, the episodic buffer to the working memory model (Figure 4.2).



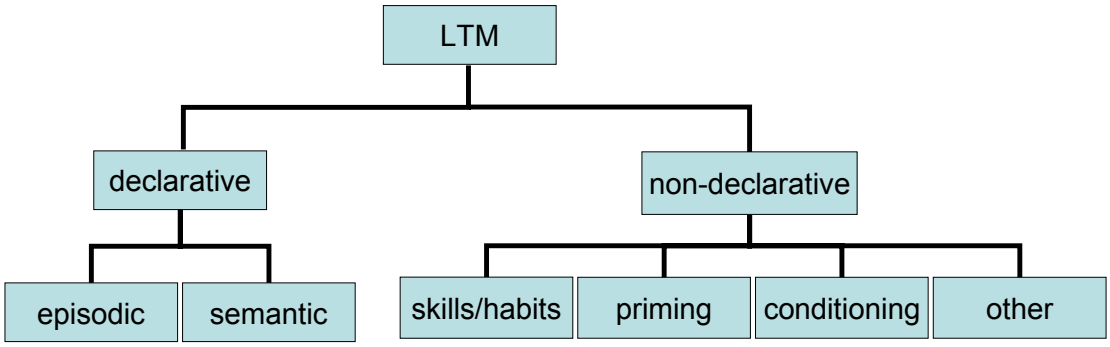
[Figure 4.2] The extended working memory model of Baddeley and Hitch. The episodic buffer comprises a limited capacity system that provides temporary storage of information held in a multimodal code, capable of binding information from the subsidiary

systems, and from long-term memory, into an episodic representational format. The episodic buffer is supposed to provide an interface to the other slave systems of working memory and to long-term memory, feeding information into and retrieving information from episodic long-term memory.

Baddeley (2000) suggests that the episodic buffer is a capacity limited system that provides temporary storage of information held in a multimodal code, which is capable of binding information from the subsidiary systems as well as from long-term memory into an episodic representational format. The episodic buffer shares some characteristics with the concept of episodic memory (Tulving, 1989) with respect to its principal mode of storing information in episodes and its integrative aspects, but differs in that it is assumed to be a temporary store. In the extraction of information from working memory a key function for the episodic buffer is integration between the different subcomponents of working memory. The episodic buffer is thought to provide an interface between the components of working memory and long-term memory. In emphasizing its short-term integrative role and its episodic format, one may hypothesize that the episodic buffer is related to the prefrontal cortex (PFC) and the medial temporal lobe (MTL) as well as the interaction between these structures. The transient early role of the MTL system in long term memory formation and sequence encoding in conjunction with the PFC makes these likely candidates (Eichenbaum, 2000; Simons & Spiers, 2003). The functional anatomical correlate of the phonological store is putatively in the left inferior parietal region (Brodmann's area [BA] 39/40) together with parts of the superior temporal cortex (Becker, MacAndrew, & Fiez, 1999; Paulesu, Frith, & Frackowiak, 1993), and the articulatory rehearsal process involving a left frontal circuit including Broca's region (BA 44) and the pre-motor cortex (BA 6, Smith & Jonides, 1998, 1999).

Human long-term memory is also commonly subdivided in different component memory systems (Tulving & Schacter, 1994) and although the concepts and terminology used to characterize these memory systems has varied, there is a consensus concerning the broad division of human memory into declarative and non-declarative memory (Figure 4.3). Declarative memory supports the capacity to encode, store, and retrieve facts and

events and is contrasted with a heterogeneous collection of non-declarative memory abilities including skills and habits (Knowlton, Mangels, & Squire, 1996), different forms of conditioning (Bechara et al., 1995), and repetition priming (e.g., facilitation of recognition, reproduction or biases in selection of stimuli that have recently been perceived, Schacter, 1994). The knowledge or information acquired by non-declarative memory systems is commonly expressed through performance changes rather than explicit retrieval. Different forms of non-declarative memory depend on the integrity of specific brain systems; for example the basal ganglia, the amygdala, and the cerebellum (Eichenbaum & Cohen, 2001).



[Figure 4.3] Taxonomy of human long-term memory (LTM) systems.

Another commonly used distinction is that between explicit and implicit memory. The terms explicit and implicit memory usually refer to forms of memory expression. In this usage, implicit memory denotes the expression of memory without awareness of its

acquisition or use; that is, behavioral expressions of what an individual has learnt without remembering how, when, or where the learning occurred. In contrast, explicit memory commonly refers to the expression of what the individual is aware of and can explicitly report if probed (Tulving, 1995). In the following we will focus on declarative memory, but we will also, briefly, mention some of the non-declarative memory system.

Explicitly retrieved declarative memories are commonly conceptualized as integrated associative structures, which are continuously updated through active re-organization and integration of new information within the context of previous experiences and previously acquired knowledge (Eichenbaum, 2000). Recollection of memories represents a re-construction (re-creation) process which is partly determined by the nature and organization of the stored information as well as previously acquired knowledge. This type of memory, declarative memory, involves the representation of episodic information within the context of general knowledge. It is thought that episodic representations encode sequences of micro-events and micro-features that compose unique, individual experiences, indexed by specific times and places. Semantic (general) knowledge, on the other hand, represents an acquired knowledge base of organized and inter-related factual information, which is independent of the specific episode(s) in which the information was acquired (Eichenbaum, 2000). General world knowledge is not tied to a specific time and place of acquisition. Declarative memory thus represents the capacity to form and retrieve episodic and semantic information. A key feature of declaratively stored information is its flexible accessibility and expressibility that can be used adaptively in novel situations in an elaborate manner (i.e., flexible memory expression, Eichenbaum & Cohen, 2001; Schacter & Tulving, 1994), for example, to solve new problems and support the inferential expression of associations that are linked across separated experiences; the medial temporal lobe (MTL) memory system might play a role in integrating overlapping experiences into general knowledge in terms of reorganization, abstraction, and re-integration of episodic information (Eichenbaum, 2000).

The declarative memory system has a well-defined neuroanatomic correlate in the MTL memory system (Squire, 1992; Squire et al., 2004; Squire & Zola-Morgan, 1991). However, it should be noted that trace-conditioning is a form of complex conditioning which depends on the MTL (in particular trace conditioning, see e.g., Takehara, Kawahara,

& Kirino, 2003; Weiss, Bouwmeestev, Power, & Disterhoft, 1999), while simple conditioning seems not to depend on the MTL. In the trace-conditioning paradigm the conditioned- and unconditioned stimulus are separated by a relatively long stimulus-free interval, and it might be the case that trace-conditioning depends on associative and temporal integration capacities of the MTL. The MTL memory system is composed of three principal components: neocortical regions, the parahippocampal region and the hippocampus (Amaral, 1993; Amaral, 1999; Squire & Zola-Morgan, 1991; Suzuki, 1996, cf. Figure 4.4). The neuroanatomic organization complements the findings from studies of amnesia, suggesting that the MTL contribute to declarative memory by altering the nature and persistence as well as organization of stored memory representations in the neocortex (Eichenbaum, 2000). In contrast, the MTL memory system is not essential for non-declarative memory, which include the acquisition of perceptual, cognitive, and motor skills, as well as acquired habits and learned response biases (cf. e.g., Packard & Knowlton, 2002). These forms of memory are expressed implicitly through performance alterations (e.g., changes in error patterns, improved response times or performance scores) on a variety of tasks (cf. e.g., Knowlton et al., 1996; Knowlton & Squire, 1996; Petersson, Forkstam, & Ingvar, 2004; Poletiek, 2002; Salmon & Butters, 1996; Squire, 1994; Squire et al., 1993; Stadler & Frensch, 1998). For example, systems that include the basal ganglia and cerebellum mediate forms of implicit learning and non-declarative (procedural) memory. It now seems clear that the basal ganglia, in particular the dorsal striatum, play a role in learning and memory (Packard & Knowlton, 2002). Moreover, recent evidence suggests that the basal ganglia and the MTL memory systems can be activated simultaneously during learning and that in some learning situations competitive interference exists between these two systems (Poldrack et al., 2001; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999). However, recent FMRI data indicate that the caudate nucleus and the MTL can interact non-competitively and that the caudate nucleus is not only engaged after repeated practice, but also after single-trial learning and thus in parallel with the hippocampus (Voermans et al., 2004).

Another prominent structure of the MTL is the amygdala and some forms of affective learning and memory rely on a system that includes the amygdala as a core structure. This memory system mediates fear conditioning as well as other forms of

emotional memory (Bechara et al., 1995; Cahill, Babinsky, Markowitsch, & McGaugh, 1995). Here emotional memory refers to the formation of affective representations that is not necessarily available for explicit retrieval but can be implicitly expressed in for example attraction- and avoidance behavior, as well as in the modulation of autonomic nervous system responses. However, the amygdala appears to have a broader role in human long-term memory. In particular, the amygdala, which is a part of the anterior MTL, has prominent recurrent connections with the hippocampus and the MTL memory system. Thus, it seems that the amygdala is well placed anatomically to modulate declarative memory. For example, a time-varying learning rate that changes with the relevance of the information being processed, opens up for the possibility to control learning rate by various relevance or 'print-now' signals. This mechanism can be used to make the memory selective and modulated by relevance (cf., appendix 2.2). Several functional neuroimaging studies have investigated the role of the amygdala in enhancing declarative memory for emotional experiences and suggested a correlation between amygdala activation during encoding and subsequent memory. For example, the degree of activity in the left amygdala during encoding was predictive of subsequent memory (Canli, Zhao, Brewer, Gabrieli, & Cahill, 2000). Furthermore, it has also been suggested that the amygdala may play a role in modulating the strength and consolidation of memories in other memory systems (Cahill et al., 1995, cf. chapter 2 and appendix 2.2).

4.2 THE MEDIAL TEMPORAL LOBE

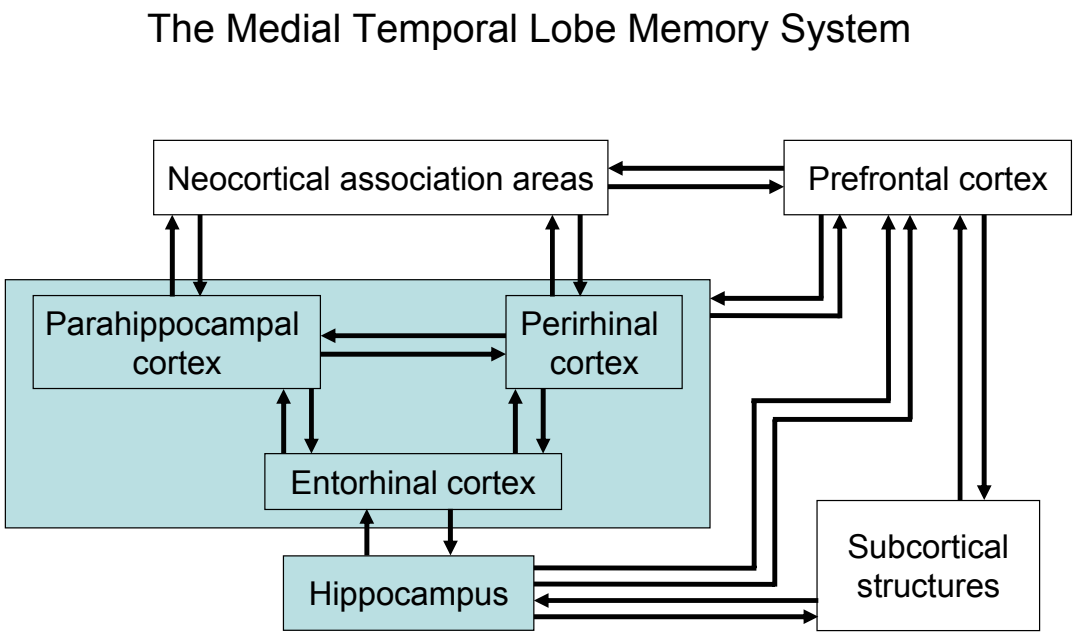
The MTL region was identified as central for declarative memory when Scoville and Milner (1957) reported severe memory loss following bilateral removal of the MTL in their patient H. M. (see also Markowitsch, 1995). The subsequent investigations of H. M. established a central role of the MTL in long-term declarative memory relatively independent of other cognitive functions (Corkin, 2002). The most prominent behavioral deficit following MTL lesions is profound forgetfulness, so-called anterograde amnesia. Anterograde amnesia refers to the incapacity to encode and store new information in long-term memory; for example, information stored in short-term memory is rapidly forgotten and not transferred into a long-term memory trace. According to Squire et al. (2004) there are three relevant aspects of this condition; (1) the impairment is multimodal; (2)

immediate short-term memory appears relatively intact; and (3) the memory impairment appears to occur against a background of intact perceptual, cognitive and motor abilities as well as intact non-declarative memory. Thus, important aspects of declarative memory are dissociated from general perceptual, cognitive, and motor function. Furthermore, the evidence suggest a particular role for the MTL memory system in encoding-storage-consolidation processes, while the role of the MTL in retrieval processes is less clear (cf., Simons & Spiers, 2003; Squire et al., 2004). One way to understand this is that the MTL creates a long-term storage format, making the stored information suitable for effective retrieval; when the MTL is damaged, immediate short-term memory representations in neocortex are not adequately processed from the perspective of long-term retrievability. If the MTL is not functional at the time of learning, declarative memory is not established in a proper way and is therefore not readily available for later retrieval. However, this does not necessarily imply that the MTL is a (permanent) repository of memory, and because remote memory is spared in patients with large MTL lesions, it appears that long-term information would have to be stored elsewhere.

In addition to the phenomenon of anterograde amnesia, damage to the MTL often results in partial loss of memory for information acquired before the damage occurred, so-called retrograde amnesia. Squire and colleagues (2004) suggest that when the MTL lesion is limited to the hippocampus, entorhinal cortex, and/or fornix, the retrograde memory impairment is temporally graded. This implies that more recently formed memories are relatively more impaired than more remotely acquired information. Temporal gradients of retrograde amnesia have also been described in patients with damage limited to the hippocampal region (Kapur & Brooks, 1999; Manns, Hopkins, & Squire, 2003). Furthermore, the remote memory for facts (semantic memory) is intact and it seems that remote episodic memory for autobiographical events can also be preserved (Bayley & Squire, 2003).

The consolidation view of temporally graded retrograde amnesia begins with the principle that long-term memory is stored as outcomes of MTL processing in interaction with the regions in the neocortex that are specialized for what is to be remembered (Squire et al., 2004). On this view, the MTL initially works together with the neocortex to allow memory to be formed into a retrievable format. Through a gradual process of integration

and re-organization (cf., Figure 2.2), it is suggested that the connections among neocortical regions are progressively strengthened until the neocortical memory can be retrieved independently of the hippocampus.



[Figure 4.4] The medial temporal lobe memory system. Adapted from Simons and Spiers 2001. Whereas the medial temporal lobe has been associated with the encoding, storage and retrieval of long-term declarative memory, the prefrontal cortex, and the neocortex more generally, also plays an important role in declarative memory (Eichenbaum & Cohen, 2001; Nyberg, Cabeza, & Tulving, 1996; Simons & Spiers, 2003; Squire et al., 2004). The prefrontal cortex has been linked to short-term working memory, language processing, and various cognitive control processes such as selection, monitoring, manipulation and inhibition of information (Baddeley, 1992, 2000, 2003; Bookheimer, 2002; Fletcher & Henson, 2001). Here working memory refers to short-term online representations of information that are explicitly available for further processing (e.g., active rehearsal or manipulation/processing). Encoding, consolidation, and retrieval from declarative memory are thought to depend on the interaction between the medial temporal

lobe and the prefrontal cortex, as well as posterior neocortical regions. Other brain regions, which are also important for long-term declarative memory, include the thalamus (e.g., the anterior and mediodorsal nuclei), the mamillary bodies, and the basal forebrain nuclei, as well as the retrosplenial cortex.

Electrophysiological studies have demonstrated the importance of the parahippocampal and perirhinal cortices in memory consolidation in animal models. In rats, perirhinal cortex activation appears to promote enhancement in cortico-cortical pathways (Ivanco, Michelin, & Racine, 1996). In monkeys, Higuchi and Miyashita (1996) demonstrated that lesions of the entorhinal and perirhinal cortices prevented the formation of neuronal memory responses to visual paired associates in the inferotemporal cortex. The parahippocampal region seems to mediate the extended persistence of these cortical representations and processing within the neocortex may take advantage of lasting parahippocampal representations, and come to reflect complex associations between events that are processed in different neocortical regions or occur sequentially in the same or different areas (Eichenbaum, 2000). An alternative proposal states that the hippocampus and related structures are always necessary for recalling the richness of detail available in autobiographical recollections (Nadel & Moscovitch, 1997).

In the following section we will outline a position on human memory systems that is closely related to that of Squire et al. (2004) and Eichenbaum and Cohen (2001). However, it should be noted that there are several alternative perspectives (cf., section 4.3), suggesting that distinct sub-regions of the MTL support dissociable functions (e.g., Aggleton & Brown, 1999; Murray & Mishkin, 1986; Murray & Bussey, 1999; Simons & Spiers, 2003; Tulving & Markowitsch, 1998; Yonelinas et al., 2002). For example, one model suggests that a system involving the hippocampus (as well as the thalamus, mamillary bodies and retrosplenial cortex) subserve recollection, while parts of the parahippocampal cortex (perirhinal cortex) support familiarity-based recognition. In addition, Squire and colleagues (2004) as well as Eichenbaum and Cohen (2001) argue for a time-limited role in remote declarative memory and a central role of the MTL in memory consolidation.

Afferent information to the MTL originates from most neocortical association areas (Suzuki & Amaral, 1994a, 1994b). These neocortical regions project to one or more of the parahippocampal subdivisions, which include the parahippocampal, the perirhinal, and the entorhinal cortices. The subdivisions of the parahippocampal region are strongly interconnected and send efferent projections to several parts of the hippocampus itself, including the dentate gyrus, the CA1-3 (Cornu Ammonis) fields, and the subiculum. Within the hippocampus, there are divergent and convergent connections, supporting plasticity mechanisms that participate in the rapid encoding of information (Amaral, 1993; Amaral & Witter, 1989; Bliss & Collingridge, 1993). In particular, the CA3 has the basic architecture of a generic recurrent network. The outcome of hippocampal processing is returned, via the parahippocampal region, to the same brain regions from where the information originated (Burwell, Witter, & Amaral, 1995; Suzuki, 1996). Several additional structures, including for example the mamillary bodies, the anterior and mediodorsal thalamic nuclei, the basal forebrain nuclei along with other subcortical nuclei, interact with the hippocampus through a major fiber bundle the fornix (Lavenex & Amaral, 2000; Squire, 1992; Squire & Zola-Morgan, 1991; Suzuki, 1994).

A prominent feature of the structures forming the neocortical-MTL loop is their organization into hierarchical association networks (Felleman & Van Essen, 1991; Lavenex & Amaral, 2000). The connections within the parahippocampal, entorhinal, and perirhinal cortices as well as the convergence/divergence of inputs at the different levels of the neocortical-MTL loop enable a significant amount of integration before information reaches the hippocampus proper. Lavenex and Amaral (2000) suggest that the level of integration and complexity of the information increases when moving from the neocortex to the hippocampal complex; unimodal information becomes polymodal/amodal, and reaches the highest level of abstraction within the hippocampal complex, before it is returned to the neocortex. This suggests a significant contribution of the neocortical-MTL loop to declarative memory formation, consolidation, and memory retrieval. Furthermore, the hippocampal output can also influence the processing of incoming information through the feedback projections from the hippocampal complex to the neocortex. The output of the MTL system is ultimately distributed, via these feedback projections, to much of the neocortex. Neocortical regions have specific perceptual, cognitive, and motor processing

functions required to complete a given memory task, including the important aspects of short-term working memory. It has been suggested that the parahippocampal region is critical in extending the persistence of information over brief periods (Eichenbaum, 2000), while the ultimate structure along the neocortical-hippocampal loop, the hippocampal complex, participates at the highest level of integration and organization of information (Felleman & Van Essen, 1991; Lavenex & Amaral, 2000). Thus, it is suggested, that the parts of the parahippocampal region, which receives convergent inputs from the neocortical association areas and return projections to all of these areas, might mediate the extended persistence of neocortical representations. If this is correct, the interactions between neocortical regions and the MTL can utilize briefly lasting parahippocampal representations, which reflect complex associations between events that are processed separately in different cortical regions or occur sequentially in the same or different areas (Eichenbaum, 2000). It is interesting to note an early suggestion of David Marr (1971), who first suggested that the hippocampal formation creates indexes or pointers of incoming information for rapid storage. These pointers are thought to participate in the consolidation and integration/reorganization of neocortical representations during sleep. In line with this suggestion, Squire and colleagues (2004) suggest that the MTL interacts with the neocortex in order to establish, maintain, and retrieve long-term memory, and that ultimately, declarative memory becomes (relatively) independent of the MTL through a process of consolidation. Finally, it is likely that the principal component of the declarative memory system, including the neocortex, contributes differentially to declarative memory and the interactions between these components are essential (Simons & Spiers, 2003). However, given the neuroanatomic characteristics of the neocortical-MTL system (weak hierarchical organization, high level of associativity, and recurrent connectivity), it might be difficult to experimentally distinguish the different functional properties of some of the sub-structures. Although, it has been suggested that neurophysiological, neuroimaging, and neuroanatomic data indicate a division of labor within the MTL (Tulving & Markowitsch, 1997), Squire and colleagues (2004) suggest that the available data do not support simple dichotomies between the functions of the hippocampus and the adjacent MTL structures (e.g., associative vs. non-associative memory, episodic vs. semantic memory, recollection vs. familiarity).

4.3 SOME ALTERNATIVE PERSPECTIVES

Since the discovery of hippocampal place cells in the rodent (O'Keefe & Dostrovsky 1971), an influential idea has been that the MTL and in particular the hippocampus forms spatial cognitive maps and that the predominant function of the MTL is to support spatial memory (O'Keefe & Nadel 1978). Consistent with this suggestion, several lines of investigation have related the MTL to learning and memory of visuo-spatial material (Eichenbaum & Cohen, 2001; Maguire, Frith, Burgess, Donnett, & O'Keefe, 1998; Nadel, 1994; O'Keefe, Burgess, Donnett, Jeffery, & Maguire, 1998). However, it has been argued that the formation of cognitive maps, and spatial memory more generally, can be viewed as a special case of declarative memory (Eichenbaum & Cohen, 2001; Squire et al., 2004). On the latter view, the role of MTL is conceived of as more general and not restricted to spatial memory only, suggesting that the MTL is central for both spatial and non-spatial declarative memory, in particular when information has to be acquired in ways that allows it to be used in a flexible manner to explicitly guide behavior (McNamara & Shelton, 2003).

Memory-impaired patients with MTL lesions sometimes show a residual capacity for fact-like learning (i.e., semantic memory, Bayley & Squire, 2002; Tulving, Hayman, & MacDonald, 1991; Vargha-Khadem et al., 1997; Westmacott & Moscovitch, 2001). Similarly, it has been suggested that patients with developmental amnesia, with damage limited to the hippocampus, provide an exception to the necessary engagement of the MTL in semantic memory (Baddeley, Vargha-Khadem, & Mishkin, 2001; Vargha-Khadem et al., 1997). This raises the question what kind of learning process or memory system is engaged in these cases? Squire and colleagues (2004) argues that when the MTL damage is extensive (complete), then acquisition is supported by the neocortex, while in non-complete cases the remaining MTL structures is responsible for successful learning. In developmental amnesia, it may be the case that early hippocampal lesions allow neocortical regions to compensate for the dysfunctional MTLs, or perhaps, through the use of acquired alternative learning strategies. Because general knowledge can be acquired through multiple learning events, and since episodic memory is unique to a single event, semantic memory is predictably better preserved than episodic memory. An alternative possibility is

that, given sufficient training and repetition, subjects with developmental amnesia are able to acquire semantic knowledge in proportion to what would be expected from their day-to-day episodic memory ability, presuming residual MTL capacities (Squire et al., 2004).

It has also been suggested that the hippocampus may have a special role in tasks that depend on relating, associating, or combining information from multiple sources, such as episodic or associative memory. It has also been suggested that when these task requirements are not present or in tasks that only require familiarity judgments, then the hippocampus is not necessary but can be supported by the neocortex adjacent to the hippocampus (Tulving & Markowitsch, 1998). However, Manns et al. (2003) reported impaired semantic memory in patients with hippocampal damage, and based on such findings, Squire and colleagues (2004) as well as Eichenbaum and Cohen (2001) argue that the hippocampus is necessary for the acquisition of general knowledge. Similarly it has been suggested that the ability to combine two or more unrelated items into a long-term (conjunctional or associative) memory depends more on the hippocampal region compared to single-item memory. However, recent findings indicate that the hippocampus is important for single-item as well as associative memory (Stark & Squire, 2003). In addition, although the perirhinal cortex has been linked to non-associative (single-item) memory, activation of this region has also been observed in relation to associative memory (e.g., recollection), and activity in the hippocampal region has been correlated with non-associative memory, for example item recognition and familiarity (for reviews see e.g., Fletcher, Frith, & Rugg, 1997; Lepage, Habib, & Tulving, 1998; Schacter & Wagner, 1999; Squire et al., 2004).

Another hypothesis suggests another division of labor within the MTL and concerns recognition memory (the capacity to identify an item previously encountered). Single-cell recordings during recognition performance in rodents, monkeys, and humans suggest that the contribution of the hippocampus is different compared to the contribution of the adjacent cortex (Suzuki & Eichenbaum, 2000). Moreover, studies of both monkeys and rodents have typically found recognition memory impairment following restricted hippocampal lesions (Eichenbaum, 2000; Eichenbaum & Cohen, 2001).

Subjectively, it appears that judgments in for example a recognition memory test can be based on a sense of familiarity or on recollection of detailed information about

previous events (Mandler, 1980). A variety of dual-process models propose that recognition reflects the products of two distinct memory processes and it is commonly suggested that recognition consists of two components: a recollective (episodic) component and a familiarity component. During recollection, information about a given episode is retrieved together with contextual information, while a familiarity signal only indicates that a stimulus has been encountered, and no additional information about the spatiotemporal or event context is retrieved. In general, dual-process models assume that recollection depends on the same or similar processes that are involved in recall tasks, while familiarity reflects a purely quantitative, ‘'strength-like’' memory signal (Rugg & Yonelinas, 2003). At a behavioral level, several methods have been used to measure the hypothetical components recollection and familiarity. These methods have indicated that the two forms of memory can be dissociated (Rugg & Yonelinas, 2003). For example, recollection benefits more than familiarity from elaborative meaning-based encoding or active generation compared with passive reading; recollection is slower and requires more attention during both encoding and retrieval than familiarity; familiarity is more sensitive to perceptual changes between study and test; and recollection is less affected by increased study-test interval compared to familiarity effects (Rugg & Yonelinas, 2003).

Several researchers have suggested that recollection depends on the hippocampus while familiarity depends on the adjacent parahippocampal cortex (e.g., Brown & Aggleton, 2001; Rugg & Yonelinas, 2003), that is, the hippocampus is more active when recognition is accompanied by recollection than when recognition is based on familiarity alone. For example, recent data suggest a process dissociation within the human MTL, successful retrieval of contextual information was accompanied by an activity increase, while it was suggested that a familiarity signal was provided by an activity decrease that was sufficient for successful item recognition (Weis et al., submitted), see also Henson et al. (2003).

The remember/know technique is commonly used to assess recollection and familiarity, respectively. Here, 'remember' is associated with recollection while 'know' is related to familiarity and a recent study on patients with lesions limited to the hippocampal formation reported similar impairment in terms of knowing and remembering (Manns, Hopkins, Reed, Kitchener, & Squire, 2003). At present, evidence does not support the view

that familiarity is preserved in amnesic patients (Rugg & Yonelinas, 2003). It may also be the case that the remember vs. know contrast is related to a difference in the amount of information retrieved (Nyberg, 1998; Nyberg, McIntosh, Houle, Nilsson, & Tulving, 1996b) rather than recollective experience per se (Petersson et al., 2001; Rugg & Yonelinas, 2003). However, Rugg and Yonelinas (2003) suggest that FMRI findings can be interpreted as support for the dual-process framework as well as the proposal that the distinction between recollection and familiarity is maintained within the MTL. They also note that, although single-process models, in which recollection is assumed to reflect the retrieval of strong content-rich memories whereas familiarity is associated with weaker less specific memories, have difficulty accounting for all behavioral dissociations, and their parsimony makes them potential important alternative models.

Furthermore, attempts to related MTL sub-components to encoding and retrieval have also been pursued (Gabrieli, Brewer, Desmond, & Glover, 1997). For example, Lepage and colleagues (1998) suggested that encoding is more related to the anterior while retrieval is more related to the posterior MTL. However, Schacter and Wagner argued that based on the available functional neuroimaging literature (Schacter & Wagner, 1999) it is difficult to deduce any large-scale functional specialization with respect to encoding and retrieval. For example, Small et al. (2001) reported hippocampal activations extending over most of the longitudinal axis of the hippocampus during both encoding and retrieval, while Stark and Okado (2003) reported encoding and recognition related activity in both the hippocampal region as well as in the perirhinal and the parahippocampal cortices (see also, Weis, Klaver, Reul, Elger, & Fernández, 2004).

An alternative hypothesis concerning the role of the MTL suggests that the hippocampal formation may subserve aspects of novelty detection (Tulving, Markowitsch, Kapur, Habib, & Houle, 1994; Tulving, Markowitsch, Craik, Habib, & Houle, 1996). Both functional neuroimaging (Dolan & Fletcher, 1997; Stern et al., 1996) and electrophysiological studies (Grunwald, Lehnertz, Heinze, Helmstaedter, & Elger, 1998) of episodic encoding and the MTL have been interpreted in line with this suggestion. Novelty detection is commonly taken to mean that new information activates the MTL and that novelty might be related to attentional effects in combination with retrieval during encoding. In other words, new attended information is automatically encoded, engaging the

MTL, whereas if the information has already been encoded this may be retrieved, indicating to the learning system that the given information was recently encoded. However, the effects observed in the MTL may be secondary to the result of novelty processing elsewhere and/or attentional effects which do not relate to the MTL per se. Moreover, both Henke et al. (1997) and Montaldi et al. (1999) suggest that the MTL effects observed in their studies cannot be explained in terms of novelty detection. In addition, several studies of the subsequent memory effect also suggest that MTL activation cannot be related to novelty detection in any simple way, since novelty is held constant in these studies and thus cannot explain the subsequent memory effect (Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Fernandez et al., 1998; Petersson et al., 1999; Wagner, Schacter et al., 1998). In the context of long-term declarative memory, it seems more natural to re-interpret the novelty detection idea in terms of familiarity and item recognition (Petersson et al., 2001). Electrophysiological studies have shown that the activity of the (anterior) parahippocampal region decreases during item recognition (Brown, Wilson, & Riches, 1987) as well as a consequence of decrease repeated exposure (Brown & Xiang, 1998). Moreover, a recent meta-analysis suggested that less anterior MTL activity is related to the degree of familiarity (Henson et al., 2003) and it has been hypothesized that the perirhinal cortex contributes to recognition memory by assessing relative familiarity, based on neuronal response decrements (Brown & Aggleton, 2001). In addition, recent results suggest that a similar mechanism might be at play in humans. These findings suggest that more neural resources may be needed for items that are processed for the first time compared to those that have been encountered before (Weis et al., submitted). Weiss and colleagues suggested that this mechanism might support item recognition by a familiarity signal which essentially is based on reduced processing demands for more familiar items. Thus, it was hypothesized that the anterior MTL region is related to both item recognition (activity decrease) as well as declarative memory formation (activity increase). This suggestion is in line with electrophysiological data recorded from within this region in epilepsy patients, where the very same event related potential is correlated negatively with item recognition and positively with encoding success (Fernández et al., 1999; Fernández et al., 2001; Fernández, Klaver, Fell, Grunwald, & Elger, 2002; Smith, Stapleton, & Halgren, 1986). Finally, the MTL has been related to retrieval success. For example, Nyberg and

colleagues reported a positive correlation between retrieval success and MTL activity (Nyberg, McIntosh, Houle, Nilsson, & Tulving, 1996a), while Eldridge et al. (2000) more recently have related hippocampal activation to the remember – know contrast.

4.4 THE FRONTAL LOBE

Cognitive neuroscience has made progress in understanding the roles of the medial temporal and frontal lobes in long-term memory. The importance of the frontal lobe in long-term memory has been recognized only recently and largely due to functional neuroimaging studies. Although lesion studies had already hinted at this possibility (Shimamura, 1995; Wheeler, Stuss, & Tulving, 1995), a common observation made in these studies is that the prefrontal cortex (PFC) is activated during both encoding and retrieval (Nyberg, Cabeza et al., 1996; Tulving, Kapur, Craik, Moscovitch, & Houle, 1994). Thus it seems that the PFC is engaged in processes important for both memory formation and memory retrieval. Whereas the MTL has been associated with memory formation, storage, and retrieval of information from long-term memory, the PFC has been related to what can be called executive aspects of working memory or control processes such as monitoring, selection, and manipulation, as well as maintenance and inhibition. However, the precise functional role of the PFC is not well-understood in general and a precise characterization of PFC functions in long-term memory has been elusive and often couched in relatively unspecific general terms (Duncan, 2001; Fletcher & Henson, 2001; Kimberg, D'Esposito, & Farah, 1997; Miller & Cohen, 2001; Simons & Spiers, 2003; Wood & Grafman, 2003); this will also become apparent as we review some current perspectives on the functional role of the PFC in the following sections.

4.4.1 SOME GENERAL PERSPECTIVES ON FRONTAL LOBE FUNCTION

Although it seems clear that the PFC is important for higher cognition, including for example, attention, language, memory, problem solving, decision making, as well as the temporal organization of behavior – the question of how the PFC subserves these cognitive functions is not well-understood and to a large degree underspecified. One view of prefrontal function – a processing specific perspective - suggests that distinct regions of prefrontal cortex are specialized for different cognitive functions (Petrides, 1995) relatively

independent of modality, while others have emphasized the adaptive nature of the prefrontal cortex (Duncan, 2001; Miller & Cohen, 2001). In addition, various versions of domain or modality specific perspectives have been put forward (cf. e.g., Fuster, 1995; Fuster, 1997; Goldman-Rakic, 1988, 1998).

The primate PFC has been investigated at the neuronal level on a wide range of tasks, including for example categorization, working memory, rule learning, rule switching, and cross-modal integration (for reviews see e.g., Duncan, 2001; Duncan & Miller, 2002). The response properties of prefrontal neurons appear surprisingly adaptable and it seems that any given neuron can be driven by several different kinds of input. This might be a result of the dense recurrent connectivity that exist within the PFC itself and/or the reciprocal connections between the PFC and many other neocortical and sub-cortical structures (Fuster, 1997; Mesulam, 2002; Stuss & Knight, 2002). The adaptive coding model of Duncan (2001) suggests that working memory, attention and cognitive control are subserved by common processing properties of PFC neurons in combination with the adaptable nature of these neurons. This, it is argued, allows the PFC to represent task-relevant information and provide a temporary, task-specific, context-dependent working memory space. Duncan (2001) suggests that this working space serves as a mechanism for selective attention and control by selecting task-relevant inputs, represented in the posterior neocortical regions, and for further elaborate processing or manipulation of the task-relevant information. It is suggested that the PFC biases or focuses processing in posterior cortical regions on task-relevant representations. This idea is similar to the integrative theory of PFC function advanced by Miller and Cohen (2001). Miller and Cohen (2001) argue that the PFC stores representations of task-specific rules, attentional templates, and task relevant goals. In their view, an important role of the PFC is to bias the activation of goal related representations that are stored, represented, and processed in the posterior neocortical regions. They propose that this form of guided or controlled activation of posterior representations is essential for rule acquisition as well as the acquisition of new information and behaviors. Miller and Cohen (2001) suggest that repeated activation of the same processing pathways creates stronger associations between posterior representations (i.e., stronger connections between posterior representational regions), while at the same time, the role of the PFC gradually diminishes in controlling posterior neocortical

processing as this becomes increasingly automatic.

Fuster (1995; 1997) proposes that the over-arching role of the PFC is to temporally organize goal-directed behavior and that this global function can be analyzed in terms of working memory, attention and inhibitory control. Fuster (1995; 1997) outlines mechanisms for monitoring, short-term memory and attentional selection that prioritize goals and task appropriate behavioral sequencing. Temporal integration is achieved by the PFC in interactions with posterior cortical regions, determined by the modalities of task-relevant sensory and motor information. Moreover, he suggests that prefrontal representations and processing are recruited in non-automatic behavior, while well-practiced tasks can be performed relatively independently of the PFC. Several other researchers have sketched similar ideas in terms of a global workspace for non-automatic cognitive processing (Cohen, Dunbar, & McClelland, 1990; Cohen, Servan-Schreiber, & McClelland, 1992; Dehaene, Kerszberg, & Changeux, 1998). Similarly, it has been suggested that the PFC serves as a working memory structure (cf. the Baddeley-Hitch model outlined in section 4.1) that keeps stimulus representations active for on-line processing (Fuster, 1995; Fuster, 1997; Goldman-Rakic, 1988). In particular, it is proposed that the PFC, being part of an integrated network of regions including temporal, parietal, and limbic, is involved in the representation of stimuli in their absence. This would allow the PFC to guide behavioral responses through internal representations (in the sense of cognitive states; cf., chapter 1 and 2).

Closely related views, emanating from investigations of language processing, suggest that the PFC is engaged in structural integration that serve to rapidly and selectively bring together information in posterior representational regions; different linguistic representations (e.g., phonological, syntactic, semantic, and pragmatic) are activated in parallel and integrated in a prefrontal workspace where incremental unification takes place (cf. e.g., Forkstam, Hagoort, Ingvar, & Petersson, in preparation; Hagoort, 2004; Petersson et al., 2004). Similar ideas have been put forward with respect to structural integration in music perception (for a review see, Patel, 2003).

Several of the perspectives outlined so far are processing oriented, but also representational perspectives on PFC function have been put forward (cf., Miller & Cohen, 2001; Wood & Grafman, 2003). For example, in the structured event complex framework

outlined by Wood and Grafman (2003), it is suggested that PFC stores representations of knowledge in the form of so-called goal-oriented sets of events. These goal-oriented sets carries a schematic sequence structure and represents various forms of knowledge, like event features, event boundaries, social rules, thematic knowledge, concepts, as well as grammars; the different aspects of a structured event complex are represented independently, they are encoded and retrieved in an episodic format (Wood & Grafman, 2003).

In summary, given the capacity for on-line maintenance as well as the monitoring, manipulation, and selection functions of the executive part of working memory and its suggested relation to the PFC, it can be hypothesized that the PFC is involved in attentional processing, cognitive and behavioral selection, the decomposition of task processing into goals and sub-goals (i.e., prioritized dynamic scheduling or planning of sub-tasks), problem solving, as well as non-automatic and flexible cognition and behavior.

4.4.2 THE FRONTAL LOBE AND LONG-TERM MEMORY

As already noted, while the importance of MTL in declarative memory has been recognized for at least half a century, the importance of the frontal lobe has been appreciated only recently. The role of PFC regions in long-term memory has been intensely investigated with functional neuroimaging and several types of regional specializations of the PFC have been suggested. In the meta-analysis of Wheeler, Stuss, and Tulving (1995), it was shown that patients with frontal lesions exhibit memory deficits for both recognition and recall, and that the impairment appeared to be greater for recall compared to recognition tests. More specifically, patients with frontal lesions show retrieval difficulties when it is necessary to retrieve contextual details or when minimal retrieval cues are provided (Gershberg & Shimamura, 1995; Shimamura, 1995). This suggests that patients with frontal lesions are not able to effectively utilize organizational retrieval strategies. In addition, it appears that patients with frontal lobe lesions are more sensitive to interference between stimuli during encoding or retrieval (Incisa della Rocchetta & Milner, 1993; Shimamura, 1995). Thus, it has been suggested that the PFC support control processes during long-term memory encoding and retrieval rather than automatic storage and retrieval processes (Fletcher & Henson, 2001). Based on this, it seems natural to suggest that the

role of the PFC in long-term memory is related to various aspects of working memory.

One of the first, empirically guided, frameworks for the role of PFC in declarative (episodic) memory was put forward by Tulving and colleagues, the so-called hemispheric encoding and retrieval asymmetry (HERA) model (Tulving, Kapur et al., 1994). According to HERA, the left PFC is more involved than the right in episodic memory encoding, whereas the right PFC is more involved than the left in episodic memory retrieval (Habib, Nyberg, & Tulving, 2003). The retrieval part of the HERA model, which also seems to be the most empirically substantiated part of the HERA generalization, was recently elaborated in terms of a specific episodic retrieval mode (Lepage, Ghaffar, Nyberg, & Tulving, 2000). Here, retrieval mode refers to a cognitive state that sets the stage for episodic remembering (Lepage et al., 2000). Several PET and FMRI studies have reported results which are consistent with the general HERA pattern, but there are also a number of exceptions (for reviews see e.g., Fletcher & Henson, 2001; Habib et al., 2003; Simons & Spiers, 2003), and based on this, others have argued that the HERA generalization is not sufficient to capture all the relevant data (Fletcher & Henson, 2001; Miller, Kingstone, & Gazzaniga, 2002; Owen, 2003). For example, there are studies of verbal retrieval that have reported bilateral or left PFC activations (Fletcher & Henson, 2001). Similarly, both left and right PFC activations have been observed during encoding of figurative and non-figurative visual material (Petersson, Sandblom, Elfgren, & Ingvar, 2003). Furthermore, the PFC has been shown to be sensitive to the type of material processed during both encoding and retrieval. For example, Kelley and colleagues (1998a) reported material specific PFC activation during both encoding and retrieval, where the left PFC was more related to verbal or verbalizable material while the right PFC was shown to be more related to non-verbalizable material (see also Wagner, Poldrack et al., 1998b). However, it has been argued that material specific effects may occur independent of process specific effects (Habib et al., 2003; Nyberg et al., 2000).

It is well-accepted that episodic encoding benefits from meaning-based elaborate processing of information. When stimulus material is processed in an elaborate meaning-based or conceptual manner, so-called deep processing, the material will be better remembered or more effectively retrieved than when the same material is processed with an emphasis on superficial or perceptual (surface) features, so-called shallow processing. This

so-called levels-of-processing (LOP) effect is a robust effect observed in human memory (Craik & Lockhart, 1972). Given that the left hemisphere is generally regarded as more related to language processing compared to the right, it seems reasonable to suggest that the left-lateralization of activation during encoding is related to language based processing (e.g., with respect to meaning, which entails among other things retrieval from semantic long-term memory), which might generate a linguistic format which is highly effective for long-term memory storage and retrieval. This hypothesis would provide a potential explanation for the left hemisphere part of the HERA generalization as well as the LOP effect.

An alternative approach to the role of PFC in long-term-memory, proposed by Nolde and colleagues (1998), suggests that the various component processes involved in encoding and episodic remembering (as well as in working memory, comprehension, problem-solving, etc.) are drawn from the same set of underlying cognitive processes. An example of such a component processing perspective is outlined by the multiple-entry-modular memory system, the so-called MEM-model of Johnson and Hirst (1993). This model distinguishes between perceptual processes (e.g., locating and identifying external targets) and reflective processes (e.g., processes that sustain/maintain, manipulate, revive, and evaluate information). Nolde and colleagues (1998) suggest that reflective processes are subserved by the PFC in interaction with other (posterior) neocortical regions. With respect to the HERA generalization they suggest that the right PFC subserves a variety of heuristic reflective processes (e.g., refresh activated information, shift between representations, register relations, such as whether an item matches a standard or criterion, and comparison of two stimuli on some relevant dimension), which are sufficient to support simple episodic memory tasks, but that more complex episodic memory tasks require additional systematic reflective processes mediated by the left PFC (e.g., rehearsing, initiating recursive strategies, and generating cues for retrieval of inactive information). Thus, Nolde and colleagues (1998) argued, the observed empirical generalization described by the HERA model arises because the encoding tasks that have been investigated so-far require on average more complex reflective processing than the retrieval tasks that has been investigated. Nolde and colleagues (1998) suggested that the specific processes supported by the left and right PFC might be best analyzed in the context of more

general component process architectures rather than as processes dedicated to any particular process, such as episodic encoding or retrieval. Fletcher and Henson (2001) reached a similar conclusion.

4.4.2.1 THE FRONTAL LOBE AND MEMORY FORMATION

Cognitive research on long-term memory indicates that several factors and modes of processing contribute to memory formation, including for example material, task and task instruction, as well as meaning-based (vs. surface based), associative relational and context processing, emotional significance and attention allocation. Moreover, functional neuroimaging studies have demonstrated that several types of processing interact to promote long-term memory formation. For example, meaning-based semantic generation, verbal working memory, and episodic memory encoding activate similar left PFC regions during active processing (Buckner & Koutstaal, 1998). One common denominator of these active on-line processes is the concept of working memory, which is likely to be engaged during retrieval of general knowledge, maintenance and further processing of the generated information (e.g., manipulation, selection, and organization), dynamically organizing different task objectives, as well as keeping encoding and retrieval strategies on-line.

Returning to the LOP effect and the framework formulated by Craik and Lockhardt (1972), which suggested that a deeper more elaborate and meaning-based processing of information yields more extensive associations with previously acquired general knowledge. Craik and Lockhardt (1972) thus hypothesized that the richness and the number of associations that results from the processing of the stimulus determine the stability (or durability) of the memory trace. Consequently, encouraging a processing strategy that leads to the formation of relatively more associations will prolong the lifetime of the memory trace, reduce its forgetting rate, and generate more associative access pathways for later retrieval. Subsequently Craik and Tulving (1975) provided data indicating that the LOP effect could not be explained in terms of task demand, that is, that the meaningful semantic encoding was simply more demanding (task difficulty) or time-consuming (time on task) compared to shallow processing. Another related hypothesis regarding the basis of the LOP effect suggests that the effect depends on the discriminability or distinctiveness of the memory trace relative to other memory traces (Baddeley, 1998; Craik & Tulving, 1975).

This reasoning suggests that recognition depends on the selection from any number of memory traces to match the retrieval cue. It follows that the likelihood of a memory trace being correctly selected is a function of its distinctiveness or discriminability. As already noted, another possibility is that the use of a linguistic representational format might be particularly efficient with respect to memory formation and subsequent retrieval.

Whether memory encoding is incidental or intentional appears to be of little consequence for the occurrence of the LOP effect (Kapur et al., 1994) and processing that engages meaning-based elaboration activates the left PFC specifically, independent of whether verbal (Kapur et al., 1994; Otten, Henson, & Rugg, 2001; Rugg, Fletcher, Frith, Frackowiak, & Dolan, 1997) or non-verbal material (Petersson et al., 2003) is used. Moreover, it seems that meaning-based instructions promote a mode of processing that enhances retrieval regardless of whether there is explicit semantic content in the stimulus material or not (Petersson et al., 2003; Vochatzer & Blick, 1989). For example, Vochatzer and Blick (1989) investigated the LOP effect using words and pseudowords and their results indicated that the LOP effect in the pseudoword condition was comparable to the effect in the word condition. Similarly, the LOP effect was similar for figurative and non-figurative line-drawings (Petersson et al., 2003).

The anterior parts of the left ventrolateral PFC have been associated with semantic working memory processes, including the retrieval, selection, maintenance, integration and evaluation of semantic knowledge, presumably represented elsewhere in neocortex (Bookheimer, 2002; Demb et al., 1995; Gold & Buckner, 2002; Hagoort, Hald, Baastiansen, & Petersson, 2004; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997), while the posterior parts of the left ventrolateral PFC have been related to phonological working memory (Bookheimer, 2002; Gold & Buckner, 2002; Wagner, 1999). Moreover, while the left PFC is more active during verbal working memory conditions, the right PFC is more active during visuo-spatial working memory conditions (D'Esposito et al., 1998). Owen and colleagues (1996) proposed a two-stage model of working memory, in which the ventrolateral and dorsolateral PFC regions mediate distinct working memory processes (cf., Simons & Spiers, 2003). They hypothesized that the ventrolateral PFC subserve maintenance and evaluation of information held on-line, while the dorsolateral PFC subserve monitoring and manipulation of the representations maintained in working

memory. Similarly, Simons and Spiers (2003) argue that the dorsolateral PFC organizes information to be remembered, while the ventrolateral PFC is related to semantic/phonological elaborative processing of MTL representations to ensure trace distinctiveness. Furthermore, memory formation is suggested to depend on perceptual processing in hierarchically organized posterior regions, resulting in more abstract representations that are integrated into a memory trace in interaction with the MTL. Thus, encoding control is supported by the PFC, involving elaborative processing of representations in the ventrolateral PFC, and information is selected, manipulated and organized in the dorsolateral PFC (Simons & Spiers, 2003). In the context of episodic memory formation, Simons and Spiers (2003) suggest that one role of the PFC is to ensure event separation (i.e., reduction of trace overlap) in order to reduce interference or cross-talk between memory traces.

Fletcher and Henson (2001), summarizes several ideas about the role of the (left) PFC in memory formation. They conceptualize the function of PFC in terms of working memory related processes regardless of whether the PFC is engaged in memory formation or retrieval; these, include: generation and retrieval of general knowledge (semantic information), maintenance and task-appropriate selection as well as organization the information to be encoded. They suggest that organization depends on selection, which depends on maintenance, and that maintenance depends on generation/retrieval from semantic memory. Hence, it is suggested that the contribution from meaning-based elaborated processing is derived from the anterior ventrolateral PFC, while selection is related to the dorsolateral and the posterior ventrolateral PFC and organization to the dorsolateral PFC (cf., Nyberg et al., 2003; Petersson et al., 1997; Petersson et al., 1999a; Petersson et al., 2001; Wagner, 1999).

4.4.2.2 THE FRONTAL LOBE AND MEMORY RETRIEVAL

Retrieving information from declarative long-term memory is a complex cognitive process that emerges from the interaction of an array of processes in order to reconstruct a representation of the retrieved information. Memory retrieval is thought to depend on the interaction between retrieval cues, supplied by the environment as in recognition or self-generated by a goal-directed retrieval attempt as in free recall, and the long-term memory

store, leading to the reconstruction of some aspects of a memory trace (Rugg & Wilding, 2000). Whether a retrieval attempt is successful or not is influenced by several factors, including the way information was encoded, the type of cues available, and the processes engaged during the retrieval attempt (Tulving, 1983). As already noted in this chapter, several investigators have emphasized the close connection between working memory, long-term memory retrieval and the PFC (Fletcher & Henson, 2001; Nyberg et al., 2003; Petersson et al., 1997; Petersson et al., 1999a; Simons & Spiers, 2003; Wagner, 1999). A general characterization of the functional role of different PFC regions in long-term memory retrieval has been outlined by Fletcher and Henson (2001, see also, Buckner & Wheeler, 2001; Simons & Spiers, 2003; Wagner, 1999), relating the ventrolateral PFC to the specification of memory search parameters, up-dating (bringing new information into working memory) and maintenance of retrieval cues; the dorsolateral PFC to monitoring, evaluation (verification), and processing (manipulation, selection) of the retrieved information; while the anterior PFC is thought to subserve control processing in terms of developing retrieval objectives, utilizing and coordinating retrieval (i.e., search) strategies as well as monitoring processes.

In order to understand the role of the frontal lobes in memory retrieval it is necessary to understand how PFC regions subserve executive control processes in general and how these processes control retrieval. The retrieval process can broadly be divided into several component processes: processing of retrieval cues; access to the memory store; re-instantiation of retrieved information in working memory, in which this information is maintained and subjected to further processing in terms of monitoring, evaluation, and selection; as well as higher-order control processing in terms of developing and utilizing retrieval strategies, specifying retrieval objectives and dynamic scheduling of different component processes as well as meta-mnemonic reasoning (cf., Buckner & Wheeler, 2001; Fletcher & Henson, 2001; Wagner, 1999). It has been suggested that the PFC interacts with posterior brain regions as well as the MTL during retrieval. By this account, the PFC support representations of retrieval cues that trigger reactivation of the cortical networks (including posterior regions) that represent the memory trace, while the MTL participates in these interactions (in particular for non-consolidated information). The MTL may be thought of as storing intermediate-term retrieval pointers, created by the rapid initial

binding of information during memory formation, which is subsequently consolidated in neocortical networks (cf., section 4.2 and 4.5). As information is being retrieved and represented in neocortical networks, it is suggested that the re-activation processes cascade backwards through the neocortical hierarchy, depending on the level of perceptual or motor detail that is required (Buckner & Wheeler, 2001; Rugg & Wilding, 2000). Thus, reactivation of the domain-specific memory contents engages different stages of perceptual and motor processing regions (Nyberg et al., 2000; Nyberg et al., 2001). Moreover, since memory content extends beyond perceptual and motor information, higher-level abstract representation (e.g., language mediated conceptual cognition, emotional significance, individual perspective, general world knowledge and model based cognition, etc.) supported by amodal/polymodal neocortical regions also become engaged. For example, Buckner and Wheeler (2001) suggest that it is likely that the PFC participates in the ongoing evaluation and integration of the information emerging during the process of retrieval attempt and that these processes are recursively engaged depending on retrieval success as well as retrieval objectives.

Several factors have been suggested to play a role in declarative memory retrieval, including retrieval mode, retrieval attempt, retrieval effort, the content of the retrieval process, retrieval objectives as well as retrieval success (cf., Rugg & Wilding, 2000). For example, Tulving (1983) suggested that a stimulus event is only treated as an episodic retrieval cue if the individual is in a particular cognitive state, the so-called retrieval mode, which thus constitutes a constantly maintained state that is necessary when there is need for episodic retrieval (Lepage et al., 2000). Retrieval effort refers to the level of processing resources utilized during retrieval attempt; retrieval success encompasses any process which depends on successfully retrieved information (Rugg & Wilding, 2000). Rugg and Wilding (2000) introduced an additional factor which they call retrieval orientation and suggest that retrieval orientation determines the specific form of the processing that is applied to a retrieval cue; retrieval orientation differ according to task requirements and the type of information to be retrieved.

The objective of retrieval attempt processes is to reconstruct previously encoded information from memory and several investigators have suggested that strategic, working memory related, aspects of a retrieval attempt is supported by the PFC. The PFC is also

likely to support conceptually oriented (language- or non-language based) processing as well as on-line integrative monitoring and maintenance related retrieval processes. Some of these proposed higher-order strategic aspects of retrieval are hypothesized to be related to the anterior frontopolar region (Buckner & Wheeler, 2001; Simons & Spiers, 2003), while posterior regions of the PFC have been related to more general-purpose cognitive processes. Buckner and Wheeler (2001) suggest the posterior PFC regions are recruited to the degree which they are engaged in retrieval, that is, as the retrieval more difficult, they will be recruited more extensively and will thus reflect the level of retrieval difficulty (so-called retrieval effort). Relative to the posterior PFC, the anterior PFC appears to be more selectively engaged in long-term memory retrieval. For example, the anterior PFC appears to be engaged in dynamic organization and scheduling of multiple sub-tasks (Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999), and based on this, Buckner & Wheeler (2001) proposed that some aspects of long-term memory retrieval might depend on dynamic navigation between retrieval cues, retrieval objectives, and reconstructions from long-term memory. This suggestion is clearly related to the idea of process complexity (Nolde et al., 1998) and context dependent retrieval processing (Wagner, Desmond, Glover, & Gabrieli, 1998). The context-dependent view of retrieval (Wagner, Desmond et al., 1998) suggests that the PFC supports several aspects of information processing during retrieval (including for example selection of retrieval strategies, initiation of retrieval search, and evaluation of information retrieved, as well as repeated initiation of retrieval attempts) and is consistent with the view that different prefrontal processing components are selected from the same set of underlying sub-processes but engaged differentially depending on task context and task complexity. However, several prefrontal regions seem to be co-activated relatively independent of the cognitive demand (Duncan & Owen, 2000; Nyberg, Forkstam, Petersson, Cabeza, & Ingvar, 2002; Nyberg et al., 2003). The PFC has also been related to retrieval success (Henson, Rugg, & Shallice, 2000; Rugg, Fletcher, Frith, Frackowiak, & Dolan, 1996); while the dorsolateral PFC regions might have a role in retrieval monitoring, the activity pattern of the anterior PFC region seems to be consistent with a retrieval success perspective (Henson et al., 2000). However, Wagner et al. (1998a) argued that their results were inconsistent with a retrieval success interpretation and Tulving et al. (1999) observed a negative correlation between recognition performance and PFC activation, both

in the anterior PFC and the posterior dorsolateral regions, casting some doubt on these suggestions.

In conclusion, as has become clear from this overview of PFC function, caution is prudent when interpreting the role of the PFC in long-term memory as well as in cognition more generally. Most conceptualizations of prefrontal functions are at present general in character, too general to specify with confidence the contribution of the PFC in cognition (Buckner & Wheeler, 2001; Simons & Spiers, 2003). These conceptualizations need to be developed into explicit accounts of the role of the various proposed PFC functions and their interplay. Moreover, these accounts need to indicate how these component processes and their interactions can be empirically characterized (i.e., measured).

4.5 NEOCORTICAL AND MEDIAL TEMPORAL LOBE INTERACTIONS

As emphasized through out this chapter, accumulating evidence suggests that long-term declarative memory formation and retrieval are supported by distributed functional networks of brain regions, including the PFC and the MTL (Eichenbaum, 2000; Simons & Spiers, 2003). Equally important to the specification of the separate contributions of these regions is an understanding of the interaction between these regions and several studies have shown that the PFC and the MTL are activated in parallel during various memory tasks (Eichenbaum, 2000; Simons & Spiers, 2003).

For information to be encoded, that is, transferred from an active working memory representation to a long-term memory trace, the information is processed and integrated in the neocortical hierarchy. Thus higher-level abstract representations are formed and subsequently bound into a MTL representation (e.g., index or pointer; cf., section 4.2). It is hypothesized that the MTL is involved in associative binding of distributed neocortical representations which are actively processed on-line and subsequently stored as a long-term declarative memory at the time of memory formation. The interaction with the PFC, presumably subserving working memory processes as outlined above, provides organization and control of the memory formation as well as the retrieval process (Buckner, Logan, Donaldson, & Wheeler, 2000; Buckner & Wheeler, 2001; Fletcher & Henson, 2001; Simons & Spiers, 2003; Wagner, 1999).

Recently Cabeza and colleagues (Cabeza, Dolcos, Graham, & Nyberg, 2002)

reported overlapping activations in the MTL related to episodic retrieval and verbal working memory. They argued, based on a review of the literature, that there are evidence linking MTL regions with working memory; for example, in both human (Holdstock, Shaw, & Aggleton, 1995; Murray & Mishkin, 1986; Owen, Sahakian, Semple, Polkey, & Robins, 1995) and nonhuman primates (Eichenbaum & Cohen, 2001; Zola et al., 2000) MTL lesions have been found to impair performance in trial unique working memory tasks with short retention intervals. There is also electrophysiological (Davachi & Goldman-Rakic, 2001; Suzuki, Miller, & Desimone, 1997), autoradiographical (Curtis, Zald, Lee, & Pardo, 2000; Sybriska, Davachi, & Goldman-Rakic, 2000), and functional neuroimaging evidence (Elliott & Dolan, 1999; O'Reilly, Braver, & Cohen, 1999; Ranganath & D'Esposito, 2001) indicating that the MTL is active during working memory tasks. However, the interpretation of these results is still unclear.

Preliminary investigations, using a network approach based on structural equations modeling (cf., chapter 3), indicate that the neocortex and the MTL might interact during short-term working memory tasks. More specifically, preliminary results indicate that the interaction between a neocortical verbal working memory network and the MTL is sensitive to task modulation (Petersson, Gisselgård, Gretzer, & Ingvar, in preparation). One possibility is to interpret this finding in the light of a recently introduced, fourth component of the Baddeley-Hitch working memory model, the so-called episodic buffer (Baddeley, 2000). The episodic buffer comprises a capacity limited system that provides temporary storage of information in a multimodal code, which is capable of binding information represented in the subsidiary systems (i.e., phonological loop, visuo-spatial sketch pad), and in long-term memory, into an episodic representational format (Figure 4.2). Similar concepts have been put forward in terms of long-term working memory (Ericsson & Kintsch, 1995) and working-with-memory (Moscovitch, 1992, 1994). Another possibility, suggested by Cabeza and colleagues (2002), is that rehearsal processes might involve the reactivation of the working memory representations transiently stored in the neocortex and that accessing these representations may engage the MTL. Similarly, it has been suggested that the MTL may serve as a convergence zone that rapidly store arbitrary associations or conjunctions of information and thus binding distributed neocortical representations that are active at the time of memory formation. For example, the PFC, and other neocortical

regions, may utilize briefly lasting parahippocampal representations to support on-line working memory processing (cf., Eichenbaum, 2000). One possibility is that such pointers, indexes or chunks (Wickelgren, 1979) are created by the MTL and might be used to access working memory traces transiently stored in neocortex. Following Miller (1991), it may thus be suggested that that working memory can be supported by MTL indexing mechanisms, when necessary, in order to access short-term memory representations. Since, the episodic buffer provide an interface between the slave systems of short-term working memory and long-term (or intermediate-term) memory, an interesting possibility is that the episodic buffer may be instantiated as an interaction between the PFC and the MTL, probably also including posterior cortical regions (Petersson et al., in preparation). This would suggest a more intimate relation between the MTL and working memory, although a prominent short-term deficit is not a typical feature in amnesic patients. However, as noted above, there are data indicating that MTL lesions can be paralleled by short-term recognition deficits in humans (Buffalo, Reber, & Squire, 1998; Holdstock et al., 1995; Owen et al., 1995) and functional neuroimaging evidence (Cabeza et al., 2002; Elliott & Dolan, 1999; O'Reilly et al., 1999; Ranganath & D'Esposito, 2001) indicating that the MTL is active during working memory tasks. A potential explanation may be that simple short-term memory tasks may not be sensitive enough to detect subtle short-term memory deficits, instead such deficits may be more pronounced if an additional distracting task or disturbing input is delivered. Consistent with this suggestion, Zarahn and colleagues (2004) speculate, based on lesion studies in human and non-human primates (Buffalo et al., 1998; Holdstock et al., 1995; Owen et al., 1995; Squire, Zola-Morgan, & Chen, 1988), that in the absence of a perceived possibility of distraction, memory over brief delays is independent of the hippocampus while this might not be the case in the presence of distraction. This hypothesis may be tested directly on patients with MTL lesions.

Along similar lines it may be suggested that the PFC–MTL interaction during retrieval to a large degree reflect the interaction between attentional processes, working memory, and long-term memory, in particular for relatively non-consolidated information. For example, Simons and Spiers (2003) suggested that the PFC-MTL interaction might be particular important for complex retrieval tasks such as recall with greater demands on retrieval organization in terms of search, generation, evaluation and possible further

processing in service of response organization. Moreover, studies of functional connectivity have provided support for the view that neocortical-MTL interaction is important for declarative memory retrieval (Köhler, McIntosh, Moscovitch, & Winocur, 1998; McIntosh, Nyberg, Bookstein, & Tulving, 1997). However, the precise characterization the patterns of neocortical-MTL interaction during both memory formation and retrieval from long-term memory remain to be worked out.

4.6 PRACTICE, WORKING MEMORY AND THE FRONTAL LOBES

Schneider and Shiffrin (1977) suggested that there is a qualitative difference between performance on novel tasks compared to well-practiced automatic tasks. Novel task performance is dependent on attentional resources and controlled processing, closely related to on-line working memory, to a greater extent than practiced automatic performance. The attentional and control processes, on which novel task performance is dependent, are characterized by flexibility, rapid establishability, and capacity limitations. With practice and increased automaticity, processing typically becomes faster, less variable, less sensitive to capacity limits, and more difficult to alter or inhibit; automaticity appears to develop gradually with practice (Cohen et al., 1990; MacLeod & Dunbar, 1988; Schneider, Pimm-Smith, & Worden, 1994). Thus, novel task performance is thought to depend more, and as performance becomes more automatic, less on attentional and working memory resources (Carr, 1992; Cohen et al., 1990; Petersson et al., 1999a; Raichle et al., 1994).

The ability to automate performance is important for complex task execution as it enables reallocation of limited attentional and control resources, and it enables learning of increasingly complex modes of processing by building upon previously acquired information and skills (Logan, 1988). Restructuring proposals suggest that the shift from controlled to automatic processing involves organizational changes in the sense of restructuring of processing pathways and that different sub-systems might be involved in automatic and controlled processing. Logan (1988) suggested that the transition from controlled to automatic processing represents a transition from algorithm-based to memory-based processing. In other words, performance will gradually come to depend on memory and adaptive changes as a result of practice. Alternatively, processing-based proposals

suggest that the processes involved in novel task performance are refined and become more effective as a result of practice. For example, the various sub-systems involved, may come to interact more efficiently as well as develop more efficient and appropriate representations for task performance (Jansma, Ramsey, Slagter, & Kahn, 2001; LaBerge & Samuels, 1974).

Presumably some aspects of the controlled processing are related to executive aspects of working memory supported by the PFC (Baddeley, 2003) and, as has already been outlined, attentional processes as well as working memory processes interact with learning and memory processes (Baddeley, 1998; Cohen et al., 1990; Schneider et al., 1994). We suggest that it is likely that the gradual transition from controlled to automatic processing is supported both by a restructuring of the functional processing architecture and the development of a more efficient processing infrastructure as well as representations of information. Functional neuroimaging studies have yielded support for both the restructuring and the processing efficiency perspective (Garavan, Kelley, Rosen, Rao, & Stein, 2000; Jansma et al., 2001; Petersson et al., 1999a; Raichle et al., 1994; Wiser et al., 2000). Overall, the general finding in these studies suggests a progressive reduction of the level of activation in attention and working memory related neocortical regions, including the dorsolateral PFC, the anterior cingulate/supplementary motor area, and posterior parietal and temporal regions, which putatively correspond to the transition from controlled to automatic processing as a consequence of practice. In addition, several studies have reported increased levels of activations with practice in domain specific posterior regions (Petersson et al., 1999a; Raichle et al., 1994; Wiser et al., 2000). This set of findings is consistent with the hypothesis, based on behavioral data, that aspects of controlled and automatic processing are supported in part by qualitatively or quantitatively different processing modes correlating with a restructuring of the functional processing architecture (cf., section 6.3).

5. CHARACTERISTICS OF ILLITERATE AND LITERATE COGNITIVE PROCESSING

Literacy and education represent essential aspects of contemporary society and subserve important aspects of socialization and cultural transmission. The study of illiterate subjects represents one approach to investigate the interactions between neurobiological and cultural factors and their influence on the outcome of cognitive development. Acquiring reading and writing skills as well as other cognitive skills during formal education can be viewed as an institutionalized cultural process and an important source for structured cultural transmission (cf., Figure 1.1). Important alternative approaches have also been explored with respect to cross-cultural variation, including the implications of transparent and non-transparent orthographies on brain function (Paulesu et al., 2000) and their consequences for the expression of dyslexia (Paulesu et al., 2001).

Reading and writing represent cognitive abilities that depend on human cultural evolution (Vygotsky, 1962). Varney (2002) emphasizes that reading and writing evolved through cultural developments and became typical acquired human abilities only within the last 200 years in Europe and America, and only after World War II in the rest of the World. In fact, reading and writing skills are still far from universal at the beginning of the 21st century. At present, it is estimated that there are close to one billion illiterate humans in the world, two thirds who are women (UNESCO, 2003, www.portal.unesco.org), while the mean educational level is only about 3–4 years of school (cf., the World Bank Report on 'Improving adult literacy outcomes: Lessons from cognitive research for developing countries' (Abadzi, 2003)). In this chapter, which is based on Petersson et al. (2005, in press; 2001), we will give an overview over some recent work comparing literate and illiterate cognition on a variety of experimental tasks. In particular we will focus on results from a series of experiments with an illiterate population and their matched literate controls living in southern Portugal and we review some recent cognitive, neuroanatomic, and functional neuroimaging results indicating that formal education influences important aspects of the human brain. Taken together this provides strong support for the idea that the brain is modulated by literacy and formal education. As a consequence, this changes the brains capacity to interact with its environment, including the individual's contemporary

culture: the individual's ability to participate in, interact with, and actively contribute to the process of cultural transmission in new ways through acquired cognitive skills.

Reading and writing skills do not represent a species wide adaptation of the kind that natural language is a paradigmatic example of. In contrast to language acquisition, which is largely a spontaneous, non-supervised, and self-organized process (cf., section 2.1.1), the learning of reading and writing skills typically requires great effort and focused training on the part of the individual. During the process of reading and writing acquisition, the child creates the ability to represent aspects of the phonological component of language by an orthographic representation and relate this to a visuo-graphic input-output code. This is commonly achieved by means of a supervised learning process (i.e., teaching; cf. section 2.1.1). Writing was a relatively late invention in human history, invented some 6,000 years ago, and it seems unlikely that specific brain structures have developed for the purpose of mediating reading and writing skills. Instead it may be suggested that these skills are supported by pre-adapted brain structures, brain structures that have evolved to serve specific functions but have come to serve as means for a different end.

Natural language is a system of knowledge, a system of representation and processing, as well as a system for communicative use (Chomsky, 1986). However, aspects of language can also be an object of cognition, so-called meta-cognition. Meta-linguistic awareness involves explicit processing and intentional control over aspects of phonology, syntax, semantics, discourse, as well as pragmatics. These processes are different from the implicit language processes used in natural language comprehension and production. In addition to the acquisition of language, children gradually develop explicit representations and acquire processing mechanisms that allow for reflecting and analyzing different aspects of language function and language use (Karmiloff-Smith, Grant, Sims, Jones, & Cuckle, 1996). Children do not learn language passively but actively construct representations on the basis of linguistically relevant constraints and abstractions of the linguistic input (Karmiloff-Smith et al., 1996). Meta-cognitive and meta-linguistic awareness develops progressively over the early years of life (Karmiloff-Smith, 1992). When children subsequently learn to read this has repercussions on the phonological representations of spoken language (Morais, 1993; Petersson et al., 2000). Rather than a simple one-way influence, several lines of research indicate that there is an intricate interplay between

meta-linguistic awareness and reading. Moreover, various types of meta-linguistic skills, including phonological awareness, correlate with literacy skills as well as levels of formal education (Ravid & Tolchinsky, 2002).

Literacy, reading and writing, as well as printed media represent extensive cultural complexes and like most cultural expressions they originate in human cognition and social interaction. Goody's work on literacy emphasizes the role which written communication has played in the emergence, development, and organization of social and cultural institutions (e.g., Goody, 2000). The emergence of writing, entailing the ability to preserve speech and knowledge in printed media, transformed human culture. This allowed societies with a literate tradition to develop and accumulate knowledge over time, and also, in general sense, greater opportunities to control the environment as well as living conditions. For example, the nature of oral communication has a considerable effect upon both the content and transmission of the cultural repertoire of a society; the content of the cultural traditions and knowledge has to be held in memory when written record is not an option. Instead, individual memory will mediate the cultural heritage between generations and new experience will be integrated with the old by a process of interpretation. The invention of new communication media have had significant impact on the way information is created, stored, retrieved, transmitted, and used, and by implication on cultural evolution as a whole. Moreover, reading and writing makes possible an increasingly articulate feedback and independent self-reflection as well as the development of other meta-cognitive skills; while auditory-verbal language use is oriented towards content, aspects of this knowledge can become explicitly available to the language user in terms of cognitive control and analytic awareness. It has thus been suggested that the acquisition of reading and writing skills, as well as formal education more generally, facilitates this by a process of representational construction and reorganization (Karmiloff-Smith, 1992). Ravid and Tolchinsky (2002) suggest that meta-linguistic development is catalyzed by the acquisition of literacy and school-based knowledge; the acquisition of written language skills promotes flexible and manipulable representations for meta-cognitive use (Karmiloff-Smith, 1992).

5.1 THE STUDY POPULATION OF SOUTHERN PORTUGAL

The fishermen village Olhão of Algarve in southern Portugal, where all of our studies have been conducted, is socio-culturally homogeneous and the majority of the population has lived most of their lives within the community. Mobility within the region has been limited and the main source of income is related to agriculture or fishing. Illiteracy occurs in Portugal due to the fact that forty or fifty years ago, it was common for older daughters of a family to be engaged in the daily household activities instead of being sent to school. However, later in life they may have started to work outside the family. In larger families, the younger children were generally sent to school when they reached the age of 6 or 7 while the older daughters typically helped out with the younger siblings at home.

Literate and illiterate subjects live intermixed in this region of Portugal and participate actively in their community; illiteracy is not perceived as a (functional) handicapped and the same socio-cultural environment influences both literate and illiterate subjects on similar terms. Some of the literate and illiterate subjects in our studies are from the same family and thus increasing the homogeneity in background variables. Furthermore, most of the literate subjects participating in our studies are not highly educated and most often they have had approximately 4 years of schooling. In the context of our research, it is important to ensure that the subjects investigated are not cognitively impaired and also that the illiterate are matched to the literate subjects in as many relevant respects as is possible (except for the consequences of not having had the opportunity to receive formal education). In our studies we have attempted to match the different literacy groups as far as possible in terms of several relevant variables, including for example age, sex, general health, socio-cultural background, and level of everyday functionality. The literacy groups are comparable along socio-economic dimensions as well. For a more detailed characterization of our study population and our selection procedures see Reis, Guerreiro, & Petersson (2003). These protocols and procedures ensures with reasonable confidence that the illiterate subjects are cognitively normal, that their lack of formal education results from specific socio-cultural reasons and not due to, for example, low intelligence, learning disability, or any other pathology potentially affecting the brain.

5.2 COGNITIVE-BEHAVIORAL FINDINGS

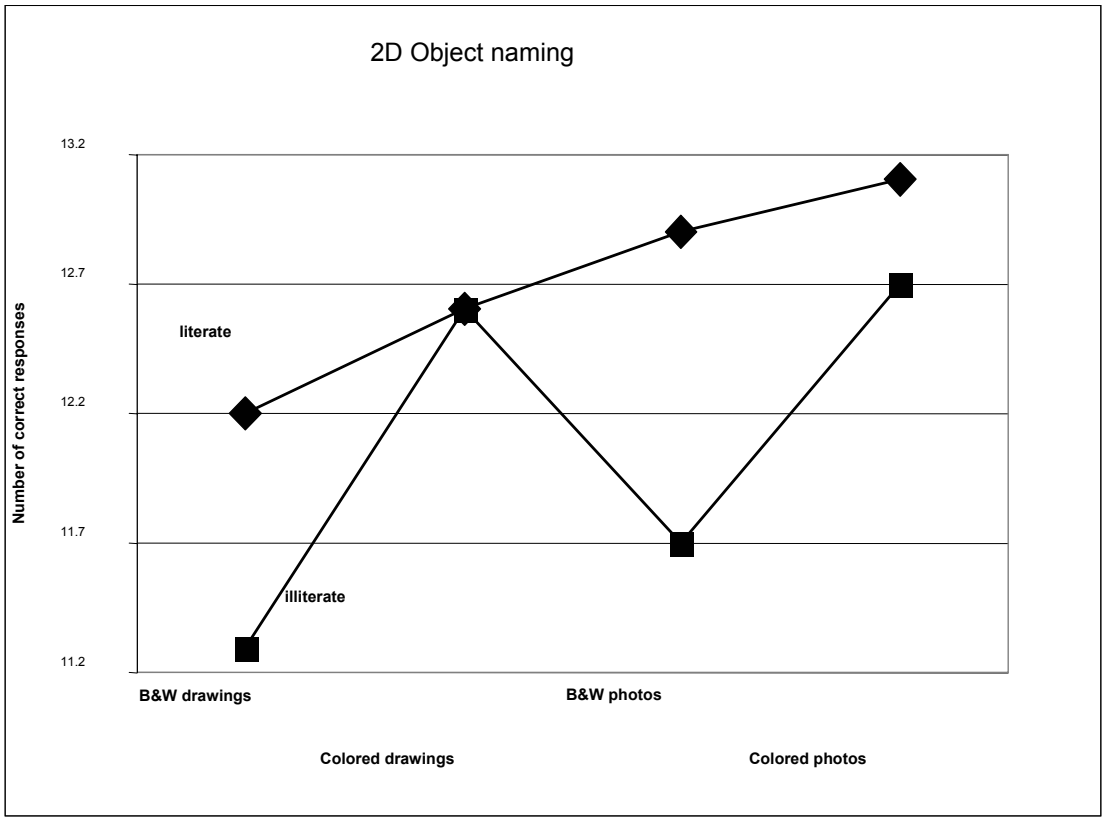
Cognitive-behavioral studies have demonstrated that literacy as well as the extent of formal education influence the performance of several behavioral tasks commonly used in neuropsychological assessment (e.g., Ardila, Rosselli, & Rosas, 1989; Lecours, Mehler, Parente, Aguiar et al., 1987; Lecours, Mehler, Parente, Caldeira et al., 1987; Manly et al., 1999; Rosselli, Ardila, & Rosas, 1990). For example, behavioral results indicate that the acquisition of written language skills significantly modulates the auditory-verbal language system (e.g., Mendonça et al., 2002; Morais, 1993; Morais & Kolinsky, 1994; Reis & Castro-Caldas, 1997; Silva et al., 2002). Additional data indicate that other cognitive functions are influenced as well, for example, visuo-spatial skills (e.g., Kremin et al., 1991; Manly et al., 1999; Ostrosky, Efron, & Yund, 1991; Reis, Guerreiro, & Castro-Caldas, 1994; Reis, Petersson, Castro-Caldas, & Ingvar, 2001; Rosselli et al., 1990). However, it is still unclear which processes and mechanisms mediate these effects of literacy and formal education. A detailed understanding of which parts of the cognitive system and which processing levels are affected is still lacking. In this section we will focus on some aspects of object naming, short-term memory, phonological processing and word awareness in spoken sentence context, as well as semantic memory organization and semantic processing. The basic idea is that literacy influences some aspects of auditory-verbal language processing related to phonological processing and verbal short-term working memory as well as visuo-motor skills related to reading and writing.

5.2.1 OBJECT NAMING

Literacy influence the performance when naming 2D pictorial representations of objects (e.g., Kremin et al., 1991; Manly et al., 1999; Reis et al., 1994; Rosselli et al., 1990). The performance on simple object naming tasks is mainly dependent on the systems for visual recognition, lexical retrieval, and the organization of articulatory speech output, as well as the interaction between these systems (Levelt, 1989). In our study population, learning and practice in interpreting schematic 2D representations most often took place in school simultaneously with the acquisition of written Portuguese. It is of course also the case that reading and writing depend on advanced visual and visuo-motor skills in coding, decoding, and generation of 2D representations. It is therefore likely that the interpretation and

production of 2D representations of real objects as well as the coding and decoding 2D material in terms of figurative/symbolic semantic content is more practiced in literate than in illiterate individuals, who generally have received little systematic practice in interpreting conventional visuo-symbolic representations. We thus speculated that there may be differences in 3D and 2D object naming skills between literate and illiterate individuals. This speculation was followed up in a simple visual object naming experiment in which the participants named common everyday objects (Reis, Petersson et al., 2001). Reis et al. (2001) reported differences between literate and illiterate subjects related to 2D object naming but found no difference when subjects named real 3D objects, both with respect to naming performance and in terms of response times. In addition, the two groups dissociated in terms of their error patterns, with the illiterate group more prone to make visually related errors (e.g., pen instead of needle), while the literate group tended to make semantically related errors (e.g., necklace instead of bracelet).

Though the results on 2D line drawings and real objects were clear in the study of Reis et al. (2001), the result on colored photos did not clearly dissociate between the literacy groups in terms of 2D vs. 3D naming skills. We therefore speculated that the semantic significance of object color might play a role, in particular for the illiterate subjects, since they are prone to be driven by semantic rather than formal aspects of stimuli or information, a theme we will return to in subsequent sections (cf., e.g., section 5.2.3). In a recent follow-up study, using a similar experimental set-up as Reis et al. (2001), we presented common everyday objects as black and white (i.e., grey scaled) as well as colored drawings and photos in an immediate 2D object naming task. Consistent with the results outlined above the literate group performed significantly better than the illiterate group on black and white items (i.e., both line drawings and photos). In contrast, there was no significant difference between literacy groups on the colored items (Figure 5.1). Interestingly, the illiterate participants performed significantly better on colored line drawings compared to black and white photos. Preliminary investigations also indicate that the color effect is related to the semantic value of the color in the sense that the effect seems more pronounced for objects with no or little consistency in the color-object-relation compared to objects with a consistent relation to its color (e.g., lemons are yellow).



[Figure 5.1] Simple immediate object naming of common everyday objects. The 2D stimuli included black and white (B&W) as well as colored line drawings and photos of common everyday objects. The literate subjects performed significantly better than the illiterate group on black and white items (B&W line drawings: $P = 0.009$; B&W photos: $P < 0.001$). In contrast, there was no significant difference between literacy groups on the colored items (colored photos: $P = .21$). The illiterate participants performed significantly better ($P = 0.02$) on colored line drawings compared to black and white photos.

In summary, the absence of group differences when naming real 3D objects, and in particular the absence of response time (RT) differences on correctly named real objects indicate that the RT differences on drawings and photos is not simply related to slower visual or language processing in general. Instead, the longer processing time in the illiterate group appears to be related to the processing of 2D visual information or the interaction between lexical retrieval and the processing of 2D visual information. The latter possibility

would suggest that the interface between the two systems is configured differently in the two literacy groups, leading to differences in the effectiveness of the necessary information transfer between the two systems. The result of the error analysis is consistent with this interpretation, since the illiterate subjects made relatively more visually related errors than language related while the pattern was the opposite for the literate group. In fact the qualitative distribution of errors was not significantly different for real object naming between groups. Taken together this interpretation is consistent with a recent suggestion that orthographic knowledge is an integral component of the general visual processing system (Patterson & Lambon Ralph, 1999) indicating that the acquisition of alphabetic orthographic knowledge may affect specific components of visual processing. A positive correlation between reading abilities and the capacity to name line drawings have also been reported (Goldblum & Matute de Duran, 2000). Recent findings also indicate that color can play an important role for the illiterate group, when naming 2D pictorial representations of common everyday objects. This seems so when the semantic value of the color of an object is prominent (Reis, Petersson, Faísca, & Ingvar, in preparation).

5.2.2 SHORT-TERM WORKING MEMORY AND PHONOLOGICAL PROCESSING

Our previous investigations of our study population have indicated that the acquisition of reading and writing skills influences aspects of the auditory-verbal language system. In particular, aspects of sub-lexical phonological processing appear to differentiate the two literacy groups. This is most prominently expressed in terms of phonological awareness, the most well-accepted difference between schooled and unschooled individuals that does not depend on educational level as such (Coppens, Parente, & Lecours, 1998). Previous results have also indicated that there are differences in phonological loop interactions between literate and illiterate subjects related to the inferior parietal cortex (Petersson et al., 2000) and it was recently suggested that the phonological loop (cf., Figure 4.2) might serve as a language learning device, with an integral role in the systems for spoken and written language acquisition (Baddeley, Gathercole, & Papagno, 1998).

The relation between literacy and so-called phonological awareness has been investigated since Morais et al. (1979) indicated that illiterate subjects have some difficulty

in dealing with tasks requiring explicit phonological processing. The results of Morais et al. (1979) showed that illiterate subjects found it more difficult to add or remove phonemes in the beginning of words and pseudowords. However, these tasks may be of different ecological validity for literate and illiterate individuals complicating the interpretation of the finding (cf., discussion in e.g., Reis & Petersson, 2003; Silva, Petersson, Ingvar, & Reis, 2001; Silva, Petersson, Faísca, Ingvar, & Reis, 2004). It is still an open question what type of relation exists between phonological processing, verbal working memory, and the acquisition of orthographic knowledge. Moreover, it may be suggested that the phonological processing difficulties in illiterate subjects are not limited to phonological awareness but involve other aspects of sub-lexical phonological processing and skills related to verbal (phonological) working memory (e.g., phonological recoding in working memory). There is some evidence indicating that these effects may be specific to alphabetic orthographies and may not necessarily generalize to non-alphabetic orthographies.

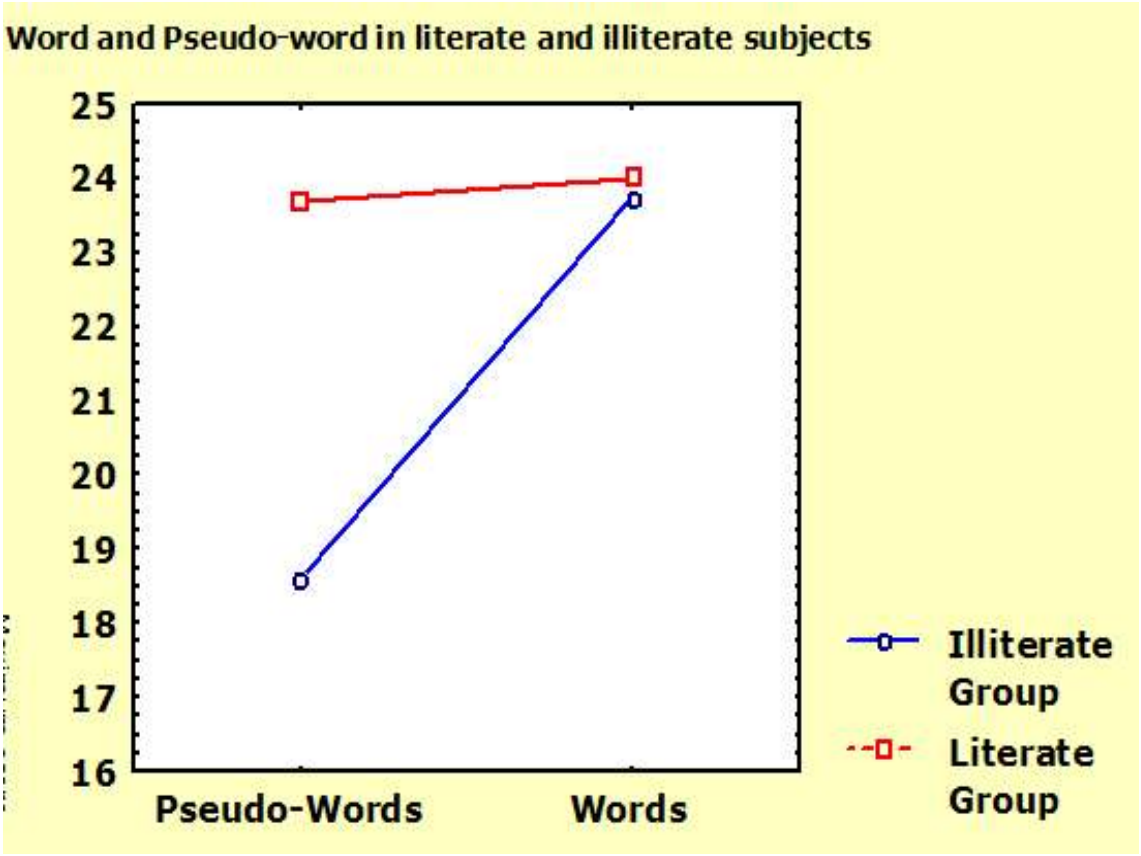
Pseudoword repetition and digit span tasks are good measures of verbal working memory capacities and these measures have also been related to reading achievements in children (Gathercole, 1995a, 1995b, 1995c; Gathercole & Baddeley, 1995). Additional research points toward a role of verbal working memory and the efficiency of phonological processing in relation to reading skills (Brady, 1991). Several studies have indicated that there is a difference in digit span between literate and illiterate individuals (e.g., Ardila et al., 1989; Garcia & Guerreiro, 1983; Reis, Guerreiro, Garcia, & Castro-Caldas, 1995). In a recent study by Reis et al. (2003) it was shown that the difference in digit span is not a simple effect of literacy as such but the digit span performance appears to be dependent on other factors as well as including the extent of formal education. In particular, illiterate participants had a mean digit span of 4.1 (± 0.9) performing significantly lower than literate participants. However, also literate subjects with 4 years of education (5.2 ± 1.4) performed significantly lower than literate subjects with 9 years of education (7.0 ± 1.8). Thus it appears that not only literacy but education more generally contribute to the observed difference (overall effect $P < 0.001$). In a recently completed follow-up study we compared 19 literate (4.4 ± 1 years of schooling) and 19 illiterate participants (mean age 66.2 ± 7 and 68.9 ± 4 years; non-significant $P = 0.1$) directly on the digit span and spatial span sub-tasks of the Wechsler Memory Scale (III revision). Consistent with the results just described

there was a significant difference between literacy groups on the digit span ($P = 0.004$) while there was no significant difference on the spatial span task ($P = 0.3$). Thus, literate and illiterate subjects appear to dissociate in terms of performance on verbal- but not on spatial span tasks. This is of interest since the illiterate group performs less well on immediate 2D object naming but not on 3D naming compared to literate subjects. These results are thus a first indication that verbal short term memory is specifically influenced by literacy and formal education, possibly related to more effective verbal working memory representations in literate individuals (e.g., chunking, cf. e.g., Olesen, Westerberg, & Klingberg, 2004).

The second task commonly used task to investigate verbal working memory capacity is pseudoword repetition (Gathercole, 1995a, 1995b, 1995c; Gathercole & Baddeley, 1995). Reis et al. (1997) concluded that illiterate performed similarly to literate subjects on word repetition, while there was a significant difference on pseudoword repetition ($P < 0.001$; Figure 5.2). We have suggested that this is related to an inability to handle certain aspects of sub-lexical phonological structure and also indicates that the phonological representations or the processing of these representations are differently developed in literate and illiterate individuals (Petersson et al., 2000; Petersson et al., 2001). Taken together these results indicate that there is a relation between the acquisition of reading and writing skills and aspects of phonological processing. Alternatively, the system for orthographic representation may support phonological processing as an auxiliary interactive network (Petersson et al., 2001).

Because several aspects of auditory-verbal language may differ between literate and illiterate subjects it is of interest to isolate the different sources contributing to these differences between literacy groups in phonological processing. In particular, it is important to study the differences in phonological processing relatively independent of lexicality effects (e.g., vocabulary size and frequency effects) as well as articulatory mechanisms. In order to do so we used an immediate auditory-verbal serial recognition paradigm (Gathercole, Pickering, Hall, & Peacker, 2001) in a recent follow-up study (Petersson et al., manuscript in preparation). In general, immediate serial recognition is independent of speech output and serial recognition of pseudowords is (relatively) independent of lexicality effects. In this experiment we compared illiterate and literate subjects on

immediate recognition of lists of 3 CVCV-syllable items (C = consonant, V = vowel). The lists varied in lexicality (words/pseudowords) and phonological similarity (dissimilar/similar) and the participants to judge whether two lists (presented one after the other) contained items presented in the same or different order.



[Figure 5.2] Immediate verbal repetition. Literate and illiterate subjects repeated common words and pseudowords constructed from the words, by changing the consonants, thus preserving the length and syllable structure of the words.

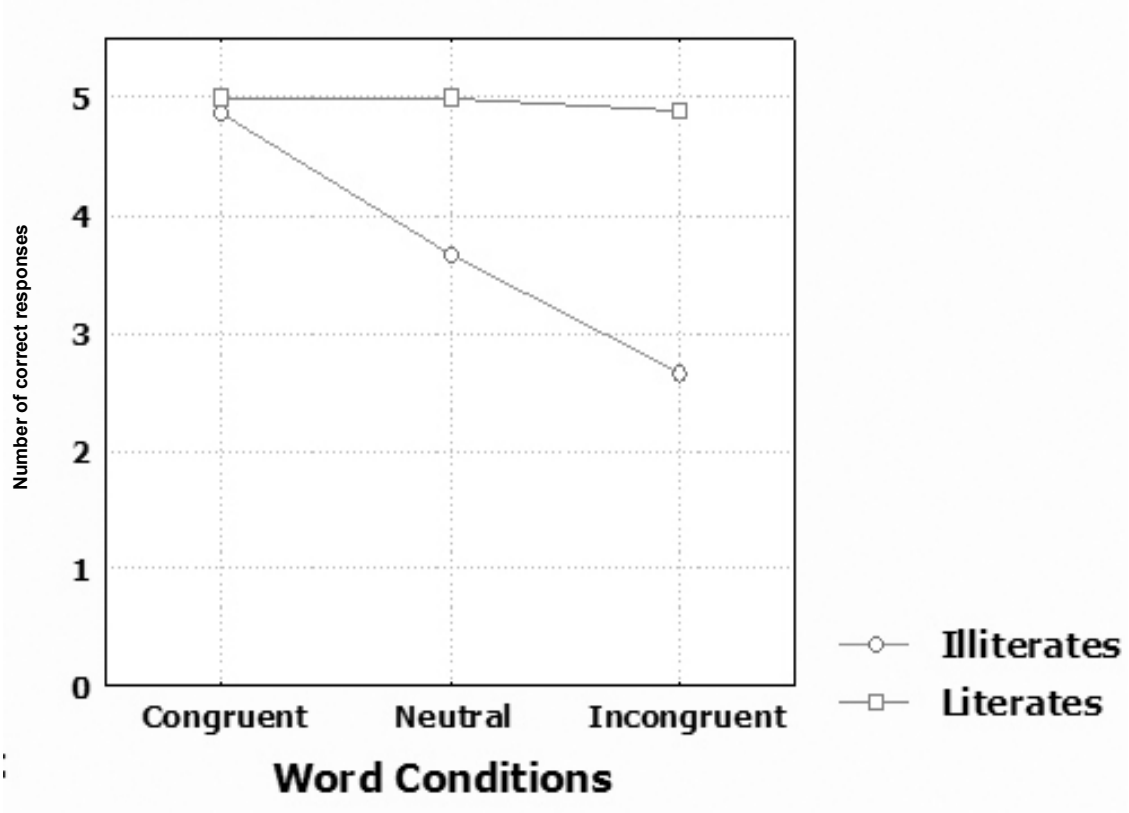
Group comparisons indicated that the literate group performed better than the illiterate in all conditions (pseudoword/dissimilar $P < 0.001$; pseudoword/similar $P = 0.03$; word/similar $P = 0.003$) except recognition of phonologically different words ($P = 0.2$). Of the four different conditions, the phonologically different word condition is of course the easiest to handle from a phonological point of view; on the one hand words are more familiar than pseudowords, and on the other hand, the phonological contrast is greater in

the different compared to the similar condition. These results are thus consistent with the differences in pseudoword repetition (literate > illiterate) and digit span performance and indicate that there are differences in verbal working memory performance between literacy groups. In addition, the results on immediate serial recognition indicate that these differences are independent of lexicality effects, articulatory organization (e.g., output phonology) or other speech output mechanisms.

5.2.3 AWARENESS OF PHONOLOGICAL FORM AND THE INTRUSION OF LEXICAL SEMANTICS

A characteristic of problem solving capabilities in illiterate individuals is their tendency to prefer semantic-pragmatic strategies if such are possible. More specifically, when an illiterate individual is confronted with a problem that can be solved by using strategies based on formal/abstract or semantic/pragmatic aspects the illiterate individual is likely to base his or her strategy on the latter type of information. For example, Kolinsky et al. (1987) investigated the notion of phonological word length in literate and illiterate subjects when asked to attend to abstract phonological properties of words the illiterate group found it difficult to ignore their semantic content. In other words, the illiterate group showed difficulty in inhibiting the intrusion of semantic information when attempting to solve the task based on form criterion. This suggests that explicit awareness of words as phonological form may depend on orthographic knowledge or more generally on formal education.

In a recent experiment literate and illiterate participants listened to words and pseudowords (Silva et al., 2002) during a phonological ('sound') length decision task, in which the participants were asked to decide which item in a pair was the longest in phonological terms. In the word condition we manipulated the relationship between word length and size of the denoted object yielding three sub-conditions: 1) *Congruent* - the longer word denoted the larger object; 2) *Incongruent* - the longer word denoted the smaller object; 3) *Neutral* - only phonological length of the words varied, denoting objects of similar size. Pseudowords pairs were constructed based on the real words pairs by changing the consonants and maintaining the vowels as well as word length. Each subject practiced each condition until the subject fully understood the task.



[Figure 5.3] The literate group performed significantly better than the illiterate group on both words ($P < 0.001$) and pseudowords ($P = 0.001$). The results between the different word conditions (i.e., congruent, neutral, vs. incongruent) showed a significant effect in the illiterate group ($P < 0.001$). There was no significant difference between word (collapsed over conditions) and pseudoword performance in the literate group ($P = 0.3$). In contrast, the illiterate group showed significantly better performance on pseudowords compared to words ($P = 0.01$).

Two effects were of interest in the results. Firstly, the literate subjects showed no effect of semantic interference while this was clearly the case in the illiterate group (Figure 5.3). Secondly, while the literates performed at similar levels on words and pseudowords, the illiterate group performed significantly better on pseudowords compared to words. In fact, the mean performance in the pseudoword condition was slightly better than in the neutral word condition. Thus, as predicted, these results indicate that the illiterate subjects show a

greater difficulty in inhibiting the influence of semantic interference, that is, the intrusion of lexical semantics in the decision process.

5.2.4 AWARENESS OF WORDS IN A SENTENCE CONTEXT

Little is known about how adult illiterate subjects perceive words in the context of a sentence. Awareness of words as independent lexical units has been investigated in children, both before and after acquiring reading skills (e.g., Barton, 1985; Hamilton & Barton, 1983; Karmiloff-Smith et al., 1996), and also in illiterate adults (Cary & Verhaeghe, 1991). The results of these studies suggest that explicit knowledge of words as independent lexical units is to some degree dependent on literacy. Cary & Verhaegh (1991) suggested that the difficulty for illiterate subjects to efficiently identify closed-class words because of their relative lack of semantic content. However, given the prominent syntactic role of closed-class words in sentence processing, including sentence comprehension, and the fact that illiterate and literate individuals acquire spoken natural language on similar terms, we were interested in whether the effects related to closed-class words could be given a phonological explanation. In two recent studies we revisited these issues (Mendonça et al., 2002). In the first study, we investigated the awareness of words in the context of sentences with the aim of clarifying the role of literacy in the recognition of words as independent lexical units and the possible relation to the known phonological processing characteristics of illiterate subjects. We used short sentences, presented in random order to the participants, that varied in their constituent structure. All articles, prepositions, pronouns, and adverbs were included in the closed-class category and we divided this class into phonologically stressed and non-stressed words, where the latter are characterized by the absence of a stressed vowel in contrast to the former. Each sentence was orally presented and subjects were instructed to listen to the sentence, to immediately repeat it after the presentation, and to identify its constituent words by enumerating them. All spontaneous corrections were considered and after the experimental session subjects were asked to correct three of the incorrect segmentations. The behavioral data were scored according to the following: A) *Global quantitative*: (1) total of correct sentence segmentation (maximum score: 18); (2) spontaneous corrections; and (3) corrections made when probed; B) *Segmentation errors*: (1) blending (a so-called ‘clitization’ phenomena)

words in the boundaries of sentence's main constituents and (2) blending words within phrases; C) *Omissions* of stressed and non-stressed closed-class words: (1) and (2) non-stressed closed-class words.

Table 1. (a) Means and standard deviations for sentence segmentation scores (maximum = 18; between-group Mann-Whitney U Test). (b) Mean and standard deviations of proportions of segmentation errors committed internal to the phrase type by the illiterate group.

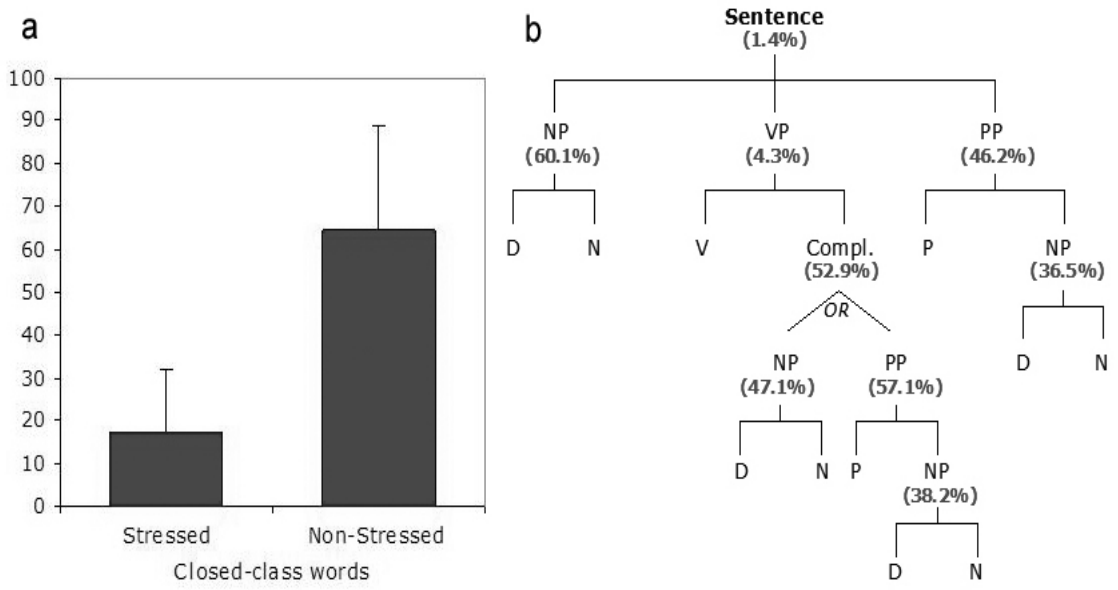
Behavioral Measure	***Illiterate***	***Literate***	***P-value***
Correct sentence segmentation	3 ± 2.9	17 ± 2.0	<0.001
Spontaneous corrections	0.1 ± 0.2	1 ± 1.3	0.001
Percentage of questions corrected	19 ± 33	80 ± 45	0.01

It is clear from Table 1, that the literate group performed significantly better compared to the illiterate on the sentence segmentation task and the results show that illiterate subjects did not spontaneously correct themselves, not even when probed. For all error types, the group comparison investigated showed significant differences.

Table 2. Mean and standard deviations of proportions of errors committed in the internal composition of phrase by the illiterate group.

Blending internal constituents of phrases	***Percentage***
Determiner + noun	51 ± 22
Preposition + determiner + noun	18 ± 19
Preposition + determiner	14 ± 13
Preposition + noun	77 ± 28
Contraction + noun	62 ± 28

Therefore, and in order to further understand the behavioral pattern of illiterate group, the subsequent error analysis focused on this group only. In order to compare the incidence of the different error types, percentage of errors was computed based on the total number of possible occurrences for each type. The illiterate group showed a specific pattern of merging or 'clitization' of words (Table 2 and Figure 5.4). There are very few mergers between the major syntactic constituents (1.4% error rate), meaning that illiterates are sensitive to the major syntactic structure of the sentence, as expected.



[Figure 5.4] (a) The proportion of errors related to closed-class words. The closed-class words were either phonologically stressed or non-stressed. The illiterate subjects committed significantly more segmentation errors related to the non-stressed (64 ± 24%) compared to the stressed closed-class words (17 ± 14%; Wilcoxon $P < 0.001$). (b) The proportion of segmentation errors related to the phrase structure of the sentence.

This was also the case for syntactic boundaries within verb phrases (4.3%). Increasing rates of merging were observed within phrase internal constituents related to noun- (NP) and prepositional phrases (PP), but this seemed to depend on the particular syntactic context or alternatively on the linear sentence position. The words of NPs in subject position were more frequently merged (60%) compared to NPs within VPs or PPs in complement position (47% and 37%, respectively). Within the PPs composed of a preposition or contraction and a noun, the illiterates committed the highest rate of mergers; in differently composed PPs mergers were less frequent.

The closed-class word analysis revealed that illiterate subjects were unable to correctly segment 50% of the instances. Comparing the stressed and the non-stressed closed-class words showed that the merging tendency was significantly more prominent for the non-stressed closed-class words (Figure 5.4). In a recent follow-up study (Mendonça et al., 2003), using a similar experimental design, these effects were replicated. In brief, while there was no significant difference in sentence repetition ($P = 0.7$), the literate sentence segmentation performance was significantly better than the illiterate ($P < 0.001$). The mergers observed in the illiterate group related to closed-class words was observed significantly more often with non-stressed compared to stressed closed-class words in the illiterate group ($P < 0.001$; Table 3).

Table 3. Mean scores and standard deviation of proportion of errors committed with closed-class words (both phonologically stressed and non-stressed) in the illiterate group (within-group comparisons Wilcoxon).

	Stressed	***Non-stressed***	***P-value***
Closed-class words	17% ± 14	64% ± 24	< 0.001

More detailed preliminary analysis indicates that the merging effect is dependent on the type of closed-class, that is, the stressed vs. non-stressed effect was most common for determiners and least common for prepositions. Overall then, the present results corroborate previous suggestions that recognition of words as independent phonological

units in sentence context depends on literacy. Cary & Verhaegh (1991) suggested that the difficulty observed in illiterate subjects is related to a difficulty in efficient identification of closed-class words due to their relative lack of semantic content. However, the present results show that this cannot serve as a unitary explanation since the segmentation failures did not distribute evenly over closed-class words, not even within sub-types, but occurred more often with phonologically non-stressed than phonologically stressed closed-class words. The illiterate subjects are thus more sensitive to phonologically stressed closed-class words which they are able to segment quite efficiently. Instead, we suggest that illiterate segmentation performance is closely related to sentence internal prosody and phonological stress. Thus, the difficulty seems to be a phonological phenomenon rather than related to lexical semantics per se. In addition, the 'clitization' phenomenon seem not to be related to phrase structure per se since the illiterate group respected phrasal boundaries, that is, blending mainly occurred within phrases and rarely across phrasal boundaries or boundaries between major sentence constituents (e.g., specifier, verb and complement). Another contributing factor to segmentation difficulties may be verbal working memory capacity, since the performance of the illiterate group increased from the start to the end of sentences. In other words, also the linear sentence position may play a role. In summary, illiterate word segmentation of sentences appears to depend on factors related to phonology, syntactic structure, and linear position, but appears to be unrelated to lexical semantic.

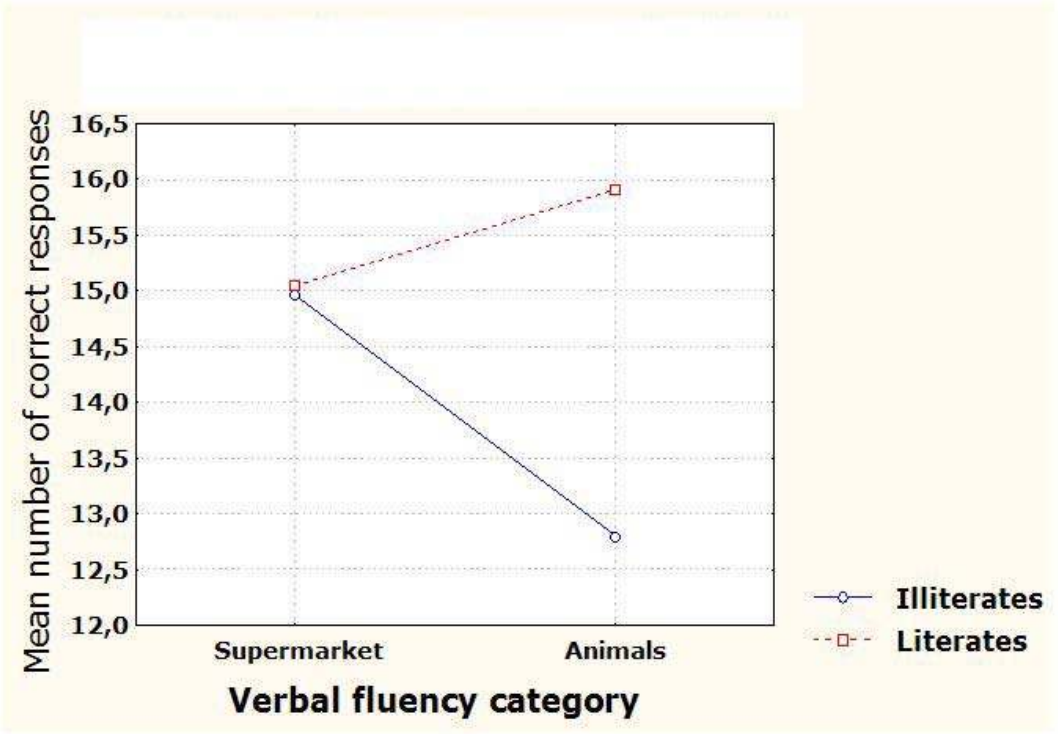
5.2.5 SEMANTIC FLUENCY AND THE IMPORTANCE OF ECOLOGICAL RELEVANCE

Literacy and formal education has also been associated with the capacity to acquire a broader base of general information as well as the capacity to process this information in a more abstract and systematic manner. Hence, literacy and formal education catalyze the development of several cognitive skills in addition to reading and writing skills. Task selection is thus of importance when investigating populations with different cultural backgrounds. In particular, when the objective is to relate differences in performance between populations it is important that the tasks investigated are of comparable ecological relevance to the study populations involved. This goes beyond matching populations for

background variables related to socio-economic status (cf. e.g., Coppens et al., 1998; Reis & Petersson, 2003). A clear illustration of this are the results reported in a recent study of semantic fluency by Silva et al. (2004).

Verbal fluency tasks (i.e., production tasks in which subjects generate as many words as possible during e.g. 1 min according to some given criteria) are commonly used in neuropsychological assessment since they are easy to administer, sensitive to brain damage and cognitive deterioration. Clear and consistent differences between literacy groups have been reported when a phonological fluency criterion was used (for a recent review see Silva et al., 2004). In contrast, several studies, comparing literate and illiterate subjects on semantic criteria have yielded contradictory results. At present the reasons for this are unclear but might be related to the specific semantic criterion used and/or the particular study populations investigated. Reis et al. (2003) suggested that the non-convergence of results could be related to the ecological or cultural relevance of semantic criterion used. In order to investigate this issue further, Reis et al. (2001; 2003) decided to use a semantic criterion of equal natural relevance to female literate and illiterate subjects and asked the participants to name things one can buy at the supermarket. The relevance of this criterion springs from the fact that almost all of these individuals do the major part of their regular shopping at supermarkets and at comparable levels over time. Reis et al. (2001; 2003) found no significant difference between illiterates, subjects with 4 years of education, and subjects with more than 4 years of education.

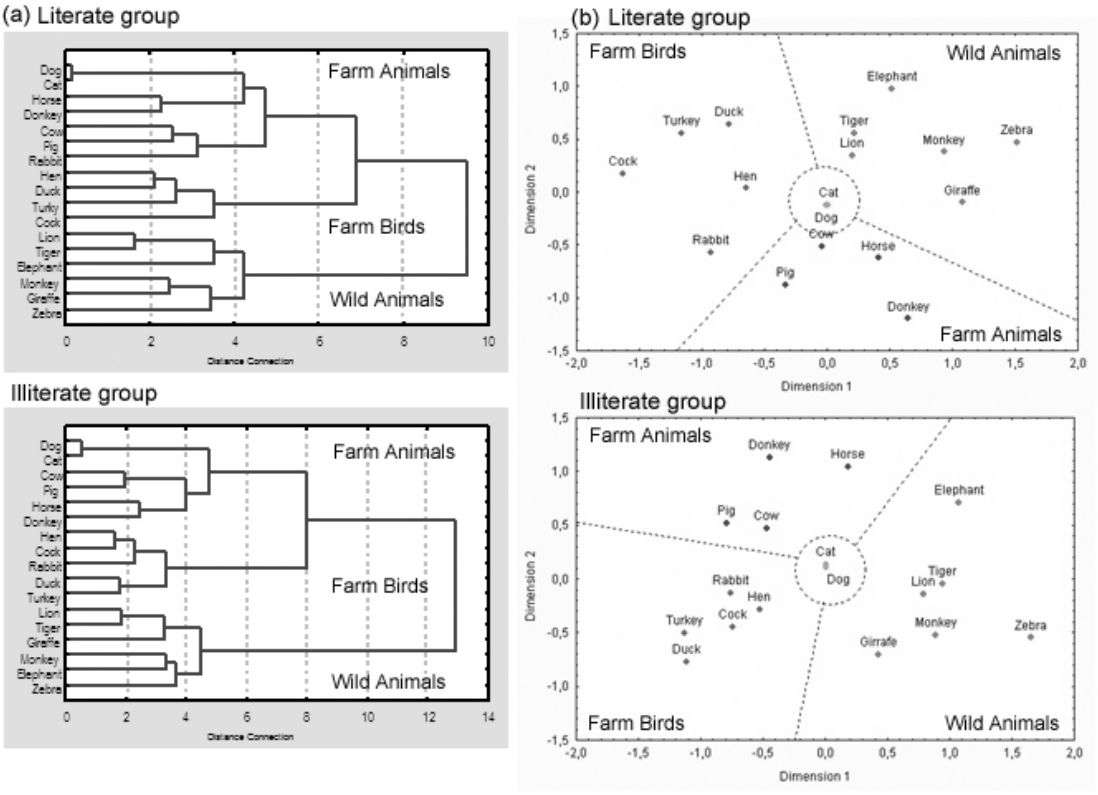
Silva et al. (2004) attempted to relate the concept of ecological relevance to the level of shared cultural background except for differences in literacy or formal education. More specifically, Silva et al. (2004) compared the performance of the same illiterate and literate subjects on two time-constrained semantic fluency tasks, the first using the semantic category of food items that can be bought at the supermarket (supermarket fluency task), and the second, animal names (animal fluency task, Figure 5.5). The equal performance on the supermarket task exclude a simple explanation for the performance differences on the animal fluency task (literate > illiterate) in terms of general factors such as cognitive speed or fluency. Instead, the interaction between literacy and semantic criterion might be explained in terms of similarities and differences in shared cultural background, that is, greater for supermarket items and lesser for animals.



[Figure 5.5] Semantic fluency. Two time-constrained semantic fluency tasks, which were identical except for the semantic criterion used, were investigated in literate and illiterate subjects. The first used the semantic category of food items that can be bought at the supermarket, and the second, animal names. The literate group performed significantly better on the animal fluency task compared to the illiterate subjects, while there was no difference on the supermarket fluency task.

One possibility is that this reflects a type of frequency of exposure effect, making lexical access less readily available in illiterate subjects on the animal fluency task. In other words, this difference may be a consequence of education or secondary effects of literacy. For example, reading skills should facilitate access to information, through printed media, thus providing an opportunity to broaden different semantic categories that transcend the shared socio-cultural background of the two literacy groups. However, it appears that it is not just that the two semantic categories used in this study are associated with differences in socio-cultural background specifically related to literacy/education; they also differ in the level of

reference to concrete knowledge and situation specificity. Thus, the observed differences between the literacy groups may not only relate to the semantic category used but potentially also to the extension (the semantic field; the potential number of available elements) of the semantic category. In other words, written language provides an opportunity to broaden the different semantic categories, and by using written language, we can access information (i.e., elements of the semantic category) that we cannot access though our direct experience. Thus, an important determinant for verbal fluency performance might relate to the type of experience we have had with the elements of a semantic category.

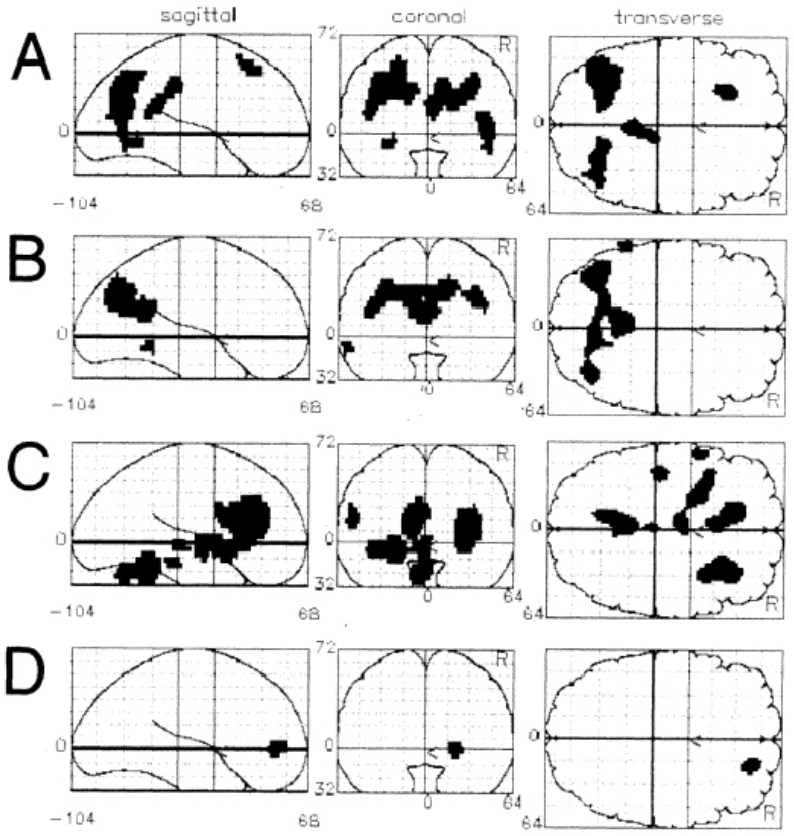


[Figure 5.6] Literate and illiterate semantic spaces related to the animal fluency task. (a) Hierarchical cluster analysis (Ward's method) of the semantic fluency responses. (b) Multi-dimensional scaling including the seventeen most frequent responses. Observe that the results are rotationally invariant so the results indicate that the aspects of semantic memory reflected in the data are similarly organized in both literacy groups

We further investigated the effects of formal schooling on the semantic organization of the responses from the animal fluency task (Faísca, Reis, & Petersson, in preparation) using a non-metric multidimensional scaling approach. This approach assumes that the item sequence in a fluency task reflects the semantic organization for a given semantic domain. The most frequent responses in both groups were selected for further analysis and the serial position was used to build a distance matrix. The matrixes for each group were analyzed and displayed in a 2D semantic space. As can be seen from Figure 5.6, the semantic organization for the common responses are similar in the two literacy groups. Both groups allocated the different exemplars according to the same sub-categories (farm birds, farm animals and wild animals). Note that the literacy group differed on the animal fluency task in terms of the number of generated exemplars, but there was little support for any differences in semantic organization on the seventeen most frequent responses, as characterized by non-metric multidimensional scaling.

In summary, the semantic fluency results shows that significant literacy effects may or may not be observed depending on the choice of semantic criterion. This emphasizes the importance of developing instruments that are free of educational and cultural biases, or alternatively, in an effective manner handles such effects (e.g., statistically), and at the same time permit the investigation of cognitive functions of interest. The multidimensional scaling results on animal category responses suggest that on the high frequency responses there is no difference between groups in terms of semantic organization, indicating that differences between groups emerged after the first items of a category had been generated. We suggest that the initial production reflects the shared cultural background, while differences related to differences in cultural background only emerge in the later phase of the production and that these later differences are associated with differences in literacy.

5.3 NEUROIMAGING STUDIES OF LITERATE AND ILLITERATE SUBJECTS



[Figure 5.7] Brain activations in the literate and illiterate brain during immediate verbal repetition. Word repetition in (A) the literate and (B) in the illiterate group. Pseudoword repetition in (C) the literate and (D) in the illiterate group.

In our first PET study of literate and illiterate subjects, we compared the two literacy groups on immediate verbal repetition (Castro-Caldas et al., 1998). The groups were matched for age and socio-cultural background (Reis et al., 2003) and the subjects in the literate group had received 4 years of schooling. The subjects were instructed to repeat words or pseudowords and to avoid any other type of speech production. Though there were performance differences between groups, these did not correlate with the pattern of brain activations in either group or condition (Petersson et al., 2000). Including the performance scores as a confounding covariate did not affect outcome and the differences

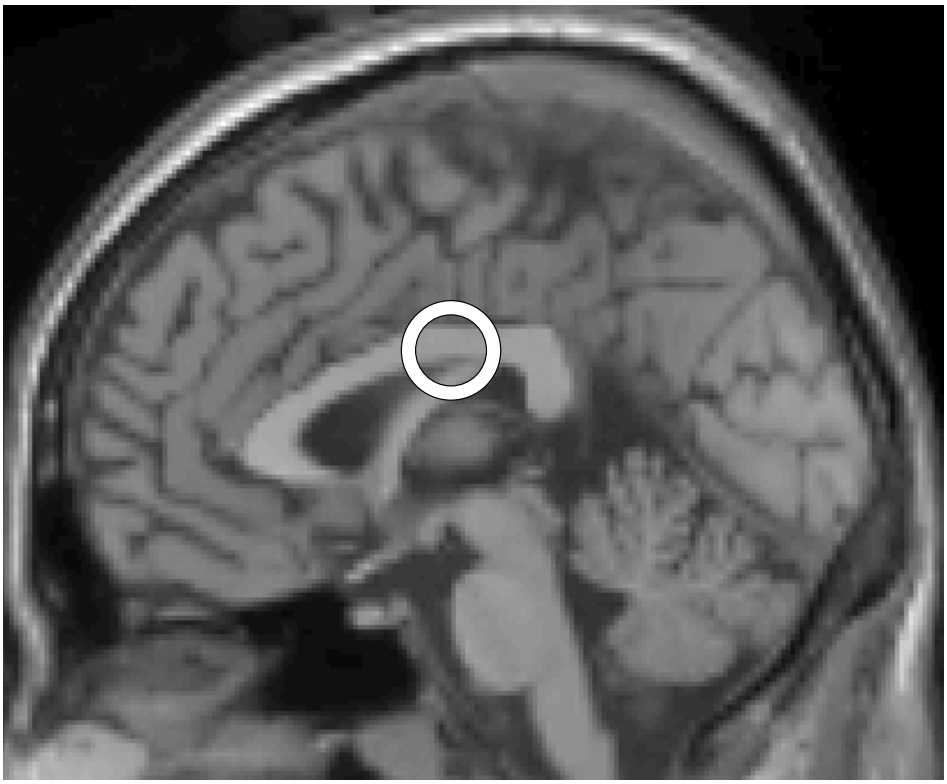
between the literate and illiterate group were generally independent of whether performance was included in the analysis or not (Petersson et al., 2000). Comparing the PET data between groups (Figure 5.7) suggested that there was a more prominent left-sided inferior parietal (BA 40) activation in the words vs. pseudowords comparison in the literate group. In the reverse comparison (pseudowords vs. words), the literate group displayed a significant activation in the anterior insular cortex (BA 14/15) bilaterally and in the right inferior frontal/frontal opercular cortices (BA 44/45/47/49), left perigenual anterior cingulate cortex (BA 24/32), left basal ganglia, midline anterior thalamus and midline cerebellum. In the illiterate group, significant activation was only observed in the middle frontal/frontopolar region (BA 10). These results represented the first indication that the functional architecture of auditory-spoken language processing is influenced by literacy. This suggests that there exists a relation between the acquisition of alphabetic orthographic knowledge and aspects of phonological processing in terms the functional brain organization, consistent with behavioral findings outlined above.

5.3.1 A NETWORK ANALYSIS OF IMMEDIATE VERBAL REPETITION IN LITERATE AND ILLITERATE SUBJECTS

Complementary to the approach and results outlined in the previous section is to take a network perspective on cognitive brain function. In general, information is thought to be represented as distributed activity in the brain while information processing, subserving cognitive brain functions, is thought to emerge from the interactions between different functionally specialized regions or neural groups. When trying to understand cognitive processing as instantiated in the brain it is therefore natural to analyze functional interactions from a network perspective (Ingvar & Petersson, 2000).

Structural equation modeling (SEM, cf., chapter 3) provides one approach to test for differences in network interactions explicitly. Petersson et al. (2000) attempted to characterize the functional organization of immediate verbal repetition in literate and illiterate subjects in terms of effective connections between regions in a given functional-anatomical model (cf., section 6.6). In terms of network interactions, the results showed no significant difference in the literate group when comparing the word and pseudoword condition. Neither was there any significant difference between the literate and illiterate

group in the word repetition condition. In contrast, there were significant differences between word and pseudoword repetition in the illiterate group and between the illiterate and literate group in the pseudoword condition. The differences between groups were mainly related to the phonological loop, in particular, the interaction between Broca's region and the inferior parietal region.



[Figure 5.8] Differences between literacy groups in the local thickness (circle) of the corpus callosum indicate that this is thinner in illiterate compare to the literate subjects ($P < 0.01$).

The absence of significant difference between word and pseudoword repetition in the literate group relates to the fact that the network interactions were similar during both word and pseudoword repetition. This indicates that the literate subjects automatically recruit the same processing network during immediate verbal repetition for words and pseudowords. In contrast, this was not the case for the illiterate group. This is consistent

with the suggestion that phonological processing is differently organized in illiterate individuals due to a different developmental background related to the acquisition of alphabetic reading and writing skills. Based on this and in conjunction with the behavioral results outlined in sections 5.2.2-4, we suggest that these differences in phonological loop interactions might represent a primary difference between the two literacy groups related to differences in sub-lexical phonological processing. This is in line with the suggestion that the parallel interactive processing characteristics of the language system differ between literate and illiterate subjects (Petersson et al. 2000).

5.3.2 NEUROANATOMICAL FINDINGS RELATED TO THE CORPUS CALLOSUM

One may wonder whether there are neuroanatomic correlates corresponding to the literacy status. It is well-known that the corpus callosum, the large fiber bundle that interconnects the two brain hemispheres, develops during childhood and into young adulthood. In particular, there is an active myelination process of the neuronal axons running through this structure in order to establish efficient communication between the brain's two hemispheres (Giedd et al., 1996). Recent results suggest that the posterior mid-body part of the corpus callosum undergoes extensive myelination during the years of reading acquisition in children, that is, during 6-10 years of age (Thompson et al., 2000) and the fibers that cross over in this region of the corpus callosum interconnect the parieto-temporal regions (for a general review see e.g., Zaidel & Iacoboni, 2003). A recent study of the morphology of the corpus callosum suggested that the posterior mid-body region of the corpus callosum (Figure 5.8), that is, the part that interconnect the left and right parieto-temporal cortices, is thinner in illiterate compare to the literate subjects (Castro-Caldas et al., 1999). Petersson et al. (1998) hypothesized that this may be related to a difference in the inter-hemispheric interactions between literacy groups with respect to the parieto-temporal cortices.

A large number of neuropsychological studies of acquired reading and writing impairment (alexia and agraphia) describe neuroanatomic lesions most prominently centered on the parieto-temporal region, including the inferior parietal cortex and the posterior portions of the superior temporal gyrus, thus suggesting this region is important for the mapping of orthographic representations onto phonologic representations

(Friedman, Ween, & Albert, 1993). Ernest Weber suggested in 1904 that the left hemispheric language dominance might depend on the acquisition of reading and writing skills and early attempts to address the issue in aphasic patients appeared to support this hypothesis (Cameron, Currier, & Haerer, 1971; Wechsler, 1976). Specific differences in language processing between literate and illiterate aphasic subjects has been reported, in particular with respect to pseudoword repetition and verbal memory tasks (Coppens et al., 1998). Lecours (1989) suggested that illiterate subjects are more likely to use processing networks that include right-hemisphere regions when performing language tasks (Coppens et al., 1998). Moreover, a reversal of ear advantage for phonetically similar words in illiterate subjects has been reported (Damásio, Damásio, Castro-Caldas, & Hamsher, 1979). However, the mechanisms influencing hemispheric specialization and the consequent interhemispheric interaction are not well-understood and both genetic as well as environmental factors appear to be relevant (Sommer, Ramsey, Mandl, & Kahn, 2002; Thompson et al., 2001). Functional hemispheric lateralization has been shown to depend on several factors, including stimulus material (Kelley et al., 1998), experimental task (Stephan et al., 2003), and a recent review concluded that hemispheric specialization for language is multi-factorial and may depend on both task as well as brain region (Josse & Tzourio-Mazoyer, 2004), and it is well accepted that both hemispheres play a role in language processing (Friederici, 2002; Knecht et al., 2002). Furthermore, computational modeling has indicated that several possible mechanisms can support hemispheric lateralization (Reggia & Schulz, 2002).

5.3.3 HEMISPHERIC DIFFERENCES RELATED TO LITERACY

In a recent study (Petersson et al., submitted) we investigated two different samples of illiterate female subjects and their matched literate controls with respect to the hemispheric lateralization of the inferior parietal cortex in two simple auditory-verbal language tasks encompassing four different conditions (cf., section 6.7). In brief, while the literate group showed a positive left–right difference, the illiterate subjects showed a negative left–right difference in the inferior parietal region (Figure 6.4 and 6.5). Detailed inspection of the results suggested that the degree of functional lateralization may be task dependent. However, what stayed constant independent of task was the relation of the left-right

differences between the two literacy groups (i.e., group x hemisphere interaction), indicating that this relation generalizes over auditory-verbal tasks. Thus, it appears that literacy influences the functional hemispheric balance in the inferior parietal region. It is interesting to note that recent experimental results have suggested a rostral to caudal myelination process of the corpus callosum during childhood and early adulthood (Thompson et al., 2000; Zaidel & Iacoboni, 2003), indicating an ongoing developmental process to establish efficient interactions between the two hemispheres. The fibers that cross over in the posterior mid-body region of corpus callosum interconnect the parieto-temporal regions and undergoes extensive myelination during the years of reading acquisition (Thompson et al., 2000) and this is the same region in which recent evidence indicate that literate subjects are thicker compared to illiterate subjects (cf., section 5.3.2). Thus, one may speculate that acquiring reading and writing skills at the appropriate age shapes not only the local morphology of the corpus callosum but also the degree of functional specialization as well as the pattern of interaction between the interconnected inferior parietal regions.

In conclusion, literacy represents an essential aspect of contemporary culture. Formal education and the educational system can be viewed as an institutionalized process of structured cultural transmission. The results outlined in this chapter indicate that formal education and its use influence important aspects of cognition and behavior as well as structural and functional properties of the brain. Taken together, the evidence provides strong support for the hypothesis that the brain is modulated by literacy and formal education.

6. EXPERIMENTAL STUDIES

In this chapter, we will briefly outline and discuss the main findings of the eight experimental studies included in this thesis. For further details and results we refer to the original papers in chapter 9.

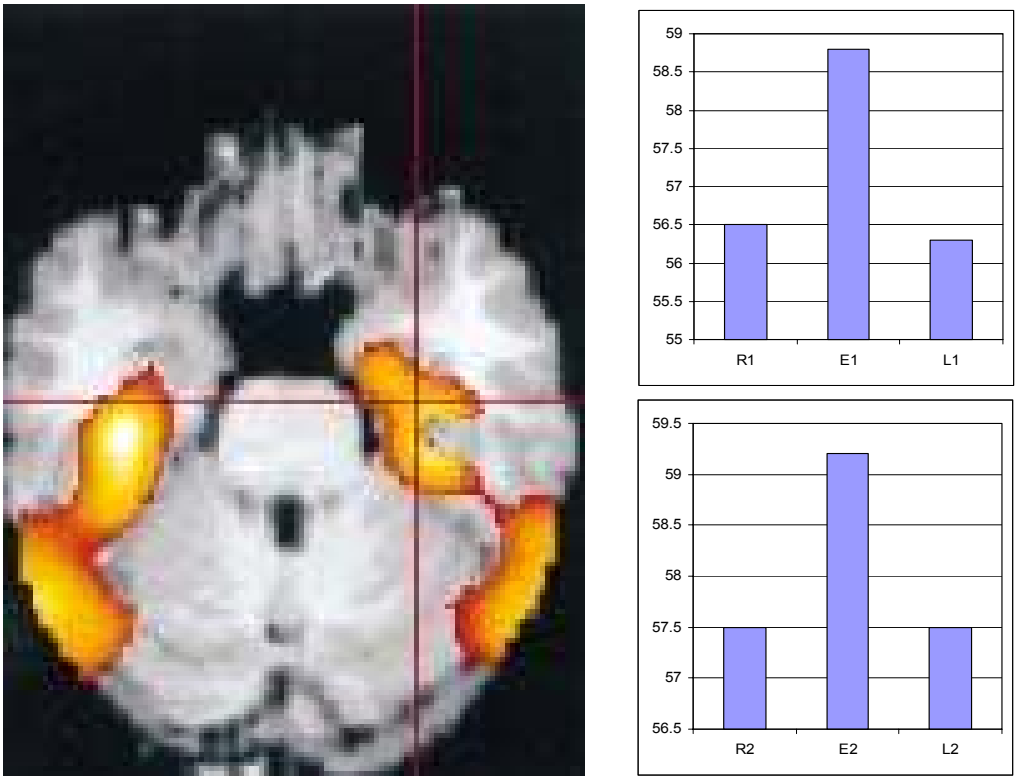
6.1 LEARNING RELATED EFFECTS AND FUNCTIONAL NEUROIMAGING

A problem in the study of learning in the human brain is that non-specific time effects not necessarily related to learning may parallel learning related changes. However, there are several ways to account for non-specific time effects. Petersson and colleagues (1999b) describe these approaches and we outlined the potential difficulties in investigating learning related effects in humans using functional neuroimaging. We also illustrated and compared two approaches by analyzing a PET dataset related to the medial temporal lobe from a previously reported learning study (Petersson et al., 1997). The first strategy for separating specific learning effects from non-specific time effects is to model time as a confounding covariate. The second strategy views learning related effects as a condition x time interaction in a linear model. The main finding of the Petersson et al. (1999b) study suggested that the two approaches yielded similar results. More specifically, the particular approach used to characterize the learning related effects influenced the outcome of the statistical analysis only weakly, in the case of statistically robust results, while it made a difference for less robust results, at least for the particular dataset investigated. We also proposed a third way to investigate learning related effects. This approach is basically a variation on a common experimental design theme, which handles the influence of non-specific time-changes by balancing the experimental design over this factor. In other words, the proposal is based on a temporally balanced experimental design, which is perhaps theoretically the most satisfying of the three approaches, but entails an additional overhead from a practical point of view in terms of stimulus material, stimulus presentation, and number of practice sessions.

6.2 A DYNAMIC ROLE OF THE MEDIAL TEMPORAL LOBE IN FREE RECALL

Learning and memory are fundamental brain functions that enable the brain to learn and adapt in its environment. Learning can be defined as the processes by which the central

nervous system functionally restructures its processing pathways or its representations of information (cf., chapter 2). Memory, or more specifically, the memory trace results from active processing of information in combination with system plasticity (cf., chapter 1 and 2). In other words, from a dynamical systems perspective, learning is a dynamical consequence of information processing and network plasticity. In the case of the human brain, learning and memory is thought to result from changes in the synaptic structure of the processing network. Such changes can be detected indirectly at a behavioral level as changes in the performance of a task as a result of practice.



[Figure 6.1] Retrieval related activity decreases as a function of practice (Petersson et al., 1997).

Petersson and colleagues (1997) investigated the role of the medial temporal lobe during free recall of simple abstract designs in a less practiced memory condition as well as in a well-practiced (well-encoded) memory condition. The results showed an increased activity of the medial temporal lobe bilaterally (Figure 6.1) during retrieval in the less

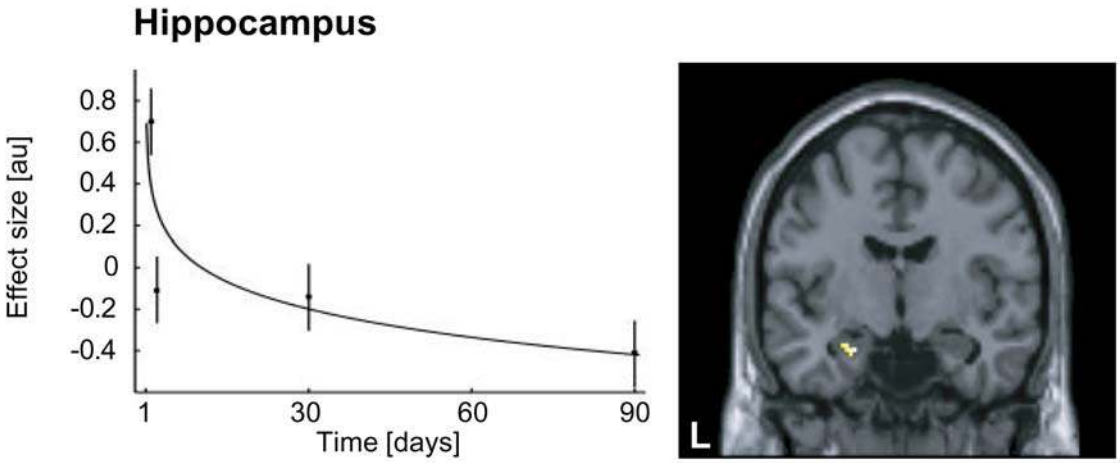
practiced memory state compared to the well practiced memory state. This result was interpreted as support for a dynamic role of the medial temporal lobe in retrieval. More specifically, the medial temporal lobe activation decreased during free recall when comparing the earlier stages of acquisition compared to the well-encoded memory condition.

The medial temporal lobe has been ascribed several functions (cf., chapter 4). For example, the medial temporal lobe may be needed to bind together distributed representations (associative conjunctive learning) supported within different neocortical association regions, to enable the rapid acquisition of declarative knowledge for long-term integrative learning, to create flexible cognitive map representations (relational representations), to reduce interference in information to be encoded with respect to information previously stored (representation separation/orthogonalization), and has also be suggested to play a role in memory consolidation (cf., chapter 4).

The phenomenon of temporally-graded retrograde amnesia has given rise to the concept of memory consolidation (Ribot, 1882; Zola-Morgan & Squire, 1990), meaning the high-level process by which declarative memory becomes independent of the medial temporal lobe memory system (Squire, 1992; Squire, Cohen, & Nadel, 1984). This concept of memory consolidation entails the idea that information is reorganized and integrated in the neocortex and should be distinguished from other concepts of memory consolidation (e.g., with a biochemical connotation), which view consolidation as a simple stabilization of the memory trace.

If distributed activity in the neocortex, subserving perception and short-term memory, is to be stored as a long-term declarative memory trace, then the medial temporal lobe structures must be engaged at the time of learning (Eichenbaum & Cohen, 2001; Squire, 1992). Presumably, the final storage site of declarative knowledge is in the neocortex (Eichenbaum & Cohen, 2001; Squire, 1992). This implies that declarative learning and memory storage is dependent on some form of interaction between the medial temporal lobe and the neocortex. Several researchers have suggested that repeated reactivations of the neocortical representations of declarative memories strengthen the neocortical interconnections so that the neocortical memory network eventually can support declarative memory retrieval independently of the medial temporal lobe (Alvarez & Squire,

1994; McClelland, 1994; Squire & Alvarez, 1995; Treves & Rolls, 1994). This suggests that the interaction between the medial temporal lobe and the neocortex is dynamic and changing as a function of repeated activations of the interacting networks. In the study of Petersson and colleagues (1997) it was assumed that practice (in this case repeated encoding) would reactivate the relevant neocortical regions and hence strengthen interconnections in the neocortical network in such a way that the neocortex eventually would support declarative retrieval, if not independent, so at least less dependent on the interaction with the medial temporal lobe.



[Figure 6.2] Declarative memory consolidation. The medial temporal lobe showed a decline over time (the figure on the right represents an overlap of the significant effects observed in the medial temporal lobe; i.e., the day 1 – day 2 effect inclusively masked by the day 30 – day 90 and the overall time x condition interaction).

Petersson and colleagues (1997) also assumed that this would translate into a decreased level of activity of the medial temporal lobe as a function of repeated encoding. Our

findings are consistent with this hypothesis and suggests an inverse relation between the strength of encoding and the activation of the medial temporal lobe during retrieval (Mesulam, 1998). This also suggests that re-activation (e.g., during sleep, rehearsal, or repeated remembering) of neocortical representations could support system-level consolidation. Besides time and rehearsal, sleep, in particular slow wave sleep, appears to play an important role in memory consolidation (Maquet, 2001; Stickgold, Hobson, Fosse, & Fosse, 2001). During sleep, evidence suggests that hippocampal assemblies reactivate neocortical representations of recent events (Hoffman & McNaughton, 2002; Skaggs & McNaughton, 1996), which may thereby strengthen and refine cortical assemblies so that they can be used for retrieval without requiring hippocampal processing. In a recent study, we investigated both the effects of slow wave sleep and time on declarative memory retrieval with event-related FMRI (Takashima, Petersson et al., submitted). Preliminary results indicate that subjects who had a short nap improved their memory performance compared to those who did not sleep in a yes-no visual recognition paradigm using landscape scenes. Moreover, the level of medial temporal lobe activity related to hit events correlated negatively with the amount of slow wave sleep. In addition, the level of medial temporal lobe activity during hit events decreased over time (cf., Figure 6.2).

6.3 DYNAMIC CHANGES IN THE FUNCTIONAL ANATOMY OF THE HUMAN BRAIN DURING RECALL OF ABSTRACT DESIGNS RELATED TO PRACTICE

In the study of Petersson and colleagues (1999a), we explored the less practiced and the well-practiced condition during free recall of abstract designs from the perspective of controlled and automatic processing. The results suggest that as an automatic processing mode develops (i.e., a decreased dependence on attentional and working memory resources), a parallel decrease of activity is observed in the prefrontal, anterior cingulate, posterior parietal regions. Furthermore, a pattern of practice related increases were observed in the auditory, posterior insular-opercular extending into perisylvian supramarginal, and right mid occipito-temporal regions.

This set of results were interpreted as providing support for the view the neural network subserving free recall includes dynamic components and that there is a functional restructuring of the processing networks during the learning process. More specifically, we

suggested that the development of automaticity is associated with decreased activity in attentional and working memory related prefrontal, anterior cingulate, posterior parietal regions, while activity increases in the auditory, posterior insular-opercular, and perisylvian supramarginal regions reflect a lower degree of attentional suppression of task irrelevant processing. This is in general agreement with several recent and closely related functional neuroimaging studies (Garavan et al., 2000; Jansma et al., 2001; Wiser et al., 2000). Moreover, the increase observed in the right occipito-temporal region was interpreted as reflecting a more well-developed representation of the acquired knowledge (i.e., the abstract designs). More specifically, as the representation of the acquired knowledge developed, this was reflected in the progressively stronger engagement of this region. This suggestion is consistent with results from studies in monkeys which indicate that the development of representations of visual paired associates in the infero-temporal cortex depends on an intact parahippocampal region (Higuchi & Miyashita, 1996). Finally, we suggested that a gradual transition from controlled to automatic processing is supported both by a restructuring of the processing architecture and the development of a more efficient representations of information (cf., chapter 4). Recent functional neuroimaging studies have provided support for both the restructuring as well as the processing efficiency perspective (Garavan et al., 2000; Jansma et al., 2001; Petersson et al., 2001; Wiser et al., 2000).

6.4 LEARNING RELATED MODULATION OF FUNCTIONAL RETRIEVAL NETWORKS

In the follow up studies reported in Petersson et al. (2001) these investigations of learning related modulation of functional retrieval networks were extended in two different experiments. The first experiment investigated episodic recognition of object-location conjunctions with the objective of comparing recognition and recall paradigms as well as achieving high performance already in the initial stages of learning. This would allow us to examine whether similar practice related changes as described in Petersson et al. (1997; 1999a) would result also during recognition as well as relatively independent of the performance level. In the first experiment we adapted the experimental paradigm of Owen et al. (1996). This experiment was divided into two sub-experiments, the first being a

replication of the study of Owen et al. (1996) and the second using a sensory-motor baseline condition without explicit encoding or retrieval demands to investigate practice related effects. Furthermore, in the second experiment, material specific effects were examined using pseudowords in a free recall paradigm and an experimental design very similar to that used by Petersson et al. (1997; 1999a).

Overall, the results of the first part of the first experiment confirmed the results of Owen et al. (1996). Owen et al. (1996) showed that object-location recognition activated the in the occipito-temporal regions and in the right medial temporal lobe relatively more compared to location recognition, while in the reverse comparison, activations were observed in the right mid-dorsolateral prefrontal and right posterior parietal regions. However, we observed bilateral medial temporal lobe activations in the object-location vs. location comparison. Furthermore, we observed learning related medial temporal lobe decreases as a function of repeated encoding; in other words, like in our previous study (Petersson et al., 1997), the practiced (repeated encoding) object-location condition showed decreased medial temporal lobe activity relative object-location. This finding again suggests that there is an inverse relation between the strength of encoding and the activation of the medial temporal lobe during retrieval (Mesulam, 1998). Results reported by Montaldi et al. (1997), related to verbal episodic retrieval, can also be interpreted along these lines. Montaldi et al. (1997) also used repeated encoding manipulation and observed that the left medial temporal lobe was more active in the less practiced compared to the well-practiced retrieval condition. Nyberg and colleagues (1996b) reported a positive correlation between retrieval success and medial temporal lobe activity. The observed increase in medial temporal lobe activity in object-location recognition compared to location recognition, both in this and the study of Owen et al. (1996), is consistent with this finding. However, the learning related medial temporal lobe decrease observed here and particularly in the study of Petersson et al. (1997), suggest that there is a positive correlation between retrieval success and medial temporal lobe activity at a given level of encoding strength.

Several alternative interpretations of the practice related decreases observed in the medial temporal lobe have been put forward (Petersson et al., 1997; Petersson et al., 2001). One speculation suggests that the initial stabilization of the memory trace is dependent on

neocortical-medial temporal lobe interaction, as described in section 6.2. Alternatively, repeated encoding and retrieval might transform an initial episodic memory to a semantic-like memory, suggesting a 'semantization' of the information. In the light of recent findings of Eldridge and colleagues (2000), this suggestion can be phrased as follows. Initial episodic retrieval of information, which is mainly based on recollective remembering, gradually looses its precise spatio-temporal context with repeated encoding experiences, which translates the stored information into a general fact. In parallel with this process, the recognition judgment is increasingly based on familiar knowing (Henson et al., 2003).

In parallel with the medial temporal lobe decrease, we also observed decreases in the prefrontal, the anterior cingulate, the posterior parietal, and parts of the inferior temporal regions, in general agreement with our previous study (Petersson et al., 1999a) as well as more recent functional neuroimaging studies (Garavan et al., 2000; Jansma et al., 2001; Wiser et al., 2000). Again, we suggested that these learning related changes might be conceptualized in terms of dynamic modulatory effects on the interaction between attentional/control processes and learning/memory. These dynamic effects were tentatively related to the different demands for attentional and working memory resources, reflecting different adaptive processes related to the transition from a non-automatic to a more automatic mode of processing. Furthermore, in Petersson et al. (1999a) it was suggested that these practice related changes could not be explained by differences in performance and the results reported in Petersson et al. (2001) are consistent with this suggestion since the performance was almost perfect (96% correct) already in the object-location condition.

In the second experiment of this study the role of material was investigated, using pseudowords instead of visuo-spatial material as in the previous studies. The general pattern of practice related decreases and increases were similar to the results observed in the previous two studies, suggesting that these effects are independent of stimulus material. The most prominent difference between this second experiment and the previously reported results concerns the medial temporal lobe and the posterior parietal cortex. In contrast to the previous studies no significant differential retrieval related medial temporal lobe activity was observed in the free-recall compared to the base-line condition, neither was any practice related changes observed in the medial temporal lobe (not even at low thresholds of significance). This suggests that, with respect to the medial temporal lobe,

non-specific time effects, novelty detection, attentional effects, retrieval performance, or retrieval effort are not likely to explain in a simple way the observations in the first experiment of this study or the study of Petersson et al. (1997). This may instead point to a material specific effect. Thus, the practice related decreases observed in the medial temporal lobe might be more strongly expressed for either meaningful verbal material, triggering meaning based associative processing, or visuo-spatial material, triggering the processing of visuo-spatial relations. Some indications that this may be the case come from the comparison of the recall of pseudowords with the base-line task (filling in contours of simple pre-drawn shapes). The activity of the medial temporal lobe was significantly greater in the base-line task compared to the free recall of pseudowords.

Finally, in the present study, the prefrontal activation was right lateralized in the recognition condition but bilateral in the free-recall condition supporting the suggestion that free-recall is dependent on bilateral prefrontal processing (Petersson et al., 1999a). The practice related prefrontal decreases were bilateral in both recognition and free-recall. However, the prefrontal decreases, similar to the pattern observed in Petersson et al. (1999), were most prominent on the left. This is consistent with the conclusion of Nolde et al. (1998) that complex retrieval tasks are dependent on bilateral prefrontal processing, particularly complex cued-recall and free-recall tasks (cf., chapter 4).

6.5 EFFECTIVE AUDITORY-VERBAL ENCODING

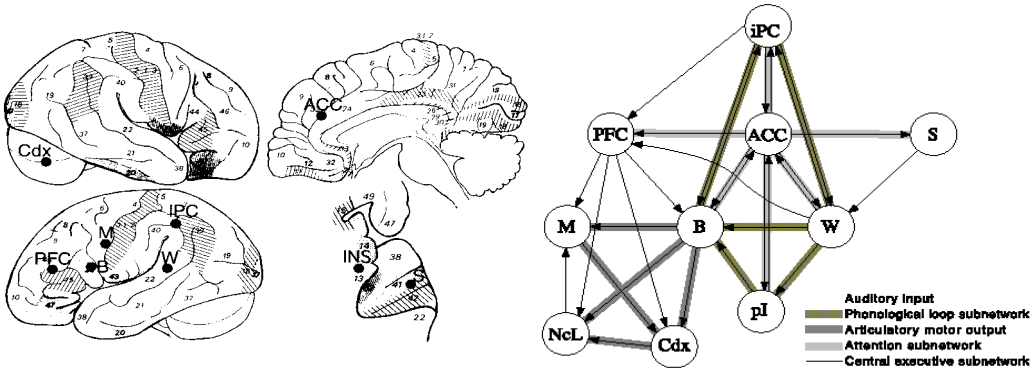
Recent event-related FMRI studies (Brewer et al., 1998; Wagner et al., 1998c) indicate that the prefrontal region and the medial temporal lobe are more active during effective encoding (i.e., when the to-be-remembered information is actually successfully retrieved) compared to ineffective encoding (i.e., when the to-be-remembered information is forgotten). The within-subject design and the use of well educated young college students in these studies make it important to replicate these results in other study populations. In the study of Petersson et al. (1999) we explored the issue of effective encoding further in a group of healthy older illiterate women. The illiterate subjects were investigated in an auditory word-pair cued-recall paradigm. In the encoding condition there was a positive correlation between the subsequent cued-recall success and the level of activation in the

inferior frontal region and the anterior medial temporal lobe, suggesting that these regions are more active during effective encoding compared to ineffective encoding.

6.6 THE ILLITERATE BRAIN

Learning specific skills during childhood may influence the functional organization of the adult brain. This hypothesis led us to investigate auditory-verbal language processing in illiterate subjects (cf., chapter 5). Here we will just restate the main findings of our first study of literate (4 years of schooling) and illiterate subjects (matched for age and socio-cultural background, cf., Reis et al., 2003), and refer to chapter 5 and the original paper for the detailed results and discussion. Instead, we move on to discuss the follow-up study in greater detail (Petersson et al., 2000). Thus, in brief, we compared the two literacy groups on immediate verbal repetition of words and pseudowords (Castro-Caldas et al., 1998). The behavioral results replicated those reported by Reis and colleagues (1997) suggesting that literate subjects performed significantly better on pseudoword repetition (cf., Figure 5.2). During word repetition, both groups performed similarly and activated similar brain regions. In contrast, illiterate subjects showed greater difficulty in repeating pseudowords correctly and did not activate the same neural structures as literate subjects. Though there was performance differences between groups, performance did not correlate with the pattern of brain activations in either group or condition and including the performance scores as a confounding covariate did not affect the results (Petersson et al., 2000).

In the follow-up study (Petersson et al., 2000), we investigated whether the parallel interactive processing characteristics of the underlying language processing network differ in literate and illiterate subjects. We therefore investigated the pattern of interactions between the regions in a large-scale functional-anatomical network for language processing with a network approach based on structural equations modeling (cf., chapter 3). The construction of the functional-anatomical language network was based on theoretical considerations as well as the empirical functional neuroimaging, neuropsychological and anatomical literature. The regions included in the functional-anatomical network were represented as spherical regions of interests in the Karolinska Computerized Brain Atlas of Greitz (1991).



Region of interest			*Brodmann's area*
S	-	primary/secondary auditory cortex	BA 41/42
W	-	Wernicke's area	BA 22 (posterior)
iPC	-	angular/supramarginal gyrus	BA 39/40
pI	-	mid-posterior insula	BA 13
B	-	Broca's area	BA 44
ACC	-	anterior cingulate cortex	BA 24/32
PFC	-	dorsolateral prefrontal cortex	BA 45/46
M	-	primary speech motor region	BA 4
NcL	-	lentiform nucleus	
Cdx	-	right lateral cerebellum	

[Figure 6.3] The simple functional network model for immediate verbal repetition used to investigate word and pseudoword repetition in literate and illiterate subjects.

The objectives of constructing the functional-anatomical network were to generate a simple model which was able to explain a sufficient part of the observed covariance structure in both groups and both conditions. At the same time we required that the network model should be both theoretically and empirically plausible. The network model (Figure 6.3) include a simplification of the Wernicke-Geschwind model represented by the Wernicke's area (W; i.e., the posterior third of the left superior temporal gyrus, BA 22) connected to the posterior part of Broca's region (B; i.e., the posterior part of the left inferior frontal gyrus BA 44; W $\rightarrow$ B) with input from the left auditory cortex (S; BA 41/42; S $\rightarrow$ W) and a simple articulatory motor output circuit (including the left lentiform nucleus, NcL, and the left articulation part of motor cortex, M, BA 4; B $\rightarrow$ M, B $\rightarrow$ NcL $\rightarrow$ M). This core was extended to include the anterior cingulate cortex (ACC), related to attention, error detection and response competition/selection; the phonological loop (Becker et al., 1999; Paulesu et

al., 1993; Paulesu et al., 1996), which introduce the inferior parietal cortex (iPC on the border between the angular and supramarginal gyrus; BA 39/40) and the mid-insular cortex (pI). The connections of the phonological loop were recurrent (W ↔ iPC ↔ B). Since the insula has been hypothesized to be a neural relay for automatic language processing the connections to and from the insula were feedforward (W → pI → B) (Damasio & Damasio, 1980; Dronkers, Redfern, & Knight, 2000; Mesulam & Mufson, 1985; Paulesu et al., 1996; Raichle, 1994; Raichle et al., 1994). The interactions between the ACC and the phonological loop were represented by recurrent connections (W ↔ ACC ↔ B and pI ↔ ACC ↔ iPC). In addition, the left middle-inferior dorsolateral prefrontal region (PFC, on the border between BA 45 and 46) suggested to subserve central executive aspects of verbal working memory (Baddeley, 2003) was included with input from the ACC, the Wernicke's area, and the inferior parietal cortex (ACC → PFC, W → PFC, iPC → PFC) and with outputs modulating the organization of the articulatory motor output (PFC → B, PFC → M, PFC → NcL, PFC → Cdx). Finally, the right lateral cerebellar region (Cdx) was included since this region has been related to certain aspects of language processing with inputs from cortical motor regions (PFC → Cdx, B → Cdx, M → Cdx → NcL).

In this network model for language processing, the results indicated that the functional interactions during word and pseudoword repetition differed between literacy groups. More specifically, while there were no significant differences in the literate group between word and pseudoword repetition, there were significant differences in terms of network interactions in the illiterate group. The differences between the two tasks in the illiterate group were interpreted in terms of differences in attentional, verbal working memory, and the articulatory organization of verbal output between conditions. Moreover, there were no significant differences between the literate and illiterate group during word repetition. In contrast, the network interactions differed between the literate and illiterate group during pseudoword repetition. In particular, these differences were prominent in the phonological loop. More specifically, these differences were related to the interaction between Broca's region and the inferior parietal cortex as well as the insular bridge between Wernicke's and Broca's region. Additional differences, similar to the between-condition differences observed in the illiterate group, were also reported. In conclusion, the results of this network analysis are consistent with the previously reported results and

support the hypothesis that learning to read and write during childhood influences the functional architecture of the adult human brain. In particular, it was suggested that the basic language network in the human brain is modified as a consequence of acquiring an alphabetic orthographic representation.

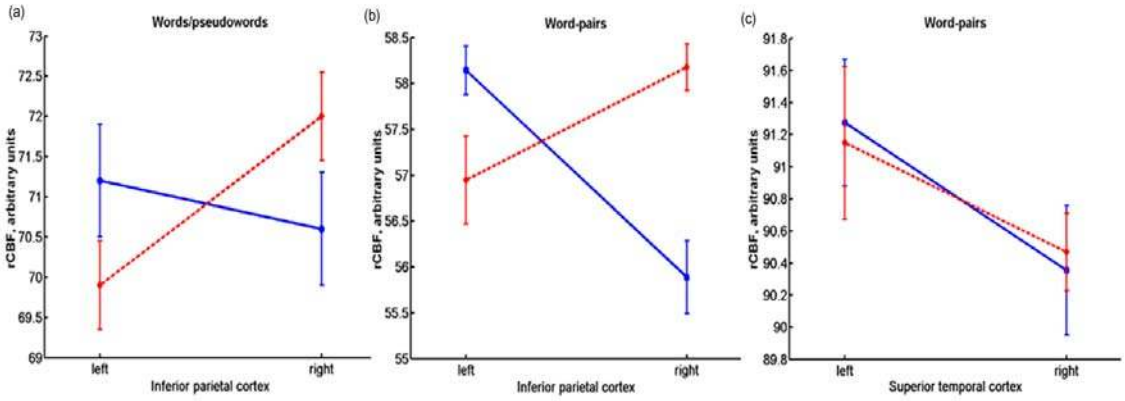
In conclusion, we suggested that the absence of significant difference between word and pseudoword repetition in the literate group is related to the fact that the literate subjects automatically recruit the same processing network for both words and pseudowords during immediate verbal repetition. In contrast, this was not the case for the illiterate group, suggesting that phonological processing is differently organized in illiterate individuals due to a different developmental background related to the acquisition of alphabetic reading and writing skills. Based on this and in conjunction with the behavioral results outlined in sections 5.2.2-4, we suggest that these differences in phonological loop interactions related to differences in sub-lexical phonological processing and reflected in the observation that the parallel interactive processing characteristics of the language system differ between literate and illiterate subjects (Petersson et al. 2000).

6.7 LITERACY: A CULTURAL INFLUENCE ON THE HEMISPHERIC BALANCE IN THE INFERIOR PARIETAL CORTEX

In the final study (Petersson et al., submitted), we investigated two different samples of illiterate female subjects and their matched literate controls with respect to the hemispheric lateralization of the inferior parietal cortex in two simple auditory-verbal language tasks encompassing four different conditions.

There are several principal reasons for investigating the inferior parietal region in literate and illiterate subjects: functionally this region has been related to reading (Friedman et al., 1993; Horwitz, Rumsey, & Donohue, 1998; Paulesu et al., 2000; Shaywitz et al., 1998) as well as to phonological processing and verbal working memory (Baddeley, 2003; Becker et al., 1999; Jonides et al., 1998; Vallar & Papagno, 1995); and recent neuroanatomic findings indicate that there are differences between literacy groups in the part of corpus callosum interconnecting the parieto-temporal cortices, as described in the previous section. Given these findings, we investigated whether the literate group would be relatively more left lateralized compared to the illiterate group, which would

show a bilateral or relatively more right lateralized pattern of results. In the first experiment (Petersson et al., 2000), in which subjects repeated words and pseudowords, task-related activation levels from regions of interest in the inferior parietal cortex (BA 39/40) were compared. The result showed a significant group difference (group x hemisphere interaction P = 0.009; Figure 6.4 a) indicating a positive left-right difference in the literate group while the illiterate subjects showed a prominent negative left-right difference. The group x hemisphere interaction was independent of whether the subjects repeated words (P = 0.017) or pseudowords (P = 0.006). Thus, while the literate subjects showed a tendency for left lateralization, the illiterate group was clearly right lateralized.



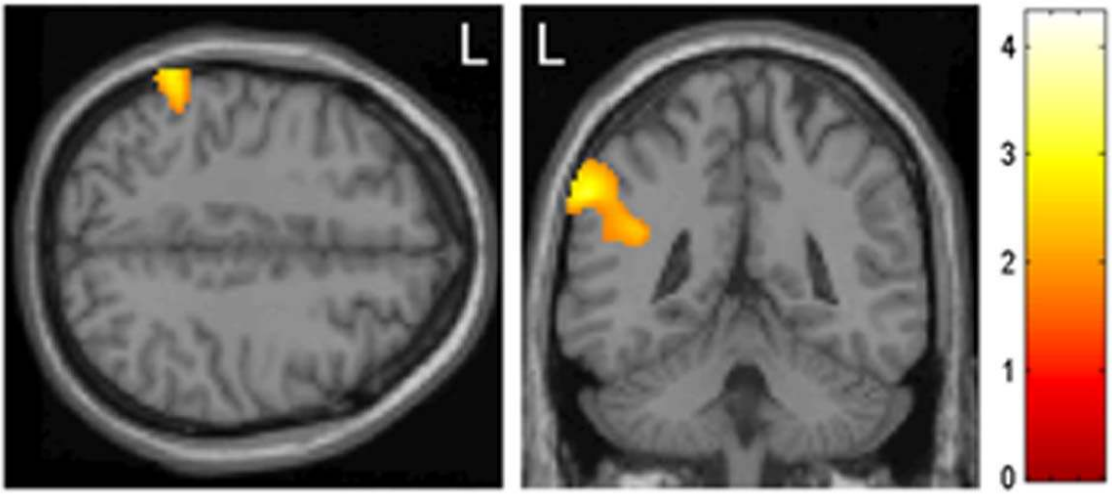
[Figure 6.4] Hemispheric differences (left-right) in activations levels between literacy groups in the inferior parietal region (Brodmann's area 39/40). (a) In experiment 1 the participants listen to and repeated words and pseudowords. The diagrams show the level of left- and right activation levels (regional cerebral blood flow, arbitrary units) as a function of literacy group (illiterate: dashed). Differences averaged over conditions P = 0.009. (b) In experiment 2 the participants were listening to and encoded word-pairs. Again we observed left-right activation differences (nearest supra-threshold cluster test, P = 0.029,

corrected) between literacy groups in the inferior parietal region (Brodmann's area 39/40). (e) To test the specificity of these left-right results with respect to the inferior parietal cortex, we also investigated the superior temporal region (BA 22/41/42) in the second experiment. The results showed that both literacy groups were similarly left lateralized in this region indicating that the functional lateralization of early speech related brain regions does not depend on literacy.

In the second experiment, in which the subjects listen to and encoded word-pairs, we examined whether the findings from the first study could be replicated in an independent sample of subjects. We thus tested for between-group left–right differences in the supra-threshold cluster nearest to the region investigated in the first experiment and observed a significant inferior parietal cluster (BA 39/40, group x hemisphere interaction P = 0.029, corrected; local maximum at [-60, -44, 38], P = 0.013, FWE-corrected; Figure 6.4 b and 6.5). Again, the group x hemisphere interaction was independent of whether the subjects listened to semantically related word-pairs (BA 39/40, cluster P = 0.015; local maximum at [-64, -42, 36], P = 0.009, FWE-corrected) or phonologically related word-pairs (BA 39/40, cluster P = 0.005; local maximum at [-52, -44, 40], P = 0.005, FWE-corrected). Thus, while the literate group showed a positive left–right difference, the illiterate subjects showed a negative left–right difference in the inferior parietal region extending inferiorly towards the temporo-parietal junction (Figure 6.5). Detailed inspection of the results suggests that the degree of functional lateralization may be task dependent. However, what stays constant independent of task is the relation of the left-right differences between the two literacy groups (i.e., group x hemisphere interaction), indicating that this relation generalizes over auditory-verbal tasks. Thus, it appears that literacy influences the functional hemispheric balance in the inferior parietal region.

It has recently been indicated that infants are left lateralized in the superior temporal gyrus when listening to speech or speech-like sounds (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002) and in order to test the specificity of our results with respect to the inferior parietal cortex, we also investigated the superior temporal region (BA 22/41/42). The results showed that both literacy groups were similarly left lateralized in this region

(Figure 6.4 c) indicating that early speech related brain regions does not depend on literacy. The present results provided evidence that a cultural factor, literacy, influences the functional hemispheric balance in the inferior parietal region. In contrast, both literacy groups showed a similar degree of left lateralization in early speech related regions of the superior temporal gyrus. The results provide further support for the idea that the degree of functional lateralization is task dependent and argues for a regional view on functional hemispheric lateralization.



[Figure 6.5] In experiment 2 the participants were listening to and encoded word-pairs. Again we observed left-right activation differences (nearest supra-threshold cluster test, P = 0.029, corrected) between literacy groups in the inferior parietal region (Brodmann's area 39/40).

It is interesting to note that recent experimental results have suggested a rostral to caudal myelination process of the corpus callosum during childhood and early adulthood

(Thompson et al., 2000; Zaidel & Iacoboni, 2003), indicating an ongoing developmental process to establish efficient interactions between the two hemispheres. The fibers that cross over in the posterior mid-body region of corpus callosum interconnect the parieto-temporal regions and undergoes extensive myelination during the years of reading acquisition (Thompson et al., 2000) and this is the same region in which recent evidence indicate that literate subjects are thicker compared to illiterate subjects (cf., section 5.3.2). One may speculate that acquiring reading and writing skills at the appropriate age shapes not only the local morphology of the corpus callosum but also the degree of functional specialization as well as the pattern of interaction between the interconnected inferior parietal regions. Thus there might be a causal connection between reading and writing acquisition, the development of the corpus callosum, and the hemispheric differences reported here, suggesting an active process of functional reconfiguration of the role of the left and right inferior parietal cortex. One possibility that cannot be addressed in the present study, which is based on comparing anatomically homotopic regions, is that the functionally relevant regions are not anatomically co-localized in the two literacy groups. However, the present findings would then indirectly reflect this underlying hemispheric difference between the literacy groups. Another issue is whether the present results reflect direct effects of acquiring reading and writing skills or reflect cumulative life-span effects.

7. ACKNOWLEDGMENTS

I wish to thank everyone who has contributed, directly or indirectly, to this thesis. In particular I would like to acknowledge the important contributions of: My supervisor Professor Martin Ingvar for his hard work, irrational optimism and permanent support. My close collaborators Dr Christina Elfgren, for collaborative work on declarative memory, and Professor Alexandra Reis, for introducing me to the field of illiteracy research and continued collaborative work on issues related to illiteracy. The Karolinska Apoteket, Professor Sharon Stone Elander for providing excellent support and a highly reliable tracer delivery process. The PET organization at the Karolinska Hospital, including Vahid Halouuli, Julio Gabriel, Monica Serrander, Göran Printz, and the late Valter Pulka for providing excellent technical support, forming the necessary foundation for functional PET investigations at the Karolinska Hospital. I am particularly grateful for the support of Gustav von Heine and Ellenor Andersson. My bothers and sisters in arms: Dr Jesper Andersson, Katrina Carlsson, Dr Peter Fransson, Christian Forkstam, Jens Gisselgård, Johan Sandblom, Dr Stefan Skare, Dr Predrag Petrovic. In particular I want to acknowledge the importance of Professor Jen-Chuen Hsieh and Dr Hamid Ghatan, two of the pioneers in the spearheading team of the Cognitive Neurophysiology Research Group. I also want to thank the late Dr Francisco Reis, Idalina Reis, Ditza Reis, and Dr Carlos Reis for their hospitality, generosity, and help with the illiteracy work. Joaquim Ferro, Professor Francisco Lacerda, Dr Manuela Guerreiro, Catarina Silva, Dr Luís Faísca, Susana Mendonça, Alexandra Mendonça, and the Olhão Friday Dinner Club for their important contribution to the illiteracy work. Finally, I also wants to thank: Professor Torgny Greitz, Professor Lars Eriksson, Docent Gunnar Blomqvist, Professor Anders Lansner, Dr Anders Sandberg, Professor Alexandre Castro-Caldas, Professor Lars Nyberg, Professor Lars-Göran Nilsson, Professor Lars Bäckman, Professor Alan Baddeley, Professor Susan Gathercole, Professor Peter Hagoort, Dr Peter Indefrey, Dr Guillen Fernandéz, Dr Ivan Toni, Professor Richard Frackowiak, Dr Cathy Price, Dr Jean-Baptiste Poline, Dr Tom Nichols, Dr Andrew Holmes, Professor Karl Friston, Dr Keith Worsley, Dr Anders Ledberg, Dr Peter Olofsson, Dr Pascal Fries, Dr Ole Jensen, and Dr Lea Hald. I thank all these people for taking time and contributing in significant ways to my education in the fields of functional neuroimaging, cognitive and computational neuroscience.

This work has been supported over the years by: The Swedish Medical Research Council; the Swedish Bank Tercentenary Foundation; the Knut and Alice Wallenberg Foundation; the Swedish Dyslexia Association; Project STRIDE (no 352/92-JNICT); Karolinska Institutet; the EU BIOMED 1 programme (BMHI CT-94-261); EU BIOMED 2 programme (IQLK6-CT-99-02140); Fundação para a Ciência e Tecnologia (FCT/POCIT/ 41669/PSI/2001), Portugal; Ministério da Ciência e do Ensino Superior, Portugal; Universidade do Algarve, Portugal; Max-Planck-Institute for Psycholinguistics, the Netherlands; and the F.C. Donders Centre for Cognitive Neuroimaging, the Netherlands.

In memory of Walter Pitts and in honour of the example set by him (cf., Heims, S. J. (1991). *The Cybernetics Group*. Cambridge, MA: MIT Press).

8. REFERENCES

Abadzi, H. (2003). *Improving Adult Literacy Outcomes: Lessons from Cognitive Research for Developing Countries*. Washington, D.C.: The World Bank, Operation Evaluation Department.

Adler, R. J. (1981). *The Geometry of Random Fields*. New York: Wiley and sons.

Adler, R. J. (1998). On excursion sets, tube formulae, and maxima of random fields. *Technical report, see http://iew3.technion.ac.il*

Aertsen, A. M. H., Gerstein, G. L., Habib, M. K., & Palm, G. (1989). Dynamics of neuronal firing correlation: Modulation of "effective connectivity". *J. Neurophysiol., 61*, 900-917.

Aertsen, A. M. H., & Preissl, H. (1991). Dynamics of activity and connectivity in physiological neuronal networks. In H. G. Schuster (Ed.), *Nonlinear Dynamics and Neuronal Networks* (pp. 281-302). New York: VHC Publishers Inc.

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behav. Brain Sci., 22*, 425–489.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics II. Spatially smoothed data collected under null-hypothesis and experimental conditions. *NeuroImage, 5*, 199-212.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998a). A critique of the use of the Kolmogorov-Smirnov (KS) statistic for the analysis of BOLD fMRI data. *Magn. Reson. Med., 39*, 500-505.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998b). The inferential impact of global signal covariates in functional neuroimaging analyses. *NeuroImage, 8*, 302-306.

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proc. Natl. Acad. Sci. USA, 91*, 7041-7045.

Amaral, D. G. (1993). Emerging principles of intrinsic hippocampal organization. *Curr. Op. Neurobiol., 3*, 225-229.

Amaral, D. G. (1999). What is where in the medial temporal lobe? *Hippocampus, 9*, 1-6.

Amaral, D. G., & Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: A review of anatomical data. *J. Neurosci., 31*, 571-591.

Amit, D. J. (1989). *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge, UK: Cambridge University Press.

Amit, D. J. (1998). Simulation in neurobiology: Theory or experiment? *Trends Neurosci., 21*, 231-237.

Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cog. Sci., 26*, 85-112.

Andersson, J. L. R. (1997). How to estimate global activity independent of changes in local activity. *NeuroImage, 6*, 237-244.

Anthony, M., & Bartlett, P. L. (1999). *Neural Network Learning: Theoretical Foundations*. Cambridge, UK: Cambridge University Press.

Arbib, M. A. (Ed.). (1995). *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press.

Arbib, M. A. (Ed.). (2003). *The Handbook of Brain Theory and Neural Networks* (2 ed.). Cambridge, MA: MIT Press.

Ardila, A., Rosselli, M., & Rosas, P. (1989). Neuropsychological assessment in illiterates: Visuospatial and memory abilities. *Brain and Cognition, 11*, 147-166.

Arndt, S. A., Cizadlo, T., O'Leary, D. S., Gold, S., & Andreasen, N. C. (1996). Normalizing counts and cerebral blood flow intensity in functional imaging studies of the human brain. *NeuroImage, 3*, 175-184.

Artola, A., & Singer, W. (1993). Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends Neurosci., 16*, 480-487.

Ashburner, J., & Friston, K. J. (1997). Spatial transformation of images. In J. C. Mazziotta (Ed.), *Human Brain Function* (pp. 43-58). San Diego: Academic Press.

Baddeley. (1998). *Human Memory: Theory and Practice*. Hove, UK: Psychology Press.

Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. (1992). Working memory. *Science, 255*, 556-559.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends Cogn. Sci., 4*, 417-423.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nat. Rev. Neurosci., 4*, 829-839.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev., 105*, 158-173.

Baddeley, A., Vargha-Khadem, F., & Mishkin, M. (2001). Preserved recognition in a case of developmental amnesia: Implications for the acquisition of semantic memory? *J. Cogn. Neurosci., 13*, 357–369.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation*. New York: Academic Press.

Barton, D. (1985). Awareness of language units in adults and children. In A. W. Ellis (Ed.), *Progress in the Psychology of Language I*: Laurence Erlbaum.

Barton, G. E., Berwick, R. C., & Ristad, E. S. (1987). *Computational Complexity and Natural Language*. Cambridge, MA: MIT Press.

Bayley, P. J., & Squire, L. R. (2002). Medial temporal lobe amnesia: gradual acquisition of factual information by nondeclarative memory. *J. Neurosci., 22*, 5741–5748.

Bayley, P. J., & Squire, L. R. (2003). The neuroanatomy of remote autobiographical memory. *Soc. Neurosci., 514.15*.

Bear, M. F., & Kirkwood, A. (1993). Neocortical long-term potentiation. *Curr. Op. Neurobiol., 3*, 197-202.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science, 269*, 1115-1118.

Beck, C., & Schlögl, F. (1993). *Thermodynamics of Chaotic Systems: An Introduction*. Cambridge, UK: Cambridge University Press.

Becker, J. T., MacAndrew, D. K., & Fiez, J. A. (1999). A comment on the functional localization of the phonological storage subsystem of working memory. *Brain and Cognition, 41*, 27-38.

Berridge, M. S., Cassidy, E. H., & Terris, A. H. (1990). A routine, automated synthesis of oxygen-15-labeled butanol for positron tomography. *J. Nucl. Med., 31*, 1727-1731.

Bickel, P. J., & Docksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Oakland, CA: Holden-Day.

Billingsley, P. (1995). *Probability and Measure* (3rd ed.). New York: Wiley and Sons.

Bilodeau, M., & Brenner, D. (1999). *Theory of Multivariate Statistics*. New York: Springer-Verlag.

Bliss, T. V., & Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature, 361*, 31-39.

Bollen, K. A. (1989). *Structural Equations with Latent Lariables*. New York: Wiley.

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci., 25*, 151-188.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometry, 50*, 229-242.

Bota, M., Dong, H. W., & Swanson, L. W. (2003). From gene networks to brain networks. *Nat. Neurosci., 6*, 795-799.

Brady, S. A. (1991). The role of working memory in reading disability. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological Processes in Literacy* (pp. 129-151). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science, 281*, 1185-1187.

Brockwell, P. J., & Davis, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag.

Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nat. Rev. Neurosci., 2*, 51–61.

Brown, M. W., Wilson, F. A., & Riches, I. P. (1987). Neuronal evidence that inferomedial temporal cortex is more important than hippocampus in certain processes underlying recognition memory. *Brain Res., 409*, 158-162.

Brown, M. W., & Xiang, J. Z. (1998). Recognition memory: Neuronal substrates of the judgment of prior occurrence. *Prog. Neurobiol., 55*, 149-189.

Buckner, R. L., & Koutstaal, W. (1998). Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. *Proc. Nat. Acad. Sci. USA, 95*, 891-898.

Buckner, R. L., Logan, J., Donaldson, D. I., & Wheeler, M. E. (2000). Cognitive neuroscience of episodic memory encoding. *Acta Psychol., 105*, 127-139.

Buckner, R. L., & Wheeler, M. E. (2001). Cognitive neuroscience of remembering. *Nat. Rev. Neurosci., 2*, 624-634.

Buechel, C., & Friston, K. J. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex, 7*, 768-778.

Buffalo, E. A., Reber, P. J., & Squire, L. R. (1998). The human perirhinal cortex and recognition memory. *Hippocampus, 8*, 330-339.

Burwell, R. D., Witter, M. P., & Amaral, D. G. (1995). Perirhinal and postrhinal cortices in the rat: A review of the neuroanatomical literature and comparison with findings from the monkey brain. *Hippocampus, 5*, 390–408.

Cabeza, R., Dolcos, F., Graham, R., & Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *NeuroImage, 16*, 317-330.

Cahill, L. F., Babinsky, R., Markowitsch, H. J., & McGaugh, J. L. (1995). The amygdala and emotional memory. *Nature, 377*, 6547-6549.

Cameron, R. F., Currier, R. D., & Haerer, A. F. (1971). Aphasia and literacy. *British Journal of Disorders of Communication, 6*, 161-163.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D. E., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *J. Neurosci., 20*, RC99: 91-95.

Carr, T. H. (1992). Automaticity and cognitive anatomy: Is word recognition "automatic" ? *Am. J. Psychol., 105*, 201-237.

Cary, L., & Verhaeghe, A. (1991). Efeito da prática da linguagem ou da alfabetização no conhecimento das fronteiras formais das unidades lexicais: Comparação de dois tipos de tarefas, *Actas das Jornadas de Estudos dos Processos Cognitivos* (pp. 33-49): Sociedade Portuguesa de Psicologia: Secção de Psicologia Cognitiva.

Castro-Caldas, A., Miranda Cavaleiro, P., Carmo, I., Reis, A., Leote, F., Ribeiro, C., & Ducla-Soares, E. (1999). Influence of learning to read and write on the morphology of the corpus callosum. *Eur. J. Neurol., 6*, 23-28.

Castro-Caldas, A., Peterson, K. M., Reis, A., Stone-Elander, S., & Ingvar, M. (1998). The illiterate brain: Learning to read and write during childhood influences the functional organisation of the adult brain. *Brain, 121*, 1053-1063.

Charniak, E. (1993). *Statistical Language Learning*. Cambridge, MA: MIT Press.

Charniak, E., & McDermott, D. (1985). *Introduction to Artificial Intelligence*. Reading, MA: Addison-Wesley.

Cherkassky, V., & Mulier, F. (1998). *Learning from Data: Concepts, Theory and Methods*. New York: Wiley and Sons.

Chomsky, N. (1957). *Syntactic Structures*. The Hague, The Netherlands: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA.: MIT Press.

Chomsky, N. (1980). On the biological basis of language capacities. In N. Chomsky, *Rules and Representations* (pp. 185-216). Oxford, UK: Blackwell.

Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.

Chomsky, N. (2000a). Language from an internalist perspective, *New Horizons in the Study of Language and Mind* (pp. 134-163). Cambridge, UK: Cambridge University Press.

Chomsky, N. (2000b). *New Horizons in the Study of Language and Mind*. Cambridge, UK: Cambridge University Press.

Chomsky, N., & Lasnik, H. (1995). The Theory of Principles and Parameters. In N. Chomsky, *The Minimalist Program* (pp. 13-128). Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (Eds.). (2001). *Connectionist Psycholinguistics*. Norwood, NJ: Ablex Publishing.

Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends Cogn. Sci., 7*, 300-307.

Churchland, P. S., & Sejnowski, T. J. (1992). *The Computational Brain*. Cambridge, MA: MIT Press.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing model of the Stroop effect. *Psychol. Rev., 99*, 45-77.

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *Am. J. Psychol., 105*, 239-269.

Coppens, P., Parente, M. A. M. P., & Lecours, A. R. (1998). Aphasia in illiterate individuals. In P. Coppens & Y. Lebrun & A. Basso (Eds.), *Aphasia in Atypical Populations* (pp. 175-202). London: Lawrence Erlbaum.

Corkin, S. (2002). What's new with the amnesic patient H. M.? *Nat. Rev. Neurosci., 3*, 153-160.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *J. Verb. Learn. Verb. Behav., 11*, 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen., 104*, 268-294.

Csete, M. E., & Doyle, J. C. (2002). Reverse engineering of biological complexity. *Science, 295*, 1664-1669.

Curtis, C. E., Zald, D. H., Lee, J. T., & Pardo, J. V. (2000). Object and spatial alternation tasks with minimal delays activate the right anterior hippocampus proper in humans. *NeuroReport., 11*(10), 2203-2207.

Cutland, N. J. (1980). *Computability: An Introduction to Recursive Function Theory*. Cambridge, Uk: Cambridge University Press.

D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., Lease, J., & Tang, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Cog. Brain Res., 7*, 1-13.

Damasio, H., & Damasio, A. R. (1980). The anatomical basis of conduction aphasia. *Brain, 103*, 337-350.

Damásio, H., Damásio, A. R., Castro-Caldas, A., & Hamsher, K. S. (1979). Reversal of ear advantage for phonetically similar words in illiterates. *J. Clin. Neuropsychol., 1*, 331-338.

Davachi, L., & Goldman-Rakic, P. S. (2001). Primate rhinal cortex participates in both visual recognition and working memory tasks: Functional mapping with 2-DG. *J. Neurophysiol., 85*, 2590-2601.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z., Schilstra, M. J., Clarke, P. J., Arnone,

M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L., & Bolouri, H. (2002). A genomic regulatory network for development. *Science, 295*, 1669-1678.

Davis, M. D., Sigal, R., & Weyuker, E. J. (1994). *Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science* (2 ed.). San Diego, CA: Academic Press.

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. USA, 95*, 14529-14534.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science, 298*, 2013-2015.

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *J. Neurosci., 15*, 5870-5878.

Devaney, R. L. (1989). *An Introduction to Chaotic Dynamical Systems* (2nd ed.). Redwood City, CA: Addison-Wesley.

Dolan, R. J., & Fletcher, P. C. (1997). Dissociating prefrontal and hippocampal function in episodic memory encoding. *Nature, 388*, 582-585.

Dronkers, N. F., Redfern, B. B., & Knight, R. T. (2000). The neural architecture of language disorders. In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2nd ed., pp. 949-958). Cambridge, MA: MIT Press.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). New York: Wiley and Sons.

Dudley, R. M. (2002). *Real Analysis and Probability* (2nd ed.). Cambridge, UK: Cambridge University Press.

Dunbar, R. (1998). The social brain hypothesis. *Evolutionary Anthropology, 6*, 178-190.

Dunbar, R. (2003). Evolution of the social brain. *Science, 14*, 1160-1161.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nat. Rev. Neurosci., 2*, 820-829.

Duncan, J., & Miller, E. K. (2002). Cognitive focus through adaptive neural coding in the primate prefrontal cortex. In R. T. Knight (Ed.), *Principles of Frontal Lobe Function* (pp. 278-292). Oxford, UK: Oxford University Press.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci., 23*, 475-483.

Edelman, G. M. (1990). *Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.

Edgington, E. S. (1995). *Randomization Tests* (3rd, revised and expanded ed.). New York: Marcel Dekker.

Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci., 1*, 41-50.

Eichenbaum, H., & Cohen, N. J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain*. Oxford, UK: Oxford University Press.

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nat. Neurosci., 3*, 1149–1152.

Elliott, R., & Dolan, R. J. (1999). Differential neural responses during performance of matching and nonmatching to sample tasks at two delay intervals. *Journal of Neuroscience., 19*, 5066-5073.

Elman, J. L. (1990). Finding structure in time. *Cogn. Sci., 14*, 179-211.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Engel, A., & Van den Broeck, C. P. L. (2001). *Statistical Mechanics of Learning*. Cambridge, UK: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychol. Rev., 102*, 211–245.

Faísca, L., Reis, A., & Petersson, K. M. (in preparation). Semantic networks and literacy: The application of the Multidimensional Scaling technique to study the “animal” semantic category.

Feldman, M. W., & Cavalli-Sforza, L. L. (1989). On the theory of evolution under genetic and cultural transmission with application to the lactose absorption problem. In M. W. Feldman (Ed.), *Mathematical evolutionary theory* (pp. 145-173). Princeton, NJ: Princeton University Press.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex, 1*, 1-47.

Fernández, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dümpelmann, M., Van Roost, D., & Elger, C. E. (1999). Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science, 285*, 1582-1585.

Fernández, G., Heitkemper, P., Grunwald, T., Van Roost, D., Urbach, H., Pezer, N., Lehnertz, K., & Elger, C. E. (2001). An inferior temporal stream for word processing with integrated mnemonic function. *Hum. Brain Map., 14*, 251–260.

Fernández, G., Klaver, P., Fell, J., Grunwald, T., & Elger, C. E. (2002). Human declarative memory formation: segregating rhinal and hippocampal contributions. *Hippocampus, 12*, 514-519.

Fernandez, G., Weyerts, H., Schrader-Bolsche, M., Tendolkar, I., Smid, H. G. O. M., Tempelmann, C., Hinrichs, H., Scheich, H., Elger, C. E., Mangun, G. R., & Heinze, H.-J. (1998). Successful verbal encoding into episodic memory engages the posterior hippocampus: A parametrically analyzed functional magnetic resonance imaging study. *J. Neurosci., 18*, 1841–1847.

Fletcher, P. C., Frith, C. D., & Rugg, M. D. (1997). The functional neuroanatomy of episodic memory. *TINS, 20*, 213-218.

Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain, 124*, 849-881.

Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Fodor, J. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.

Fodor, J. A., & Pylyshyn, Z. W. (1990). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3-71.

Forkstam, C., Hagoort, P., Ingvar, I., & Petersson, K. M. (in preparation). Neural correlates of artificial syntactic processing.

Fox, P. T., & Mintun, M. A. (1989). Non-invasive functional brain mapping by change distribution analysis of averaged PET images of H15O2 tissue activity. *J. Nucl. Med., 30*, 141-149.

Fox, P. T., Mintun, M. A., Reiman, E. M., & Raichle, M. E. (1988). Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. *J. Cereb. Blood Flow Metab., 8*, 642-653.

Fox, P. T., & Pardo, J. V. (1991). Does intersubject variability in cortical functional organization increase with neural "distance" from the periphery? *Ciba Found. Symp., 163*, 125-144.

Fox, P. T., & Raichle, M. E. (1984). Stimulus rate dependence of regional cerebral blood flow in human striate cortex demonstrated with positron emission tomography. *J. Neurophysiol., 51*, 1109-1121.

Frackowiak, R. S. J., Friston, K. J., Frith, C., Dolan, R., Price, C., Zeki, S., Ashburner, J., & Penny, W. (Eds.). (2004). *Human Brain Function* (2nd ed.). San Diego, CA: Academic Press.

Frackowiak, R. S. J., Friston, K. J., Frith, C. D., Dolan, R. J., & Mazziotta, J. C. (1997). *Human Brain Function*. San Diego: Academic Press.

Fransen, E., & Lansner, A. (1998). A model of cortical associative memory based on a horizontal network of connected columns. *Network, 18*, 115-124.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci., 6*, 78-84.

Friedman, R. F., Ween, J. E., & Albert, M. L. (1993). Alexia. In K. M. Heilman & E. Valenstein (Eds.), *Clinical Neuropsychology* (3rd ed., pp. 37-62). New York: Oxford University Press.

Friston, K. (1994). Functional and effective connectivity. *Hum. Brain Map., 2*, 56-78.

Friston, K., Holmes, A., Worsley, K., Poline, J., Frith, C., & al., e. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Map., 2*, 189-210.

Friston, K. J. (1995). Commentary and opinion: II. Statistical parametric mapping - Ontology and current Issues. *J. Cereb. Blood Flow Metab., 15*, 361-370.

Friston, K. J., Frith, C. D., Liddle, P. F., Dolan, R. J., Lammertsma, A. A., & Frackowiak, R. S. (1990). The relationship between global and local changes in PET scans. *J. Cereb. Blood Flow Metab., 10*, 458-466.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. J. (1991). Comparing functional (PET) images: The assessment of significant change. *J. Cereb. Blood Flow Metab., 11*, 690-699.

Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Applications. *NeuroImage, 16*, 484-512.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage, 19*, 1273-1302.

Friston, K. J., & Penny, W. (2003). Posterior probability maps and SPMs. *NeuroImage, 19*, 1240-1249.

Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage, 16*, 465-483.

Fuster, J. M. (1995). *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*. Cambridge, MA: MIT Press.

Fuster, J. M. (1997). *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe* (3rd ed.). New York: Lippincott-Raven.

Gabrieli, J. D. E., Brewer, J. B., Desmond, J. E., & Glover, G. H. (1997). Separate neural bases of two fundamental memory processes in the human medial temporal lobe. *Science, 276*, 264-266.

Garavan, H., Kelley, D., Rosen, A., Rao, S. M., & Stein, E. A. (2000). Practice-related functional activation changes in a working memory task. *Microscopy Res. Tech., 51*, 54–63.

Garcia, G., & Guerreiro, M. (1983). Pseudo-dementia from illiteracy. *6th European Meeting of the International Neuropsychological Society, Lisbon.*

Gathercole, S. E. (1995a). The assessment of phonological memory skills in preschool children. *British Journal of Educational Psychology, 65*, 155-164.

Gathercole, S. E. (1995b). Is nonword repetition a test of phonological memory or long term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*, 83-94.

Gathercole, S. E. (1995c). Nonword repetition: More than just a phonological output task. *Cog. Neuropsychol., 12*, 857-861.

Gathercole, S. E., & Baddeley, A. D. (1995). *Working Memory and Language*. Hillsdale: Lawrence Erlbaum Associates.

Gathercole, S. E., Pickering, S. J., Hall, M., & Peacker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Quart. J. Exp. Psychol., 54A*, 1-30.

Gazzaniga, M. S. (Ed.). (1999). *The New Cognitive Neurosciences: Second Edition*. Cambridge, MA: MIT Press.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (1998). *Cognitive Neuroscience: The Biology of the Mind*. New York: Norton.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation, 4*, 1-58.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelligence, 6*, 721-741.

Genovese, C. L., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage, 15*, 870-878.

Gershberg, F. B., & Shimamura, A. P. (1995). Impaired use of organizational strategies in free recall following frontal lobe damage. *Neuropsychologia, 33*, 1305-1333.

Gerstain, G. L., Bedenbaugh, P., & Aertsen, A. M. H. J. (1989). Neuronal assemblies. *IEEE Trans. Biomed. Engineer., 36*, 4-14.

Gerstner, W., & Kistler, W. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge, UK: Cambridge University Press.

Giedd, J. N., Rumsey, J. M., Castellanos, F. X., Rajapakse, J. C., Kaysen, D., Vaituzis, A. C., & al., e. (1996). A quantitative MRI study of the corpus callosum in children and adolescents. *Devel. Brain Res., 91*, 274–280.

Gold, B., & Buckner, R. L. (2002). Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron, 35*, 803-812.

Gold, E. M. (1967). Language identification in the limit. *Information and Control, 10*, 447-474.

Goldblum, M.-C., & Matute de Duran, E. (2000). Are illiterate people deep dyslexic? *J. Neurolinguistics, 2*, 103-111.

Goldman-Rakic, P. S. (1988). Topography of cognition: parallel distributed networks in primate association cortex. *Annu. Rev. Neurosci., 11*, 137-156.

Goldman-Rakic, P. S. (1998). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. In A. C. Roberts & T. W. Robbins & L. Weiskrantz (Eds.), *The Prefrontal Cortex: Executive and Cognitive Functions* (pp. 87-102). Oxford, UK: Oxford University Press.

Goody, J. (2000). *The Power of the Written Tradition*: Smithsonian Institution Press.

Grabowski, T. J., Frank, R. J., Brown, C. K., Damasio, H., Boles Ponto, L. L., Watkins, G. L., & Hichwa, R. D. (1996). Reliability of PET activation across statistical methods, subject groups and sample sizes. *Hum. Brain Map., 4*, 23-46.

Greitz, T., Bohm, C., Holte, S., & Eriksson, L. (1991). A computerized brain atlas: Construction, anatomical content and some applications. *J. Comput. Assist. Tomogr., 15*, 26-38.

Grossberg, S. (1988). *Neural Networks and Natural Intelligence*. Cambridge, MA: MIT Press.

Grunwald, T., Lehnertz, K., Heinze, H.-J., Helmstaedter, C., & Elger, C. E. (1998). Verbal novelty detection within the human hippocampus proper. *Proc. Natl. Acad. Sci. USA, 95*, 3193-3197.

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nat. Rev. Neurosci., 2*, 685-694.

Habib, R., Nyberg, L., & Tulving, E. (2003). Hemispheric asymmetries of memory: The HERA model revisited. *Trends Cog. Sci., 7*, 241-245.

Hagoort, P. (2004). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *NeuroImage, 20*, S18-S29.

Hagoort, P., Hald, L., Baastiansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438-441.

Hamilton, M. E., & Barton, D. (1983). Adult's definition of "word": The effects of literacy and development. *J. Pragmatics, 7*, 581-594.

Hasnain, M. K., Fox, P. T., & Woldorff, M. G. (1998). Intersubject variability of functional areas in the human visual cortex. *Hum. Brain Map., 6*, 301-315.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Hayduk, L. A. (1987). *Structural Equation Modeling with LISREL: Essentials and Advances*. Baltimore: Johns Hopkins University Press.

Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Hebb, D. O. (1949). *Organization of Behavior*. New York: Wiley.

Henson, R. N., Cansino, S., Herron, J., Robb, W., & Rugg, M. D. (2003). A familiarity signal in human anterior medial temporal cortex? *Hippocampus, 13*, 301–304.

Henson, R. N. A., Rugg, M. D., & Shallice, T. (2000). Confidence in recognition memory for words: Dissociating right prefrontal roles in episodic retrieval. *J. Cog. Neurosci., 12*, 913-923.

Herholz, K., Kessler, J., Slansky, I., Mielke, R., & Heiss, W. D. (1993). A model for separation of regional from global metabolic activation during continuous visual recognition in Alzheimer's disease. In I. Kanno (Ed.), *Quantification of Brain Function: Tracer Kinetics and Image Analysis in Brain PET* (pp. 555-560). Amsterdam: Excerpta Medica.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*. San Diego, CA: Addison-Wesley.

Higuchi, S.-I., & Miyashita, Y. (1996). Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proc. Natl. Acad. Sci. USA, 93*, 739–743.

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple Comparisons Procedures*. New York: Wiley and Sons.

Hoffman, K. L., & McNaughton, B. L. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science, 297*, 2070-2073.

Holdstock, J. S., Shaw, C., & Aggleton, J. P. (1995). The performance of amnesic subjects on test of delayed matching-to-sample and delayed matching-to-position. *Neuropsychologia, 33*, 1538-1596.

Holmes, A. P. (1994). Statistical Issues in Functional Brain Mapping, Ph.D. thesis, University of Glasgow.

Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2000). *Introduction to Automata Theory, Languages, and Computation* (2nd ed.). Reading, MA: Addison Wesley.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA, 79*, 2554-2558.

Hoppensteadt, F. C., & Izhikevich, E. M. (1997). *Weakly Connected Neural Networks*. New York: Springer-Verlag.

Horgan, T., & Tienson, J. (1996a). *Connectionism and the Philosophy of Psychology*. Cambridge, MA: The MIT Press.

Horgan, T., & Tienson, J. (1996b). What is wrong with classical cognitive science? In T. Horgan & J. Tienson, *Connectionism and the Philosophy of Psychology* (pp. 31-44). Cambridge, MA: MIT Press.

Horwitz, B. (1998). Using functional brain imaging to understand human cognition. *Complexity, 3*, 39-52.

Horwitz, B., McIntosh, A. R., Haxby, J. V., & Grady, C. L. (1995). Network analysis of brain cognitive function using metabolic and blood flow data. *Behav. Brain Res., 66*, 187-193.

Horwitz, B., Rumsey, J. M., & Donohue, B. C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc. Natl. Acad. Sci. USA, 95*, 8939-8944.

Horwitz, B., Soncrant, J. V., & Haxby, J. V. (1992). Covariance analysis of functional interactions in the brain using metabolic and blood flow data. In F. Gonzalez-Lima & T. Finkenstaedt & H. Scheich (Eds.), *Advances in Metabolic Mapping Techniques for Brain Imaging of Behavioral and Learning Functions* (pp. 189-217). Dordrecht, The Netherlands: Kluwer Academic Publishing.

Horwitz, B., Tagamets, M.-A., & McIntosh, A. R. (1999). Neural modeling, functional brain imaging, and cognition. *Trends Cog. Sci., 3*, 91-98.

Huang, K. (2001). *Introduction to Statistical Physics*. New York: Taylor & Francis.

Hunton, D. L., Miezin, F. M., Buckner, R. L., van Mier, H. I., Raichle, M. E., & Petersen, S. E. (1996). An assessment of functional-anatomical variability in neuroimaging studies. *Hum. Brain Map., 4*, 122-139.

Hutterlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia, 28*, 517-527.

Incisa della Rocchetta, I., & Milner, B. (1993). Strategic search and retrieval inhibition: The role of the frontal lobes. *Neuropsychologia, 31*, 503-524.

Ingvar, M., & Petersson, K. M. (2000). Functional maps - cortical networks. In A. W. Toga & J. C. Mazziotta (Eds.), *Brain Mapping: The Systems* (pp. 111-140). San Diego, CA: Academic Press.

Isidori, A. (1995). *Nonlinear Control Systems* (3rd ed.). New York: Springer Verlag.

Ivanco, T. L., Michelin, M., & Racine, R. J. (1996). Long-term potentiation in the neocortex: evidence suggesting that perirhinal cortex may play a role in memory and preconsolidation memory storage. *Soc. Neurosci), 22:1508*.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford, UK: Oxford University Press.

Jain, S., Osherson, D., Royer, J. S., & Sharma, A. (1999). *Systems That Learn*. Cambridge, MA: MIT Press.

Jansma, J. M., Ramsey, N. F., Slagter, H. A., & Kahn, R. S. (2001). Functional anatomical correlates of controlled and automatic processing. *J. Cog. Neurosci., 13*, 730-743.

Johnson, M. H. (1997). *Developmental cognitive neuroscience*. Oxford, UK: Blackwell.

Johnson, M. K., & Hirst, W. (1993). MEM: Memory subsystems as processes. In M. A. Conway & P. E. Morris (Eds.), *Theories of Memory* (pp. 241–286). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Johnston, D., & Wu, S. M.-S. (1995). *Foundations of Cellular Neurophysiology*. Cambridge, MA: MIT Press.

Jonides, J., Schumancher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reurter-Lorenz, P. A., Marchuetz, C., & Willis, C. R. (1998). The role of parietal cortex in verbal working memory. *J. Neurosci., 18*, 5026-5034.

Josse, G., & Tzourio-Mazoyer, N. (2004). Hemispheric specialization for language. *Brain Res. Rev., 44*, 1-12.

Jöreskog, K., & Sörbom, D. (1996). *LISREL 8 User's Reference Guide*. Chicago, IL: Scientific Software International.

Kanno, I., Hatazawa, J., Shimosegawa, E., Ishii, K., & Fujita, H. (1996). Proportionality of reaction CBF to baseline CBF with neural activation and deactivation. In T. Jones (Ed.), *Quantification of Brain Function Using PET* (pp. 362-362). San Diego: Academic Press.

Kapur, N., & Brooks, D. J. (1999). Temporally-specific retrograde amnesia in two cases of discrete bilateral hippocampal pathology. *Hippocampus, 9*, 247–254.

Kapur, S., Craik, F. I. M., Tulving, E., Wilson, A. A., Houle, S., & Brown, G. M. (1994). Neuroanatomical correlates of encoding in episodic memory: Levels of processing effect. *Proc. Natl. Acad. Sci. USA, 91*, 2008-2011.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Karmiloff-Smith, A. (1993). Self-organization and cognitive change. In M. H. Johnson (Ed.), *Brain development and cognition* (pp. 592-618). Oxford, UK: Blackwell.

Karmiloff-Smith, A. (1994). Précis of Beyond modularity: A developmental perspective on cognitive science. *Behav. Brain Sci., 17*, 693-745.

Karmiloff-Smith, A., Grant, J., Sims, K., Jones, M. C., & Cuckle, P. (1996). Rethinking metalinguistic awareness: representing and accessing knowledge about what counts as a word. *Cognition, 58*, 197-219.

Kelley, W. M., Miezin, F. M., McDermott, K. B., Buckner, R. L., Raichle, M. E., Cohen, N. J., Ollinger, J. M., Akbudak, E., Conturo, T. E., Snyder, A. Z., & Peterson, S. E. (1998). Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. *Neuron, 20*, 927-936.

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Frontal lobes: Cognitive neuropsychological aspects. In M. J. Farah (Ed.), *Behavioral Neurology and Neuropsychology* (pp. 409-418). New York: McGraw-Hill.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*, 671-680.

Kitano, H. (2002). Systems biology: A brief overview. *Science, 295*, 1662-1664.

Knecht, S., Floel, A., Drager, B., Breitestein, C., Sommer, J., Henningsen, H., Ringelstein, E. B., & Pascual-Leone, A. (2002). Degrees of language lateralization determine susceptibility to unilateral brain lesions. *Nat. Neurosci., 5*, 695-699.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273*, 1399–1401.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *J. Exp. Psychol.: Learn. Mem. Cog., 22*, 169-181.

Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons*. New York: Oxford University Press.

Koch, C., & Davis, J. L. (1994). *Large-Scale Neuronal Theories of the Brain*. Cambridge, MA: MIT Press.

Koch, C., & Laurent, G. (1999). Complexity and the nervous system. *Science, 284*, 96-98.

Koch, C., & Segev, I. (1998). *Methods in Neuronal Modeling: From Ions to Networks* (2nd ed.). New York: Oxford University Press.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature, 399*, 148-151.

Kolinsky, R., Cary, L., & Morais, J. (1987). Awareness of words as phonological entities: the role of literacy. *Applied Psycholinguistics, 8*, 223-237.

Kremin, H., Deloche, G., Metz-Lutz, M.-N., Hannequin, D., Dordain, M., Perrier, D., Cardebat, D., Ferrand, I., Larroque, C., Naud, E., Pichard, B., & Bunel, G. (1991). The effect of age, educational background and sex on confrontation naming in normals; principles for testing naming ability. *Aphasiology, 5*, 579-582.

Köhler, S., McIntosh, A. R., Moscovitch, M., & Winocur, G. (1998). Functional interactions between the medial temporal lobes and posterior neocortex related to episodic memory retrieval. *Cerebral Cortex, 8*, 451–461.

LaBerge, D., & Samuels, S. J. (1974). Towards a theory of automatic information processing in reading. *Cog. Psychol., 6*, 293–323.

Lansner, A., & Ekeberg, Ö. (1989). A one-layer feedback artificial neural network with a Bayesian learning rule. *Int. J. Neural Systems, 1*, 77-87.

Lasota, A., & Mackey, M. C. (1994a). *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. New York: Springer-Verlag.

Lasota, A., & Mackey, M. C. (1994b). Dynamical and semidynamical systems. In A. Lasota & M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics* (pp. 191-194). New York: Springer-Verlag.

Lavenex, P., & Amaral, D. G. (2000). Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus, 10*, 420-430.

Lecours, A. R. (1989). Literacy and acquired aphasia. In A. M. Galaburda (Ed.), *From Reading to Neurons* (pp. 27-39). Cambridge, MA: MIT Press.

Lecours, A. R., Mehler, J., Parente, M. A., Aguiar, L. R., Silva, A., B., Caetano, M., Camarotti, H., Castro, M. J., Dehaut, F., Dumais, C., Gauthier, L., Gurd, J., Leitao, O., Maciel, J., Machado, S., Melaragno, R., Oliveira, L. M., Paciornik, J., Sanvito, W., Silva, E. S., Silifrandi, M., & Torn,, C. H. (1987). Illiteracy and brain damage: 2. Manifestations of unilateral neglect in testing "Auditory comprehension" with iconographic materials. *Brain & Cognition, 6*, 243-265.

Lecours, A. R., Mehler, J., Parente, M. A., Caldeira, A., Cary, L., Castro, M. J., Dehaut, F., Delgado, R., Gurd, J., Karmann, D. F., Jakubovicz, R., Osorio, Z., Cabral, L. S., & Junqueira, A. M. S. (1987). Illiteracy and brain damage: 1. Aphasia testing in culturally contrasted populations (control subjects). *Neuropsychologia, 25*, 231-245.

Ledberg, A., Fransson, P., Larsson, J., & Petersson, K. M. (2001). A 4D approach to the analysis of functional brain images: Applications to FMRI data. *Hum. Brain Map., 13*, 185-198.

Leff, H. S., & Rex, A. F. (Eds.). (1990). *Maxwell's Demon: Entropy, Information, Computing*. Princeton, NJ: Princeton University Press.

Lepage, M., Ghaffar, O., Nyberg, L., & Tulving, E. (2000). Prefrontal cortex and episodic memory retrieval mode. *Proc. Natl. Acad. Sci. USA, 97*, 506-511.

Lepage, M., Habib, R., & Tulving, E. (1998). Hippocampal PET activations of memory encoding and retrieval: The HIPER model. *Hippocampus, 8*(4), 313-322.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Lewis, H. R., & Papadimitriou, C. H. (1981). *Elements of the Theory of Computation*. Englewood Cliffs, NJ: Prentice-Hall.

Li, S.-C. (2003). Biocultural orchestration of developmental plasticity across levels: The interplay of biology and culture in shaping the mind and behavior across the life span. *Psych. Bull., 129*, 171-194.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychol. Rev., 95*, 492-527.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*, 150-157.

Macdonald, C., & Macdonald, G. (1995). *Connectionism: Debates on Psychological Explanation*. Oxford, UK: Blackwell Publishers.

Mackey, M. C. (1992). *Time's Arrow: The Origins of Thermodynamic Behavior*. Berlin: Springer-Verlag.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *J. Exp. Psychol. Learn. Mem. Cogn., 14*, 126-135.

Maguire, E. A., Frith, C. D., Burgess, N., Donnett, J. G., & O'Keefe, J. (1998). Knowing where things are parahippocampal involvement in encoding object locations in virtual large-scale space. *J. Cog. Neurosci., 10*, 61-76.

Maitra, R. (1997). Estimating precision in functional images. *J. Comp. Graph. Stat., 6*, 132-142.

Mandl, F. (1988). *Statistical Physics* (2nd ed.). New York: Wiley and Sons.

Mandler, G. (1980). Recognising: The judgment of previous occurrence. *Psychol. Rev., 87*, 252-271.

Manly, J. J., Jacobs, D. M., Sano, M., Bell, K., Merchant, C. A., Small, S. A., & Stern, Y. (1999). Effect of literacy on neuropsychological test performance in nondemented, education-matched elders. *J. Int. Neuropsychol. Soc., 5*, 191-202.

Manns, J. R., Hopkins, R. O., Reed, J. M., Kitchener, E. G., & Squire, L. R. (2003). Recognition memory and the human hippocampus. *Neuron, 37*, 1–20.

Manns, J. R., Hopkins, R. O., & Squire, L. R. (2003). Semantic memory and the human hippocampus. *Neuron, 37*, 127–133.

Maquet, P. (2001). The role of sleep in learning and memory. *Science, 294*, 1048-1052.

Markowitsch, H. J. (1995). Anatomical basis of memory disorders. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 765-779). Cambridge, Massachusetts: A Bradford Book, The MIT Press.

Marr, D. (1971). Simple memory: A theory for archicortex. *Phil. Trans. Roy. Soc. Series B, 262*, 23-81.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman and Company.

McCauley, J. L. (1993a). *Chaos, Dynamics, and Fractals: An Algorithmic Approach to Deterministic Chaos*. Cambridge, UK: Cambridge University Press.

McCauley, J. L. (1993b). From flows to automata: Chaotic systems as completely deterministic machines. In J. L. McCauley, *Chaos, Dynamics, and Fractals: An Algorithmic Approach to Deterministic Chaos*. Cambridge, UK: Cambridge University Press.

McClelland, J. L. (1994). The organization of memory: A parallel distributed processing approach. *Rev. Neurol. (Paris), 150*, 570-579.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev., 102*, 419-437.

McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* (Vol. 2, Psychological and Biological Models). Cambridge, MA: MIT Press.

McColl, J. H., Holmes, A. P., & Ford, I. (1994). Statistical methods in neuroimaging with particular application to emission tomography. *Stat. Meth. Med. Res., 3*, 63-86.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys., 5*, 115-133.

McIntosh, A. R., & Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Map., 2*, 2-22.

McIntosh, A. R., Grady, C. L., Haxby, J. V., Maisog, J. M., Horwitz, B., & Clark, C. M. (1996). Within-subject transformation of PET regional cerebral blood flow data: ANCOVA, ratio, and Z-score adjustment on empirical. *Hum. Brain Map., 4*, 93-102.

McIntosh, A. R., Nyberg, L., Bookstein, F. L., & Tulving, E. (1997). Differential functional connectivity of prefrontal and medial temporal cortices during episodic memory retrieval. *Hum. Brain Map., 5*, 323–327.

McNamara, T. P., & Shelton, A. L. (2003). Cognitive maps and the hippocampus. *Trends Cogn. Sci., 7*, 333–335.

Mendonça, A., Mendonça, S., Reis, A., Faisca, L., Ingvar, M., & Petersson, K. M. (2003, November 10-12). *Reconhecimento de unidades lexicais em contexto frásico: O efeito da literacia.* Paper presented at the Congresso em Neurociências Cognitivas, Évora, Portugal.

Mendonça, S., Faisca, L., Silva, C., Ingvar, M., Reis, A., & Petersson, K. M. (2002). The role of literacy in the awareness of words as independent lexical units. *J. Int. Neuropsychol. Soc., 8*, 483.

Mesulam, M. M. (1998). From sensation to cognition. *Brain, 121*, 1013-1052.

Mesulam, M.-M. (2002). The human frontal lobes: Transcending the default mode through contingent encoding. In R. T. Knight (Ed.), *Principles of Frontal Lobe Function*. Oxford, UK: Oxford University Press.

Mesulam, M. M., & Mufson, E. F. (1985). The insula of Reil in man and monkey: Architectonics, connectivity and function. In E. G. Jones (Ed.), *Cerebral Cortex* (Vol. 4, pp. 179-226). New York: Plenum press.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci., 24*, 167-202.

Miller, M. B., Kingstone, A., & Gazzaniga, M. S. (2002). Hemispheric encoding asymmetry is more apparent than real. *J. Cog. Neurosci., 14*, 702-708.

Miller, R. (1991). *Cortico-Hippocampal Interplay and the Representation of Contexts in the Brain*. Berlin: Springer-Verlag.

Minsky, M. L. (1967). *Computation: Finite and Infinite Machines*. Englewood Cliffs, NJ: Prentice-Hall.

Montaldi, D., Mayes, A. R., Barnes, A., Pirie, H., Hadley, D., Patterson, J., & Wyper, D. (1997). Medial temporal lobe activations are produced by visual associative encoding and auditory verbal retrieval. *NeuroImage, 5*, S614.

Moore, C. (1991a). Generalized one-sided shifts and maps of the interval. *Nonlinearity, 4*, 727-745.

Moore, C. (1991b). Generalized shifts: Unpredictability and undecidability in dynamical systems. *Nonlinearity, 4*, 199-230.

Morais, J. (1993). Phonemic awareness, language and literacy. In R. M. Joshi & C. K. Leong (Eds.), *Reading Disabilities: Diagnosis and Component Processes* (pp. 175-184). Dordrecht, NL: Kluwer Academic Publishers.

Morais, J., & Kolinsky, R. (1994). Perception and awareness in phonological processing: The case of the phoneme. *Cognition, 50*, 287-297.

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *J. Cog. Neurosci., 4*, 249-252.

Moscovitch, M. (1994). Memory and working with memory: Evaluation of a component process model and comparisons with other models. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 269-310). Cambridge, MA: MIT Press.

Murray, A. C., & Mishkin, M. (1986). Visual recognition in monkeys following rhinal cortical ablations combined with either amygdalectomy or hippocampectomy. *J. Neurosci., 6*, 1991-2003.

Murray, E. A., & Bussey, T. J. (1999). Perceptual-mnemonic functions of the perirhinal cortex. *Trends Cog. Sci., 3*, 142-151.

Nadal, J., Toulouse, G., Changeux, J., & Dehaene, S. (1986). Networks of formal neurons and memory palimpsests. *Europhys. Lett., 1*, 535-542.

Nadel, L. (1994). Multiple memory systems: What and why, an update. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 39-64). Cambridge, MA: MIT Press.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr. Opin. Neurobiol., 7*, 217–227.

Newell, A. (1990). *Unified Theories of Cognition*: Harvard University Press.

Nieuwenhuys, R., Voogd, J., & van Huijzen, C. (1988). *The Human Central Nervous System: A Synopsis and Atlas* (3rd revised ed.). Berlin: Springer-Verlag.

Nishimori, H. (2001). *Statistical Physics of Spin Glasses and Information Processing*. Oxford, UK: Oxford University Press.

Nolde, S. F., Johnson, M. K., & Raye, C. L. (1998). The role of the prefrontal cortex during tests of episodic memory. *Trends Cog. Sci., 2*, 399-406.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature, 417*, 611-617.

Nyberg, L. (1998). Mapping episodic memory. *Behav. Brain Res., 90*, 107-114.

Nyberg, L., Cabeza, R., & Tulving, E. (1996). PET studies of encoding and retrieval: The HERA model. *Psychonom. Bull. Rev., 3*, 135-148.

Nyberg, L., Forkstam, C., Petersson, K. M., Cabeza, R., & Ingvar, I. (2002). Brain imaging of human memory systems: Between-systems similarities and within-system differences. *Cog. Brain Res., 13*, 281-292.

Nyberg, L., Marklund, P., Persson, J., Cabeza, R., Forkstam, C., Petersson, K. M., & Ingvar, I. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia, 41*, 371-377.

Nyberg, L., McIntosh, A. R., Houle, S., Nilsson, L. G., & Tulving, E. (1996a). Activation of medial temporal structures during episodic memory retrieval. *Nature, 380*(6576), 715-717.

Nyberg, L., McIntosh, A. R., Houle, S., Nilsson, L. G., & Tulving, E. (1996b). Activation of medial temporal structures during episodic memory retrieval. *Nature, 380*, 715-717.

Nyberg, L., Persson, J., Habib, R., Tulving, E., McIntosch, A. R., Cabeza, R., & Houle, S. (2000). Large scale neurocognitive networks underlying episodic memory. *J. Cog. Neurosci., 12*, 163-173.

Nyberg, L., Petersson, K. M., Nilsson, L.-G., Sandblom, J., Åberg, C., & Ingvar, M. (2001). Reactivation of motor brain areas during explicit memory for actions. *NeuroImage, 14*, 521-528.

O'Keefe, J, & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. Brain Ress 34, 171-175.

O'Keefe, J, & Nadel, L. (1979). Precis of O'Keefe and Nadel's The Hippocampus as a Cognitive Map. Behav. Brain Sci. Vol 2, 487-533.

O'Keefe, J., Burgess, N., Donnett, J. G., Jeffery, K. J., & Maguire, E. A. (1998). Place cells, navigational accuracy, and the human hippocampus. *Phil. Trans. Roy. Soc. London B, 353*, 1333-1340.

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nat. Neurosci., 7*, 75-79.

Oppenheim, A. V., Willsky, A. S., Hamid, S., & Hamid Nawab, S. (1996). *Signals and Systems* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In P. Shah (Ed.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge, UK: Cambridge University Press.

Ostrosky, F., Efron, R., & Yund, E. W. (1991). Visual detectability gradients: effect of illiteracy. *Brain and Cognition, 17*, 42-51.

Ott, E. (1993). Dynamical systems. In E. Ott, *Chaos in Dynamical Systems* (pp. 6-9). Cambridge, UK: Cambridge University Press.

Ott, E. (2002). *Chaos in Dynamical Systems* (2nd ed.). Cambridge, UK: Cambridge University Press.

Otten, L. J., Henson, R. N. A., & Rugg, M. D. (2001). Depth of processing effects on neural correlates of memory encoding: Relationship between findings from across- and within-task comparisons. *Brain, 124*, 399-412.

Owen, A., Evans, A., & Petrides, M. (1996). Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: a positron emission tomography study. *Cerebral Cortex, 6*, 31-38.

Owen, A. M. (2003). HERA today, gone tomorrow? *Trends Cog. Sci., 7*, 383-384.

Owen, A. M., Milner, B., Petrides, M., & Evans, A. C. (1996). A specific role for the right parahippocampal gyrus in the retrieval of object-location: A positron emission tomography study. *J. Cog. Neurosci., 8*, 588-602.

Owen, A. M., Sahakian, B. J., Semple, J., Polkey, C. E., & Robins, T. W. (1995). Visuo-spatial short-term recognition memory and learning after temporal lobe excision, frontal lobe excision or amygdalo-hippocampectomy in man. *Neuropsychologia, 33*, 1-24.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci., 25*, 563-593.

Papadimitriou, C. H. (1994). *Computational Complexity*. Reading, MA: Addison Wesley.

Partee, B. H., ter Meulen, A., & Wall, R. E. (1990). *Mathematical Methods in Linguistics*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Passingham, R. E., Stephan, K. E., & Kötter, R. (2002). The anatomical basis of functional localization in the cortex. *Nat. Rev. Neurosci., 3*, 606-616.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci., 6*, 674 - 681.

Paterson, S. J., Brown, J. H., Gsödl, M. K., Johnson, M. H., & Karmiloff-Smith, A. (1999). Cognitive modularity and genetic disorders. *Science, 286*, 2355-2358.

Patterson, K., & Lambon Ralph, M. A. (1999). Selective disorders of reading? *Curr. Op. Neurobiol., 9*, 235-239.

Paulesu, E., Démonet, J.-F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., Cappa, S. F., Cossu, G., Habib, M., Frith, C. D., & Frith, U. (2001). Dyslexia: Cultural diversity and biological unity. *Science, 291*, 2165-2167.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*, 342-345.

Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiak, R. S. J., & Frith, C. D. (1996). Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain, 119*, 143-157.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith, C. D., & Frith, U. (2000). A cultural effect on brain function. *Nat. Neurosci., 3*, 91-96.

Petersen, C. C., Malenka, R. C., Nicoll, R. A., & Hopfield, J. J. (1998). All-or-none potentiation at CA3-CA1 synapses. *Proc. Natl. Acad. Sci. USA, 95*, 4732–4737.

Petersson, K. M. (1998). Comments on a Monte Carlo approach to the analysis of functional neuroimaging data. *NeuroImage, 8*, 108-112.

Petersson, K. M. (2004). The human brain, language, and implicit learning. *Impuls, Journal of Psychology (Norwegian), 58*(3), 62-72.

Petersson, K. M. (2004, in press). On the relevance of the neurobiological analogue of the finite-state architecture. *Neurocomputing*.

Petersson, K. M., Elfgren, C., & Ingvar, M. (1997). A dynamic role of the medial temporal lobe during retrieval of declarative memory in man. *NeuroImage, 6*, 1-11.

Petersson, K. M., Elfgren, C., & Ingvar, M. (1999a). Dynamic changes in the functional anatomy of the human brain during recall of abstract designs related to practice. *Neuropsychologia, 37*, 567-587.

Petersson, K. M., Elfgren, C., & Ingvar, M. (1999b). Learning-related effects and functional neuroimaging. *Hum. Brain Map., 7*, 234-243.

Petersson, K. M., Forkstam, C., & Ingvar, M. (2004a). Artificial syntactic violations activate Broca's region. *Cognitive Science, 28*, 383-407.

Petersson, K. M., Grenholm, P, & Forkstam, C. (in preparation). Artificial grammar learning and neural networks.

Petersson, K. M., Gisselgård, J., Gretzer, M., & Ingvar, M. (in preparation). Interaction between a verbal working memory network and the medial temporal lobe.

Petersson, K. M., Nichols, T. E., Poline, J.-B., & Holmes, A. P. (1999a). Statistical limitations in functional neuroimaging I: Non-inferential methods and statistical models. *Phil. Trans. R. Soc. Lond. B, 354*, 1239-1260.

Petersson, K. M., Nichols, T. E., Poline, J.-B., & Holmes, A. P. (1999b). Statistical limitations in functional neuroimaging II: Signal detection and statistical inference. *Phil. Trans. R. Soc. Lond. B, 354*, 1261-1282.

Petersson, K. M., & Reis, A. (2005, in press). Characteristics of illiterate and literate cognitive processing: implications for brain-behavior co-constructivism. In P. B. Baltes & F. Rösler & P. A. Reuter-Lorenz (Eds.), *Lifespan Development and the Brain: The Perspective of Biocultural Co-Constructivism*. New York: Cambridge University Press.

Petersson, K. M., Reis, A., Askelöf, S., Castro-Caldas, A., & Ingvar, M. (1998). Differences in inter-hemispheric interactions between literate and illiterate subjects during verbal repetition. *NeuroImage, 7*, S217.

Petersson, K. M., Reis, A., Askelöf, S., Castro-Caldas, A., & Ingvar, M. (2000). Language processing modulated by literacy: A network-analysis of verbal repetition in literate and illiterate subjects. *J. Cog. Neurosci., 12*, 364-382.

Petersson, K. M., Reis, A., Castro-Caldas, A., & Ingvar, M. (1999). Effective auditory-verbal encoding activates the left prefontal and the medial temporal lobes: A generalization to illiterate subjects. *NeuroImage, 10*, 45-54.

Petersson, K. M., Reis, A., Castro-Caldas, A., & Ingvar, M. (submitted). Literacy: A cultural influence on the hemispheric balance in the inferior parietal cortex.

Petersson, K. M., Reis, A., & Ingvar, M. (2001). Cognitive processing in literate and illiterate subjects: A review of some recent behavioral and functional data. *Scand. J. Psychol., 42*, 251-167.

Petersson, K. M., Sandblom, J., Elfgren, C., & Ingvar, M. (2003). Instruction specific brain activations during episodic encoding: A generalized levels of processing effect with visuo-spatial material. *NeuroImage, 20*, 1795-1810.

Petersson, K. M., Sandblom, J., Gisselgård, J., & Ingvar, M. (2001). Learning related modulation of functional retrieval networks in man. *Scand. J. Psychol., 42*, 197-216.

Petrides, M. (1995). Functional organization of the human frontal cortex for mnemonic processing. Evidence from neuroimaging studies. *Ann. N. Y. Acad. Sci., 769*, 85-96.

Pinker, S. (1991). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *J. Clin. Exp. Neuropsychol., 17*, 291-331.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546-550.

Poldrack, R. A., Prabhakaran, V., Seger, C., & Gabrieli, J. D. E. (1999). Striatal activation during cognitive skill learning. *Neuropsychologia, 13*, 564-574.

Poletiek, F. H. (2002). Implicit learning of a recursive rule in an artificial grammar. *Acta Psychol., 111*, 323–335.

Poline, J.-B., Vandenberghe, R., Holmes, A. P., Friston, K. J., & Frackowiak, R. S. J. (1996). Reproducibility of PET Activation Studies: Lessons from a Multi-Center European Experiment. *NeuroImage, 4*, 34-54.

Poline, J. B., Worsley, K. J., Holmes, A. P., Frackowiak, R. S., & Friston, K. J. (1995). Estimating smoothness in statistical parametric maps: Variability of p values. *J. Comp. Assist. Tomogr., 19*, 788-796.

Posner, M. I. (Ed.). (1989). *The Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Quartz, S. R., & Sejnowski, T. J. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behav. Brain Sci., 20*, 537-596.

Raichle, M. E. (1994). Images of the mind: Studies with modern imaging techniques. *Annu. Rev. Psychol., 45*, 333-356.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A.-M. K., Pardo, J. V., Fox, P. T., & Peterson, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex, 4*, 8-26.

Ramsay, S. C., Murphy, K., Shea, S. A., Friston, K. J., Lammertsma, A. A., Clark, J. C., Adams, L., Guz, A., & Frackowiak, R. S. (1993). Changes in global cerebral blood flow in humans: Effect on regional cerebral blood flow during a neural activation task. *J. Physiol., 471*, 521-534.

Ramsey, N. F., Kirby, B. S., Gelderen, P. V., Berman, K. F., Duyn, J. H., Frank, J. A., Mattay, V. S., Van Horn, J. D., Esposito, E., Moonen, C. T. W., & Weinberger, D. R. (1996). Functional mapping of human sensorimotor cortex with BOLD fMRI correlates highly with H2O[15] PET rCBF. *J. Cereb. Blood Flow Metab., 16*, 755-764.

Ranganath, C., & D'Esposito, M. (2001). Medial temporal lobe activity associated with active maintenance of novel information. *Neuron, 31*, 865-873.

Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *J. Child Lang., 29*, 417-447.

Rees, G., Friston, K. J., & Koch, C. (2000). A direct, quantitative relationship between the functional properties of human & macaque V5. *Nat. Neurosci., 3*, 716-723.

Reggia, J. A., & Schulz, R. (2002). The role of computational modeling in understanding hemispheric interactions and specialization. *Cog. Sys. Res., 3*, 87-94.

Reif, F. (1965). *Fundamentals of Statistical and Thermal Physics*. New York: McGraw-Hill.

Reis, A., & Castro-Caldas, A. (1997). Illiteracy: A bias for cognitive development. *J. Int. Neuropsychol. Soc., 3*, 444-450.

Reis, A., Guerreiro, M., & Castro-Caldas, A. (1994). Influence of educational level of non brain-damaged subjects on visual naming capacities. *J. Clin. Exp. Neuropsychol., 16*, 939-942.

Reis, A., Guerreiro, M., Garcia, C., & Castro-Caldas, A. (1995). How does an illiterate subject process the lexical component of arithmetics? *J. Int. Neuropsychol. Soc., 1*, 206.

Reis, A., Guerreiro, M., & Petersson, K. M. (2001). Educational level on a neuropsychological baterry. *J. Int. Neuropsychol. Soc., 7*, 422-423.

Reis, A., Guerreiro, M., & Petersson, K. M. (2003). A socio-demographic and neuropsychological characterization of an illiterate population. *Applied Neuropsychology, 10*, 191-204.

Reis, A., & Petersson, K. M. (2003). Educational level, socioeconomic status and aphasia research: A comment on Connor et al. (2001) - Effect of socioeconomic status on aphasia severity and recovery. *Brain and Language, 87*, 1795-1810.

Reis, A., Petersson, K. M., Castro-Caldas, A., & Ingvar, M. (2001). Formal schooling influences two-but not three-dimensional naming skills. *Brain and Cognition, 47*, 394-411.

Reis, A., Petersson, K. M., Faísca, L., & Ingvar, M. (in preparation). Color makes a difference: Two-dimensional object naming skills in literate and illiterate subjects.

Ribot, R. (1882). *Diseases of Memory*. New York: Appleton.

Rieke, F., Warland, D., van Steveninck, R. R. D., & Bialek, W. (1996). *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press.

Robins, A. (1995). Catastrophic forgetting, rehearsal, and pseudorehearsal. *Connection Science, 7*, 123-146.

Robins, A. (1996a). Consolidation in neural networks and in the sleeping brain. *Connection Science, 8*, 259-275.

Robins, A. (1996b). Transfer in cognition. *Connection Science, 8*, 185-203.

Rogers, H. (2002). *Theory of Recursive Functions and Effective Computability, 5th print*. Cambridge, MA: MIT Press.

Rosenbleuth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose, and teleology. *Philos. Sci., 10*, 18-24.

Rosenfeld, A., & Kak, A. C. (1982). *Digital Picture Processing*. Orlando, Fla.: Academic Press.

Rosselli, M., Ardila, A., & Rosas, P. (1990). Neuropsychological assessment in illiterates: II. Language and praxic abilities. *Brain and Cognition, 12*, 281-296.

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R., & Dolan, R. J. (1996). Differential activations of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain, 119*, 2073-2083.

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1997). Brain regions supporting intentional and incidental memory: a PET study. *Neuroreport, 8*, 1283-1287.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends Cogn. Sci., 4*, 108-115.

Rugg, M. D., & Yonelinas, A. P. (2003). Human recognition memory: A cognitive neuroscience perspective. *Trends Cogn. Sci., 7*, 313–319.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* (Vol. 1, Foundations). Cambridge, MA: MIT Press.

Russel, S., & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall.

Sag, I. A., Wasow, T., & Bender, E. M. (2003). *Syntactic Theory: A Formal Introduction* (2nd ed.). Stanford, CA: Center for the Study of Language and Information.

Salmon, D. P., & Butters, N. (1996). Neurobiology of skill and habit learning. *Curr. Opin Neurobiol., 5*, 184–190.

Sandberg, A., Lansner, A., & Petersson, K. M. (2001). Selective enhancement of recall through plasticity modulation in an autoassociative memory. *Neurocomputing, 38-40*, 867-873.

Sandberg, A., Lansner, A., Petersson, K. M., & Ekeberg, Ö. (2000). A palimpsest memory based on an incremental Bayesian learning rule. *Neurocomputing, 32-33*, 987-994.

Sandberg, A., Lansner, A., Petersson, K. M., & Ekeberg, Ö. (2002). Bayesian attractor networks with incremental learning. *Network: Computation in Neural Systems, 13*, 179-194.

Savage, J. E. (1998). *Models of Computation: Exploring the Power of Computing*. Reading, MA: Addison-Wesley.

Scannell, J. W., & Young, M. P. (1999). Neuronal population activity and functional imaging. *Proc. R. Soc. London B, 266*, 875-881.

Schacter, D. L. (1994). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In E. Tulving & D. L. Schacter (Eds.), *Memory Systems 1994*. Cambridge, MA: MIT Press.

Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In E. Tulving & D. L. Schacter (Eds.), *Memory Systems 1994* (pp. 1-38). Cambridge, MA: MIT Press.

Schacter, D. L., & Wagner, A. D. (1999). Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus, 9*, 7-24.

Schneider, W., Pimm-Smith, M., & Worden, M. (1994). Neurobiology of attention and automaticity. *Curr. Op. Neurobiol., 4*, 177-182.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search,and attention. *Psychol. Rev., 84*, 1–66.

Scholz, B. C., & Pullum, G. K. (2002). Searching for arguments to support linguistic nativism. *Ling. Rev., 19*, 185-223.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiat., 20*, 11-21.

Senda, M., Ishii, K., Oda, K., Sadato, N., Kawashima, R., Sugiura, M., Kanno, I., Ardekani, B., Minoshima, S., & Tatsumi, I. (1998). Influence of ANOVA Design and Anatomical Standardization on Statistical Mapping for PET Activation. *NeuroImage, 8*, 283-301.

Shastri, L. (1995). Structured connectionist models. In M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks* (pp. 949-952). Cambridge, MA: MIT Press.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist encoding of rules, variables, and dynamic bindings using temporal synchrony. *Behav. Brain Sci., 16*, 417-494.

Shaywitz, S. E., Shaywitz, B. A., Pugh, K. R., Fulbright, R. K., Constable, R. T., Mencl, W. E., Shankweiler, D. P., Liberman, A. M., Skudlarski, P., Fletcher, J. M., Katz, L., Marchione, K. E., Lacaide, C., Gatenby, C., & Gore, J. C. (1998). Functional disruption in the organization of the brain for reading in dyslexia. *Proc. Natl. Acad. Sci. USA, 95*, 2636-2641.

Shepherd, G. M. (1997). *The Synaptic Organization of the Brain* (4th ed.). New York: Oxford University Press.

Shimamura, A. P. (1995). Memory and frontal lobe function. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 803-813). Cambridge, MA: MIT Press.

Shinohara, M., Dollinger, B., Brown, G., Rapoport, S., & Sokoloff, L. (1979). Cerebral glucose utilization: Local changes during and after recovery from spreading cortical depression. *Science, 203*, 188-190.

Shinohara, T. (1994). Rich classes inferable from positive data: Length-bounded elementary formal systems. *Information and Computation, 108*, 175-186.

Shinohara, T., & Arimura, H. (2000). Inductive inference of unbound unions of pattern languages from positive data. *Theoretical Computer Science, 241*, 191-209.

Siegelmann, H. T. (1999). *Neural Networks and Analog Computation: Beyond the Turing Limit*. Basel, Switzerland: Birkhäuser.

Siegelmann, H. T., & Fishman, S. (1998). Analog computation with dynamical systems. *Physica D, 120*, 214-235.

Siegelmann, H. T., & Sontag, E. D. (1994). Analog computation via neural networks. *Theoretical Computer Science, 131*, 331-360.

Siesjö, B. K. (1978). *Brain Energy Metabolism*. Chichester: Wiley.

Silva, C., Faisca, L., Mendonça, S., Ingvar, M., Petersson, K. M., & Reis, A. (2002). Awareness of words as phonological entities in an illiterate population. *J. Int. Neuropsychol. Soc., 8*, 483.

Silva, C., Petersson, K. M., Ingvar, M., & Reis, A. (2001). The effects of formal education on the qualitative aspects of a verbal fluency task. *J. Int. Neuropsychol. Soc., 7*, 419.

Silva, C. G., Petersson, K. M., Faísca, L., Ingvar, M., & Reis, A. (2004). The effects of formal education on the quantitative and qualitative aspects of verbal semantic fluency. *J. Clin. Exp. Neuropsychol., 26*, 266-277.

Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nat. Rev. Neurosci., 4*, 637-648.

Skaggs, W. E., & McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science, 271*, 1870-1873.

Small, S., Arun, A., Perera, G., Delapaz, R., Mayeaux, R., & Stern, Y. (2001). Circuit mechanisms underlying memory encoding and retrieval in the long axis of the hippocampal formation. *Nat. Neurosci., 4*, 442–449.

Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proc. Natl. Acad. Sci. USA, 95*, 12061-12068.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*, 1657-1661.

Smith, M. E., Stapleton, J. M., & Halgren, E. (1986). Human medial temporal lobe potentials evoked in memory and language tasks. *Electroencephalogr. Clin. Neurophysiol., 63*, 145-159.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behav. Brain Sci., 11*, 1-74.

Sommer, I. E. C., Ramsey, N. F., Mandl, R. C. W., & Kahn, R. S. (2002). Language lateralization in monozygotic twin pairs concordant and discordant for handedness. *Brain, 125*, 2710-2718.

Sontag, E. D. (1998). *Mathematical Control Theory: Deterministic Finite Dimensional Systems* (2nd ed.). New York: Springer Verlag.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol. Rev., 99*, 195-231.

Squire, L. R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 203-231). Cambridge, MA: MIT Press.

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: A neurobiological perspective. *Curr. Op. Neurobiol., 5*, 169-177.

Squire, L. R., Cohen, N. J., & Nadel, L. (1984). The medial temporal region and memory consolidation: A new hypothesis. In H. Weingartner & E. Parker (Eds.), *Memory consolidation* (pp. 185-210). Hillsdale, New Jersey: Erlbaum.

Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annu. Rev. Psychol., 44*, 453-495.

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.*

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science, 253*, 1380-1386.

Squire, L. R., Zola-Morgan, S., & Chen, K. S. (1988). Human amnesia and animal models of amnesia - performance of amnesic patients on tests designed for the monkey. *Behav. Neurosci., 102*, 210-221.

Stadler, M. A., & Frensch, P. A. (Eds.). (1998). *Handbook of Implicit Learning*. Thousand Oaks, CA: Sage Publications.

Stark, C. E. L., & Okado, H. (2003). Making memories without trying: Medial temporal lobe activity associated with incidental memory formation during recognition. *J. Neurosci., 23*, 6748–6753.

Stark, C. E. L., & Squire, L. R. (2003). Hippocampal damage equally impairs memory for single items and memory for conjunctions. *Hippocampus, 13*, 281–292.

Stephan, K. E., Hilgetag, C.-C., Burns, G. A. P. C., O'Neill, M. A., Young, M. P., & Koetter, R. (2000). Computational analysis of functional connectivity between areas of primate cerebral cortex. *Phil.Trans. R. Soc. Lond. B, 355*, 111-126.

Stephan, K. E., Marshall, J. C., Friston, K. J., Rowe, J. B., Ritzl, A., Zilles, K., & Fink, G. R. (2003). Lateralized cognitive processes and lateralized task control in the human brain. *Science, 301*, 384-386.

Stern, C. E., Corkin, S., Gonzalez, R. G., Guimaraes, A. R., Baker, J. R., Jennings, P. J., Carr, C. A., Sugiura, R. M., Vedantham, V., & Rosen, B. R. (1996). The hippocampal formation participates in novel picture encoding: Evidence from functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA, 93*, 8660-8665.

Stickgold, R., Hobson, J. A., Fosse, R., & Fosse, M. (2001). Sleep, learning, and dreams: Off-line memory reprocessing. *Science, 294*, 1052-1057.

Stillings, N. A., Weisler, S. E., Chase, C. H., Feinstein, M. H., Garfield, J. L., & Rissland, E. L. (1995). *Cognitive Science: An Introduction* (2nd ed.). Cambridge, MA: MIT Press.

Stuss, D. T., & Knight, R. T. (Eds.). (2002). *Principles of Frontal Lobe Function*. Oxford, UK: Oxford University Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Suzuki, W. A. (1994). What can neuroanatomy tell us about the functional components of the hippocampal memory system. *Behav. Brain Sci., 17*, 496-498.

Suzuki, W. A. (1996). Neuroanatomy of the monkey entorhinal, perirhinal, and parahippocampal cortices: Organization of cortical inputs and interconnections with amygdala and striatum. *Semin. Neurosci., 8*, 3-12.

Suzuki, W. A., & Amaral, D. G. (1994a). Perirhinal and parahippocampal cortices of the Macaque monkey: Cortical afferents. *J. Comp. Neurol., 350*, 497-533.

Suzuki, W. A., & Amaral, D. G. (1994b). Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *J. Neurosci., 14*, 1856-1857.

Suzuki, W. A., & Eichenbaum, H. (2000). The neurophysiology of memory. *Ann. NY. Acad. Sci., 911*, 175–191.

Suzuki, W. A., Miller, E. K., & Desimone, R. (1997). Object and place memory in macaque entorhinal cortex. *J. Neurophysiol., 78*, 1062-1081.

Sybriska, E., Davachi, L., & Goldman-Rakic, P. S. (2000). Prominence of direct entorhinal-CA1 pathway activation in sensorimotor and cognitive tasks revealed by 2-DG functional mapping in non-human primate. *J. Neurosci.*, 20,5827-5834.

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., McNaughton, B. L., & Fernandez, G. (submitted). Prospective tracking of system level correlates of declarative memory consolidation in humans.

Takehara, K., Kawahara, S., & Kirino, Y. (2003). Time-dependent reorganization of the brain components underlying memory retention in trace eyeblink conditioning. *J. Neurosci., 23*, 9897-9905.

Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. Stuttgart: George Thieme Verlag.

Tanenbaum, A. S. (1990). *Structured Computer Organization* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Thompson, P. M., Cannon, T. D., Narr, K. L., van Erp, T., Poutanen, V. P., Huttunen, M., Lonnqvist, J., Standertskjold-Nordenstam, C. G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C. I., & Toga, A. W. (2001). Genetic influences on brain structure. *Nat. Neurosci., 4*, 1253-1258.

Thompson, P. M., Giedd, J. N., Woods, R. P., MacDonald, D., Evans, A. C., & Toga, A. W. (2000). Growth patterns in the developing brain detected by using continuum mechanical tensor maps. *Nature, 404*, 190–193.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc. Natl. Acad. Sci. USA, 94*, 14792-14797.

Toga, A. W., Mazziotta, J. C., & Frackowiak, R. S. J. (Eds.). (2000). *Brain Mapping*. San Diego, CA: Academic Press.

Trappenberg, T. P. (2002). *Fundamentals of Computational Neuroscience*. Oxford, UK: Oxford University Press.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus, 4*, 374-391.

Tulving, E. (1983). *Elements of Episodic Memory*: Calderon.

Tulving, E. (1989). Memory: performance, knowledge and experience. *Eur. J. Cog. Psychol., 1*, 3-26.

Tulving, E. (1995). Organization of memory: Quo vadis? In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 839-847). Cambridge, MA: MIT Press.

Tulving, E., Hayman, C. A. G., & MacDonald, C. A. (1991). Long-lasting perceptual priming and semantic learning in amnesia: A case experiment. *J. Exp. Psychol. Learn. Mem. Cogn., 17*, 595–617.

Tulving, E., Hebib, R., Nyberg, L., Lepage, M., & McIntosh, A. R. (1999). Positron emission tomography correlations in and beyond the medial temporal lobes. *Hippocampus, 9*, 71-82.

Tulving, E., Kapur, S., Craik, F. I. M., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron Emission tomography findings. *Proc. Natl. Acad. Sci. USA, 91*, 2016-2020.

Tulving, E., Markowitsch, H., Kapur, S., Habib, R., & Houle, S. (1994). Novelty encoding networks in the human brain: Positron emission tomography data. *NeuroReport, 5*, 2525-2528.

Tulving, E., & Markowitsch, H. J. (1997). Memory beyond the hippocampus. *Curr. Op. Neurobiol., 7*, 209-216.

Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: Role of the hippocampus. *Hippocampus, 8*, 198–204.

Tulving, E., Markowitsch, H. J., Craik, F. I. M., Habib, R., & Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cerebral Cortex, 6*, 71-79.

Tulving, E., & Schacter, D. L. (Eds.). (1994). *Memory Systems 1994*: The MIT Press.

Wagner, A. D. (1999). Working memory contributions to human learning and remembering. *Neuron, 22*, 19-22.

Wagner, A. D., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). Prefrontal cortex and recognition memory: Functional-MRI evidence for context-dependent retrieval processes. *Brain, 121*, 1985-2002.

Wagner, A. D., Poldrack, R. A., Eldridge, L. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998b). Material-specific lateralization of prefrontal activation during episodic encoding and retrieval. *NeuroReport, 9*, 3711-3717.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutsmal, W., Maril, A., Dale, A., Rosen, B. R., & Buckner, R. L. (1998). Building memories: Remembering and forgetting verbal experiences as predicted by brain activity. *Science, 281*, 1188-1191.

Wahlgren, N., & Lansner, A. (2001). Biological evaluation of a Hebbian–Bayesian learning rule. *Neurocomputing, 38–40*, 433–438.

Vallar, G., & Papagno, C. (1995). Neuropsychological impairments of short-term memory. In A. D. Baddeley & B. A. Wilson & W. F. N. (Eds.), *Handbook of Memory Disorders*. New York: Wiley.

Vapnik, V. (1998). *Statistical Learning Theory*. New York: Wiley and Sons.

Varela, F., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nat. Rev. Neurosci., 2*, 229-239.

Vargha-Khadem, F., Gaffan, D., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science, 277*, 376–380.

Wasow, T. (1989). Grammatical theory. In M. Posner (Ed.), *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Wechsler, A. F. (1976). Crossed aphasia in an illiterate dextral. *Brain and Language, 3*, 164-172.

Weiss, C., Bouwmeestev, H., Power, J. M., & Disterhoft, J. F. (1999). Hippocampal lesions prevent trace eyeblink conditioning in the freely moving rat. *Behav. Brain Res., 99*, 123-132.

Weis, S., Klaver, P., Reul, J., Elger, C. E., & Fernández, G. (2004). Identical temporal and cerebellar regions support both declarative memory formation and retrieval. *Cerebral Cortex, 14*, 256-267.

Weis, S., Specht, K., Klaver, P., Tendolkar, T., Willmes, K., Ruhlmann, J., Elger, C. E., & Fernández, G. (submitted). Contextual retrieval and item recognition are related to distinct processes within the human medial temporal lobe.

Westmacott, R., & Moscovitch, M. (2001). Names and words without meaning: incidental postmorbid semantic learning in a person with extensive bilateral medial temporal lobe damage. *Neuropsychologia, 15*, 586–596.

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1995). Frontal lobe damage produces episodic memory impairment. *J. Int. Neuropsychol. Soc., 1*, 525-536.

Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychol. Rev., 86*, 44-60.

Wiener, N. (1948). *Cybernetics or Control and Communication in the Animal and the Machine*. New York: Technology Press.

Wilson, R. A., & Keil, F. C. (Eds.). (2001). *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA.

Wiser, A. K., Andreasen, N., O'Leary, D. S., Crespo, F. B., Boles-Ponto, L. L., Watkins, G. L., & Hichwa, R. D. (2000). Novel vs. well-learned memory for faces: A positron emission tomography study. *J. Cog. Neurosci., 12*, 255–266.

Vochatzer, K. G., & Blick, K. A. (1989). Levels of processing and the retention of paired-associates. *Perceptual and Motor Skills, 69*, 349-350.

Voermans, N. C., Petersson, K. M., Daudey, L., Weber, B., van Spaendonck, K. P., Kremer, H. P. H., & Fernández, G. (2004). Interaction between the human hippocampus and caudate nucleus during route recognition. *Neuron, 43*, 427-435.

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nat. Rev. Neurosci., 4*, 139-147.

Worsley, K., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Map., 4*, 58-73.

Worsley, K. J. (1996). The geometry of random images. *Chance, 9*, 27-40.

Worsley, K. J., Andermann, M., Koulis, T., MacDonald, D., & Evans, A. C. (Abstract presented at HBM99). Detecting changes in non-stationary images via statistical flattening. *NeuroImage*.

Worsley, K. J., Evans, A. C., Marrett, S., & Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab., 12*, 900-918.

Worsley, K. J., Marrett, S., Neelin, P., & Evans, A. C. (1996b). Searching scale space for activations in PET images. *Hum. Brain Map., 4*, 74-90.

Vygotsky, L. S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.

Yonelinas, A. P., Kroll, N. E. A., Quamme, J. R., Lazzara, M. M., Sauvé, M.-J., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nat. Neurosci., 5*, 1236-1241.

Young, M. P., Hilgetag, C.-C., & Scannell, J. W. (2000). On imputing function to structure from behavioral effects of brain lesions. *Phil. Trans. R. Soc. Lond. B, 355*, 147-161.

Zaidel, E., & Iacoboni, M. (Eds.). (2003). *The Parallel Brain: The Cognitive Neuroscience of the Corpus Callosum*. Cambridge, MA: MIT Press.

Zarahn, E., Aguirre, G. K., & D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage, 5*, 179-197.

Zarahn, E., Rakitin, B., Abela, D., Flynn, J., & Stern, Y. (2004). Positive evidence against human hippocampal involvement in working memory maintenance of familiar stimuli. *Cerebral Cortex*,

Zola, S. M., Squire, L. R., Teng, E., Stefanacci, L., Buffalo, E. A., & Clark, R. E. (2000). Impaired recognition memory in monkeys after damage to the hippocampal region. *J. Neurosci., 20*, 451-463.

Zola-Morgan, S., & Squire, L. R. (1993). Neuroanatomy of memory. *Annu. Rev. Neurosci., 16*, 547-563.

Zola-Morgan, S. M., & Squire, L. R. (1990). The primate hippocampal formation: Evidence for a time-limited role in memory storage. *Science, 250*, 288-290.

Øksendal, B. (2000). *Stochastic Differential Equations: An Introduction with Applications* (5 ed.). Berlin: Springer-Verlag.

9. ORIGINAL PAPERS

Paper 1.

Petersson, K. M., Elfgren, C., & Ingvar, M. (1999). Learning-related effects and functional neuroimaging. Hum. Brain Map., 7, 234-243.

Paper 2

Petersson, K. M., Elfgren, C., & Ingvar, M. (1997). A dynamic role of the medial temporal lobe during retrieval of declarative memory in man. NeuroImage, 6, 1-11.

Paper 3

Petersson, K. M., Elfgren, C., & Ingvar, M. (1999). Dynamic changes in the functional anatomy of the human brain during recall of abstract designs related to practice. Neuropsychologia, 37, 567-587.

Paper 4

Petersson, K. M., Sandblom, J., Gisselgård, J., & Ingvar, M. (2001). Learning related modulation of functional retrieval networks in man. Scand. J. Psychol., 42, 197-216.

Paper 5

Petersson, K. M., Reis, A., Castro-Caldas, A., & Ingvar, M. (1999). Effective auditory-verbal encoding activates the left prefontal and the medial temporal lobes: A generalization to illiterate subjects. NeuroImage, 10, 45-54.

Paper 6

Castro-Caldas, A., Petersson, K. M., Reis, A., Stone-Elander, S., & Ingvar, M. (1998). The illiterate brain: Learning to read and write during childhood influences the functional organization of the adult brain. Brain, 121, 1053-1063.

Paper 7

Petersson, K. M., Reis, A., Askelöf, S., Castro-Caldas, A., & Ingvar, M. (2000). Language processing modulated by literacy: A network-analysis of verbal repetition in literate and illiterate subjects. J. Cog. Neurosci., 12, 364-382.

Paper 8

Petersson, K. M., Reis, A., Castro-Caldas, A., & Ingvar, M. (submitted). Literacy: A cultural influence on the hemispheric balance in the inferior parietal cortex.