

The Neural Bases of Regularity Learning

Dissertation

zur Erlangung des Grades eines Doktors der Philosophie
der Philosophischen Fakultäten der Universität des Saarlandes

vorgelegt von Dipl-Psych. Christian Döller

geboren am 10.12.1973 in Würzburg

Dekan:

Prof. Dr. Rainer Krause, Universität des Saarlandes

Berichterstatter:

Prof. Dr. Axel Mecklinger, Universität des Saarlandes

PD Dr. Hubert Zimmer, Universität des Saarlandes

Tag der Disputation:

14.02.2005

Für Tanja♥

We must suppose a very delicate adjustment whereby the circulation follows the needs of the cerebral activity. Blood very likely may rush to each region of the cortex according as it is most active, but of this we know nothing.

William James, *The Principles of Psychology* (1890)

Cognitive neuroscience can move forward with greater confidence in the knowledge that changes in blood flow and oxygen levels do represent definable alterations in neuronal activity.

Marcus E. Raichle, *Nature*, 412 (2001)

Abstract

The existence of two separate learning and memory systems has been proposed in experimental psychology and cognitive neuroscience (Chapter 1), specialized for complementary functions, namely (1) fast learning unique episodes (first system; Chapter 2) and (2) gradual learning regularities across multiple episodes (second system; Chapter 3). The present thesis aimed at specifying the neural and cognitive bases of the second system. Within the framework of this thesis, the term ‘regularity learning’ was introduced to define the learning process associated with the second system, namely the extraction of regularities, i.e. overlapping, invariant features of multiple episodes.

The neural and cognitive processes underlying regularity learning were investigated in three experiments. Two brain regions were of main interest, the hippocampus and the prefrontal cortex, PFC (Chapter 4). Functional magnetic resonance imaging (fMRI) was used as the main method (Chapter 5). In all experiments, volunteers had to learn object-position conjunctions in several experimental trials and blocks. The design of the experiments included two conditions, a context-specific (CS) and an invariant learning (IL) condition (Chapter 6). In the CS condition, objects and positions were variably mapped across trials. In contrast, in the IL condition positions (Experiment 1A and 3B) or objects (Experiment 1B, 2, and 3A) were held constant within blocks, enabling subjects to extract regularities across trials, i.e. invariant positions or objects in object-position conjunctions, respectively.

In Experiment 1A (Chapter 7) and Experiment 1B (Chapter 8), performance increased across trials within blocks of the IL condition, but not in the CS condition. In both experiments, hippocampal activity decreased as a function of learning in the IL condition. Conversely, the hippocampus was activated continuously in the CS condition. In contrast to a right-lateralized hippocampal activation decrease during learning spatial regularities (Experiment 1A), learning object regularities (Experiment 1B) was associated with a decrease of bilateral hippocampal activity. In both experiments, left lateral prefrontal and right striatal brain regions showed an increase of activity as a function of learning in the IL condition. In Experiment 2 (Chapter 9), the cognitive processes underlying learning object regularities were investigated in more detail. In six behavioral studies, the influence of object distinctiveness (Experiment 2A-2B), memory load and learning duration (Experiment 2C-2E), and the effect of simultaneous object-position bindings (Experiment 2F) on learning invariant objects in object-position conjunctions was examined. The behavioral

data indicate that the increase of object distinctiveness mainly affects overall memory performance, whereas manipulation of the learning duration and the memory load affects the learning process. Moreover, a dissociation of two learning mechanisms could be observed: a process operating within learning blocks and a process operating across blocks. Finally, Experiment 3 (Chapter 10) aimed at specifying whether separable brain regions mediate these two learning mechanisms. In contrast to Experiment 1A and 1B, trials were blocked by condition to minimize the probability that subjects based their judgment on a common strategy for both conditions. On a behavioral level, subjects showed a performance increase within and across learning blocks in an object IL condition (Experiment 3A) and in a spatial IL condition (Experiment 3B), but not in the respective CS conditions. On a neural level, within-block learning was associated with a learning-related decrease of hippocampal and a learning-related increase of prefrontal-striatal activity in both experiments, by this replicating the results of Experiment 1A and 1B. In contrast, distinct prefrontal-striatal regions were selectively involved in across-block learning.

To summarize, on a cognitive level regularity learning within the framework of the present thesis can be characterized by the following criteria: (1) the extraction of regularities across input pattern, (2) the gradual nature of the extraction process, (3) the robust maintenance of extracted regularities over time, and (4) the aggregated representational format of the extracted information. Furthermore, the distinction between a within-block and an across-block learning effect might indicate that the former effect is associated with an instance-based learning process, whereas the latter effect might reflect a rule-based learning process, i.e. the transfer of knowledge of invariant features to new instances. On a neural level, regularity learning was associated with a decrease of hippocampal and an increase of prefrontal-striatal activity. The reduced learning-related hippocampal activation presumably reflects lower binding requirements in the IL condition, since variable objects can be bound to invariant positions (Experiment 1A and 3) or vice versa (Experiment 1B, 2, and 3). Furthermore, the imaging results of Experiment 1A and 1B suggest a domain-specific hemispheric specialization of the hippocampus during regularity learning, reflecting hippocampal sensitivity to perceptual stimulus attributes of invariant episodic features (bottom-up mechanism). The learning-related prefrontal modulation seems to reflect the requirement to extract and maintain regularities across trials and the adjustment of object-position conjunctions on the basis of the extracted knowledge, possibly mediated by rule-like prefrontal representations (top-down mechanism). Finally, the striatum might encode the increased predictability of invariant features as a function of learning and possibly provides an internal reinforcement signal to the PFC.

In conclusion, the present results provide new insights into the neural basis of regularity learning and point to a transition of the relative roles of distinct neural systems during the time-course of regularity learning, i.e. learning is accompanied by a shift from a hippocampal to a prefrontal-striatal brain system.

Contents

0	Overview	1
I	Introduction	4
1	Introduction	5
II	Theory	10
2	Fast Learning Unique Episodes: The Hippocampus and Relational Binding	11
2.1	Introduction	12
2.2	Relational Binding and the Hippocampus: The Eichenbaum Model and the O'Reilly Model	14
2.2.1	The Eichenbaum Model	14
2.2.2	The O'Reilly Model	15
2.3	Studies from Different Areas of Behavioral and Cognitive Neuroscience . . .	16
2.3.1	Single Cell Recordings and Lesion Studies in Animals	17
2.3.2	Amnesia Research	20
2.3.3	Electrophysiological Studies in Humans	23
2.3.4	Imaging Studies	26
2.4	Representation of Regularities across Multiple Episodes	31
2.4.1	Assumptions of the Eichenbaum Model	31
2.4.2	Assumptions of the O'Reilly Model	32
2.4.3	Concluding Remarks	33
3	Gradual Learning Regularities across Multiple Episodes: Basic Principles and Neural Mechanisms	35
3.1	Introduction	36
3.2	Reinforcement Learning	38
3.2.1	Basic Principles	38
3.2.2	Neural Mechanisms	39

3.2.3	Summary	46
3.3	Category Learning	46
3.3.1	Basic Principles	46
3.3.2	Neural Mechanisms	47
3.3.3	Summary	51
3.4	Artificial Grammar Learning	51
3.4.1	Basic Principles	51
3.4.2	Rule-based vs Similarity-based Learning	54
3.4.3	Neural Mechanisms	55
3.4.4	Summary	57
3.5	Sequence Learning	57
3.5.1	Basic Principles	57
3.5.2	Neural Mechanisms	62
3.5.3	Summary	65
4	Neuroanatomical Basis: The Medial Temporal Lobe and the Prefrontal Cortex	66
4.1	Medial Temporal Lobe	67
4.1.1	Main Structures	67
4.1.2	Main Pathways	68
4.2	Prefrontal Cortex	69
4.2.1	Main Structures	69
4.2.2	Main Pathways	70
4.3	Medial Temporal-prefrontal Interconnections	72
III	Methods	74
5	Functional Neuroimaging	75
5.1	Principles of Magnetic Resonance Imaging	76
5.1.1	Nuclear Spin Systems in the Magnet	76
5.1.2	Excitation of the Spin System	77
5.1.3	Magnetic Resonance Imaging	80
5.1.4	Imaging Parameters and Sequences	81
5.2	The BOLD Signal	83
5.2.1	Physical Basis	83
5.2.2	Physiological Basis	84
5.2.3	Main Characteristics	85
5.3	Analysis of fMRI Time Series	88
5.3.1	Spatial Preprocessing	88
5.3.2	Modeling Hemodynamic Responses	90
5.3.3	Statistical Inference	93

5.4	Designing	94
5.4.1	Event-related Functional Imaging	95
5.4.2	Design Efficiency	97
IV	Experiments	99
6	Rationale of the Present Experiments	100
6.1	The Aim of the Present Studies	100
6.2	The Paradigm	102
6.3	Hypotheses and Predictions	105
6.4	Methods	106
7	Experiment 1A: Prefrontal-Hippocampal Dynamics Involved in Learning Regularities across Episodes	108
7.1	Introduction	109
7.2	Materials and Methods	112
7.3	Results	119
7.4	Discussion	125
8	Experiment 1B: Domain-specific Hemispheric Specialization of the Hippocampus during Learning Regularities across Episodes	135
8.1	Introduction	136
8.2	Materials and Methods	138
8.3	Results	139
8.4	Discussion	141
9	Experiment 2: Cognitive Mechanisms Subserving Regularity Learning	146
9.1	Introduction	148
9.2	General Methods	150
9.3	Experiment 2A: Distinctiveness of Objects I: Object Features	152
9.3.1	Introduction	152
9.3.2	Experimental Procedures	153
9.3.3	Results	153
9.3.4	Discussion	154
9.4	Experiment 2B: Distinctiveness of Objects II: Object Types	155
9.4.1	Introduction	155
9.4.2	Experimental Procedures	155
9.4.3	Results	156
9.4.4	Discussion	157
9.5	Experiment 2C: Memory Load and Learning Phase I	157
9.5.1	Introduction	157
9.5.2	Experimental Procedures	158

9.5.3	Results	159
9.5.4	Discussion	159
9.6	Experiment 2D: Memory Load and Learning Phase II	160
9.6.1	Introduction	160
9.6.2	Experimental Procedures	161
9.6.3	Results	161
9.6.4	Discussion	162
9.7	Experiment 2E: Distinctiveness of Objects Revisited	162
9.7.1	Introduction	162
9.7.2	Experimental Procedures	163
9.7.3	Results	163
9.7.4	Discussion	164
9.8	Experiment 2F: Simultaneous Object-Position Bindings	164
9.8.1	Introduction	164
9.8.2	Experimental Procedures	165
9.8.3	Results	165
9.8.4	Discussion	166
9.9	General Discussion	166
10	Experiment 3: Differential Hippocampal and Prefrontal-striatal Contributions to Instance-based and Rule-based Learning	169
10.1	Introduction	171
10.2	Materials and Methods	173
10.3	Results	178
10.4	Discussion	187
V	General Discussion	192
11	General Discussion	193
11.1	Summary	194
11.2	Two Separate Learning and Memory Systems: Fast Trial-unique vs Gradual Regularity Learning	198
11.3	A Definition of Regularity Learning	199
11.3.1	Regularity Extraction across Multiple Episodes	200
11.3.2	The Gradual Nature of the Extraction Process	201
11.3.3	Maintenance of Extracted Regularities	203
11.3.4	Abstract Knowledge Base	203
11.4	A Tentative Model of Regularity Learning	205
11.4.1	Hippocampus and Relational Binding	206
11.4.2	Striatum and Predictability Coding	209
11.4.3	Prefrontal Cortex and the Representation of Rule-like Information	211

11.4.4 Possible Functional Interactions between the Hippocampus, the Striatum, and the Prefrontal Cortex	212
11.5 Conclusions and Open Issues	217
References	219
List of Figures	253
List of Tables	255
Symbols and Abbreviations	256
Index	259
Zusammenfassung	264
Danksagung	278

Chapter 0

Overview

The structure of this thesis is as follows: In Part I (Introduction; Chapter 1), the aim of the present thesis will be outlined. In particular, the distinction of two separate learning and memory systems, i.e. fast learning unique episodes vs gradual learning regularities across multiple episodes, will be introduced. In Part II (Theory; Chapter 2-4), the principles of the two learning and memory systems are highlighted from the perspective of experimental psychology, animal studies, cognitive neuroscience, and computational neuroscience. Chapter 2 is concerned with the first system, i.e. learning unique episodes. Here, the crucial role of the hippocampus for this kind of learning will be described. Evidence from single cell recording and lesion studies in animals as well as from amnesia research and from electrophysiological and imaging studies in humans will be resumed. Furthermore, two influential theories about hippocampal function will be described in more detail, namely the Eichenbaum model (Eichenbaum, 2000) and the O'Reilly model (O'Reilly & Norman, 2002). Chapter 3 deals with the second learning and memory system, i.e. regularity learning. Results from studies using different learning tasks will be described. The basic principles and the neural basis of reinforcement learning, category learning, sequence learning, and artificial grammar learning will be addressed in more detail. The theoretical part ends with a brief introduction to the neuroanatomy of the medial temporal lobe, the prefrontal cortex and the main anatomical interconnections between both structures (Chapter 4). The subject of Part III (Methods; Chapter 5) is the basis

of functional magnetic resonance imaging (fMRI). Here, an overview of the physical principles of magnetic resonance imaging, the BOLD signal, the analysis of fMRI time series, and the designing of fMRI experiments will be given.

In Part IV (Experiments), the three experiments of this thesis will be described. First, Chapter 6 is concerned with the rationale of the present experiments. More precisely, the aim of the present thesis and the core paradigm will be described. Furthermore, the main hypotheses and predictions will be outlined. Experiment 1A (Chapter 7) addresses the question which brain structures support regularity learning in the spatial domain. In the critical condition of Experiment 1A, spatial regularities across episodes were introduced. To investigate, whether regularity learning is based on domain-specific or domain-general neural mechanisms, in Experiment 1B (Chapter 8) the same experimental procedures were used as in Experiment 1A, with the critical exception that object regularities, not spatial regularities were introduced. The topic of Chapter 9 (Experiment 2) is the cognitive basis of regularity learning. In a series of six behavioral studies, learning mechanisms were investigated in more detail. Based on the behavioral results of Experiment 2, the final fMRI study (Experiment 3; Chapter 10) aimed at specifying whether two separate mechanisms involved in regularity learning are based on distinct brain structures. Using a modified paradigm, the neural correlates of learning object regularities (Experiment 3A) and spatial regularities (Experiment 3B) were investigated. This thesis ends with a General Discussion. Here, the experimental findings will be discussed and a tentative model of regularity learning will be outlined (Part V, Chapter 11).

This thesis contains the following articles/manuscripts:¹

Doeller C. F., Opitz B., Krick, C. M., Mecklinger, A., & Reith, W. (in press).
Prefrontal-hippocampal dynamics involved in learning regularities across episodes.
Cerebral Cortex.²

Doeller C. F., Opitz B., Krick, C. M., Mecklinger, A., & Reith, W. (in preparation).
Domain-specific hemispheric specialization of the hippocampus during learning
regularities across episodes.³

Doeller, C. F. & Opitz, B. (2004). Cognitive and neural mechanisms subserving
regularity learning. In A. Mecklinger, H. Zimmer, & U. Lindenberger (Eds.),
Bound in memory: insights from behavioral and neuropsychological studies
(pp. 129-157). Aachen: Shaker Verlag.⁴

Doeller C. F., Opitz B., Krick, C. M., Mecklinger, A., & Reith, W. (in preparation).
Differential hippocampal and prefrontal-striatal contributions to instance-based
and rule-based learning.⁵

¹Please note that this thesis was submitted in October 2004.

²Experiment 1A (Chapter 7).

³Experiment 1B (Chapter 8).

⁴Experiment 2 (Chapter 9). Please note that the six behavioral experiments described in this article are referred to as Experiment 2A-2F in the following.

⁵Experiment 3 (Chapter 10). Please note that the two fMRI experiments described in this manuscript are referred to as Experiment 3A and 3B in the following.

Part I

Introduction

Chapter 1

Introduction

Remembering personally experienced events is a hallmark of cognition. To exemplify this, imagine the situation when you met a particular friend in a bar last Wednesday, watching a football game on television. A few days later you can recollect specific details of this episode, e.g. the location (a specific bar in your city), the date and time (Wednesday, 8.30-11.00 p.m.), the persons (your friend and other people), the football match (Bayern Munich vs Real Madrid), Roberto Carlos' free kick goal, your drinks (two glasses of beer), and so forth. Encoding and retrieval of such specific, unique episodes depends on the integrity of the hippocampus (Squire, 1992). But how does the brain represent repeated overlapping features of episodes, i.e. regularities across episodes? Think about similar episodes like the one described above. For instance, watching another UEFA Champions League match (VfB Stuttgart vs Manchester United), a specific German Bundesliga match, or the match Portugal vs England in the quarter final of the European Championships at different dates and times. In some situations you drunk water or wine instead of beer. In some situations you were in the bar together with your wife, in other situations you met colleagues from your lab. But you were always in the same bar watching a football match. All these situations comprise distinct episodes, however they also share some common features, i.e. regularities across multiple episodes, namely the specific bar and watching a football match. The present thesis aims at specifying the neural cor-

relates of learning such overlapping features of episodes, i.e. the gradual acquisition of knowledge of regularities across multiple episodes.

In the above scenario, two fundamentally different learning and memory processes were described. First, learning a unique episode, including its singular episodic features and second, the gradual acquisition of knowledge of regularities across multiple episodes. In experimental psychology and cognitive neuroscience, a large number of researchers have proposed different distinctions between two separable learning and memory systems, such as declarative vs non-declarative/procedural (Squire, 1992; Squire & Zola, 1996, cf. Cohen & Eichenbaum, 1993), fast vs slow (McClelland, McNaughton & O'Reilly, 1995; O'Reilly & Norman, 2002), memory vs habit (Mishkin, Malamut & Bachevalier, 1984), locale vs taxon (O'Keefe & Nadel, 1978), and System II vs System I (Sherry & Schacter, 1987, see also Packard & Knowlton, 2002; Poldrack & Packard, 2003; Rolls, 2000; White & McDonald, 2002, for other multiple memory/learning systems approaches).¹ Despite several differences between these theories (e.g. with respect to the research field [experimental psychology, neuropsychology, animal neuroscience, imaging neuroscience] and the proposed neural basis of both systems), *on a broad level* these taxonomies have in common that both learning and memory systems are supposed to be specialized for complementary functions: learning unique episodes and learning regularities across episodes (i.e. the acquisition of knowledge about generalities), respectively. Thus, different kinds of information are acquired by both systems. These models further assume that both systems differ with respect to operational characteristics, i.e. the first and second system are characterized by fast vs gradual learning processes, respectively. Moreover, it is assumed that the first system is more flexible than the second system, i.e. in contrast to the second system, knowledge acquired by the first system could be expressed by multiple response systems (Squire & Zola, 1996). Finally, some au-

¹Some of these models put the main emphasis on learning processes, whereas others mainly stress memory operations. In the following, both systems will be referred to as “first learning and memory system (first system)” and “second learning and memory system (second system)”, respectively.

thors argue that the information acquired by the first system is explicit, whereas the acquired knowledge of the second system is mainly not accessible to awareness, i.e. implicit (Squire, 1992). In the present thesis, the term ‘regularity learning’ will be used to refer to the learning process underlying the second system (see Chapter 3).

In general, the knowledge acquired by the first system is referred to as episodic memory (Tulving, 1983). The importance of the hippocampus during the formation of episodic memories has been highlighted by studies using a variety of neuroscientific methods in animals (Fortin, Agster & Eichenbaum, 2002; Huxter, Burgess & O’Keefe, 2003; Wirth et al., 2003; Zola et al., 2000) and humans (Davchi, Mitchell & Wagner, 2003; Fell et al., 2001; Fernández et al., 1999; Yonelinas et al., 2002). Two influential models of hippocampal function have been proposed in the last years (Eichenbaum, 2000; O’Reilly & Norman, 2002). In his relational memory theory based on rodent studies, Eichenbaum (2000) suggests that the hippocampus is critically involved in binding different features and events that compose episodic memories (see also Eichenbaum, Otto & Cohen, 1994; Eichenbaum, Schoenbaum, Young & Bunsey, 1996; Eichenbaum, 1997; Eichenbaum, Dudchenko, Wood, Shapiro & Tanila, 1999). For example, it has been shown that hippocampal cells in the rat bind singular conjunctions of episodic features, such as a specific odor occurring at a particular position (Wood, Dudchenko & Eichenbaum, 1999). Consistent with the Eichenbaum model, the hippocampus is involved in representing the temporal order of events (Dusek & Eichenbaum, 1997; Fortin, Agster & Eichenbaum, 2002; Wallenstein, Eichenbaum & Hasselmo, 1998).

In the last years, O’Reilly and coworkers have developed an influential computational model of learning and memory, the complementary learning systems approach (McClelland, McNaughton & O’Reilly, 1995; Norman & O’Reilly, 2003; O’Reilly & Munakata, 2000; O’Reilly & Rudy, 2000, 2001; O’Reilly & Norman, 2002). This model assumes two qualitatively different learning systems, a fast hippocampal and a slow learning system associated with the rhinal and parahippocampal cortex (the MTL cortex system in the model of O’Reilly and Norman, 2002). It is assumed that the hippocampus is involved in learning conjunctions of features and events of

episodes by means of pattern-separated representations, an essential prerequisite for the avoidance of interference between distinct episodes.

However, less is known about the role of the hippocampus during learning regularities across multiple episodes, i.e. overlapping features of episodes, the core function of the second learning and memory system. In the O'Reilly model (O'Reilly & Norman, 2002), the rhinal and parahippocampal cortex is assumed to represent regularities in the environment by assigning overlapping distributed representations to similar stimuli. Here, a sharpening mechanism selectively strengthens representations of frequent stimuli. In contrast, Eichenbaum (2000) posits that particular hippocampal neurons are specialized to represent features that are common across many experiences. By this, overlapping episodes could be linked together and semantic knowledge structures gradually evolve (see also Agster, Fortin & Eichenbaum, 2002; Eichenbaum, in press).

In the last decades, the cognitive mechanisms associated with the second learning and memory system have gained considerable interest. A variety of learning tasks have been developed, such as artificial grammar learning (Gomez & Schvaneveldt, 1994; Meulemans & Van der Linden, 1997; Perruchet & Pacteau, 1990; Reber, 1967), category learning (Ashby & Perrin, 1988; Erickson & Kruschke, 1998; Nosofsky, 1986; Posner & Keele, 1968, 1970), and sequence learning tasks (Cleeremans & McClelland, 1991; Cohen, Ivry & Keele, 1990; Nissen & Bullemer, 1987; Stadler, 1995). Taken together, despite the differences between tasks, there is converging evidence that humans acquire a (fragmentary) knowledge of regularities across input pattern in these learning tasks.

Tightly coupled with the debate about the exact nature of the learning process is the question which brain structures might subserve regularity learning. In the last years, functional imaging studies in humans and single cell recordings in monkeys have begun to unravel the network of brain areas involved in regularity learning, using mainly the aforementioned learning tasks (Aizenstein et al., 2004; Bischoff-Grethe, Goedert, Willingham & Grafton, 2004; Fletcher, Büchel, Josephs, Friston & Dolan, 1999; Freedman, Riesenhuber, Poggio & Miller, 2001; Hazeltine, Grafton & Ivry, 1997; Opitz & Friederici, 2003, 2004; Reber, Gitelman, Parrish & Mesulam,

2003; Schendan, Searl, Melrose & Stern, 2003; Strange, Henson, Friston & Dolan, 2001). Despite differences in experimental procedures, a consistent pattern of results is beginning to emerge. Overall, these studies converge to suggest that the lateral prefrontal cortex plays a pivotal role during learning task-relevant regularities in the environment.

In this thesis, the neural and cognitive processes underlying regularity learning were investigated in several fMRI and behavioral experiments. A paradigm was developed which allowed to separate (1) learning unique episodes (the core function of the first system) and (2) learning regularities across multiple episodes (the core function of the second system). By this, two main research fields in cognitive neuroscience, namely episodic memory and rule learning research, were integrated.

Part II

Theory

Chapter 2

Fast Learning Unique Episodes: The Hippocampus and Relational Binding

Contents

2.1	Introduction	12
2.2	Relational Binding and the Hippocampus: The Eichenbaum Model and the O'Reilly Model	14
2.2.1	The Eichenbaum Model	14
2.2.2	The O'Reilly Model	15
2.3	Studies from Different Areas of Behavioral and Cognitive Neuroscience	16
2.3.1	Single Cell Recordings and Lesion Studies in Animals	17
2.3.2	Amnesia Research	20
2.3.3	Electrophysiological Studies in Humans	23
2.3.4	Imaging Studies	26
2.4	Representation of Regularities across Multiple Episodes	31
2.4.1	Assumptions of the Eichenbaum Model	31
2.4.2	Assumptions of the O'Reilly Model	32
2.4.3	Concluding Remarks	33

2.1 Introduction

As outlined in Chapter 1, a dichotomous distinction between two separable learning and memory systems has been proposed by several influential models (McClelland et al., 1995; Mishkin et al., 1984; O’Keefe & Nadel, 1978; O’Reilly & Norman, 2002; Sherry & Schacter, 1987; Squire, 1992; Squire & Zola, 1996, cf. Packard & Knowlton, 2002; Poldrack & Packard, 2003; Rolls, 2000; White & McDonald, 2002): learning unique episodes (first system) vs learning regularities across multiple episodes (second system). In the present chapter, the cognitive and neural basis of the first system will be described. The cognitive and neural processes underlying the second system will be outlined in Chapter 3.

According to the aforementioned taxonomy, episodic memory formation is the core function of the first system. Episodic memory refers to the ability to remember specific personally experienced events set in a spatio-temporal context (Tulving, 1983).¹ This ability includes the formation of new, stable memory traces during encoding and processes involved in recollecting these memory traces. During encoding, representations of incoming stimuli are build. Subsequently, these representations are bound together into enduring memory traces (Paller & Wagner, 2002). During retrieval, partial elements of the episode lead to a reinstantiation of the original memory trace, i.e. the interaction between a retrieval cue and a memory representation entails the reactivation of the past episode (Tulving, 1983).

The formation of episodic memory critically depends on the integrity of the hippocampus. The hippocampal formation is comprised of the hippocampus proper (hippocampal cell fields and the dentate gyrus), the subicular complex and the entorhinal cortex (Amaral & Insausti, 1990, see Chapter 4 for a detailed neuroanatomical description of the hippocampus). Damage to the hippocampus and adjacent medial temporal lobe (MTL) structures produces an episodic memory loss (Scoville & Milner, 1957). In the recent years, supporting evidence for the crucial role of the hippocampus during episodic memory formation has been provided by studies

¹It should be noted that several authors use the term context-specific memory instead of episodic memory. In the following these terms will be used as equivalent.

using a wide variety of neuroscientific methods. For instance, converging evidence comes from lesion studies (Fortin, Agster & Eichenbaum, 2002, Zola et al., 2000) and single cell recordings (Wirth et al., 2003; Wood, Dudchenko & Eichenbaum, 1999) in animals. Furthermore, the view that the hippocampus plays an essential role during episodic memory formation has been substantiated by studies with amnesics (Mishkin, Vargha-Khadem & Gadian, 1998, Yonelinas et al., 2002), intracranial EEG studies (Fell et al., 2001; Fernández et al., 1999), and functional imaging studies in humans (Brewer, Zhao, Desmond, Glover & Gabrieli, 1998; Davachi, Mitchell & Wagner, 2003; Eldrige, Knowlton, Furmanski, Bookheimer & Engel, 2000; Gabrieli, Brewer, Desmond & Glover, 1997; Otten, Henson & Rugg, 2001). These studies further indicate that the hippocampus is involved during encoding as well as during retrieval of episodic memories.

Several prominent theories of hippocampal function have been proposed (e.g. Aggleton & Brown, 1999; Eichenbaum, 2000; McClelland, McNaughton & O'Reilly, 1995; Nadel, Samsonovich, Ryan & Moscovitch, 2000; Norman & O'Reilly, 2003; O'Keefe & Nadel, 1978; Shastri, 2002; Squire, 1992). In the scope of the present thesis, two models are most important, namely the Eichenbaum model (Eichenbaum, 2000) and the O'Reilly model (O'Reilly & Norman, 2002), since both models make clear predictions about hippocampal involvement during learning unique episodes and learning regularities across episodes, respectively. The O'Reilly model is a computational neuroscience model, whereas the Eichenbaum model is based on findings from rodent lesion and single cell recording studies. Both models address hippocampal memory operations and storage principles of the hippocampus. In brief, both models assume that the hippocampus is involved in binding different features that compose episodic memories. This aspect of the models will be described in more detail in the following Section 2.2. Evidence for this view will be provided in Section 2.3. Here, results of neurophysiological, lesion, human electrophysiological, and neuroimaging studies investigating the role of the hippocampus during episodic memory formation will be reviewed. With respect to the scope of the present thesis, Section 2.4 will outline the assumptions of both models of how the brain represents regularities across episodes.

2.2 Relational Binding and the Hippocampus: The Eichenbaum Model and the O'Reilly Model

2.2.1 The Eichenbaum Model

In his relational memory theory, Eichenbaum (2000) suggests that the hippocampus is critically involved in binding (cf. Zimmer, Mecklinger & Lindenberger, in press) of different features and events that compose episodic memories (Eichenbaum, Otto & Cohen, 1994; Eichenbaum, Schoenbaum, Young & Bunsey, 1996; Eichenbaum, 1997, in press, see also Cohen et al., 1999). In contrast, it is assumed that the parahippocampal region is involved in representing single stimuli or single features of complex stimuli. Moreover, the parahippocampal region is associated with the maintenance of single stimuli in task situations without interference and represents an index of stimulus familiarity. The hippocampus is supposed to bind together those parahippocampal representations to form an enduring representation of an episode. For example, it has been shown that hippocampal cells in the rat bind singular conjunctions of episodic features, such as a specific odor occurring at a particular position (Wood et al., 1999). Moreover, Eichenbaum proposes that the hippocampal binding processes are relational by nature. Mathematically, a relation is defined for instance by the characteristics of symmetry and transitivity. Eichenbaum and coworkers (Bunsey & Eichenbaum, 1996; Dusek & Eichenbaum, 1997) could demonstrate that rats with hippocampal lesions show no evidence for symmetry and transitivity in inferential tasks. In contrast, rats with intact hippocampi could infer symmetry and transitivity in these tasks. For instance, in the Dusek and Eichenbaum (1997) study, both groups of rats learned to prefer on odor over another (e.g. $A > B$, $B > C$) in a learning phase. In a subsequent test phase, these pairs of odors and new pairs (e.g. $A > C$) were presented. Rats with intact hippocampi showed transitive inference, i.e. these rats selected odor A over C, whereas rats with hippocampal lesions showed no capacity for transitive inference. Further supporting evidence for relational binding in the hippocampus is provided by the aforementioned study of Fortin et al. (2002). This study demonstrated that the hippocampus is involved in

representing the orderliness of events (see also Eichenbaum, in press; Wallenstein, Eichenbaum & Hasselmo, 1998). In addition, Eichenbaum argues that the hippocampus links representations of episodes to representations of already stored episodes, thus providing a network of memories (Agster et al., 2002; Eichenbaum, 2000, in press, cf. Section 2.4). To summarize, the Eichenbaum model assumes that the hippocampus binds features and events of complex episodes, reflecting the temporal and spatial order of these events and is involved in linking these representations to stored representations of other episodes and by this mediates associations between distinct episodes (Eichenbaum, 2000, in press).

2.2.2 The O'Reilly Model

In the last years, O'Reilly and coworkers have developed an influential computational model of learning and memory, the complementary learning systems approach (McClelland et al., 1995; Norman & O'Reilly, 2003; O'Reilly & Munakata, 2000; O'Reilly & Rudy, 2000, 2001; O'Reilly & Norman, 2002). This model proposes two qualitatively different learning systems, a fast hippocampal and a slow learning system associated with the adjacent MTL cortex, comprising the perirhinal, entorhinal, and parahippocampal cortex. It is assumed that the hippocampus is involved in learning conjunctions of features and events defining specific episodes. In contrast, the MTL cortex represents statistical summaries of multiple input pattern. O'Reilly and coworkers implemented both systems in different networks (Figure 2.1). The MTL cortex network has an input and a hidden layer (see Figure 2.1a). The architecture of the hippocampal network takes into account the neuroanatomy of the medial temporal lobe (see Chapter 4). Accordingly, the network has different layers representing the entorhinal cortex, the dentate gyrus, and hippocampal fields CA1 and CA3 (see Figure 2.1c). In accordance with the neuroanatomical projections, different connections between layers exist. The input layer projects to the entorhinal layer which projects to the dentate gyrus, CA1, and CA3 layer. The CA3 layer projects back to the entorhinal output layer via the CA1 layer.

O'Reilly and coworkers implemented two different learning principles in the MTL

cortex and the hippocampal model, respectively. The hippocampal system represents episodes by means of pattern-separated, sparse representations which have the advantage of minimizing interference between representations of different episodes. It is assumed that these non-overlapping pattern-separated representations are built during encoding in hippocampal region CA3. During retrieval, partial cues lead to a reinstatement of the original CA3 representation by means of pattern completion (see O'Reilly & McClelland, 1994, for a detailed description of pattern separation and pattern completion). A pattern-separated representation of specific episodes is a necessary prerequisite for successful episodic memory encoding and storage, since representations of unique episodes have to be kept separate from other episodes to avoid interference between distinct episodes (O'Reilly & Rudy, 2000). In a similar way, this pattern separation mechanism has to be fast, since episodes have to be encoded directly when they occur. Moreover, the O'Reilly model assumes that the hippocampal representation is sparse and conjunctive (O'Reilly & Rudy, 2001). The sparseness entails fast encoding as well as minimizing interference between representations of distinct episodes. A conjunctive representation is the result of binding of different features defining a unique episode (O'Reilly & Rudy, 2000). This conjunctive unitary representation is later reinstated by the hippocampus upon presentation of a retrieval cue (see Nakazawa et al., 2002, for a possible molecular mechanism). The MTL cortex model, in contrast, represents the general statistical structure of multiple episodes (see Section 2.4).²

2.3 Studies from Different Areas of Behavioral and Cognitive Neuroscience

A large number of studies using different neuroscientific methods in animals as well as in humans suggests that the hippocampus plays a crucial role during the forma-

²Based on these different learning operations it is proposed that the hippocampus supports recollection-based recognition, whereas the MTL cortex is involved in familiarity-based recognition (see Section 2.3.2 for a description of recollection-based and familiarity-based recognition).

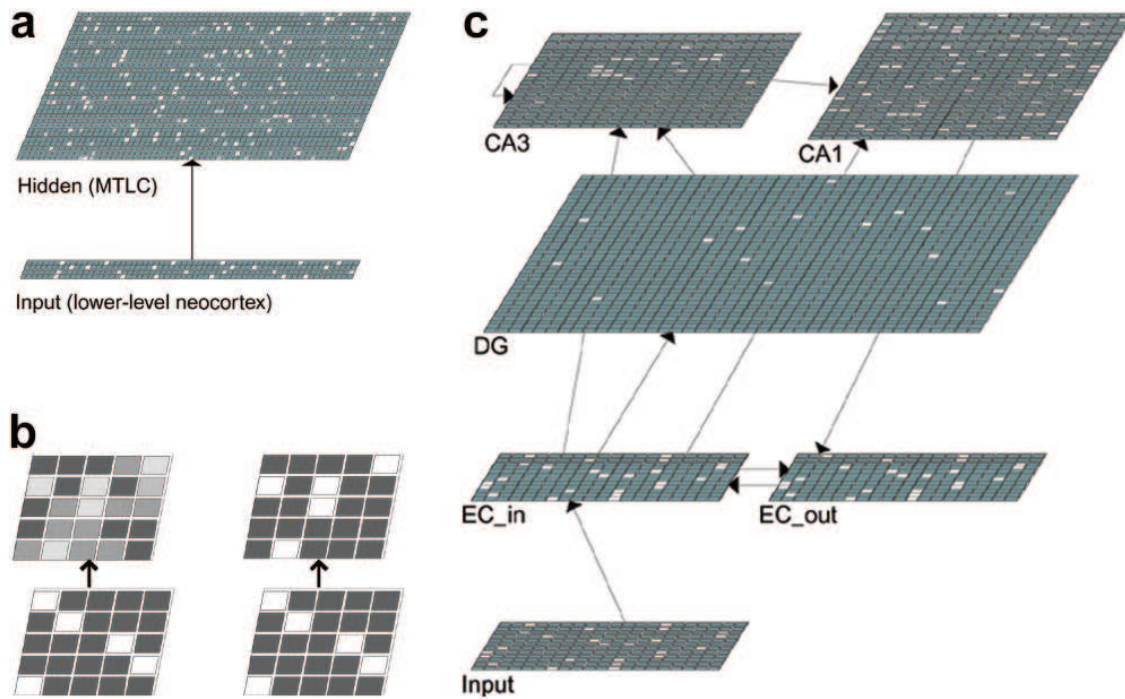


Figure 2.1. Network Architecture of the MTL Cortex and the Hippocampal Model (Complementary Learning Systems Approach; O'Reilly & Norman, 2002). (a) The MTL cortex network with the input and hidden MTL cortex layer. (b) The dynamics of the sharpening mechanism implemented in the MTL cortex model. In an early phase, the representation of an input is unspecific (left). After multiple presentations of overlapping input pattern, the sharpening mechanism entails more differentiated representation (right). (c) The hippocampal network. The different layers constitute the input, the entorhinal (EC), the dentate gyrus (DG), the CA1, and CA3 layer. Arrows denote projections between different layers. [Adapted from Norman and O'Reilly (2003). Copyright (2003) American Psychological Association, Inc. (Psychological Review).]

tion of episodic memory. In this section an overview about findings from single cell recordings and lesion studies in animals (2.3.1), neuropsychological investigations in amnesic patients (2.3.2), and electrophysiological (2.3.3) and functional imaging experiments (2.3.4) in humans will be given.

2.3.1 Single Cell Recordings and Lesion Studies in Animals

In the following, some core memory studies in animals using single cell recording techniques and different lesion approaches will be reviewed. To give a broad overview, I will focus on earlier studies using the delayed nonmatch to sample task (DNMS), on studies investigating spatial memory in rodents, as well as on recent studies using more complex memory paradigms. For a more detailed review of the animal

literature, please see Aggleton and Brown, 1999, Eichenbaum, Dudchenko, Wood, Shapiro, and Tanila, 1999, Eichenbaum, Alvarez, and Ramus, 2000, O'Keefe and Nadel, 1978, O'Keefe, 1999, and Suzuki, 1999.

Early studies in monkeys revealed ambiguous results with respect to hippocampal involvement during memory formation. In these studies, effects of hippocampal lesions on memory performance have been widely investigated using the DNMS task. In each trial of this task, the animal is presented with a sample stimulus. After a delay, the sample stimulus is presented along with a new stimulus and the choice of this new stimulus is rewarded (nonmatch). Several studies could show that monkeys with lesions including the hippocampus and the surrounding MTL cortex are impaired in the DNMS task, especially in the trial-unique version of this task, in which new, unfamiliar items are used in each trial (e.g. Mishkin, 1978; Zola-Morgan, Squire & Mishkin, 1982; Zola-Morgan & Squire, 1985, see also the reviews of Aggleton and Brown, 1999, and Eichenbaum et al., 2000). A review of monkey lesion data provides evidence for the notion that memory impairments in the DNMS task depend on the extent of the medial temporal lobe lesion (Zola-Morgan, Squire & Ramus, 1994). Lesions restricted to the hippocampus proper and the subicular complex were associated with mild memory impairments, whereas severe memory impairments could be observed in monkeys with lesions also including the surrounding entorhinal, perirhinal, and parahippocampal cortex. In line with this observation, selective damage to the hippocampus using modern ablation techniques revealed no impairments or only impairments after long delays in the DNMS task (Alvarez, Zola-Morgan & Squire, 1995; Murray & Mishkin, 1998). Based on these results, it has been proposed that object recognition, as measured in the DNMS task does not depend on an intact hippocampus, but rather on an intact MTL cortex, most of all on the perirhinal cortex (Aggleton & Brown, 1999; Brown & Aggleton, 2001), specialized for familiarity/recency discrimination (see Section 2.3.2). Brown and Aggleton (2001) argue further that the hippocampus is crucial for representing associations between stimuli (see Wan, Aggleton & Brown, 1999, for a direct test of the model).

The role of the rodent hippocampus in memory has been examined using tests of allocentric spatial memory, e.g. the Morris water maze task and the radial maze

task (see Eichenbaum et al., 1999, for a review). In these tasks, rats have to learn routes in an experimental environment. Rats with selective hippocampal lesions showed impaired memory performance in these tasks (e.g. Morris, Garrud, Rawlins & O'Keefe, 1982). In addition, single cell recording studies in rats revealed specialized hippocampal neurons, the so-called place cells (O'Keefe & Dostrovsky, 1971, see Muller, 1996, for a review) which provide a representation of the current location of an animal by encoding topological information of the environment (Muller, Kubie & Ranck, 1987; O'Keefe & Dostrovsky, 1971; O'Keefe & Burgess, 1996). Based on such findings, it has been proposed that the hippocampus represents a cognitive map of the environment (O'Keefe & Nadel, 1978; O'Keefe, 1999).

In addition, there is a growing number of single cell recording and lesion studies in rodents (Bunsey & Eichenbaum, 1996; Ferbinteanu & Shapiro, 2003; Hampson et al., 1999; Wood et al., 1999, 2000) and monkeys (Brasted, Bussey, Murray & Wise, 2003, Wirth et al., 2003; see Suzuki, 1999, for a review of the monkey literature and possible explanations for the ambiguous results of single unit recordings in monkeys) showing hippocampal involvement in memory for non-spatial stimuli. For instance, in a study by Wood et al. (1999) rats performed a DNMS task while neuronal activity was recorded from cells in hippocampal fields CA1 and CA3. In this experiment, different odors could appear at different locations and the rat was rewarded when selecting the odor that was different from that in the previous trial (nonmatch). Wood et al. (1999) observed hippocampal neurons that encode single features of episodes, e.g. a specific position, and more interestingly cells representing specific conjunctions of episodic features, such as a specific odor occurring at a particular position as a match trial. These data indicate that beyond the representation of single episodic features, the hippocampus is critically involved in binding of different spatial and non-spatial features that compose episodic memories (Eichenbaum, 2000).

Supporting evidence for this view is provided by a recent lesion study (Fortin et al., 2002). In a first trial phase different odors were presented in temporal sequence with a 2.5 min delay between presentations. In a second phase after a 3-min delay either a sequential order or a recognition test was conducted (see Figure 2.2a). In the se-

quential order task, rats had to select one of two presented odors which occurred earlier in the previous sequence. In the recognition task, rats were required to indicate the odor not presented in the sequence. For one group of rats the hippocampus was removed using radiofrequency lesions. In the recognition task, no performance differences could be observed between the lesion group and a healthy control group (Figure 2.2c), i.e. both groups of rats correctly selected roughly 80-90% of the new odors. In contrast, performance significantly decreased for lesioned rats in the sequential order task as compared to the controls (Figure 2.2b). These results show that the hippocampus represents the sequential order of events, an essential prerequisite for building episodic memories (cf. Dusek & Eichenbaum, 1997).

Taken as a whole, the lesion and single unit recording studies reviewed here are consistent with the idea that the hippocampus plays an important role in representing spatial and non-spatial features (especially object and temporal information) of episodes (Eichenbaum et al., 1999, see Huxter, Burgess, and O'Keefe, 2003, for a possible neurophysiological explanation) and by this provides a basis for building episodic memories (Eichenbaum, 2000).

2.3.2 Amnesia Research

Bilateral removal of the medial temporal lobe, including the hippocampus results in severe anterograde amnesia (Scoville & Milner, 1957). Patients suffering from anterograde amnesia are not able to learn new information after the onset of the disease. Retrograde amnesia refers to impairments of memories acquired before the onset of the disease (Squire, Clark & Knowlton, 2001). Episodic memory impairments in amnesics following medial temporal lobe lesions have been described extensively in the last decades, while other cognitive functions like working memory and language skills are preserved (see Aggleton & Brown, 1999; Gabrieli, 1998; Spiers, Maguire & Burgess, 2001; Squire, 1992; Squire & Zola, 1997, for an overview). Due to space limitations, in the following only a few particularly influential recent studies with amnesic patients will be described in more detail. These studies were selected, since they address three main topics in memory research, namely the distinction between

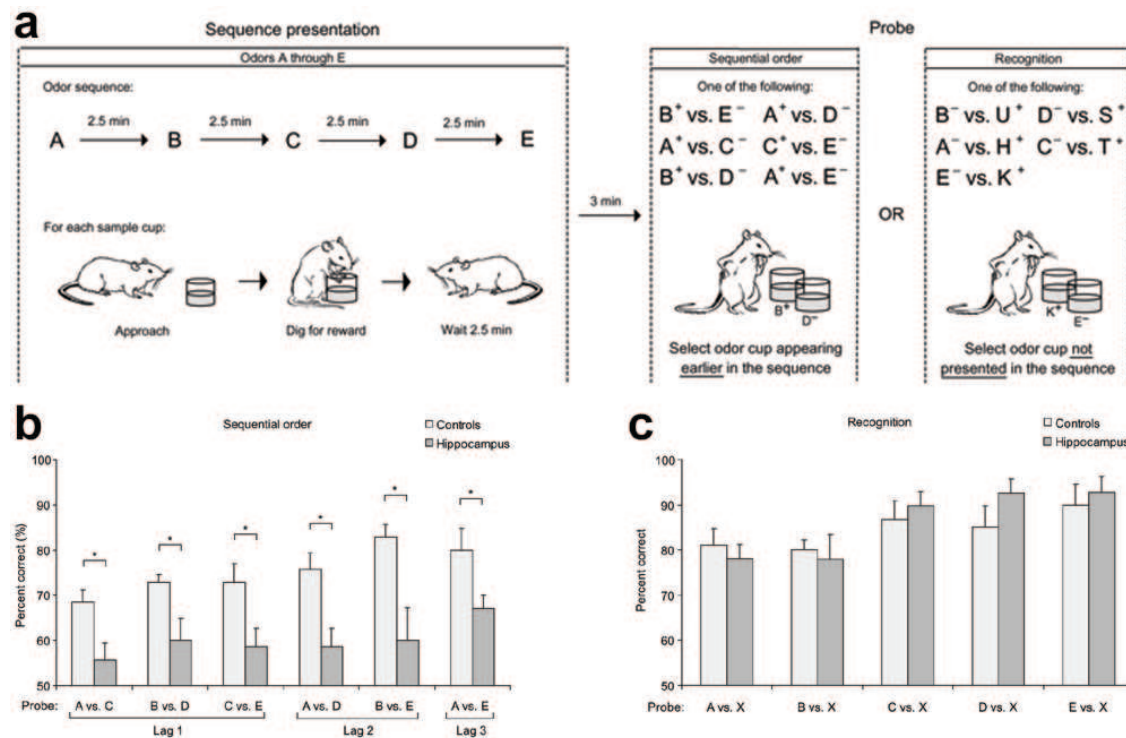


Figure 2.2. Design and Results of the Fortin et al. (2002) Study. (a) Different odors A-E are presented in sequential order (left). After a 3-min delay either a sequential order or a recognition task is presented (right). (b) Results for the sequential order task. For all probes and lags (i.e. the number of intervening odors) hippocampal lesioned animals (grey) showed reduced performance as compared to controls (white). (c) Results for the recognition task revealed no significant differences between both groups. [Adapted from Fortin et al. (2002). Copyright (2002) Macmillan Magazines Ltd (Nature Neuroscience).]

episodic vs semantic memory, recollection vs familiarity, and single item vs associative memory, respectively.

Vargha-Khadem et al. (1997) described three patients with anterograde amnesia. All three patients had selective bilateral hippocampal pathology since early childhood, i.e. approximately 50% hippocampal volume reduction. Neuropsychological tests revealed that verbal, pictorial, and episodic memory performance was significantly reduced as compared to controls. Most interestingly, all three patients had normal intelligence and normal semantic memory. This study gives evidence for the notion that only episodic memory and not semantic memory crucially depends on intact hippocampi (but see Manns, Hopkins & Squire, 2003b). It is still controversial, whether the hippocampus also subserves semantic memory (Squire & Zola, 1998; Tulving & Markowitsch, 1998). Squire and Zola (1998, see also Squire, 1992)

suggest that the hippocampus and the surrounding MTL cortex is crucial for episodic and semantic memory (declarative memory) but not for non-declarative memory.³ In contrast, Tulving and Markowitsch (1998) argue that the acquisition of episodic memory and semantic memory can occur independently. Furthermore, Tulving proposes that episodic memory depends on the integrity of the hippocampus, whereas the acquisition of semantic memory is based on the rhinal and parahippocampal cortex.

Furthermore, it is a topic of current interest whether the hippocampus mediates recollection-based or familiarity-based recognition memory. In recent years, several memory researchers have proposed that recognition memory is based on two independent, qualitatively different processes, recollection-based and familiarity-based recognition (so-called dual-process models; e.g. Aggleton & Brown, 1999; Brown & Aggleton, 2001; Mecklinger, 2000; O'Reilly & Norman, 2002; Rugg & Yonelinas, 2003; Yonelinas, 2001). These models assume that the hippocampus is crucial for recollection-based recognition but not for familiarity-based recognition.⁴ The influence of hippocampal damage on recollection and familiarity measures of recognition memory was investigated in a recent study with amnesic patients (Yonelinas et al., 2002). In the study of Yonelinas et al. (2002), memory performance was measured in a hypoxic-ischemic patient group. The patients had severe atrophy limited to

³Following Squire's taxonomy of memory, declarative memory refers to the conscious recollection of episodes (episodic memory) and general facts (semantic memory), whereas non-declarative memory denotes different types of implicit memory, like procedural memory, priming, classical conditioning, and non-associative memory (cf. Squire, 1992; Squire & Zola, 1996).

⁴These models suggest that recollection is based on the retrieval of specific details of the study episode, whereas familiarity describes an unspecific memory signal, indicating the prior occurrence of an item without recollecting specific details of the episode, e.g. the context. Familiarity is supposed to be faster than recognition, but also to decrease faster. Furthermore, recollection but not familiarity seems to be independent of changes to the stimulus domain between study and test phase (see Yonelinas, 2002, for a detailed description of familiarity and recollection). Furthermore, it is assumed that recollection-based recognition memory depends on the hippocampus, whereas familiarity-based recognition is associated with the surrounding MTL cortex, especially the perirhinal cortex (Brown & Aggleton, 2001)

the hippocampi. Patients showed greater deficits in free recall than in recognition, indicating that recollection but not familiarity was impaired.⁵ Furthermore, using structural covariance modeling the authors could show that the degree of hypoxic severity was correlated with recognition performance, but not with measures of familiarity. In an additional experiment using the remember know procedure (Tulving, 1985, see Section 2.3.4), the hypoxic-ischemic patient group showed solely recollection deficits. In contrast, a group with extensive medial temporal lobe lesions, including the hippocampus and the parahippocampal gyrus, showed pronounced recollection and familiarity deficits. Taken together, the results of this study indicate that recollection-based, but not familiarity-based recognition memory requires an intact hippocampus (however, see Manns et al., 2003a, for a different view).

Another topic of current interest is whether the hippocampus subserves memory for single items or memory for associations of items. Supporting evidence for the notion that the hippocampus is important for associative memory, i.e. binding different features of an episode was provided by a recent neuropsychological investigation of patient YR, who suffered a bilateral lesion of the hippocampus (Mayes et al., 2004). Results of several recognition tests revealed preserved single item memory and preserved recognition memory for intra-item associations (e.g. composite words), but impaired recognition memory for inter-item associations in patient YR. For instance, the patient's memory performance was significantly reduced for associations between objects and locations, the temporal order of items, and picture-sound associations. Taken together, these patient studies with amnesic patients suggest that episodic memory, recollection-based memory, and associative memory crucially depends on the integrity of the hippocampus.

2.3.3 Electrophysiological Studies in Humans

In recent years, the neural correlates of episodic memory have also been investigated using intracranial recordings in humans (Cameron et al., 2001; Fell et al.,

⁵It is assumed that free recall is based on recollection, whereas recognition is based on both recollection and familiarity.

2001; Fernández et al., 1999, 2002; Fried et al., 1997; Grunwald et al., 2003).⁶ For instance, Fernández and coworkers (Fell et al., 2001; Fernández et al., 1999, 2002) recorded event-related potentials (ERP) directly from the human medial temporal lobe using intracerebral electrodes in patients suffering from pharmacoresistant temporal lobe epilepsy. In these studies, patients performed a free recall task for words. ERPs were measured in the study phase allowing signals to be classified based on subjects' subsequent memory performance (subsequent memory paradigm). In comparison to later forgotten words, correctly remembered words elicited a negativity between 400 and 500 ms after stimulus onset in the anterior parahippocampal gyrus (rhinal cortex; AMTL-N400) in the Fernández et al. (1999) study. In a time window between 500 and 2000 ms, ERPs measured in the hippocampus were more positive for subsequently remembered than forgotten items. These data suggest that both MTL structures, the rhinal cortex and the hippocampus, are important for the formation of episodic memories. Furthermore, the authors argue that memory formation seems to be subdivided into two sequential processes, an early rhinal and a late hippocampal process. To investigate these subprocesses in more detail, Fell et al. (2001) analyzed phase synchronization in the γ -frequency range between rhinal and hippocampal electrodes in a similar study. Subsequently remembered words elicited an early increase (100-300 and 500-600 ms after stimulus onset) and a late decrease (1000-1100 ms) in phase synchronization between both MTL structures as opposed to later forgotten items (cf. Figure 2.3a,b). These data indicate that successful memory encoding seems to be based on a coupling between the rhinal cortex and the hippocampus. In another depth EEG study, Fernández et al. (2002) presented either high frequency or low frequency words in the study phase. Recall performance was greater for high frequency than for low frequency words. The hippocampal subsequent memory effect, i.e. more positive ERPs for subsequently remembered than

⁶It is noteworthy that the Fried group (e.g. Cameron et al., 2001; Fried et al., 1997) directly records single cell activity (action potentials) by means of microelectrodes (see Fried et al., 1997, for recording details), whereas the Fernández group (e.g. Fell et al., 2001; Fernández et al., 1999, 2002) measures local field potentials by means of multicontact depth electrodes (see Fernández et al., 1999, for recording details).

forgotten items, was similar for both low and high frequency words. However, the rhinal AMTL-N400 was solely observed for high frequency words but not for low frequency words. The authors argue that the rhinal cortex is involved in semantic processing of to be remembered items, by this strengthening the memory representation indirectly as reflected in more effective memory formation for high frequency as compared to low frequency words. In contrast, Fernández and colleagues propose that the hippocampus seems to be characterized by exclusively mnemonic processes. Taken together, the results of the Fernández studies indicate that episodic memory formation depends on both MTL structures, the rhinal cortex and the hippocampus with temporal and functional specializations of either structure.

Single cell activity in the medial temporal lobe of epilepsy patients was recorded by Fried and colleagues (Cameron, Yashar, Wilson & Fried, 2001; Fried, MacDonald & Wilson, 1997). In the study by Fried et al. (1997), patients had to encode and recognize pictures of faces and objects. MTL neurons responded differentially to specific items as well as to specific feature conjunctions, i.e. a specific face with a specific facial expression during encoding and retrieval. These data indicate that the hippocampus and associated MTL cortices are involved in encoding conjunctions of stimulus features. In the Cameron et al. (2001) study, the authors investigated single cell activity during encoding and retrieval of word paired associates. In accordance with the results of Fernández et al. (1999), hippocampal cell activity predicted later memory performance.

A single case scalp ERP study with one of the patients with early hippocampal damage in the aforementioned study (Vargha-Khadem et al., 1997) was conducted by Düzel, Vargha-Khadem, Heinze, and Mishkin (2001). The authors investigated ERP old-new effects associated with recognition memory, i.e. more positive ERP waveforms for correctly classified old items (hits) as compared to correctly rejected new items during a recognition phase. An early (frontal) old-new effect is associated with familiarity-based recognition, whereas a late (parietal) old-new effect reflects recollection-based recognition (Mecklinger, 2000). In this recognition memory study, the patient showed a normal early ERP old-new effect (300 - 500 ms after stimulus onset), but the late (500 - 700 ms) old-new effect was absent. These

data parallel the behavioral data of Vargha-Khadem et al. (1997), suggesting that episodic memory and recollection-based recognition depends on the integrity of the hippocampus, whereas an intact parahippocampal region (perirhinal, entorhinal, and parahippocampal cortex) is crucial for the formation of semantic memory as well as for familiarity-based recognition.

Taken together, these single cell recording and ERP studies in humans provide further, converging evidence for the notion that the hippocampus is crucial for memory binding. These studies further indicate hippocampal-rhinal interactions during memory formation and underscore the importance of the hippocampus for recollection-based recognition.

2.3.4 Imaging Studies

In the last years functional neuroimaging research has begun to study the neural correlates of episodic memory (e.g. Alkire et al., 1998; Brewer et al., 1998; Gabrieli et al., 1997; Wagner et al., 1998, for a review see Brewer and Moghekar, 2002). Research concerning encoding-related neural processes revolved mainly around the issue whether neural processes are modulated by the specific material of the items to be remembered, e.g. verbal vs pictorial material (cf. Paller & Wagner, 2002). Studies using the subsequent memory paradigm could show a material-specific lateralization in the hippocampus/MTL. Subsequent memory for incidentally learned items was associated with left hippocampus/MTL activation for verbal material (Golby et al., 2001; Kirchhoff et al., 2000; Otten et al., 2001; Wagner et al., 1998, but see Fernández et al., 1998), with bilateral hippocampus/MTL activation for figural material (Golby et al., 2001; Kelley et al., 1998, see also Stern et al., 1996), and with right (and bilateral) activation for spatial material (Golby et al., 2001; Kelley et al., 1998; Kirchhoff et al., 2000). Furthermore, recent fMRI studies reported right hippocampal involvement during spatial navigation and spatial memory (Burgess, Maguire, Spiers & O'Keefe, 2001; Grön, Wunderlich, Spitzer, Tomczak & Riepe, 2000; Hartley, Maguire, Spiers & Burgess, 2003; Iaria, Petrides, Dagher & Bohbot, 2003, Maguire et al., 1998; Maguire, Valentine, Wilding & Kapur, 2003; see Burgess,

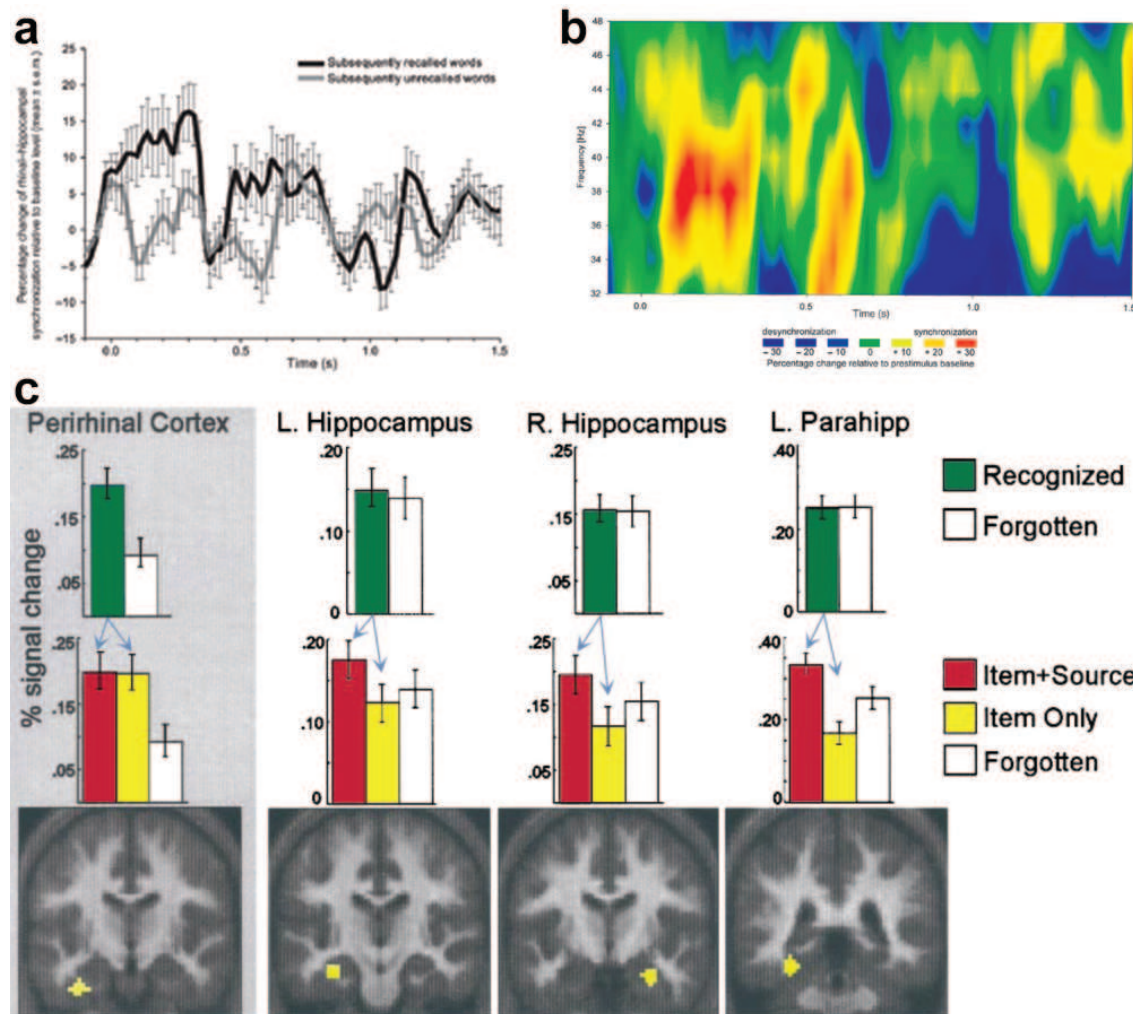


Figure 2.3. Results of the Fell et al. (2001) and Davachi et al. (2003) Study. (a) Changes of phase synchronization between rhinal cortex and hippocampal electrode sites for subsequently recalled (black) and unrecalled (gray) words in the γ -frequency range (32-48 Hz) in the Fell et al. (2001) study. These data show early coupling (100-300 and 500-600 ms after stimulus onset) and late decoupling (1000-1100 ms) between both MTL structures. (b) Graphical depiction of the difference phase synchronization values (synchronization of the recalled minus synchronization of the unrecalled words) from (a) as a function of the frequency range and time after stimulus onset. (c) Peak BOLD signal change and activated subregions of the MTL during the encoding phase in the Davachi et al. (2003) study. Signal change is plotted for subsequently recognized (green) and forgotten (white) items (upper row). Signal change for recognized items is subdivided into signal change for correctly recognized item and source information (red) and item information only (yellow) (middle row). [(a-b) adapted from Fell et al. (2001). Copyright (2001) Macmillan Magazines Ltd (Nature Neuroscience). (c) adapted from Davachi et al. (2003). Copyright (2003) National Academy of Sciences, USA (Proceedings of the National Academy of Sciences of the United States of America).]

Maguire, and O'Keefe, 2002, for a recent review). For instance, Golby et al. (2001) could show that encoding of words produced a left lateralized MTL activation, encoding of scenes and faces was associated with a bilateral MTL activation pattern, whereas the right MTL was involved during encoding abstract visuo-spatial pattern (see Kelley et al., 1998, for similar findings). These results indicate that hippocampal memory operations during encoding are sensitive to the specific domain of the to-be-remembered items. Several fMRI and PET studies have observed hippocampus/MTL involvement during retrieval from episodic memory (Cabeza et al., 2001; Gabrieli et al., 1997; Heckers et al., 2002; Ryan et al., 2001; Schacter et al., 1996; Tsukiura et al., 2002, see the review of Schacter and Wagner, 1999), indicating that the hippocampus is also involved during the reinstatement of prior memory representations.⁷

Recent imaging studies investigated hippocampal memory operations in more detail (Davachi & Wagner, 2002; Düzel et al., 2003; Giovanello, Schnyder & Verfaellie, 2004; Henke, Buck, Weber & Wieser, 1997; Jackson & Schacter, 2004; Preston, Shrager, Dudukovic & Gabrieli, 2004; Ranganath, Cohen, Dam & D'Esposito, 2004a; Zeineh, Engel, Thompson & Bookheimer, 2003). Results of these studies converge to suggest that the hippocampus is specialized for associative and particularly relational memory binding. For instance, Henke et al. (1997) could show that the hippocampus exhibits more pronounced activity during encoding and retrieval of associations between items (photographs of the interior/exterior of houses and persons) as compared to single items. Davachi and Wagner (2002) presented word triplets and contrasted two verbal encoding conditions. In an item-based condition, subjects rehearsed the three items in the order presented on the screen, whereas in a relational condition subjects were required to internally reorder the items as a function of desirability. Encoding-related activity in the hippocampus was greater in the relational as compared to the item-based condition, whereas the reversed comparison revealed activation in the entorhinal and parahippocampal cortex. Moreover,

⁷However, it should be noted that the evidence for hippocampal involvement during retrieval is not as consistent as for encoding (e.g. Henson et al., 1999b; see Rugg, Otten & Henson, 2002, for a discussion).

subsequent memory effects in the hippocampus were restricted to stimuli encoded in a relational manner. In a similar vein, in the study by Jackson and Schacter (2004), hippocampal activation during encoding word pairs predicted subsequent memory for associations of words (in contrast to memory for single items). In another associative memory study (Giovanello et al., 2004), subjects had to encode word pairs. In comparison to single items, the retrieval of word-pairs was associated with hippocampal activity (see also Düzel et al., 2003; Ranganath et al., 2004a, for further examples for hippocampal involvement during associative memory retrieval). Using new fMRI analysis methods, Zeineh and colleagues (Zeineh et al., 2003) could further disentangle hippocampal subregions during associative memory encoding and retrieval of face-name pairs. Encoding of face-name conjunctions was associated with enhanced activity in hippocampal cell fields CA2 and CA3 and the dentate gyrus, whereas pronounced activity in the subiculum could be observed during retrieval of conjunctions. In a further associative memory study (Preston et al., 2004), subjects had to learn sets of face-house (sets A-B and C-B) and face-face pairs (D-E). Critically, the two sets of face-house pairs comprised the same houses, i.e. different faces (A and C) were associated with the same houses (B). During retrieval, the hippocampus showed stronger BOLD responses for related face-face pairs (A-C) than for initially learned face-face pairs (D-E). These data indicate that the hippocampus is critically involved in the formation of relations between episodes, i.e. linking together separated episodes (cf. Bunsey & Eichenbaum, 1996, for similar results in rats). To summarize, the results of these studies indicate that the hippocampus is preferentially involved in memory for associations and relations among elements of episodes (Cohen & Eichenbaum, 1993; Cohen et al., 1999; Eichenbaum, 2000) by means of a conjunctive representation of episodes (Davachi, 2004; O'Reilly & Rudy, 2001).

Another topic of current interest is whether the hippocampus is differentially involved during familiarity-based and recollection-based recognition. Eldridge and colleagues (Eldridge, Knowlton, Furmanski, Bookheimer & Engel, 2000) used the remember know procedure (Tulving, 1985). In a first phase, subjects studied words outside the scanner. Subsequently, old and new items were presented during fMRI measure-

ments and subjects performed a recognition memory task. After their recognition judgment, subjects had to indicate whether they remembered the item (remember judgment) or whether their judgment was solely based on familiarity (know judgment). As mentioned before, dual-process models of recognition memory assume that recognition memory is supported by two independent processes, recollection and familiarity (Brown & Aggleton, 2001; Mecklinger, 2000; O'Reilly & Norman, 2002; Rugg & Yonelinas, 2003; Yonelinas, 2001). It is supposed that remember judgments are associated with recollection, whereas know judgments are based on familiarity (Tulving, 1985). Only correctly recognized items which were classified as remembered were associated with bilateral hippocampal activation.

Furthermore, functional imaging studies addressed the question whether the hippocampus subserves source memory. Source memory is a specific form of associative memory which requires building of associations between an item and a particular source (e.g. a specific color background, a specific voice associated with an item). For instance, Davachi, Mitchell, and Wagner (2003) investigated the function of the hippocampus during the encoding phase of a source memory task. In this study, subjects either had to encode items based on mental imagery or solely through reading. Subsequently, subjects performed a recognition memory task. After classifying items as new or old, subjects had to indicate the encoding task associated with old items, i.e. imagery or reading encoding (source memory task). The bilateral hippocampus and the left parahippocampal cortex showed greater activity for remembered items, when also the source (encoding task) was remembered (source memory) relative to remembered items, for which subjects could not correctly indicate the source (item memory). In contrast, perirhinal cortex activity predicted later item but not source memory (see Figure 2.3c). In accordance with these results, two recent fMRI studies observed retrieval-related hippocampal activation for items with correct source judgments (Cansino, Maquet, Dolan & Rugg, 2002; Dobbins, Rice, Wagner & Schacter, 2003). By combining a source memory task with confidence ratings, Ranganath et al. (2004b) could disentangle encoding-related activity in the rhinal cortex and the hippocampus, respectively. Rhinal cortex activity was associated with subsequent familiarity-based recognition, whereas hippocampal activity

was associated with subsequent recollection-based recognition. The results of these studies are compatible with a dual-process account of recognition memory and underscore the importance of the hippocampus during recollection-based recognition. Combining findings from the reviewed functional imaging studies, it appears that the hippocampus plays a critical role during encoding and retrieval of episodic memories. More precisely, the hippocampus is preferentially involved when relations among items have to be bound in memory, i.e. associations between distinct stimuli and associations between an item and a source. These hippocampal memory operations seem to be a prerequisite for recollection-based recognition of prior episodes.

But what about the representation of regularities across multiple episodes? Both, the Eichenbaum as well as the O'Reilly model propose that medial temporal lobe structures are involved in this function. The assumptions of both models will be outlined in the next section.

2.4 Representation of Regularities across Multiple Episodes

2.4.1 Assumptions of the Eichenbaum Model

Eichenbaum (2000, in press) posits that particular hippocampal neurons are specialized to represent features that are common across many experiences. By this, overlapping episodes could be linked together and semantic knowledge structures gradually evolve (see also Wallenstein et al., 1998). Eichenbaum argues that the hippocampus is involved in detecting overlapping features of distinct episodes and in binding these overlapping features into a common representation. Importantly, Eichenbaum proposes that “(...) (t)he representation of these general regularities constitutes semantic ‘knowledge’ that is not bound to the particular episode or context in which the information was encoded” (Eichenbaum, 2004, p. 5). These mechanisms entail the development of context-free generalized knowledge structures. Representations of context-specific memories are linked to these generalized knowledge structures and provide a relational memory network (Eichenbaum, in press).

Evidence for this notion is provided by a recent lesion study (Agster et al., 2002). In this study rats with hippocampal lesions and rats with intact hippocampi (control group) learned two odor sequences in separate sessions. These odor sequences comprised six pair-wise odor choices. Both sequences shared two common odor choices in the middle of both sequences. In a test phase, where both sequences were presented in randomized order, rats in the control group could disambiguate both sequences, whereas rats in the lesion group showed impaired performance. Based on these results, the authors argue that the hippocampus is involved in learning overlapping odor choice sequences and in disambiguating these sequences. Moreover, it has recently been shown (Wood et al., 1999) that in addition to hippocampal cells which encode specific feature conjunctions of unique episodes, particular hippocampal cells increase their firing rate selectively for events that are common across many episodes. Based on these results, Eichenbaum (2000, in press) argues that the hippocampus is also involved in the representation of overlapping features of multiple episodes.

2.4.2 Assumptions of the O'Reilly Model

In the O'Reilly model (Norman & O'Reilly, 2003; O'Reilly & Norman, 2002), the rhinal and parahippocampal cortices are assumed to represent regularities in the environment by assigning overlapping distributed representations to similar stimuli.⁸ As outlined above, the MTL cortex model is composed of an input and a hidden layer (Figure 2.1a). The hidden layer corresponds to the MTL cortex, whereas the input layer represents neocortical input into the MTL cortex. The input layer projects to the MTL cortex layer. During learning these connections are strengthened or weakened via Hebbian learning mechanisms. A sharpening mechanism selectively strengthens representations of frequent stimuli (see Figure 2.1b). More precisely,

⁸It is noteworthy that in earlier versions of this model (especially McClelland, McNaughton & O'Reilly, 1995), this function, i.e. the representation of regularities/generalities across multiple episodes was ascribed to the neocortex. In the most recent version of the model (Norman & O'Reilly, 2003; O'Reilly & Norman, 2002) which is more precise in neuroanatomical terms, the medial temporal lobe cortex (comprising the rhinal and parahippocampal cortices) is supposed to mediate this function.

when stimuli are presented for the first time they are represented in an unspecific way by a large number of weakly activated MTL cortex layer units. After repetition of familiar items, the number of activated units decreases but the activity of these few units is increased. O'Reilly and coworkers argue that the MTL cortex system is a slow learning system, since this system encodes "(...) the general statistical structure of the environment, abstracted away from the specifics of individual events which enables generalization to novel situations" (O'Reilly & Rudy, 2000, p. 390). To summarize, the O'Reilly model proposes that the rhinal and parahippocampal cortices are crucial for the representation of repeated overlapping features of episodes.

2.4.3 Concluding Remarks

Both models, the O'Reilly model (based on computational modeling) and the Eichenbaum model (based on rodent lesion and single unit recording studies) propose that the hippocampus plays a crucial role during the formation of episodic memories. Both models assume that the hippocampus is involved in relational memory binding, i.e. the hippocampus binds together features defining a unique episode. Both models make different predictions how regularities across episodes, i.e. overlapping features of multiple episodes are represented in the brain. O'Reilly assumes that the parahippocampal and the rhinal cortex represent these regularities, whereas Eichenbaum posits that specialized hippocampal neurons support the representation of overlapping features. Especially the O'Reilly model proposes that learning unique episodes vs learning regularities across episodes comprise two fundamentally different functions. However, it is not clear which learning mechanisms exactly support the representation of regularities across episodes. In the scope of the framework of two distinct learning and memory systems, learning unique episodes is the core function of the first system, whereas learning regularities across multiple episodes is supported by the second system. The term 'regularity learning' was introduced to describe the learning process of the second system to emphasize the importance of regularities across multiple episodes during learning. In the last three decades, the learning mechanisms associated with the second learning system have gained

considerable interest. Several learning tasks investigated regularity learning, e.g. reinforcement learning, category learning, artificial grammar learning, and sequence learning tasks. Moreover, neuroscientific studies have begun to unravel the network of brain areas involved during regularity learning. Consequently, in the following Chapter 3 the basic principles and neural mechanisms underlying learning regularities across multiple episodes will be outlined in more detail.

Chapter 3

Gradual Learning Regularities across Multiple Episodes: Basic Principles and Neural Mechanisms

Contents

3.1	Introduction	36
3.2	Reinforcement Learning	38
3.2.1	Basic Principles	38
3.2.2	Neural Mechanisms	39
3.2.3	Summary	46
3.3	Category Learning	46
3.3.1	Basic Principles	46
3.3.2	Neural Mechanisms	47
3.3.3	Summary	51
3.4	Artificial Grammar Learning	51
3.4.1	Basic Principles	51
3.4.2	Rule-based vs Similarity-based Learning	54
3.4.3	Neural Mechanisms	55
3.4.4	Summary	57
3.5	Sequence Learning	57
3.5.1	Basic Principles	57

3.5.2	Neural Mechanisms	62
3.5.3	Summary	65

3.1 Introduction

According to the taxonomy of two separable learning and memory systems (McClelland et al., 1995; Mishkin et al., 1984; O’Keefe & Nadel, 1978; O’Reilly & Norman, 2002; Sherry & Schacter, 1987; Squire, 1992; Squire & Zola, 1996, cf. Packard & Knowlton, 2002; Poldrack & Packard, 2003; Rolls, 2000; White & McDonald, 2002), in this chapter the second learning system associated with learning regularities across multiple episodes will be outlined. This chapter starts with a brief introduction of reinforcement learning (3.2). Subsequently, relevant aspects of category learning (3.3), artificial grammar learning (3.4), and sequence learning research (3.5) will be reviewed. In each of these sections, the basic principles and the neural mechanisms of these forms of learning will be outlined.

It is proposed that these different learning tasks - on a broad level - have three main characteristics in common. *First*, subjects extract regularities across multiple input pattern, i.e. overlapping features of episodes. *Second*, given the feature overlap during reinforcement learning, category learning, artificial grammar learning, and sequence learning it is proposed that learning in these tasks is supported by the detection of contingencies between elements of episodes, i.e. relations among the elements are built due to the co-occurrence of the elements (cf. Shanks, 1995). *Third*, learning in these different tasks is incremental by nature, since learning evolves across multiple episodes. The purpose of the present chapter is to review important aspects of the different types of learning (tasks) and to discuss them according to the three characteristics.¹ As mentioned above, the generic term ‘regularity learning’ will be used to describe the learning process underlying the second learning system. It is supposed that this term is appropriate to characterize learning in these distinct tasks which will be reviewed in the present chapter. More precisely, the existence

¹See the ‘Summary’ in the following sections.

of regularities across multiple episodes is supposed to be the critical component which allows learning to occur. It is noteworthy that the term ‘regularity learning’ is not intended to substitute terms like reinforcement learning, artificial grammar learning, category learning, sequence learning, implicit learning, gradual learning, or incremental learning. In fact, the term ‘regularity learning’ is introduced as a superordinate concept which emphasizes the importance of regularities across multiple episodes during learning in distinct learning tasks.

It is worth noting that it is a topic of current interest whether learning in these tasks is implicit or explicit.² For instance, Cleeremans et al. (1998) argues that especially artificial grammar learning and sequence learning is mainly based on implicit processes. In contrast, others (e.g. Shanks & St. John, 1994) challenge the existence of implicit learning. Shanks and St. John (1994) define critical criteria for implicit learning (cf. Reingold & Merikle, 1988) and cannot find evidence for implicit learning in a wide variety of learning tasks, including classical and instrumental conditioning, category learning, artificial grammar learning, and sequence learning. Thus, it appears that a clear distinction between pure implicit and pure explicit processes is rather difficult. It is beyond the scope of the present work to investigate the implicit/explicit basis of learning. Thus, this issue will not be further addressed in the present thesis.

²Different definitions of implicit learning have been proposed by several authors. On a broad level, implicit learning is defined as the “(..) ability to learn without awareness (...)” (Cleeremans, Destrebecqz & Boyer, 1998, p. 406), whereas during explicit learning subjects are (fully) aware of the learning process and the contents of learning. However, there is still controversy with respect to the exact nature and an overall definition of implicit learning and the best-suited methods to investigate implicit learning (Cleeremans et al., 1998; Frensch, 1998; Reingold & Merikle, 1988; Shanks & St. John, 1994).

3.2 Reinforcement Learning

3.2.1 Basic Principles

Classical Conditioning During Pavlovian or classical conditioning, an unconditioned stimulus (US) which produces an unconditioned reaction (UR) is paired with a conditioned stimulus (CS). After learning, the CS on its own is able to produce a response, the conditioned reaction (CR) which is highly similar to the original UR (Pavlov, 1927). For instance, in Pavlov’s (1927) original experimental procedure, a tone (CS) was paired with the presentation of food (US). When food was presented, the dog salivated (UR). After sufficient CS-US pairings, the dog started to salivate immediately after hearing the tone before the food was delivered, i.e. an initial neutral stimulus became a conditioned stimulus after learning.

The Rescorla-Wagner Model Learning during classical conditioning can be described by means of the Rescorla-Wagner formula (Rescorla & Wagner, 1972). An experimental procedure is defined in terms of several cues $C_0, \dots, C_n$ and an outcome O (cf. Danks, 2003). The change of associative strength ΔV_i between a cue C_i and an outcome O is defined by

$$\Delta V_i = \alpha_i \beta (\lambda - \sum_{j=1}^n V_j) \quad (3.1)$$

In Equation 3.1, $\sum_{j=1}^n V_j$ denotes the associative strength of all available cues n , λ represents the maximal associative strength, and α and β are learning parameters associated with the cue and the outcome, respectively.³ As apparent from Equation 3.1, learning (ΔV_i) is a function of the discrepancy between the actual outcome (λ) and the expected outcome ($\sum_{j=1}^n V_j$). At the beginning of learning, the outcome is highly unexpected, i.e. unpredicted by the cue, thus ΔV_i increases. In subsequent learning trials (suppositional that the outcome follows the cue), unexpectedness decreases, i.e. the degree of predictability of the outcome increases. As a result, the

³The parameter α denotes the salience of the cue C_i , whereas β is a measure of the salience of the learning situation.

learning increase is reduced and even diminishes during later learning phases, when the outcome is fully predicted by the cue.

The Temporal Difference Model The temporal difference (TD) model (Sutton & Barto, 1990, cf. Sutton & Barto, 1981) is an extension of the Rescorla-Wagner model. Similar to the Rescorla-Wagner model, the TD model proposes that reinforcement learning can be formalized by means of the difference between the expected and the actual outcome, the so-called prediction error (PE). In extension of the Rescorla-Wagner model, the TD model assumes that within a single learning trial, the sum of all future rewards is computed (see Schultz, Dayan & Montague, 1997). Moreover, in contrast to the Rescorla-Wagner model the TD model accounts for the intratrial temporal structure of classical conditioning. To put it simply, imagine a learning experiment where a specific cue is followed by a reward. In each learning trial, the PE represents the difference between the actual outcome (presentation or absence of a reward) and the expected outcome. During initial learning trials, a reward following a cue is highly unexpected, i.e. the PE is positive. After several experiences of a specific cue-reward pairing the PE decreases, since the expectation of reward delivery increases for a specific cue. Thus, the PE converges to zero, since the cue becomes a valid predictor of the reward. When the cue-reward contingency is changed, e.g. no reward is delivered after cue presentation (extinction), the PE becomes negative (reward expected, no reward delivered).

3.2.2 Neural Mechanisms

Animal Studies In accordance with the assumptions of the TD model (Sutton & Barto, 1990), it has been proposed that midbrain dopamine neurons represent predictions of future reward delivery and by this drive learning associations between stimuli (Schultz, Dayan & Montague, 1997; Schultz, 2000, 2002). In more formal terms, the activity of dopamine neurons reflects the PE, i.e. the difference between the actual and the predicted outcome. As apparent from Figure 3.1a, activity of a single dopamine neuron increases after unexpected reward delivery (top panel). Here, according to the TD model, the PE is high. After the stimulus-reward association

has been learned, the reward becomes highly predictive. Thus, the dopamine response during reward delivery (and the PE) decreases (middle panel). Interestingly, the dopamine response is shifted backwards in time from the primary reward to the stimulus predicting future reward delivery (middle panel; cf. Schultz et al., 1997). When reward delivery is omitted, the dopamine response is depressed (and the PE becomes negative; bottom panel). For instance, Hollerman and Schultz (1998) could show that dopamine responses to a rewarding stimulus gradually decrease during the time course of learning as a function of reward predictability.

Recently, further evidence for neuronal coding of reward prediction errors was provided by a seminal study (Waelti, Dickinson & Schultz, 2001). The authors recorded activity from single midbrain dopamine neurons of macaques (neurons in the substantia nigra and the ventral tegmental area, VTA). The animals performed a classical conditioning experiment (blocking paradigm). In a first phase of the experiment (pretraining), two different types of visual stimuli were presented, one stimulus predicted reward delivery (A^+), whereas another stimulus was never followed by a reward (B^-). Subsequently, stimuli A and B were paired with two novel stimuli X and Y , resulting in pairs AX and BY (compound learning). Both compounds were rewarded. In terms of formal learning theories, stimulus Y becomes a valid predictor of the reward, since reward delivery following the presentation of the stimulus pair BY generates a positive PE. In contrast, stimulus X does not generate a PE, since stimulus A already fully predicts reward delivery. During a subsequent test phase, stimuli X and Y were presented alone. The authors observed learning (duration of licking behavior) in Y trials, but not in X trials, indicating that X was blocked relative to Y . Neuronal activity of dopamine neurons during stimulus processing paralleled the behavioral results and the assumptions of formal learning theories. Dopamine neurons showed enhanced activity for stimulus A^+ (positive PE), but not for stimulus B^- (no PE) during pretraining. In the compound learning phase, both compounds elicited increased neuronal activity (positive PE). Crucially, during the test phase increased neuronal activity could only be observed in Y , but not in X trials (blocking effect).

Several other target regions of midbrain dopamine neurons are involved in reward-

related learning (Schultz, 2000, 2002, cf. Figure 3.1b). Both the orbitofrontal cortex and the basal ganglia are associated with reward expectation (Schultz, Tremblay & Hollerman, 2000). It has been shown that the orbitofrontal cortex is specifically involved in the discrimination between different rewards according to the preference of the animal (Tremblay & Schultz, 1999), whereas basal ganglia structures have been implicated in reinforcement learning by means of including reward expectations into the preparation of actions (Lauwereyns, Watanabe, Coe & Hikosaka, 2002b; Schultz, Tremblay & Hollerman, 2003). In a similar manner, reward-related dorsolateral prefrontal activity has been associated with the preparation and execution of goal-directed behavior (Leon & Shadlen, 1999; Watanabe, 1996). Furthermore, the anterior cingulate cortex (Shidara & Richmond, 2002) and the amygdala (Schoenbaum, Chiba & Gallagher, 1998) also process reward-related information. Schultz (2002) argues that different brain regions contribute to specialized aspects of reward processing and by this might interact in the service of goal-directed behavior (see Figure 3.1b). For instance, Schultz proposes that basic mesolimbic reinforcement signals provided to the prefrontal cortex might play a crucial role in an efficient response execution towards rewarded stimuli in the environment (cf. Cohen, Braver & Brown, 2002; Miller, 2000).

Imaging Studies The neural basis of reinforcement learning in humans was investigated in several recent fMRI studies (Berns, McClure, Pagnoni & Montague, 2001; Gottfried, O'Doherty & Dolan, 2003; McClure, Berns & Montague, 2003; O'Doherty, Kringelbach, Rolls, Hornak & Andrews, 2001; O'Doherty, Deichmann, Critchley & Dolan, 2002; O'Doherty, Dayan, Friston, Critchley & Dolan, 2003; Pagnoni, Zink, Montague & Berns, 2002, see Braver & Brown, 2003). In accordance with the aforementioned evidence provided by animal studies, O'Doherty et al. (2001) demonstrated a preferential role of the orbitofrontal cortex in reward discrimination. The authors could show that separable orbitofrontal regions code the magnitude of reward and punishment. Furthermore, it has been shown that the orbitofrontal cortex,

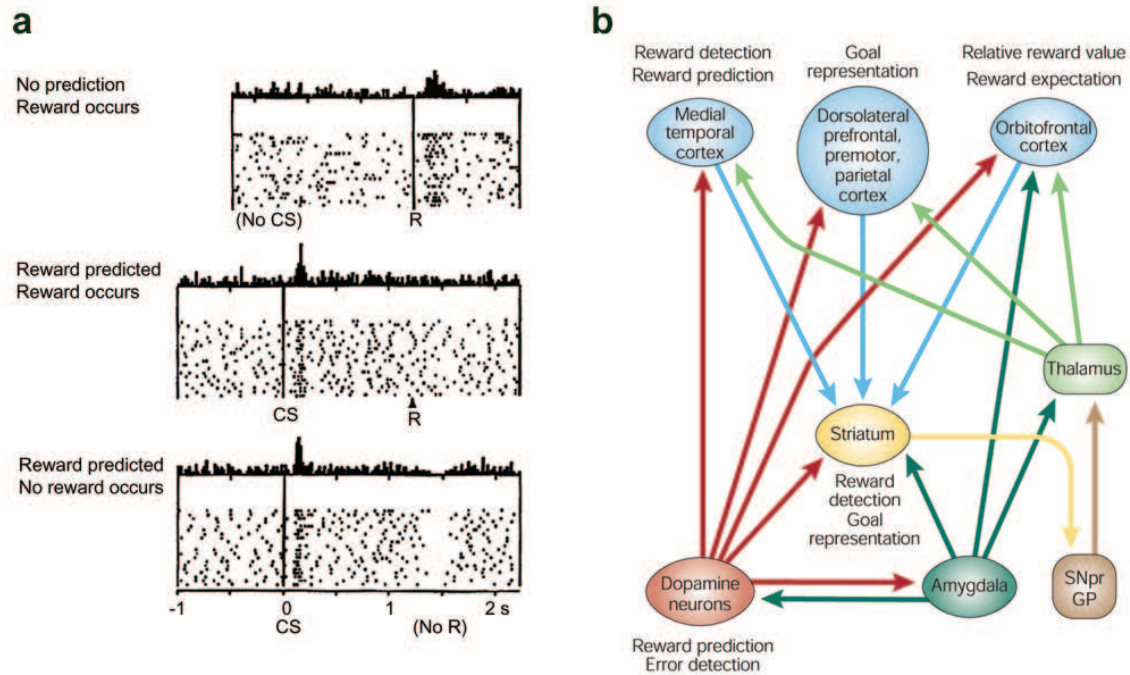


Figure 3.1. Learning-related Activity of a Dopamine Neuron and Brain Regions Involved in Reward Processing. (a) Prediction error-related activity of a single midbrain dopamine neuron. CS, conditioned stimulus, R, reward. (b) Brain regions associated with reward-related learning. See text for details. [(a) Adapted from Schultz (1997). Copyright (1997) American Association for the Advancement of Science (Science). (b) Adapted from Schultz (2000). Copyright (2000) Macmillan Magazines Ltd (Nature Reviews Neuroscience).]

the amygdala, the striatum,⁴ and dopaminergic midbrain regions are involved during the anticipation of reward (Gottfried et al., 2003; O'Doherty et al., 2002). Two recent studies provide evidence for prediction error-related brain responses in the human striatum (putamen). McClure et al. (2003) and O'Doherty et al. (2003) used visual cues and juice rewards in a classical conditioning paradigm (see Seymour et al., 2004, for recent evidence for striatal involvement in aversive conditioning). McClure et al. (2003) manipulated the delay between the cue and reward delivery. After presenting 50 trials with a constant cue-reward delay of 6 s (learning phase), 6 catch trials with a 10 s delay were introduced (test phase). First, a comparison

⁴It should be noted that most animal researchers use the term dorsal striatum to describe the caudate nucleus and the putamen, whereas the ventral striatum comprises the nucleus accumbens and portions of the olfactory tubercle (Packard & Knowlton, 2002). In contrast, others (e.g. O'Doherty et al., 2003, 2004) use the term dorsal striatum to refer to the caudate nucleus and the term ventral striatum to refer to the putamen.

between reward delivery in the test phase with reward delivery in the training phase revealed activation of the putamen. The authors argue that this striatal activation pattern is due to a positive PE elicited by unexpected reward delivery in the test phase (vs no PE for expected reward in the learning phase). Moreover, the authors contrasted brain activity during time windows of no reward delivery in both experimental phases, i.e. a time window in the (1) test phase, where juice was normally delivered in the learning phase (6 s after cue offset) and (2) a time window around 10 s after cue offset in the learning phase (the time window in which reward was delivered in the test phase). Based on the experience in the learning phase, during (1) reward was expected, but not delivered, whereas during (2) reward was not expected and not delivered. A deactivation of the putamen could be observed for (1) relative to (2), indicating that the putamen also codes a negative PE in situations when an expected reward is omitted. These results are compatible with the assumptions of formal learning theories.

In the O’Doherty et al. study (2003), assumptions of the TD learning model were tested directly by incorporating the output of a TD algorithm into the fMRI regression model. The authors modeled two critical events in their classical conditioning study, the presentation of the CS and the UCS (reward), respectively. In brief, the TD model output predicted the activation pattern in several brain regions, especially the striatum (putamen) and the orbitofrontal cortex. In these regions, PE-related neural activity could be observed, i.e. when unexpected reward was delivered, the BOLD signal increased (positive PE), whereas a negative BOLD signal could be observed when an expected reward was omitted (negative PE). Most interestingly, the authors could observe a learning-related backward shift of activity from the reward to the reward-predicting CS in the striatum, i.e. in initial trials activity was most pronounced during reward delivery, whereas in later stages of learning, striatal activity was restricted to the CS (cf. Schultz et al., 1997).

Moreover, the human striatum is also involved in instrumental conditioning, where subjects learn to select appropriate actions to increase the probability of reward delivery. Most recently, O’Doherty et al. (2004) could disentangle the contributions of

separate striatal regions during instrumental conditioning.⁵ Activity in the putamen correlated with the PE signal during both, classical and instrumental conditioning. In contrast, correlation between caudate nucleus activity and the PE signal was restricted to instrumental conditioning (see also Tricomi, Delgado & Fiez, 2004, for functional imaging evidence for caudate nucleus involvement during reinforcement of actions by means of coding action-reward contingencies).

To summarize, single unit activity and BOLD responses in several brain regions, especially the ventral tegmental area, the substantia nigra, the striatum, and the orbitofrontal cortex are crucially involved in classical and instrumental conditioning. Furthermore, activity in these regions comply with the assumptions of the temporal difference learning model.

Probabilistic Learning Beyond conditioning tasks, several recent studies used probabilistic learning tasks to examine the neural basis of reinforcement learning. In these tasks, specific cues are followed by specific outcomes in a probabilistic manner. During learning, subjects acquire knowledge about cue-outcome relations. In one study (Fletcher et al., 2001), subjects learned associations between artificial drugs and syndromes on the basis of a trial-by-trial feedback. Fletcher et al. (2001) directly tested several predictions of the Rescorla-Wagner theory. In brief, activity in the dorsolateral PFC complied with the predictions of the formal learning theory, e.g. an initial activation increase was followed by a decrease after cue-outcome pairs have been learned. Moreover, the introduction of violations of learned expectations exhibited increased dorsolateral PFC activity.

In another study (Poldrack et al., 2001), subjects performed a feedback-based probabilistic categorization task (weather prediction), i.e. specific visual pattern were associated with 'sunshine' or with 'rain', respectively. Here, subjects classified items in each trial and learned cue-outcome relations on the basis of the provided feedback. This task was contrasted with a paired associate task, in which subjects encoded

⁵The authors used the output of a computational model, the so-called actor-critic model (cf. Dayan & Balleine, 2002), a two-process account of instrumental conditioning to model learning-related activation pattern.

stimuli and corresponding categories for a later classification task (stimuli and category labels were presented simultaneously). In the feedback-based task, learning was accompanied by a shift from a medial temporal lobe to a striatal brain system. The MTL was involved at the beginning of learning. Subsequently, MTL activity decreased, whereas an activation increase of the caudate nucleus could be observed during the time course of the experiment. The comparison between the feedback-based task and the paired associate task revealed stronger BOLD responses in the MTL for the latter task, whereas in the reversed comparison pronounced activity in the caudate nucleus could be observed. The authors argue that the activation pattern in the basal ganglia and the MTL is due to the differential recruitment of declarative and non-declarative memory processes in both tasks. The feedback-based task requires non-declarative memory processes (basal ganglia), whereas the paired associate task is based on declarative memory processes (MTL). These fMRI results are in accordance with earlier reports of functional dissociations during probabilistic learning in patients with MTL and striatal dysfunction, respectively (Knowlton, Mangels & Squire, 1996). Amnesic patients with MTL dysfunction exhibited preserved learning, however their memory performance for specific details of the training phase (e.g. information about the cues and the layout of the computer screen) was impaired. Parkinsonian patients (midbrain/striatal dysfunction) showed the reversed pattern, i.e. impaired probabilistic learning, but intact memory (cf. Squire & Zola, 1996). These data demonstrate a complementary involvement of MTL and striatal regions during the time-course of feedback-based learning (see the reviews of Packard & Knowlton, 2002; Poldrack & Packard, 2003). In a follow-up study with patients showing striatal dysfunction (Parkinson disease), Poldrack and colleagues (Shohamy et al., 2004) could show that patients were impaired in a feedback-based version of a probabilistic categorization task, but not in a non-feedback version of the task, thus providing further evidence for the notion that the basal ganglia play a crucial role during feedback-based learning. In sum, the results of these studies suggest that the lateral PFC and especially the basal ganglia subserve probabilistic learning.

3.2.3 Summary

Taken together, during classical/instrumental conditioning and probabilistic learning subjects are exposed to regularities across learning episodes. Here, regularities comprise cue-outcome pairings which are repeated several times. Learning in these tasks is based on the detection of contingencies between specific cues and specific outcomes (Shanks, 1995). Moreover, learning is gradual since several contingent cue-outcome pairings are required to establish robust learning effects. Finally, the reviewed studies suggest that reinforcement learning is mainly supported by the midbrain dopamine system, the striatum as its main target area and the lateral prefrontal cortex.

3.3 Category Learning

3.3.1 Basic Principles

With respect to regularity learning, another learning task seems to be highly important, namely category learning. According to Ashby (Ashby & Ell, 2001; Ashby & Casale, 2003), there are three main experimental tasks used in category learning research: (1) rule-based tasks, (2) information-integration tasks, and (3) prototype-distortion tasks. In rule-based category learning tasks features are presented on several dimensions, e.g. circles and squares (object form) in small and large size (object size) in red and blue color (object color) and subjects have to extract the relevant rule for correct responses, e.g. to respond ‘true’, if large objects are presented, and ‘false’ in all other cases. Rules in information-integration category learning tasks are more complex. Here, a rule is defined according to a specific feature conjunction, e.g. to select objects that are small *and* red *and* circles. In contrast to rule-based tasks, in information-integration tasks subjects normally cannot describe the rule explicitly. The most prevalent category learning task is the prototype-distortion task (Posner & Keele, 1968, 1970). In this task, distortions of a prototype random dot pattern are presented in a study phase (see Figure 3.2c). In a subsequent test phase, the prototype or several distortions of the prototype and random pattern not be-

longing to the prototype category are presented. Subjects are required to classify the test items as category members or non-members and normally their classification performance lies above chance.

Two major accounts have been proposed for human categorization processes: (1) exemplar-based theories, e.g. the Generalized Context Model (Nosofsky, 1986) or the connectionistic model ALCOVE (*attention learning covering map*; Kruschke, 1992) and (2) rule-based theories⁶ (e.g. the General Recognition Theory, Ashby & Townsend, 1986; Ashby & Perrin, 1988). According to exemplar-based models, subjects represent category exemplars and their categorization decision is based on a comparison process between the current stimulus and the stored exemplars. In contrast, rule-based models assume that subjects respond to stimuli according to decision rules which partition the potential decision space. In addition, prototype models (e.g. Posner & Keele, 1970) have been proposed which posit that subjects create prototypes representing the average member of a specific category (cf. Homa, Dunbar & Nohre, 1981; Metcalfe & Fisher, 1986).

Rouder and Ratcliff (2004) directly compared both exemplar-based and rule-based models using a probabilistic assignment task, in which category membership of stimuli follows a probability distribution. The results of this study indicate that exemplar-based models account for tasks in which stimuli are distinct, whereas rule-based models can explain subjects' performance when stimuli are confusable. In a similar vein, recent models, e.g. ATRIUM (*attention to rules and instances in a unified model*; Erickson & Kruschke, 1998, 2002) suggest that category learning depends on a continuous interplay between both exemplar-based and rule-based learning processes.

3.3.2 Neural Mechanisms

Based on a review of the neuropsychological and neuroimaging literature on category learning, Ashby (Ashby & Ell, 2001; Ashby & Casale, 2003) proposes that there are different category learning systems in the human brain. Ashby suggests that cate-

⁶Also referred to as decision-bound models.

gory learning in rule-based tasks is subserved by the prefrontal cortex and the basal ganglia. Moreover, the basal ganglia are involved in information-integration tasks. Furthermore, patient and neuroimaging data suggest that the medial temporal lobe is crucial for learning categories in information-integration tasks with few categories. Finally it is proposed that the visual cortex is involved in prototype-distortion category learning. In the following, an overview of findings from single-unit recordings in monkeys and human amnesia and imaging studies will be given.

Animal Studies By measuring single cell activity in the prefrontal cortex, Freedman and colleagues (Freedman, Riesenhuber, Poggio & Miller, 2001, 2002) investigated the neurophysiological basis of category learning in monkeys. The authors used three species of cats and dogs respectively as prototype stimuli and applied a morphing algorithm to generate several linear blends between prototype pairs (Figure 3.2a). This procedure resulted in a continuum along the categories ‘cat’ and ‘dog’. In each trial, a sample stimulus was presented, followed by a delay and a probe stimulus. The monkeys’ task was to decide whether the probe belongs to the same or a different category as compared to the sample. Neuronal activity in the lateral prefrontal cortex was sensitive to the category of the stimuli, i.e. specific neurons responded selectively to the category ‘cat’, whereas other neurons encoded the category ‘dog’ even when stimuli were located near the category boundary (Figure 3.2b). Most surprisingly, after retraining one monkey using the same stimulus set but defining new category boundaries orthogonal to the original boundary the same neurons which represented the original categories now encoded the new categories. These data indicate that the lateral PFC subserves flexible category learning in monkeys.

Amnesia Research Category learning was also examined in amnesic patients (Knowlton & Squire, 1993; Squire & Knowlton, 1995). In these studies, the authors could demonstrate preserved classification ability in patients using the prototype distortion task. However, patients’ recognition performance of specific training exemplars was severely impaired. Based on these results, Squire argues that “(...)

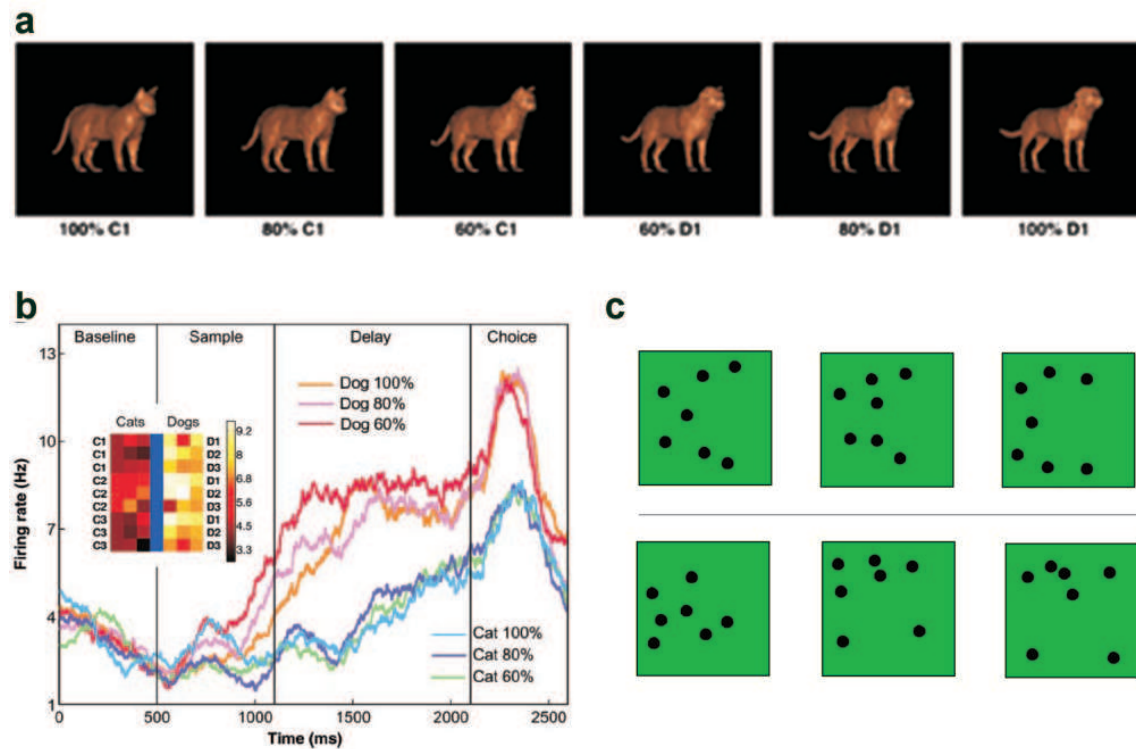


Figure 3.2. Stimuli and Results of the Freedman et al. (2001) Study and Example Stimuli of the Prototype Distortion Task. (a) Examples of six prototype morphs from '100% cat' (left) to '100% dog' (right). (b) Lateral prefrontal single cell activity for six different morph blends. (c) Example stimuli from the prototype distortion paradigm. Dot patterns belonging to a category are presented in the upper row, non-category members in the lower row. [(a-b) Adapted from Freedman et al. (2001). Copyright (2001) American Association for the Advancement of Science (Science). (c) Adapted from Ashby and Ell (2001). Copyright (2001) by Elsevier Science Ltd (Trends in Cognitive Sciences).]

declarative memory can be acquired and retained about each of the training items, and this ability depends on the medial temporal lobe and diencephalic structures that are damaged in amnesia and that are essential for declarative memory. (...) (R)epeated experience leads to knowledge about the category to which the training items belong. Category-level knowledge might be acquired by abstracting information across encounters with specific training exemplars, (...) independently of and in the absence of normal declarative memory for the items presented during learning" (Squire & Zola, 1996, p. 13517).

Imaging Studies In several fMRI studies, Reber and colleagues (Reber, Stark & Squire, 1998a,b; Reber, Gitelman, Parrish & Mesulam, 2003) identified brain re-

gions involved in category learning using the prototype distortion task. In one study (Reber et al., 1998b), categorial stimuli during the test phase elicited an activation decrease in occipital areas, whereas activity in anterior and inferior lateral prefrontal regions increased. The authors suggest that the occipital activation pattern reflects facilitated processing of learned categories. In contrast, the anterior and lateral PFC seems to be involved during retrieval of exemplars of learned categories. In another study (Reber et al., 1998a), brain activity during categorization and recognition was contrasted. In the categorization task, subjects had to judge whether items belong to the same category as the study items, whereas in the recognition task participants had to indicate whether the test item was presented during the study phase. Similar to the aforementioned study, categorization was associated with decreased occipital cortex activity, whereas prefrontal regions (middle and superior frontal gyrus) were involved during recognition. Moreover, contrary to the categorization task occipital cortex activity increased in the recognition task. Recently, Reber et al. (2003) examined the neural correlates of implicit and explicit category learning. Subjects studied items either incidentally or intentionally. Processing of categorial stimuli was associated with decreased occipital activity in the implicit condition and with increased activity in the hippocampus and the anterior PFC in the explicit condition (see Aizenstein et al., 2000, for similar results). The authors argue that implicit category learning relies on brain regions associated with non-declarative memory processes (e.g. occipital cortex), whereas explicit category learning is subserved by brain regions involved during declarative memory formation (e.g. hippocampus, PFC).

Rule- and exemplar-based categorization was examined in a recent fMRI study (Strange, Henson, Friston & Dolan, 2001). During the experiment, subjects had to indicate whether four-letter strings follow abstract rules defining category membership (e.g. ‘The first two letters are always identical’). During experimental blocks, rule changes and exemplar changes were introduced. In contrast to exemplar changes, subjects’ performance significantly decreased immediately after rule changes. These rule changes were associated with an increase of bilateral anterior prefrontal cortex activity, whereas exemplar changes modulated activity in the left anterior hippocampus. These data support the notion that prefrontal regions are involved in abstract

rule learning. Moreover, the authors propose that the hippocampus is recruited during similarity-based categorization (cf. Section 3.4.2).

Taken as a whole, these data indicate that category learning is subserved by lateral prefrontal brain regions, is associated with facilitated processing of category members in occipital areas, and seems to be (partially) independent of the medial temporal lobe memory system.

3.3.3 Summary

With respect to the characteristics proposed above, during category learning regularities comprise the repeated members of particular categories. The reviewed studies suggest that learning seems to be based on the detection of contingencies between category members and the respective category (Shanks, 1995). Similar to reinforcement learning, several exposures of category exemplars are required to entail learning. Finally, category learning seems to be mainly supported by lateral prefrontal (and occipital) brain regions.

3.4 Artificial Grammar Learning

3.4.1 Basic Principles

In the last two decades, artificial grammar learning (AGL) has become one of the most intensively investigated learning tasks. Typically there are two phases in AGL tasks. First, during an incidental phase subjects are exposed to a set of letter strings which, unknown to the subjects, follow the rules of a finite-state grammar (Chomsky & Miller, 1958, see Figure 3.3a). In a second phase, subjects are required to classify letter strings as grammatical or non-grammatical. Finite state grammars can be described in terms of a graph representation (Figure 3.3a). Grammatical strings are valid transitions from the start to the end node of the graph, whereas non-grammatical strings violate the transition rules. Reber (1967) first observed that performance in the test phase lies above chance, suggesting that participants learned the principles of the grammar.

What are the underlying learning mechanisms? In recent years, several accounts have been developed. Reber (1989) proposes that subjects learn abstract rules of the artificial grammar system (rule-based learning). More precisely, he argues that learning is implicit and yields an abstract knowledge base, i.e. the grammatical rules are not accessible to awareness.⁷ In the same vein, Dulany, Carlson, and Dewey (1984, 1985) argue that performance in AGL tasks is based on the acquisition of a set of micro-rules, approximations of the rules underlying the finite-state grammar. Moreover, supporting evidence for rule-based learning is provided by studies using AGL transfer tests (Gomez & Schvaneveldt, 1994; Mathews et al., 1989). After subjects had studied a set of letter strings generated from a finite-state grammar, in both studies a transfer test was conducted. Test items were constructed with a different set of strings, however the transition rules remained unchanged between study and transfer period. Subjects' grammaticality judgments in the Gomez and Schvaneveldt (1994) and Mathews et al. (1989) study were above chance in the transfer test. Thus, these findings suggest that participants acquire knowledge of the artificial grammar rules. In contrast, models of fragment-based learning (or chunk-based learning) propose that learning is based on knowledge of fragments/chunks of grammatical strings, e.g. letter bigrams and trigrams and not on the acquisition of an abstract rule system. For instance, Servan-Schreiber and Anderson's (1990) theory of competitive chunking assumes that every input string is parsed into hierarchical organized chunks. Servan-Schreiber and Anderson (1990) define chunks as "(...) traces in long memory, hierarchical structures whose elements are themselves chunks" (Servan-Schreiber & Anderson, 1990, p. 600). The bottom level comprises the letters of the string. Further levels are proposed: word, sentence, and phrase chunks. One critical parameter in the theory of competitive chunking is the chunk strength which reflects the activation strength of a specific chunk. How is the hierarchical chunks system build during learning? Servan-Schreiber and Anderson (1990) argue that the perception of incoming strings is a recursive process. In more detail, "(...) (t)he ele-

⁷Others (e.g. Dienes, Broadbent & Berry, 1991) have shown that subjects could acquire fragmentary explicit knowledge of the grammar.

mentary percept is formed by matching elementary chunks to the stimulus. Then the next percept is formed by elaborating on the elementary percept with more complex chunks. Then the next percepts are formed by elaborating on the current percept with increasingly complex chunks until no more chunks are available to elaborate on the current percept” (Servan-Schreiber & Anderson, 1990, p. 601). A measure of the complexity at the final stage is the variable *nchunks*, reflecting the number of chunks required to process a current input pattern. This parameter decreases during learning, i.e. a more dense, complex knowledge base is build. To summarize, the theory of competitive chunking defines learning as the organization of an input pattern into compact chunks. During the test phase, the grammaticality judgment is based on this chunk knowledge, i.e. if a test string can be parsed by the learned chunk structure it will be judged as grammatical. Furthermore, results from Perruchet and Pacteau (1990, cf. Perruchet, 1994) are in agreement with a fragment-based learning account. The authors could show that the presentation of bigrams in the learning phase of an AGL experiment revealed the same learning effects as the presentation of whole strings, suggesting that knowledge about (the overall distribution of) fragments of strings is sufficient to account for subjects’ performance.

An integration of both rule- and fragment-based learning theories was provided by Meulemans and Van der Linden (1997). The results of their study revealed one critical parameter which determines whether subjects’ grammaticality judgments are mainly based on knowledge of fragments or on abstract knowledge of the grammatical rules: the duration of the learning phase. More precisely, when participants are exposed only to a few grammatical strings during learning, fragment-based performance is dominant. In contrast, when learning duration is increased, i.e. most of the legal strings of a grammar are presented during learning, subjects seem to acquire abstract rule knowledge.⁸ In the same vein, previous studies suggest that both mechanisms, abstract rule learning as well as concrete exemplar learning are crucial for the acquisition of artificial grammars (Gomez & Schvaneveldt, 1994; Knowlton

⁸But see Johnstone and Shanks, 1999, who provided an alternative interpretation. These authors suggest that Meulemans and Van der Lindens’ (1997) subjects do not acquire abstract rule knowledge, but rather knowledge about the distribution of chunk locations.

& Squire, 1996)

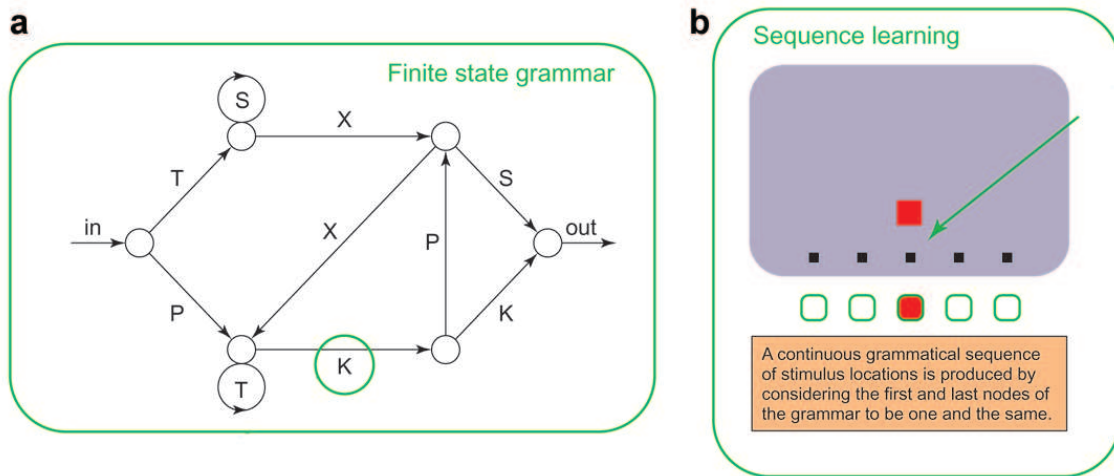


Figure 3.3. Finite State Grammar and Sequence Learning. (a) In the graph representation of a finite state grammar arrows denote valid transitions between nodes. A grammatical string is generated by traversing from the ‘in’ to the ‘out’ node. (b) The structured sequence of positions in sequence learning tasks also follows a finite-state grammar. Small black squares represent stimulus positions, white squares denote corresponding response buttons. The current stimulus-response pair is denoted in red. [Adapted from Cleeremans et al. (1998). Copyright (1998) Elsevier Science Ltd (Trends in Cognitive Sciences).]

3.4.2 Rule-based vs Similarity-based Learning

As outlined above, contrary accounts of learning exist. These theories mainly differ with respect to the form of knowledge acquired during artificial grammar and category learning. Several authors propose that learning is based on the acquisition of abstract rules, generating structured sequences of stimuli (Reber, 1989) or microrules, approximations of the original rules (Dulany et al., 1984, 1985). Evidence for this notion was provided by studies using transfer tests (Gomez & Schvaneveldt, 1994; Mathews, Buss, Stanley, Blanchard-Fields, Cho & Druhan, 1989). Here, subjects showed above-chance classification even for novel artificial strings. Supporting evidence also comes from studies which show that rule knowledge acquired in one domain (e.g. visual) can be expressed in another domain (e.g. auditory; Altmann, Dienes & Goode, 1995). In a similar manner, rule-based accounts of category learning assume that subjects learn rules which define category membership (Ashby & Perrin, 1988).

In contrast to these rule abstraction accounts, exemplar-based/instance-based (Vokey & Brooks, 1992, 1994) and fragment-based models (Perruchet & Pacteau, 1990; Perruchet, 1994) assume that learning is based on the comparison between specific training exemplars and test items and learning about the frequency of item fragments, respectively (cf. exemplar-based categorization models for similar arguments, e.g. Nosofsky, 1986). These latter accounts have been denoted similarity-based, since learning is supposed to be due to the detection of superficial similarities between stimuli (Pothos & Bailey, 2000; Pothos, in press).

To overcome this apparent discrepancy, Pothos (in press) proposed a unitary view of human learning. On the basis of categorization, he proposes that rule-based learning is prevalent, when learning “(...) is determined by a small subset of the relevant object properties (...)” (Pothos, 2004, p. 4). Contrary, similarity-based learning processes are involved when “(...) most of the relevant object properties (are) broadly equally weighted (...)” (Pothos, 2004, p. 4). In conclusion, Pothos argues that rule-based learning can be understood as special case of similarity-based learning.

3.4.3 Neural Mechanisms

Amnesia Research In several behavioral studies, Knowlton and Squire (Knowlton, Ramus & Squire, 1992; Knowlton & Squire, 1994, 1996) examined artificial grammar learning in amnesic patients. In accordance with their research on category learning, patients’ accuracy was greater than chance and their performance was similar to healthy controls in these experiments. However, when patients were required to recognize specific training exemplars (Knowlton et al., 1992) or letter chunks (bigrams and trigrams from the training exemplars; Knowlton & Squire, 1996), their performance was impaired relative to control subjects. Thus, Knowlton and Squire argue that artificial grammar learning is based on non-declarative memory processes, i.e. independent of the hippocampal memory system (cf. Squire & Zola, 1996). In contrast, correct recognition of items depends on intact (hippocampal) declarative memory operations (but see Kinder & Shanks, 2001).

Imaging Studies In the last years, several functional imaging studies tried to elucidate the neural correlates of artificial grammar learning. For example, it has been shown that lateral prefrontal regions are differentially involved in item and rule learning (Fletcher, Büchel, Josephs, Friston & Dolan, 1999). In this study, subjects were required to judge the grammatical status of four-letter strings. Whereas the right middle frontal gyrus was associated with learning individual items in an early stage of the experiment, the left middle frontal gyrus was sensitive to rule learning across experimental blocks. Furthermore, effective connectivity analyses revealed an increase in left fronto-parietal and a decrease in right fronto-parietal connectivity during the time-course of rule learning. In accordance with these data, Tettamanti and colleagues (2002) showed that Broca's area in the left inferior frontal gyrus was activated during the time-course of learning artificial grammar rules. Moreover, subjects who had a higher proficiency in grammatical rule usage showed greater activation in Broca's areas as compared to participants with lower proficiency.

In contrast to the aforementioned studies, Opitz and Friederici (2003) used an artificial grammar system that mimics structures of natural languages (see Figure 3.4a). The authors observed a learning-related increase of activity in Broca's area and a learning-related decrease of left hippocampal activation during learning (see Figure 3.4b,c). Opitz and Friederici (2003) argue that the left hippocampus is involved at an early stage of learning based on superficial features of strings, whereas the left prefrontal cortex mediates rule-based learning at a later stage of the acquisition of the finite state grammar. In a follow-up study, Opitz and Friederici (2004) could show that the left hippocampus is selectively engaged when superficial changes of grammatical sentences are introduced and that solely the left prefrontal cortex subserves learning new grammatical rules, supporting the notion that the hippocampus is associated with similarity-based learning and the left prefrontal cortex with abstract rule-learning.

Although not an AGL study, a recent fMRI experiment (Musso et al., 2003) provides new insights into the neural basis of grammar learning. The authors examined second language acquisition. In two experiments, subjects had to learn grammatical rules of either Italian or Japanese, respectively (both groups of participants

were unfamiliar with Italian or Japanese). In addition, both groups learned artificial rules of unreal languages which resulted from selectively manipulating Italian and Japanese grammatical rules, respectively. In contrast to the unreal languages, Italian and Japanese follow the principles of an universal grammar. Left inferior frontal gyrus (Broca's area) activity increased as a function of learning real languages. In contrast, no modulation of inferior frontal gyrus activity could be observed during learning grammatical rules of the unreal languages. These data indicate that the left inferior frontal gyrus subserves learning natural language rules, "(...) suggesting a universal syntactic specialization of this area among 'real' languages" (Musso et al., 2003, p. 778). To summarize, the reviewed fMRI studies indicate a preferential role of the lateral prefrontal cortex, especially Broca's area during artificial (and natural) grammar learning.

3.4.4 Summary

During artificial grammar learning, elements of grammatical strings are repeated several times, i.e. these repeated string elements comprise the regularities across learning episodes. Similar to reinforcement and category learning, the contingency between elements of artificial grammar strings is an important aspect of artificial grammar learning. The contingent sequence of grammatical strings is defined by the finite state grammar. Furthermore, the presentation of several exemplars of the grammar is a prerequisite of learning. On a neural level, artificial grammar learning seems to be mediated mainly by the left lateral PFC.

3.5 Sequence Learning

3.5.1 Basic Principles

With respect to regularity learning, another learning task seems to be relevant: sequence learning. Nissen and Bullemer (1987) first introduced the serial reaction time (SRT) task to investigate learning complex sequential structures. In this task, stimuli are presented at different locations and subjects are required to press a particular

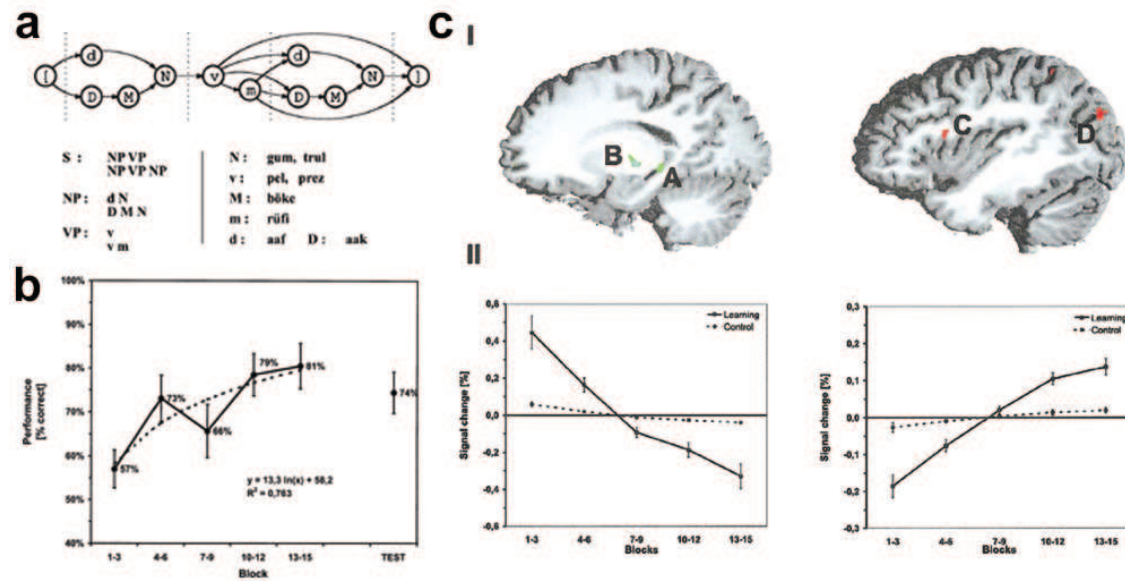


Figure 3.4. The Finite State Grammar and Results of the Opitz and Friederici (2003) Study. (a) Graph representation of the finite state grammar BROCANTO. (b) Subjects' performance for grammaticality judgements increases as a function of time. (c) Learning-related decrease in left hippocampus (A) and learning-related increase of activation in Broca's area (C). [Adapted from Opitz & Friederici, 2003. Copyright (2003) Elsevier Science Ltd (NeuroImage).]

button corresponding to a specific location (Figure 3.3b). Unknown to the subjects, the sequence of locations (and responses) follows a fixed rule. In comparison to a random sequence⁹ the rule sequence leads to decreased reaction times during the experiment, indicating that subjects learn temporal contingencies between stimuli defining a sequence (Cleeremans & McClelland, 1991). Similar to artificial grammar learning, the transition between stimuli defining a structured sequence can be described by a finite-state grammar (cf. Figure 3.3), i.e. a structured sequence is a legal transition from the start to the end node of the finite state grammar. In the following, two topics of main interest in sequence learning research, namely the attentional vs non-attentional and the stimulus vs response distinction will be reviewed.

Several studies addressed the question whether attention is a prerequisite for the

⁹Normally, several task blocks with the structured sequence are followed by a few (1-3) blocks including random sequences. Mostly, the SRT experiment ends with a few blocks of structured sequences. Learning is quantified by comparing reaction times (RT) in the random sequence blocks with the RT in the last structured sequence block(s) before and/or after the random block(s).

acquisition of sequence knowledge. It has been shown that learning effects diminish when subjects are required to perform a tone-counting task concurrently (Nissen & Bullemer, 1987). Based on these results, Nissen and Bullemer (1987) assumed that sequence learning requires attention. However, there is empirical evidence that sequence learning is preserved even under dual-task conditions depending on the specific structure of the underlying sequence (Cohen, Ivry & Keele, 1990). More precisely, when each element in a sequence has a unique association with the following sequence element, subjects exhibit decreased RT for structured sequences when performing an attentional demanding second tone-counting task in parallel. This was not the case for ambiguous sequences, i.e. each sequence element was followed by different elements at different positions in the sequence. Cohen et al. (1990) argue that sequence learning may depend on two mechanisms, a non-attentional mechanism which is responsible for building associations between sequential stimuli, and a second mechanism which is involved in the representation of hierarchical structures of input pattern. This latter mechanism is supposed to require attention. Another sequence learning study (Frensch, Buchner & Lin, 1994) investigated the influences of short term memory processes on sequence learning in more detail (cf. Frensch & Miner, 1994). The data of Frensch et al. (1994) indicate that subjects are able to learn ambiguous sequences under dual-task conditions when they have enough time to retain sequence elements in working memory.

Extending the aforementioned results (Cohen et al., 1990), Curran and Keele (1993) introduced a dual-process model of sequence learning (cf. Keele, Ivry, Mayr, Hazeltine & Heuer, 2003). This model proposes two independent and parallel learning modules, namely an attentional and a non-attentional module. The authors could show that subjects benefit from full awareness of the sequential structure (in contrast to unawareness) and from full attention (in contrast to divided attention) in an initial phase of learning, respectively. However, when the same task has to be performed subsequently under dual-task conditions, participants could not transfer their knowledge to the new experimental situation. Moreover, when the task structure was reversed, i.e. participants first learned under dual-task conditions and under single-task conditions afterwards, no improvement could be observed in the

second phase of the experiment. These results suggest that “(...) the attentional form of learning seems to require attention for its expression as well its learning. When distraction was added following such learning, expression of the attentionally dependent knowledge disappeared” (Curran & Keele, 1993, p. 199). Finally, Curran and Keele (1993) argue that “(...) non-attentional learning operated primarily by simple associations, whereas attentional learning has a mechanism for encoding the position of an event within the sequence” (Curran & Keele, 1993, p. 200), possibly based on hierarchical chunking operations (cf. Cohen et al., 1990).

In contrast, an alternative interpretation of dual-task interference effects on sequence learning was given by Stadler (1995). Stadler’s results suggest that the consistent temporal organization of input sequences and not the attentional capacity is disrupted under dual task conditions (see Frensch & Miner, 1994, for similar findings). Briefly, in this study a SRT task was paired with (a) a tone-counting task, (b) a tone-counting task in a so-called inconsistent grouping condition, or (c) a memory retention task under dual task conditions. In condition (a), a constant response-to-stimulus interval (RSI) was used, whereas in condition (b) either a short or a long RSI was used. Importantly, “(...) the probability of a long RSI was matched to the probability of a high (target) tone in the tone-counting task” (Stadler, 1995, p. 676), leading to an inconsistent grouping of the sequence elements. In this condition the temporal organization of input sequences was disrupted, but attentional capacity remained unaffected. In contrast, in condition (c) attentional demands were increased, but temporal organization of input pattern was unaffected. Performance decrease in the typical tone-counting dual-task situation (a) resembled the impairments in the inconsistent grouping condition (b), but not the impairments in the condition including the memory retention task as a second task (c). These data indicate that the temporal organization of serial transitions seems to be the critical mechanism during sequence learning (see Dominey, 1998, for a computational model simulating the temporal organization process).

Another topic of current interest is whether sequence learning is based on the representation of stimulus features or on the representation of responses (cf. Clegg, DiGirolamo & Keele, 1998). One study (Mayr, 1996) provides evidence for a stimulus-

based account. The author presented different objects at different positions and subjects had to discriminate the objects by button press in each trial. The succession of objects as well as of positions followed a structured sequence (Experiment 1). Both sequences were uncorrelated. In contrast to the object sequence, the spatial sequence was independent of response requirements. Interestingly, both sequences were learned simultaneously, as indexed by a significant increase in reaction times when a random sequence was introduced. These data indicate that knowledge about response-unrelated sequential structures can be acquired. Furthermore, Mayr (1996) could show that simultaneous learning both sequences is as effective as learning both sequences separately (Experiment 2), thus providing evidence for independent spatial and non-spatial sequence learning modules. In extending these results, Koch and Hoffmann (2000) demonstrated that both stimulus and response sequences are learned. Most importantly, the learning effect was noticeably more pronounced for spatial than for symbolic regularities. The authors systematically manipulated stimulus and response sequences. Sequences were presented in four conditions: (1) spatial stimulus and spatial response sequence, (2) spatial stimulus and object response sequence, (3) object stimulus and object response sequence, and finally, (4) object stimulus and object response sequence.¹⁰ Koch and Hoffmann (2000) argue that the cognitive mechanisms responsible for sequence learning are best suited for extracting spatial rather than symbolic regularities, irrespective whether the structured sequence refers to stimuli or responses.

To summarize, the behavioral literature reviewed so far suggests that subjects are able to learn structured sequences. Furthermore, sequence learning seems to be (partially) independent from attentional resources and from the selection of motor responses.

¹⁰Digits were used as objects in this study. Subjects in both object response conditions reported digits verbally.

3.5.2 Neural Mechanisms

In recent years, functional neuroimaging provided useful insights into the functional neuroanatomy of sequence learning. Results from studies using the serial reaction time task and other sequence learning tasks will be reviewed in the following.

Results from the Serial Reaction Time Task In their PET study, Hazeltine, Grafton, and Ivry (1997) used both a spatial and a color version of the SRT task. In the spatial condition, the critical sequence was defined by means of the temporal succession of locations, whereas in the color condition a fixed sequence of six colors was repeated during the sequence blocks. Moreover, these two tasks were presented under single-task as well as under dual-task conditions. Subjects were aware of the sequence in the single-task (explicit learning), but not in the dual-task condition (implicit learning). In the latter condition, subjects were required to perform a tone-counting task concurrently. In brief, in the condition with attentional interference activation of the supplementary motor area (SMA) and the basal ganglia was observed for both the color and spatial version of the task. In contrast, under single task conditions ventrolateral (color version) and dorsolateral (spatial version) prefrontal regions were activated. Based on these results, the authors argue that implicit and explicit sequence learning relies on different neural systems. Several recent fMRI studies also contrasted implicit and explicit sequence learning (Aizenstein et al., 2004; Schendan et al., 2003; Willingham et al., 2002). For instance, Schendan et al. (2003) found a network of medial temporal lobe (hippocampus), basal ganglia (putamen, caudate nucleus), and dorsolateral prefrontal regions during both implicit and explicit sequence learning. The exact activation foci differed only slightly between both learning conditions. Hippocampal activation was more pronounced at the beginning of learning, whereas lateral prefrontal cortex activity increased during the time course of learning. In line with the relational memory account (Cohen & Eichenbaum, 1993; Eichenbaum, 2000), Schendan and colleagues argue that the hippocampus is involved in the acquisition of higher-order associations, i.e. relations among temporally discontinuous events, most pronounced at the beginning of learning. In accordance with these results, Willingham et al. (2002) observed highly

overlapping brain regions during implicit and explicit sequence learning (basal ganglia, lateral prefrontal, and parietal brain regions). Moreover, dorsolateral prefrontal activity increased as a function of learning in the explicit condition. Learning-related increase of lateral PFC activity during explicit sequence learning has also been reported in two recent fMRI studies (Aizenstein et al., 2004; Bischoff-Grethe et al., 2004).

Finally, sequence learning within the SRT framework has been studied in amnesic patients (Curran, 1997; Reber & Squire, 1994, 1998). In brief, these studies revealed preserved sequence learning in patients in the implicit version of the SRT task, indicating that learning occurs independently of (hippocampal) declarative memory operations (cf. Squire & Zola, 1996). However, declarative knowledge of learned sequences (Reber & Squire, 1994), sequence learning in an explicit version of the SRT task (Reber & Squire, 1998), and higher-order associative learning (Curran, 1997, see Curran and Keele, 1993 and Section 3.5.1) was impaired in amnesics, suggesting that the hippocampus is presumably involved in sequence learning, whenever relational binding of stimuli is required (cf. Schendan et al., 2003; see Chapter 2).

Results from Other Sequence Learning Tasks Beyond the SRT task, different learning tasks have been used to investigate the brain network involved during learning structured sequences (Bor, Duncan, Wiseman & Owen, 2003; Huettel, Mack & McCarthy, 2002; Koechlin, Danek, Burnod & Grafman, 2002; Rose, Haider, Weiller & Büchel, 2002). For instance, in the Bor et al. (2003) study, subjects were required to encode and recall a sequence of positions in each trial. In one condition, the sequences followed structured visuo-spatial configurations (e.g. square), whereas in a control condition, sequences were unstructured. In contrast to the unstructured sequence condition, performance increase in the structured sequence condition was paralleled by an increase of lateral prefrontal activity. Based on these results, the authors propose that the lateral PFC is involved in integrating regular input pattern into higher level chunks. Further evidence for the role of the lateral PFC in sequence learning has been provided by a recent fMRI study (Huettel et al., 2002), showing activation of lateral prefrontal (and basal ganglia) brain regions whenever a regular

sequence of visual stimuli (and responses) was violated. In another study, Koehlin et al. (2002) contrasted brain activation during cognitive and motor sequence learning. Similar to studies using the SRT task, subjects had to respond to specific visual color stimuli by pressing corresponding buttons in the motor sequence condition. In the cognitive sequence condition, subjects performed a sequence of cognitive tasks, depending on the presence of specific color cues (e.g. green stimulus: ‘Is the letter the same as the previous one?’, blue stimulus: ‘Is the letter a T?’). In both conditions, either structured or random sequences were presented. The striatum was associated with motor sequence learning, whereas the anterior medial PFC was involved in learning cognitive sequences. Finally, in the study by Rose and colleagues (2002), subjects were required to learn stimulus-response sequences explicitly. Unknown to the subjects, a hidden structure was embedded in the task (the so-called number reduction task). Participants learned this hidden rule implicitly. The authors could dissociate medial temporal and striatal brain regions during explicit and implicit learning, respectively. A learning-related activation increase in the striatum (putamen) could be observed during explicit learning, whereas implicit learning the hidden rule was associated with perirhinal cortex activation.

Learning behavioral sequences has also been examined in macaques. A recent study revealed intriguing findings. Fujii and Graybiel (2003) recorded neuronal activity in the prefrontal cortex of macaques during a sequential saccade task. The authors observed a phasic neuronal signal, time-locked to the end of learned oculomotor sequences. This signal was not related to other task-relevant responses (e.g. visual and motor) or the duration of the sequence. Fujii and Graybiel (2003) argue that these phasic neuronal signals “(…) could represent a neural correlate of action-sequence chunking, a segmentation of complex behavior into action subsets (…)” (Fujii & Graybiel, 2003, p. 1249).

To summarize the reviewed literature of sequence learning, including research within the framework of the serial reaction time task as well as other sequence learning tasks, sequence learning seems to be mainly subserved by lateral prefrontal and basal ganglia structures. Evidence for medial temporal lobe involvement during sequence learning is less consistent.

3.5.3 Summary

Taken as a whole, the critical sequences in a SRT task are defined by means of a finite state grammar, similar to grammatical strings in artificial grammar learning. During sequence learning, the repeated elements of the critical regular sequence comprise the regularities across learning episodes. Similar to artificial grammar learning, the elements of the sequence are characterized by a contingent relationship. In agreement with all reviewed learning tasks, sequence learning requires the exposure of multiple repetitions of the regular sequence. On a neural level, the lateral PFC as well as basal ganglia structures play an important role during sequence learning.

Chapter 4

Neuroanatomical Basis: The Medial Temporal Lobe and the Prefrontal Cortex

Contents

4.1	Medial Temporal Lobe	67
4.1.1	Main Structures	67
4.1.2	Main Pathways	68
4.2	Prefrontal Cortex	69
4.2.1	Main Structures	69
4.2.2	Main Pathways	70
4.3	Medial Temporal-prefrontal Interconnections	72

In the scope of this thesis, two brain structures are of main interest, the hippocampus as part of the medial temporal lobe (MTL) and the prefrontal cortex (PFC).¹ In the following chapter, a brief overview of the neuroanatomy of the medial temporal lobe

¹Is is noteworthy that additional brain structures, e.g. the basal ganglia, are important with respect to regularity learning. However, due to space limitations, the neuroanatomy of the basal ganglia will not be described in this chapter. The interested reader can find a detailed neuroanatomical description of the basal ganglia in Ohye, Kimura, and McKenzie (1996).

(4.1) and the prefrontal cortex (4.2), and the main projections of both structures will be given. Finally, the main pathways between the medial temporal lobe and the prefrontal cortex will be described (4.3).

4.1 Medial Temporal Lobe

4.1.1 Main Structures

The hippocampus, the amygdala, the entorhinal cortex, the perirhinal cortex, the parahippocampal cortex, and the fusiform gyrus constitute the main components of the human medial temporal lobe (see Amaral, 1999; Insausti, Herrero & Witter, 1997; Suzuki & Amaral, 2003; Swanson & Petrovich, 1998, for reviews and detailed cytoarchitectonic analysis of these MTL structures). The amygdala or amygdalar complex refers to a set of several cell groups and makes up the most anterior portion of the MTL superior to the anterior hippocampus (Figure 4.1e). The amygdalar complex consists of the basolateral complex, the adjacent olfactory cortex, the medial amygdalar nucleus, the central amygdalar nucleus, and the bed nuclei of the stria terminalis (Swanson & Petrovich, 1998).

The hippocampus proper comprises the hippocampal cell fields and the dentate gyrus (see Figure 4.1). The term hippocampal formation (Amaral & Insausti, 1990) refers to the hippocampus proper, the subicular complex (the subiculum, the pre-subiculum, and the parasubiculum), and the entorhinal cortex.² In the long-axis of the hippocampus, the parahippocampal gyrus makes up the inferior boundary of the hippocampus (Amaral, 1999). The collateral sulcus makes up the border between the parahippocampal gyrus and the fusiform gyrus. The entorhinal and perirhinal cortices constitute the rostral part of the parahippocampal gyrus whereas the parahippocampal cortex makes up the caudal portion of the parahippocampal gyrus

²Some authors introduced the term hippocampal region which comprises the cell fields of the hippocampus proper, the dentate gyrus, and the subicular complex (Squire & Zola-Morgan, 1991). According to Squire and Zola-Morgan (1991), the term hippocampal system normally refers to the hippocampal region and the surrounding entorhinal, perirhinal, and parahippocampal cortices.

(Amaral, 1999). Furthermore, a medial-lateral distinction within the anterior inferior MTL can be made. The entorhinal cortex is located on the medial aspect of the anterior inferior MTL, whereas the perirhinal cortex constitutes the lateral portion of the anterior inferior MTL (Figure 4.1b and 4.1c).

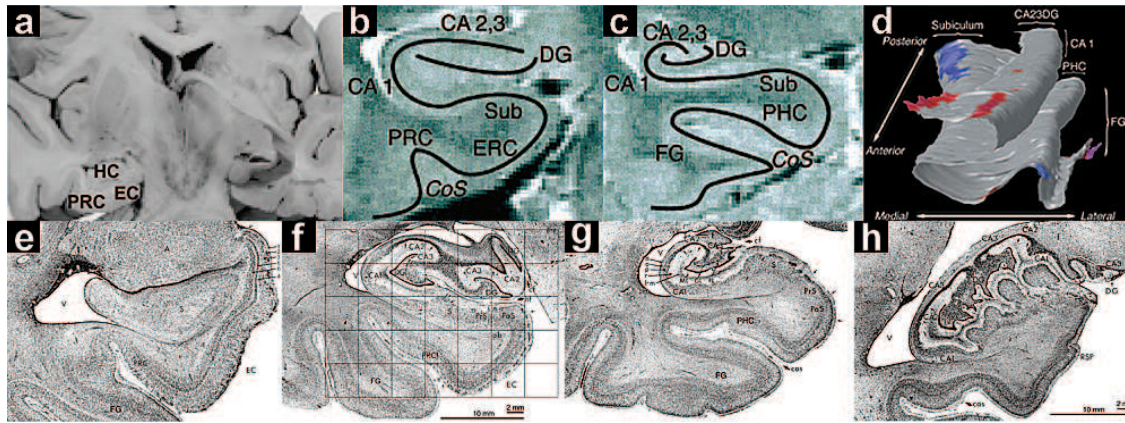


Figure 4.1. Neuroanatomy of the Medial Temporal Lobe (MTL). (a) Coronal MR section through the anterior hippocampus (HC), the perirhinal cortex (PRC, inferior-lateral) and the entorhinal cortex (EC, inferior-medial). (b) Coronal section through the anterior MTL, comprising hippocampal cell fields CA1, CA2, and CA3 (CA, cornu ammonis), the dentate gyrus (DG), the subiculum (Sub), the PRC, the ERC, and the collateral sulcus (CoS). (c) Coronal section through the posterior MTL, showing the replacement of the PRC and ERC with the parahippocampal cortex (PHC); in addition the fusiform gyrus (FG) is shown. (d) Reconstruction of the MTL, showing the folded surface of the MTL. (e-h) Nissl-stained coronal sections of the human MTL along the rostral-caudal axis (A, amygdala). [(a) Adapted from Brewer and Moghekar (2002). Copyright (2002) Elsevier Science Ltd (Trends in Cognitive Sciences). (b-c) Adapted from Zeineh et al. (2001). Copyright (2001) Wiley Liss, Inc. (The Anatomical Record (New Anat.)). (d) Adapted from Zeineh et al. (2003). Copyright (2003) American Association for the Advancement of Science (Science). (e-h) Adapted from Amaral (1999). Copyright (1999) Wiley Liss, Inc. (Hippocampus).]

4.1.2 Main Pathways

Both the parahippocampal and perirhinal cortices receive inputs from several unimodal and polymodal areas (Suzuki & Amaral, 1994a). The main input to the perirhinal cortex comes from visual areas (Murray & Bussey, 1999). In addition, the perirhinal cortex receives projections from the insula, the superior temporal gyrus, the cingulate cortex, and the orbitofrontal cortex (Murray & Richmond, 2001), and has interconnections with the amygdala (Murray & Richmond, 2001; Suzuki, 1996) and the medial dorsal thalamic nucleus (Brown & Aggleton, 2001). The parahip-

hippocampal cortex, in contrast, receives inputs from caudal visual areas, the retrosplenial cortex, the inferior parietal lobule, several prefrontal areas (Suzuki & Amaral, 1994a, cf. Section 4.3), and is interconnected with the amygdaloid complex (Stefanacci, Suzuki & Amaral, 1996). Both cortices provide the main input to the entorhinal cortex (Suzuki & Amaral, 1994b) which projects to several structures, e.g. frontal, temporal, parietal, occipital, insular, cingulate, and retrosplenial cortices (Insausti, Herrero & Witter, 1997). The entorhinal cortex constitutes the main gateway of input to the hippocampus proper (Suzuki & Amaral, 1994b, see Figure 4.3b). It projects to the dentate gyrus and the CA3 region via the perforant path. Moreover, the CA3 region receives input from the dentate gyrus via mossy fibers. Via the Schaffer collateral pathway the CA3 field in turn projects to the CA1 region which projects back to the entorhinal cortex via the subiculum (see Lisman, 1999, for a detailed description of the hippocampal circuitry). In addition, the hippocampus projects to the anterior thalamic nuclei and the mammillary bodies via the fornix (Aggleton & Brown, 1999, see Lavenex and Amaral, 2000, for a detailed description of medial temporal-neocortical projections).

4.2 Prefrontal Cortex

4.2.1 Main Structures

The human prefrontal cortex (PFC) can be subdivided into the following substructures (approximate Brodmann area [BA] given in parentheses; see Figure 4.2 and 4.3): the dorsolateral PFC (9/46), the ventrolateral PFC (44/45/47), the anterior/frontopolar PFC (10), the medial PFC (25/32), and the orbitofrontal PFC (11/12/14).

The dorsolateral PFC comprises roughly Brodman areas 9 and 46. Area 9 is located at the middle portion of the superior frontal gyrus (SFG), extending into the middle frontal gyrus (MFG; Rajkowska & Goldman-Rakic, 1995a,b). Area 46 covers the central portion of the MFG. Based on cytoarchitectonic and myeloarchitectonic studies, Petrides and Pandya (1994, 1999) subdivided area 9 into area 9 and area

9/46. Area 9/46 occupies the MFG and is bordered ventrally by areas 44, 45, and 47, dorsally by area 9, caudally by area 8,³ and rostrally by areas 46 and 10. Furthermore, area 9/46 was divided into a ventral (9/46v) and dorsal (9/46d) portion (Petrides & Pandya, 1994, 1999). Area 9 lies on the superior frontal gyrus. Areas 44, 45, and 47 make up the ventrolateral part of the PFC. Area 44 is located on the opercular part of the inferior frontal gyrus (IFG) and is bordered caudally by the ascending branch of the Sylvian fissure. Area 45 is located on the triangular part of the IFG (cf. Amunts et al., 1999; Goldman-Rakic, 1987). In the classification of Petrides and Pandya (1994), area 45 is subdivided into an anterior (45A) and posterior part (45B). Finally, area 47 is equivalent to the orbital part of the IFG. Area 10 at the frontal pole is bordered laterally by areas 9, 9/46, 46, and 47 and medially by areas 9, 32, and 14 (Petrides & Pandya, 1994) and comprises anterior parts of the SFG, the frontopolar gyrus, and the frontomarginal gyrus. Areas 24, 25, and 32 comprise the medial PFC. Area 24 occupies the anterior cingulate gyrus, and is surrounded by area 32 which lies on the paralimbic gyrus. Area 25 is located on the subcallosal gyrus (Petrides & Pandya, 1994). Finally, areas 11 and 12 can be subsumed to the orbitofrontal PFC. Petrides and Pandya (1994) designated areas 11 and 12 as area 14 which lies on the gyrus rectus and the orbital gyrus.

4.2.2 Main Pathways

Following the excellent review of Petrides and Pandya (2002), in this section the main projections of the prefrontal cortex will be briefly described. First, it should be emphasized that almost all PFC subregions have bidirectional projections to all other parts of the PFC (Simons & Spiers, 2003). The frontal lobe is linked with the parietal cortex by the superior longitudinal fasciculus (SLF). Petrides and Pandya (2002) subdivided the SLF in three different projection pathways. In brief, (1) the superior parietal lobe (SPL) is reciprocally interconnected with the (pre)frontal areas

³For the sake of completeness, area 8 lies on the posterior parts of the SFG and MFG and can further be subdivided (cf. Petrides & Pandya, 1994, 1999) into area 8A (MFG; including a ventral [8Av] and a dorsal [8Ad] portion) and 8B (SFG).

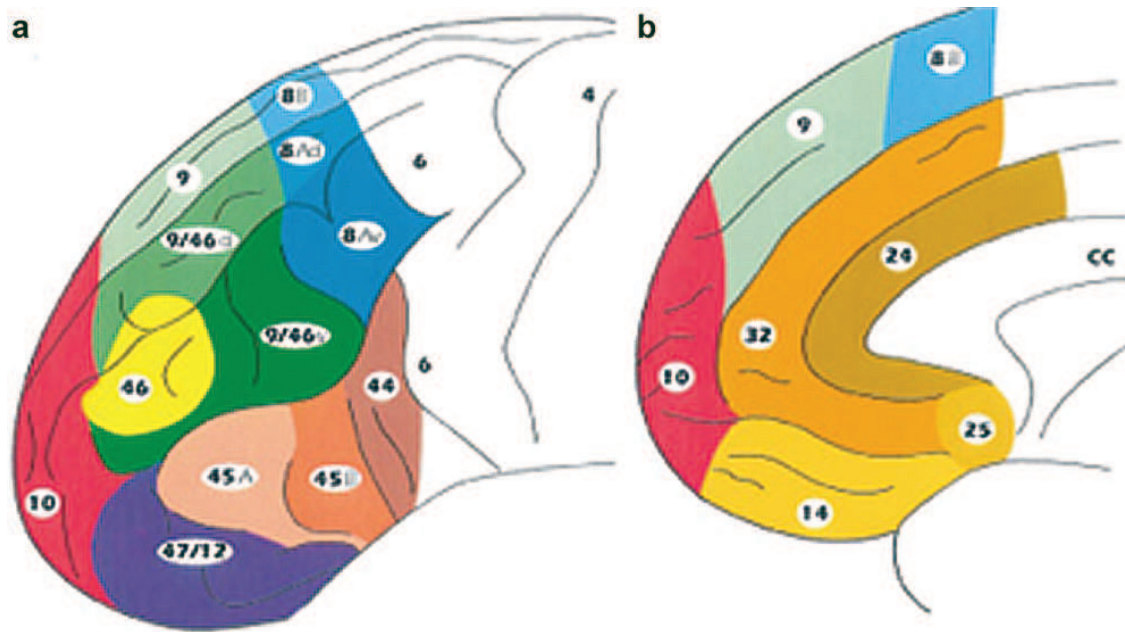


Figure 4.2. Cytoarchitectonic Maps of the Prefrontal Cortex (PFC) According to the Notation of Petrides and Pandya (1999). (a) Lateral and (b) medial view the PFC. [Adapted from Petrides and Pandya (1999). Copyright (1999) European Neuroscience Association (European Journal of Neuroscience).]

6, 8, and the cingulate sulcus, the premotor area, and the supplementary motor area (SMA). Furthermore, (2) the inferior parietal lobule has interconnections with the (pre)frontal areas 6, 8, and 46. Finally, (3) the supramarginal gyrus projects to the (pre)frontal regions 6, 44, and 9/46 (Petrides & Pandya, 2002).

Furthermore, the (a) arcuate fasciculus, the (b) extreme capsule, and the (c) uncinate fasciculus link the superior temporal cortex with prefrontal areas. More precisely, these fiber bundles originate in the (a) caudal, (b) middle, and (c) rostral portion of the superior temporal gyrus (STG) and terminate mainly in (a) prefrontal area 8, (b) areas 8, 9, 45, 46, 10, and (c) orbitofrontal areas 11,12 and medial areas 25 and 32.

Finally, two main pathways interconnect the occipital cortex with the prefrontal cortex: First, dorsal and medial regions of the occipital lobe project directly to prefrontal regions 8 and 9/46 via the occipitofrontal fasciculus. Second, the orbitofrontal and medial PFC receive projections indirectly from occipital BA 19 via two fiber bundles, the inferior longitudinal fasciculus which interconnects BA 19 with the inferotemporal cortex, and the uncinate fasciculus, originating in the most anterior

part of the inferotemporal cortex which projects to the orbitofrontal and medial PFC (see above).

4.3 Medial Temporal-prefrontal Interconnections

The medial temporal lobe and the prefrontal cortex are heavily interconnected (Simons & Spiers, 2003, cf. Figure 4.3b).⁴ The medial and orbitofrontal PFC, especially areas 10, 11, and 13 receive inputs from anterior parts of the parahippocampal region via the uncinate fasciculus (Petrides & Pandya, 2002). In contrast, a caudal pathway, originating in the caudal portion of the parahippocampal region projects to dorsolateral PFC regions 9, 46, and 9/46 via the extreme capsule. These pathways constitute the ventral limbic pathway (Petrides & Pandya, 2002). In contrast, the dorsal limbic pathway comprises parts of the cingulum bundle, originating in both anterior (BA 24) and posterior (BA 23) cingulate cortex and the retrosplenial cortex (BA 30). This fiber bundle mainly projects to dorsolateral PFC areas 9, 46, and 9/46, area 32 as part of the medial PFC, as well as parts of orbitofrontal area 11 (Petrides & Pandya, 2002).⁵ More precisely, the hippocampal cell field CA1 and the subiculum project mainly to medial and orbitofrontal aspects of the PFC (Thierry, Gioanni, Dégénétais & Glowinski, 2000). In contrast, cell fields CA2 and CA3 and the dentate gyrus do not innervate the PFC (Laroche, Davis & Jay, 2000). Furthermore, the entorhinal cortex receives inputs mainly from the orbitofrontal PFC (Lavenex & Amaral, 2000) and projects to dorsolateral and orbitofrontal PFC areas (Insausti, Herrero & Witter, 1997). Finally, the perirhinal cortex has efferent projections to medial and orbitofrontal PFC regions (mainly areas 11 and 12) and is innervated by prefrontal regions 45, 46, 11, 12, and 13, whereas the parahippocam-

⁴See particularly the Special Issue of *Hippocampus*, 2000, “The Nature of Hippocampal-Cortical Interaction: Theoretical and Experimental Perspectives”, *Hippocampus*, 10, pp. 351-499.

⁵Although the posterior cingulate cortex and the retrosplenial cortex are not part of the medial temporal lobe, both areas are heavily interconnected with several regions of the MTL (see Section 4.1). Accordingly, both structures are proposed to pertain to the extended hippocampal diencephalic system (Aggleton & Brown, 1999).

pal cortex has reciprocal projections with the dorsolateral (9, 46), ventrolateral (45), and orbitofrontal (11, 12) PFC (Lavenex, Suzuki & Amaral, 2002).

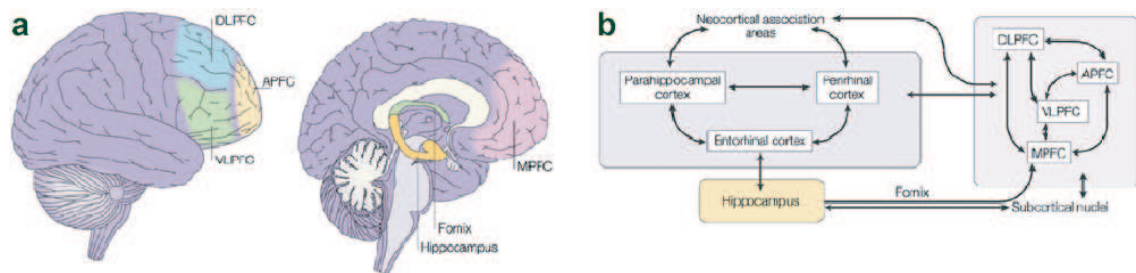


Figure 4.3. Medial Temporal-prefrontal Interconnections. (a) Lateral and medial view of the human brain. (b) Main projections within and between the medial temporal lobe and the prefrontal cortex. See text for further details. [Adapted from Simons and Spiers (2003). Copyright (2003) Macmillan Magazines Ltd (Nature Reviews Neuroscience).]

Part III

Methods

Chapter 5

Functional Neuroimaging

Contents

5.1	Principles of Magnetic Resonance Imaging	76
5.1.1	Nuclear Spin Systems in the Magnet	76
5.1.2	Excitation of the Spin System	77
5.1.3	Magnetic Resonance Imaging	80
5.1.4	Imaging Parameters and Sequences	81
5.2	The BOLD Signal	83
5.2.1	Physical Basis	83
5.2.2	Physiological Basis	84
5.2.3	Main Characteristics	85
5.3	Analysis of fMRI Time Series	88
5.3.1	Spatial Preprocessing	88
5.3.2	Modeling Hemodynamic Responses	90
5.3.3	Statistical Inference	93
5.4	Designing	94
5.4.1	Event-related Functional Imaging	95
5.4.2	Design Efficiency	97

In this chapter, the basics of functional neuroimaging will be briefly reviewed. More precisely, in Section 5.1 the physical principles of magnetic resonance imaging will be described. Section 5.2 deals mainly with the physical and physiological basis of

the BOLD response. The subject of Section 5.3 is the analysis of fMRI time series, including spatial preprocessing, modeling hemodynamic responses, and statistical inference. Finally, in Section 5.4 important aspects of event-related functional imaging and design efficiency will be outlined.¹

5.1 Principles of Magnetic Resonance Imaging

This section comprises a brief introduction to MR physics and is mainly based on the excellent reviews by Cohen (1996), Mandeville and Rosen (2002), and especially Liang and Lauterbur (2000). It gives a brief overview of the nuclear spin system (5.1.1) and the basics of the excitation of the spin system (5.1.2), the principles of magnetic resonance imaging (5.1.3), and the main imaging parameters and imaging sequences (5.1.4).

5.1.1 Nuclear Spin Systems in the Magnet

A biological sample is composed of atoms and molecules which in turn are assemblies of protons, neutrons, and electrons. The electrons orbit a nucleus which includes protons and neutrons. Nuclei with odd atomic weights and/or odd atomic numbers exhibit the fundamental property of an angular momentum $\vec{J}$, termed spin (cf. Liang & Lauterbur, 2000). Due to this angular velocity the nucleus (which has electrical charges) creates a magnetic field around it (Figure 5.1a). The nuclear magnetic dipole moment or magnetic moment $\vec{\mu}$ is related to the angular momentum by

$$\vec{\mu} = \gamma \vec{J} \quad (5.1)$$

where γ is the gyromagnetic ratio. The total or bulk magnetization of a spin system is given by

$$\vec{M} = \sum_{n=1}^{N_s} \vec{\mu}_n \quad (5.2)$$

¹Please note that the behavioral methods of this thesis will be described in more detail in Chapter 6.

where $\vec{\mu}_n$ represents the magnetic moment of the n th nuclear spin and N_s denotes the total number of spins in the measured object. When exposing the object to a magnetic field with strength B_0 , the protons can have two possible states, a low or high energy state. Accordingly, the spin vectors line up in two possible orientations, with or against the magnetic field (Figure 5.1b). The net polarization of the sample depends upon the small orientation-dependent energy splitting of a magnetic dipole moment (cf. Liang & Lauterbur, 2000). In a 1.5 Tesla (T) magnet, the net polarization is about 10^{-5} , i.e. for every 1 million protons there will be about 10 more in the lower than in the higher energy state. Moreover, the nuclear magnetic moment vectors $\vec{\mu}_n$ exhibit the important property of precession around the B_0 field. This angular frequency, the Larmor frequency is given by the Larmor equation

$$\omega_0 = \gamma B_0 \quad (5.3)$$

describing the natural frequency of a spin system (cf. Liang & Lauterbur, 2000). As apparent from Equation 5.3, the Larmor frequency ω_0 is a function of the gyro-magnetic ratio γ and the field strength B_0 . At a field strength of 1.5 T, the Larmor frequency constitutes roughly 63.8 MHz. Each spinning vector can be considered as a magnetic dipole precessing about the applied magnetic field. Due to the net polarization, the alignment of the spins results in a longitudinal magnetization M_0 , pointing along the direction of the external magnetic field B_0 (cf. Liang & Lauterbur, 2000).

5.1.2 Excitation of the Spin System

The acquisition of MRI data is based on the emission of radio frequency (RF) pulses in the presence of the B_0 field (normally defined as z -axis).

RF Pulse The RF pulse generates an oscillating $B_1(t)$ field in the x-y-direction perpendicular to the B_0 field. It is important to note that the spin system can only be excited at certain frequencies. More specifically, the RF pulse must be applied at the Larmor frequency of the precessing spins, i.e. the magnetic resonance phenomenon critically depends on an exact match between the excitation frequency and the

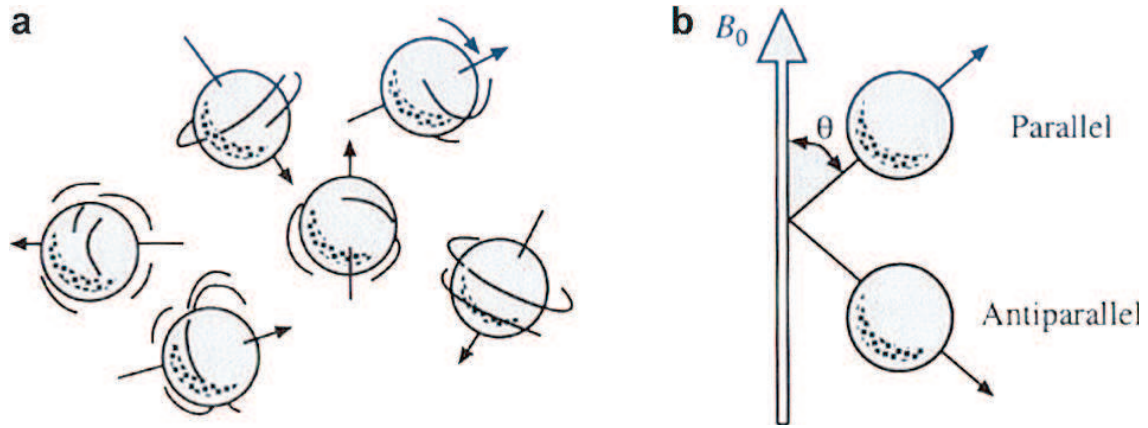


Figure 5.1. Nuclear Spin Systems in the Magnet. (a) The nuclear spin phenomenon. (b) Spins align parallel or anti-parallel to an external magnetic field. [Adapted from Liang and Lauterbur (2000). Copyright (2000) Institute of Electrical and Electronics Engineers, Inc..]

proton spin frequency of the excited spin system. After excitation, the spin system will return to its thermal equilibrium according to the laws of thermodynamics by precession of $\vec{M}$ about the B_0 field (free precession; cf. Liang & Lauterbur, 2000). This phenomenon is characterized by two subprocesses. The longitudinal relaxation describes the recovery of the longitudinal magnetization M_z , whereas the destruction of the transverse magnetization M_{xy} is denoted transverse relaxation (cf. Liang & Lauterbur, 2000). As apparent from Figure 5.2, both processes follow an exponential function.

Longitudinal Relaxation The spin system exchanges energy with the lattice. The time constant which describes how M_z returns to its equilibrium value is called the longitudinal relaxation or spin-lattice relaxation time T_1 and is given by

$$M_z = M_z^0(1 - e^{-t/T_1}) \quad (5.4)$$

where M_z^0 denotes the longitudinal magnetization at thermal equilibrium.

Transverse Relaxation After excitation, the spins precess at slightly different rates. This leads to a dephasing of the spins and by this to the decay of the transverse magnetization M_{xy} . The time constant which describes the return to equilibrium of the transverse magnetization M_{xy} is called the transverse relaxation or spin-spin

relaxation time T_2 and is given by

$$M_{xy} = M_{xy}(0)e^{-t/T_2} \quad (5.5)$$

Both time parameters are not defined as the times when M_z is fully recovered and M_{xy} has reached maximal equilibrium, respectively, but rather at the following cut-off points

$$M_z(T_1) \approx 63\%M_z^0 \quad (5.6)$$

and

$$M_{xy}(T_2) \approx 37\%M_{xy} \quad (5.7)$$

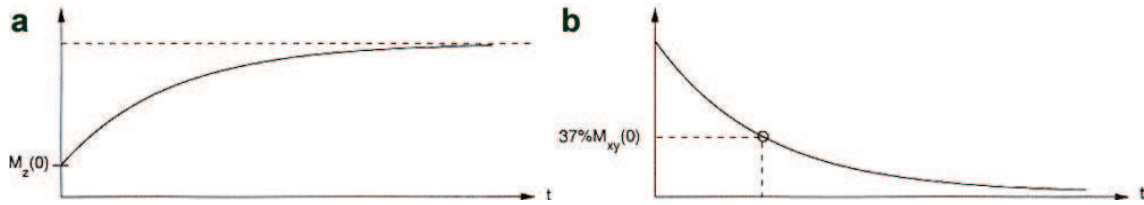


Figure 5.2. T_1 and T_2 Relaxation Curves. (a) The spin-lattice relaxation time T_1 describes the recovery of the longitudinal magnetization, whereas (b) the spin-spin relaxation time T_2 describes the decay of the transverse magnetization after an RF perturbation. Both processes follow an exponential function. [Adapted from Liang and Lauterbur (2000). Copyright (2000) Institute of Electrical and Electronics Engineers, Inc..]

In addition to the longitudinal relaxation T_2 based on spin-spin dephasing, another component contributes to the destruction of the transverse magnetization M_{xy} , namely the T_2' component which results from factors such as magnetic field inhomogeneity and magnetic susceptibility variations (cf. Liang & Lauterbur, 2000). The combined time constant is called T_2^* and is given by

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'} \quad (5.8)$$

The time constant T_2^* constitutes the critical parameter with respect to fMRI (see Section 5.2.1 for more details).

In summary, after a spin system has been disturbed from its thermal equilibrium state by conducting a RF pulse, the time-dependent behavior of $\vec{M}$ can be described

by the Bloch equation²

$$\frac{d\vec{M}}{dt} = \gamma \vec{M} \times \vec{B} - \frac{M_x \vec{i} + M_y \vec{j}}{T_2} - \frac{(M_z - M_z^0) \vec{k}}{T_1} \quad (5.9)$$

5.1.3 Magnetic Resonance Imaging

The purpose of MRI is the acquisition of 3D images. Spatial encoding of MR signals requires slice-selection, phase encoding, and frequency encoding gradients (cf. Liang & Lauterbur, 2000).

Slice Selection Gradient As apparent from Equation 5.3, the Larmor frequency is proportional to the magnetic field strength B_0 . Moreover, excitation is only possible using the specific Larmor frequency of the spin system. To selectively excite the slice of interest, the RF pulse is transmitted at a narrow band of frequencies in the presence of a z -gradient during the excitation period, the slice-selection gradient which makes the Larmor frequency position-dependent.

Phase Encoding Gradient After slice-selection, the introduction of an y -gradient results in different phases of the spins in y -direction. These phase differences depend on the duration and strength of the y -gradient and on the position of the spin. By this, each row within an excited slice can be identified by its phase.

Frequency Encoding Gradient The final step in spatial encoding includes the application of an x -gradient used for frequency encoding along the x -axis. Frequency encoding makes the oscillation frequency ω of the spins linearly dependent on its position in the x -direction.

k-Space Images are yielded by the two-dimensional Fourier transform of the signal from a selected slice, encoded by phase and frequency encoding gradients. The two-dimensional Fourier transform yields an image of the location of the MR signal. The space in which the signal is measured at discrete coordinates (k_x, k_y) is denoted

²Felix Bloch was awarded the Nobel Prize in physics in 1952 for discovering the magnetic resonance phenomenon.

k -space. It is the mathematical space in which the Fourier transform of a spatial function is represented.

5.1.4 Imaging Parameters and Sequences

Several critical parameters affect the acquisition of MR images.

T_R , T_E , Flip Angle The time between two successive RF pulses is called repetition time (T_R) and controls the effect of T_1 relaxation, whereas the time between excitation and readout of the MR signal is termed echo time (T_E). The flip angle is the amount of rotation of the bulk magnetization vector produced by an RF pulse. All three parameters determine the image contrast. The choice of the flip angle depends on T_R and T_1 of the tissue of interest.³ More precisely, when $T_R \gg T_1$, a flip angle of 90° entails the maximal signal (Mandeville & Rosen, 2002). However, at small $\frac{T_R}{T_1}$ ratios, the flip angle should be reduced appropriately to the Ernst angle given by

$$\cos(\alpha_E) = e^{-\frac{T_R}{T_1}} \quad (5.10)$$

Image Contrasts Three parameters influence the MR image: spin density and the time parameters T_1 and T_2 . Thus, three main different contrasts exist: T_1 -weighted contrasts are characterized mainly by a short T_E , whereas a long T_R is used for T_2 -weighted contrasts (see Figure 5.3 for examples). Finally, spin density-weighted contrasts are mainly characterized by a short T_E and a long T_R (cf. Liang & Lauterbur, 2000).

T_1 -weighted Sequences In this thesis, a Siemens SONATA MR scanner (Erlangen, Germany) operating at 1.5 T with a standard circularly polarized whole head coil was employed to acquire both T_1 -weighted structural images and T_2^* -weighted BOLD-sensitive functional images. In Experiment 1A (Chapter 7), Experiment 1B

³At a field strength of 1.5 T, T_1 and T_2 parameters [ms] for different brain tissues are as follows: white matter: $T_1 = 510$, $T_2 = 67$; gray matter: $T_1 = 760$, $T_2 = 77$; cerebrospinal fluid: $T_1 = 2650$, $T_2 = 280$ (cf. Liang & Lauterbur, 2000).

(Chapter 8), and Experiment 3 (Chapter 10), high-resolution structural images were acquired using an isometric magnetization-prepared rapid gradient-echo (3-D MP-RAGE) sequence ($T_R = 1900$ ms, $T_E = 3.93$ ms, IT = 1100 ms, flip angle = 8° , slice thickness = 1 mm, in-plane resolution = 1×1 mm, FoV = 256 mm^2 , 176 slices; Mugler & Brookeman, 1990, cf. Figure 5.3a). An additional T_1 -weighted structural sequence was used in Experiment 3 to improve the normalization procedure during preprocessing (cf. Section 5.3). This structural sequence ($T_R = 600$ ms, $T_E = 13$ ms, flip angle = 80° , slice thickness = 4 mm, interslice gap = 1 mm, in-plane resolution = 0.9×0.9 mm, FoV = 224 mm^2 , 20 axial slices parallel to AC-PC plane) was measured in-plane with respect to the functional sequence and applied during the coregistration procedure (see Chapter 10).

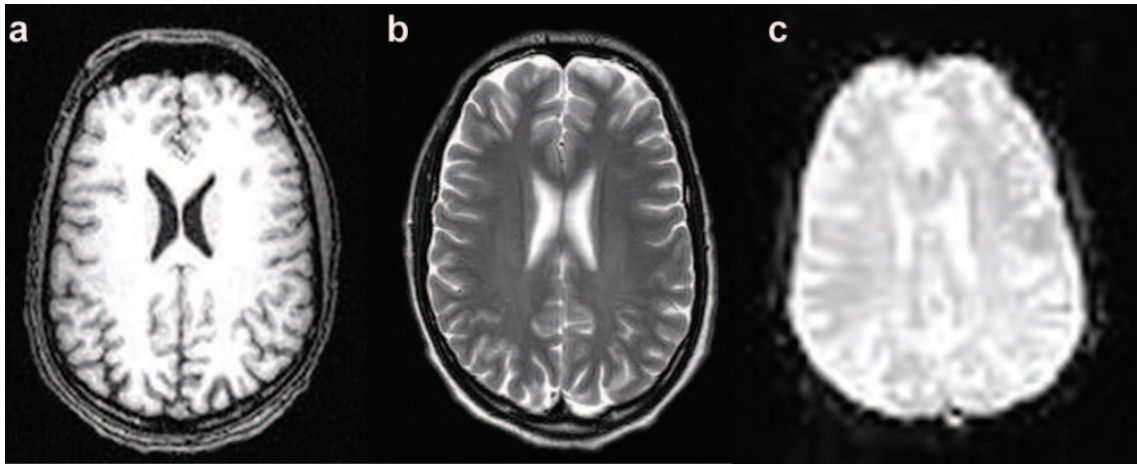


Figure 5.3. Examples of MR Sequences. Axial MR sections from an individual brain measured with (a) a T_1 -weighted, (b) a T_2 -weighted, and (c) an EPI sequence.

Echo-planar Imaging Mansfield⁴ (1977) first introduced echo-planar imaging (EPI), a fast MR scan technique which performs all spatial encoding procedures following a single RF excitation. EPI allows image acquisition in less than 100 ms. After excitation of a selected slice, a rapidly switched field gradient is applied. By this, a series of MR echoes is generated which are phase-encoded by a second field gradient (cf. Cohen, 2000; Howseman & Bowtell, 1999). Several EPI contrast variants

⁴Sir Peter Mansfield and Paul C. Lauterbur were awarded the Nobel Prize in medicine in 2003 for their discoveries concerning magnetic resonance imaging.

exist. The most common variant is the spin-echo sequence, i.e. a 90° pulse is followed by a 180° echo-forming pulse. This second pulse entails a Hahn echo during the readout period. Another variant is the gradient-echo EPI. Here the second 180° echo-forming pulse is omitted and the signal is refocussed only by the gradients (cf. Howseman & Bowtell, 1999). In contrast to spin-echo EPI which is widely applied in clinical settings, gradient-echo EPI has the advantage of a shorter T_R and therefore is commonly used in basic research (see Figure 5.3c). In Experiment 1A, 1B, and 3, T_2^* -weighted BOLD-sensitive functional images were acquired using a gradient-echo EPI pulse sequence, with the following parameters: $T_R = 1.8$ s, $T_E = 50$ ms, flip angle = 85° , slice thickness = 4 mm, interslice gap = 1 mm, in-plane resolution = 3.5×3.5 mm, FoV = 224 mm^2 , 20 axial slices parallel to AC-PC plane.

5.2 The BOLD Signal

The blood oxygen level dependent (BOLD) effect was first described by Ogawa and coworkers in 1990 (Ogawa & Lee, 1990; Ogawa, Lee, Nayak & Glynn, 1990a; Ogawa, Lee & Tank, 1990b). BOLD fMRI employs hemoglobin as an endogenous contrast agent, relying on the magnetization differences between oxy- and deoxyhemoglobin. In the following section, first the physical (5.2.1) and physiological (5.2.2) basis of the BOLD signal will be described. Second, a brief overview of the main characteristics (5.2.3) of the BOLD signal, particularly the spatial and temporal resolution will be given.

5.2.1 Physical Basis

Introducing any substance into a magnetic field alters this field to some extent. The degree of this effect is referred to as magnetic susceptibility χ (Cohen, 1996). Deoxygenated hemoglobin has a greater susceptibility effect relative to oxygenated hemoglobin. This fact is based on the observation that deoxygenated hemoglobin is more paramagnetic than oxygenated hemoglobin (Pauling & Coryell, 1936). During neuronal activation, regional cerebral blood flow (rCBF) increases and is accompanied by a relative smaller increase in oxygen extraction (Fox & Raichle, 1986; Fox,

Raichle, Mintun & Dence, 1988, see Section 5.2.2). As a result, the net deoxyhemoglobin (dHb) content of venous blood decreases and the susceptibility of the capillary contents becomes closer to that of the surrounding neural parenchyma (Cohen, 1996). According to its paramagnetic properties, the dHb concentration affects the T_2^* value by inducing a local magnetic field inhomogeneity (Howseman & Bowtell, 1999). It is assumed that the susceptibility difference is maximal for fully deoxygenated blood vessels. Ugurbil et al. (2000) formally described the influence of the magnetization susceptibility difference $\Delta\chi$ between the interior and exterior of a dHb-containing blood vessel on the T_2^* parameter:

$$\frac{1}{T_2^*} = \alpha(\Delta\chi\omega)(1 - Y)b_v \quad (5.11)$$

where α is a constant, ω the Larmor frequency as defined in Equation 5.3, Y the fraction of oxygenated blood and b_v the local cerebral blood volume in which dHb is present. As apparent from Equation 5.11, if the frequency shift (caused by the susceptibility difference) increases, the T_2^* value decreases (Figure 5.4b). In contrast, when the fraction between oxygenated and deoxygenated blood increases as during brain activation, T_2^* increases. Thus, during neuronal activity (associated with a decrease of the net dHb content) the intensity of the BOLD-sensitive MR signal increases. To summarize, the BOLD contrast originates from the local magnetic field inhomogeneity induced by paramagnetic dHb.

5.2.2 Physiological Basis

During the initial phase of brain activity, oxygen consumption is increased (Vanzetta & Grinvald, 1999; Thompson, Peterson & Freeman, 2003). During neuronal activity, the fractional change in rCBF is at least twice as large as the fractional change in oxygen metabolism (Heeger & Ress, 2002). This results in an oversupply of oxygenated blood. This increased fraction between oxy- and deoxyhemoglobin is the basis of the BOLD effect. An explanation of the neurophysiological origins of the BOLD effect is given by Heeger and Rees (2002): There is empirical evidence that neuronal activity results in increased blood flow. It is assumed that blood flow is related to glucose consumption. Furthermore, neuronal activity is associated with glucose

metabolism. More specifically, increased neuronal activity is supported by glycolysis (Fox & Raichle, 1986; Fox et al., 1988), i.e. non-oxidative glucose metabolism. In particular, astrocytes, involved in transmitter recycling rely on glycolysis during clearing glutamate from the extracellular space and the conversion of glutamate to glutamin. These processes account for the oversupply of oxygenated blood during neuronal activity (Figure 5.4a).⁵ Despite the fact that BOLD fMRI measures neuronal activity indirectly via hemodynamic correlates, in a recent seminal study (Logothetis, Pauls, Augath, Trinath & Oeltermann, 2001) combining intracortical recordings and fMRI measurements simultaneously from the macaque visual cortex, it has been shown that the BOLD response directly reflects an increase in neuronal activity. More precisely, the BOLD signal is correlated with local field potentials (LFP) which represent a weighted average of synchronized synaptic inputs of a neural population.⁶

5.2.3 Main Characteristics

Figure 5.4c shows the BOLD response. At high field strengths, the BOLD signal shows a slight decrease after neuronal activity (initial dip), possibly due to an uncoupling of blood flow and oxygen consumption (Röther et al., 2002, but see Buxton, 2001). Following the oversupply of oxy- relative to deoxyhemoglobin described in Section 5.2.2, the hemodynamic response increases and shows a pronounced peak between 6 and 10 s after stimulus onset. Furthermore, the response takes a con-

⁵According to Heeger and Rees (2002), an alternative explanation exists for the uncoupling of rCBF increase and oxygen metabolism (Buxton & Frank, 1997). Buxton and Frank (1997) suggest that the decrease of the oxygen extraction fraction during increased blood flow is due to the fact that extraction (by passive diffusion) is less efficient at high flow rates (see Magistretti and Pellerin, 1999, for a detailed description of the neurophysiological processes underlying the BOLD response and particularly the cellular mechanism of coupling synaptic activity with energy metabolism).

⁶It should be noted that several influential mathematical models of the origins of the BOLD response have been described, namely the Balloon model (Buxton, Wong & Frank, 1998), the Windkessel model (Mandeville et al., 1999), and the Volterra hemodynamic model (Friston, Mechelli, Turner & Price, 2000b; Friston, 2002a).

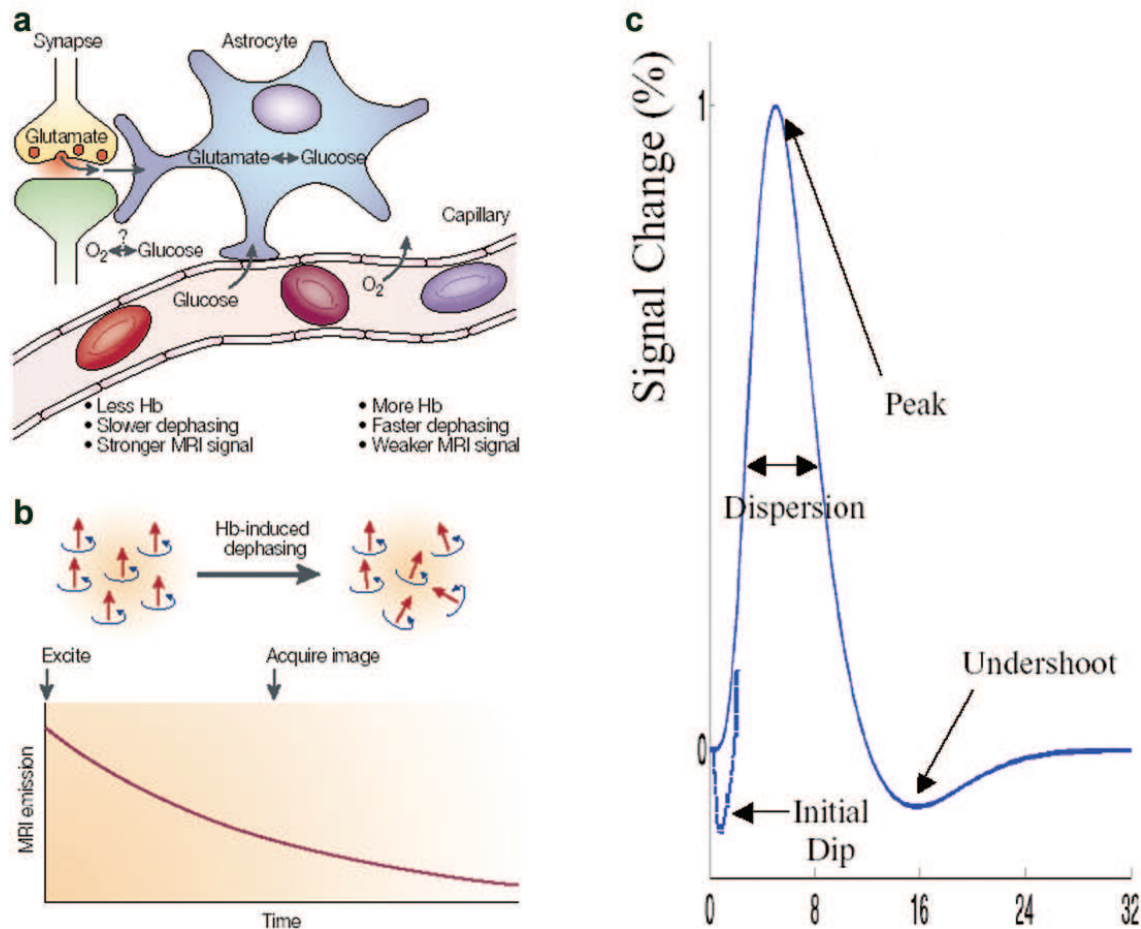


Figure 5.4. The Physiological Basis and the Main Characteristics of the BOLD Response. (a) Neurophysiological processes at the synapse. Astrocytes are involved in neurotransmitter recycling by means of non-oxidative glycolysis (uptake of glucose from blood and release of lactate). (b) Deoxyhemoglobin (dHb) entails faster dephasing of nuclear spins and by this entails a reduction of the T_2^* -weighted MR signal. (c) Main characteristics of the BOLD function. [(a-b) Adapted from Heeger and Rees (2002). Copyright (2000) Macmillan Magazines Ltd (Nature Reviews Neuroscience). (c) Adapted from Henson (in press). Copyright (2004) Elsevier.]

siderable amount of time to reach baseline at roughly 20 s. Often, a post-stimulus undershoot below baseline can be observed between 15 and 20 s after stimulation.

Spatial Resolution In most human fMRI studies at 1.5-3 T, the spatial resolution is between 2 and 5 mm in-plane with 3-10 mm slice thickness. In contrast, other non-invasive imaging techniques, e.g. positron emission tomography (PET) or single photon emission computerized tomography (SPECT) have a significantly poorer spatial resolution. In fMRI studies at higher field strengths, a spatial resolution

of roughly $150 \times 150 \mu\text{m}$ in-plane is possible. For example, at 9.4 T it is possible to map iso-orientations columns in the primary visual cortex of the cat (Kim, Duong & Kim, 2000a). However, the spatial specificity depends on the size of the local vasculature supplying the activated neuronal assembly (capillaries: $< 10 \mu\text{m}$, venules/draining veins: a few mm; Menon & Kim, 1999, see also Turner, 2002). The advantage of high magnetic fields mainly originates from two observations. First, at higher field strengths an increased contrast-to-noise ratio results in an increase of task-induced signal changes (Kim, Lee, Goodyear & Silva, 2000b). Second, at field strengths between of 1.5-3 T large vessels mostly contribute to the measured BOLD signal. In contrast, the contribution of small vessels in the generation of the BOLD signal increases at a field strength > 3 T. Hence, the spatial resolution increases substantially (Bandettini, 2000; Ugurbil, 2002), however at the cost of increased susceptibility to artifacts (see Norris, 2003, for a recent review of high field MRI).

Temporal Resolution As described in Section 5.2.3, the BOLD response peaks between 6 and 10 s after stimulus onset due to the sluggish changes in local vasculature. However, reliable relationships between stimulus properties and BOLD signal parameters exist. For example, it has been shown that stimulus duration correlates with the duration of the BOLD response and the contrast of a visual stimulus correlates with the magnitude of the BOLD signal in human primary visual cortex (Boynton, Engel, Glover & Heeger, 1996). Moreover, BOLD onset latencies correlate with behavioral parameters such as reaction times in a visuomotor task (Menon, Luknowsky & Gati, 1998). However, differences in the local vascular architecture provide constraints in comparing timing parameters between different brain areas (Miezin, Maccotta, Ollinger, Petersen & Buckner, 2000; Robson, Dorosz & Gore, 1998). Despite these limitations, several procedures for analyzing temporal properties of the BOLD signal have been developed. For instance, Henson et al. (2002) used the temporal derivative of the canonical hemodynamic response function (HRF) to estimate statistical parametric maps of HRF latency differences. These latency differences were estimated by the ratio of derivative and canonical parameter estimates. A similar approach was introduced using optimally chosen basis functions

(Liao et al., 2002).

Most recently, Belgovan, Saad, and Bandettini (2003) used a modeling approach based on a convolution of a γ -function with a function of varying onset and duration to estimate delay and width of the BOLD response in different brain regions. The authors presented words rotated 0° , 60° , or 120° during a lexical decision task and showed prolonged BOLD latencies in the inferior frontal gyrus as a function of rotation degree. In summary, these studies underscore the importance of integrating latency measures into conventional fMRI analysis.

5.3 Analysis of fMRI Time Series

The analysis of fMRI time series is the topic of this section.⁷ In more detail, this section includes several aspects of spatial preprocessing of fMRI data (5.3.1), a brief introduction to modeling hemodynamic responses in terms of the general linear model (5.3.2), and finally statistical inference (5.3.3). In this thesis, all time series analyses were performed with Statistical Parametric Mapping (SPM2; Wellcome Department of Imaging Neuroscience, London, UK; <http://www.fil.ion.ucl.ac.uk/spm/spm2.html>), implemented in Matlab6 (The Mathworks, Inc., U.S.A.).

5.3.1 Spatial Preprocessing

Spatial preprocessing usually comprises the following steps: slice timing, realignment, coregistration, normalization, and smoothing. Spatial transformations of images are an essential prerequisite for statistical analysis of fMRI time-series, e.g. move-

⁷It is noteworthy that it is also possible to model effective connectivity between brain regions involved in a particular cognitive task, so-called psychophysiological interactions (Friston et al., 1997). Several accounts have been developed for this purpose, e.g. techniques based on structural equation modeling (Büchel & Friston, 1997; Büchel, Coull & Friston, 1999), Volterra series approximations (Friston & Büchel, 2000), and Dynamic Causal Modelling, a new technique to examine effective connectivity based on a nonlinear input-state-output model within a Bayesian framework (Friston, Harrison & Penny, 2003). Due to space limitations, these accounts will not be reviewed. The interested reader will find a detailed description of connectivity analysis techniques in Friston (2002b) and especially in Frackowiak et al. (in press).

ments of the subject have to be corrected and individual brain images have to be transformed into a common standard space. In brief, registering two images together comprises two main steps: determining the warping parameters that map a source image into a target image and the transformation of the source image according to the set of warping parameters (Ashburner & Friston, 2000).

To account for the different sampling times of the slices, voxel time series have to be corrected. For example, when acquiring an EPI sequence with n slices in an ascending direction the top slice is measured $\frac{n-1}{n}T_R$ later than the bottom slice.⁸ In this thesis, the following slice timing procedure was applied to the fMRI data. Time series were (1) corrected using sinc interpolation and (2) resampled using the middle slice as a reference point.⁹

Time series are often confounded by subjects' movements during an experiment. Therefore, functional volumes have to be motion-corrected to reduce the error variance of the time series. During the realignment procedure (cf. Experiment 1A, 1B, and 3), first realignment parameters are estimated using a least squares approach and a 6-parameter rigid body spatial transformation. According to these determined parameters, all volumes are resliced to the first volume of the time-series using B-spline interpolation. Beyond the aforementioned realignment procedure, an additional unwarping algorithm (Andersson, Hutton, Ashburner, Turner & Friston, 2001) was applied to the time-series in Experiment 3 to account for movement-by-magnetization inhomogeneity interactions and by this to minimize the movement-related residual variance (see Andersson et al., 2001, for details).

After slice timing and adjustment for movement-related effects, all scans have to be transformed into a stereotactical space, normally the Talairach-space (Talairach & Tournoux, 1988) or the MNI-space (Cocosco, Kollokian, Kwan & Evans, 1997)¹⁰

⁸To be more exact, $\frac{n-1}{n}T_A$ with the acquisition time T_A given by $T_A = T_R - T_R/n$.

⁹An alternative approach was introduced by Henson et al. (Henson, Büchel, Josephs & Friston, 1999) using a Fourier basis set to account for different BOLD onset latencies.

¹⁰The MNI-space is used as reference brain in SPM2. In all imaging experiments of this thesis, MNI-coordinates have been transformed to the canonical Talairach space (see <http://www.mrc-cbu.cam.ac.uk/imaging/mnispace.html>).

to allow for between-subject and between-study comparisons. During coregistration, anatomical images (3D MP-RAGE in Experiment 1A and 1B) are coregistered to the mean functional image using a normalized mutual information cost function and a trilinear reslice interpolation method. In Experiment 3, coregistration included two steps. First, images of the 2D T_1 -weighted sequence which was measured in-plane with respect to the EPI sequence were coregistered to the mean functional image. Second, images of the 3D MP-RAGE sequence were coregistered to the images of the resliced 2D T_1 -weighted sequence.

During normalization, coregistered anatomical images (3D MP-RAGE) are then normalized (Ashburner & Friston, 1999) to the standard T_1 template based on the MNI reference brain (Cocosco et al., 1997), using a 12-parameter affine transformation along with a nonlinear transformation (cosine basis functions). Based on the determined parameters, the normalization algorithm is then applied to the functional volumes. Finally, the normalized images are resampled into 2 mm isotropic voxels and spatially smoothed with an isotropic 8-mm (Experiment 1A and 1B) and 7-mm FWHM Gaussian kernel (Experiment 3), respectively to compensate for residual between-subject variability after spatial normalization and to permit application of Gaussian random field theory (smoothing).

5.3.2 Modeling Hemodynamic Responses

The GLM is the basis for statistical analysis of fMRI time-series (see Friston et al., 1995c; Friston, in press; Henson, in press; Josephs et al., 1997; Josephs & Henson, 1999).¹¹ It provides unbiased parameter estimates and enables statistical inference by means of Gaussian random field theory (see Section 5.3.3).

¹¹Alternatively, an approach for statistical analysis of neuroimaging data based on an empirical Bayesian framework (parametrical empirical Bayes, PEB) has been developed (Friston et al., 2002a,b). In brief, the PEB procedure estimates the posterior probability distribution of the parameters given the data $p(\theta|y) \propto p(y|\theta)p(\theta)$, where $p(y|\theta)$ is the likelihood for the data given the parameters θ , whereas $p(\theta)$ denotes the Gaussian priors. These priors are parameterized via estimating hyperparameters from the data.

The General Linear Model The general linear model (GLM) at each voxel is given by (cf. Henson, in press; Josephs et al., 1997; Josephs & Henson, 1999)

$$\mathbf{Y}(t) = \mathbf{X}(t)\boldsymbol{\beta} + \boldsymbol{\epsilon}(t) \quad (5.12)$$

or in matrix formulation

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_m \\ \vdots \\ Y_M \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1n} & \dots & x_{1N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} & \dots & x_{mN} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{M1} & \dots & x_{Mn} & \dots & x_{MN} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \\ \vdots \\ \epsilon_N \end{pmatrix} \quad (5.13)$$

The data vector $\mathbf{Y}$ for each image $m = 1, \dots, M$ can be expressed as the product of $\mathbf{X}$, the $M \times N$ design matrix, and $\boldsymbol{\beta}$, the N -dimensional parameter vector plus the error vector $\boldsymbol{\epsilon}$. $\mathbf{X}$ has one column for every modeled effect and one row for each image. It represents the explanatory variables. The error vector is normally distributed with $\epsilon \sim N(0, \sigma^2 \boldsymbol{\Sigma})$. $\boldsymbol{\Sigma}$ denotes the noise autocorrelation matrix. If $(\mathbf{X}^T \mathbf{X})$ is invertible, i.e. $\mathbf{X}$ is of full rank, least-squares parameter estimates $\mathbf{b}$ for $\boldsymbol{\beta}$ in Equation 5.12 are given by (cf. Henson, in press; Josephs et al., 1997; Josephs & Henson, 1999)

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (5.14)$$

Following the notation of Henson (in press), the modeling of stimulus occurrence, neuronal activity, and hemodynamic responses will be briefly described.

Stimulus Function The stimulus function $s(t)$ represents the occurrence of stimuli as a function of time for each experimental condition and is modeled as a Kronecker δ function.¹² For $i = 1, \dots, C$ experimental conditions, each consisting of $j = 1, \dots, T$ onset times o_{ij} the stimulus model is given by

$$s_i(t) = \sum_{j=1}^T \delta(t - o_{ij}) \quad (5.15)$$

¹²In addition, the δ function can be weighted by a scaling factor, representing a parametric modulation of a particular condition, e.g. a learning function (cf. Henson, in press). See Experiment 1A (Chapter 7), Experiment 1B (Chapter 8), and Experiment 3 (Chapter 10) for applications.

Neural Activation Function For each experimental condition, the neural activation u_i can be expressed as a function of s_i . The neural activation function (Henson, in press) is given by

$$u_i(t) = s_i(t) \otimes r(t) \quad (5.16)$$

with

$$r(t) = \sum_{b=1}^B \beta_b g_b(\tau) \quad (5.17)$$

for epoch designs,¹³ where g_b denotes the $b = 1, \dots, B$ temporal basis functions and τ a finite peri-stimulus time (PST). In SPM, several temporal basis functions can be selected for modeling neural activity in epoch designs, e.g. boxcar, half-sine, and discrete cosine transform (DCT; cf. Henson, in press).

Hemodynamic Response Function According to Henson (in press), the hemodynamic response $x_i(t)$ can be modeled by convolving the neural activation function $u_i(t)$ with a hemodynamic response function $h(\tau)$ (HRF; Friston et al., 1998a)

$$x_i(t) = u_i(t) \otimes h(\tau) \quad (5.18)$$

Equation 5.18 is based on the assumption that the BOLD signal is the output of a linear time-invariant system (LTI; Boynton et al., 1996; Dale & Buckner, 1997), i.e. the hemodynamic response summates in a roughly linear fashion over time.¹⁴ The canonical HRF in SPM is composed of two γ functions (Friston, Jezzard & Turner, 1994). Beyond the canonical HRF, it is possible to span the space of likely responses using $p = 1, \dots, P$ temporal basis functions g_p (Friston, Frith, Turner & Frackowiak, 1995a; Josephs, Turner & Friston, 1997) given by

$$h(\tau) = \sum_{p=1}^P \beta_p g_p(\tau) \quad (5.19)$$

¹³In event-related designs $u_i(t) := s_i(t)$ (cf. Henson, in press).

¹⁴When stimuli are delivered at high frequency (interstimulus interval, ITI < 2 s), departures from linearity can be observed (Friston, Josephs, Rees & Turner, 1998b, cf. Friston, Mechelli, Turner, and Price, 2000b; Friston, 2002, providing a nonlinear input-output model based on Volterra series as one possibility to deal with nonlinearities).

Examples of temporal basis functions implemented in SPM are depicted in Figure 5.5: FIR basis set, Fourier basis set, and γ functions (for further explanations see Friston et al., 1998a; Henson, in press).

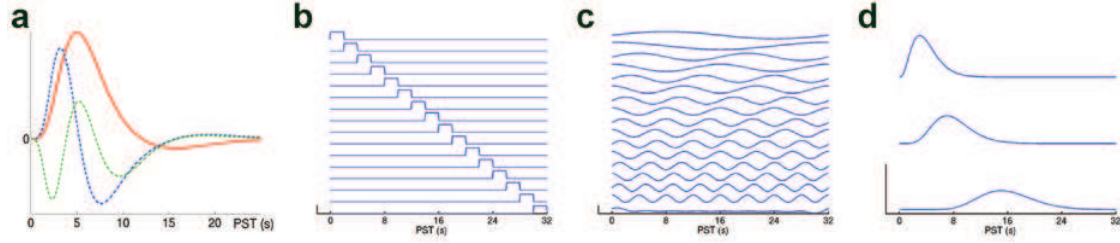


Figure 5.5. Modeling the BOLD Response by Different Temporal Basis Functions. (a) Canonical HRF (red) and its temporal (blue) and dispersion (green) derivatives. (b) FIR basis set. (c) Fourier basis set. (d) γ functions. [(Adapted from Henson (in press). Copyright (2004) Elsevier.]

5.3.3 Statistical Inference

After estimating the parameters, a priori hypotheses can be tested using contrasts, i.e. the design matrix $\mathbf{X}$ is weighted with a matrix $\mathbf{C}$ containing each contrast vector $\mathbf{c}$. Under the normality assumption (cf. Petersson, Nichols, Poline & Holmes, 1999a)

$$\mathbf{c}^T \mathbf{b} \sim \mathcal{N}(\mathbf{c}^T \boldsymbol{\beta}, \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}) \quad (5.20)$$

the significance of each contrast can be assessed with the t -statistic (i.e. Student's t distributed under the null hypothesis) at each voxel resulting in a statistical parametric map $\text{SPM}\{t\}$ which is usually transformed into a Gaussian $\text{SPM}\{z\}$. Alternatively, the omnibus F -statistic can be used (see http://www.fil.ion.ucl.ac.uk/wpenny/publications/rik_anova.pdf for examples).

Theory of Gaussian Random Fields This voxel-by-voxel hypothesis testing results in a huge amount of statistical tests, giving rise to the so-called multiple comparison problem. Thus, it is necessary to control the false-positive rate. A standard method is the Bonferroni correction which adjusts the significance level α to $\frac{\alpha}{k}$, assuming that k statistical tests are independent. However, this procedure is too conservative with respect to fMRI time-series analysis, because the z scores at each

voxel are spatially correlated. In contrast, an approach based on Gaussian random field theory (Friston, Frith, Liddle & Frackowiak, 1991; Worsley, Evans, Marrett & Neelin, 1992) controls the false positive rate of activated regions. The search volume is expressed in terms of resolution elements, i.e. the number of independent observations in a volume (Worsley et al., 1996). In brief, this method assesses the probability that the maximum z score exceeds a given threshold u (Brett, Penny & Kiebel, in press). This is equivalent to the probability of getting one region above u (Friston et al., 1995b). This probability is approximated by the expected Euler characteristic. For high thresholds, the Euler characteristic approximately equals the number of local maxima in the statistical field (Worsley et al., 1996, see Brett et al., in press; Friston, 2003, and Petersson et al., 1999b for a detailed overview of random field approaches).

Alternative Approaches Several alternative approaches for statistical inference have been developed. A method for controlling the false discovery rate (FDR) was introduced by Genovese, Lazar, and Nichols (2002). The FDR is defined as the ratio between the number of false-positive tests and the total amount of tests, for which the null hypothesis is rejected. The advantage of this method is that it is adaptive, i.e. the statistical thresholds are automatically determined from the data.

Classical statistical inference assumes that the data are normally distributed. Moreover, the distribution of the test statistic under the null hypothesis is well known. In contrast, non-parametric approaches (Statistical Non-Parametric Mapping, SnPM) determine the distribution under the null hypothesis empirically by creating all possible permutations of experimental conditions (randomization tests; cf. Holmes, Blair, Watson & Ford, 1992; Nichols & Holmes, 2001).

5.4 Designing

This section deals with the concept of event-related fMRI (5.4.1) and several aspects of design efficiency (5.4.2). Two main design types are used in fMRI studies: blocked

designs and event-related designs. In blocked designs, several trials of one condition are presented sequentially in one block. In contrast, event-related fMRI allows the analysis of single trial-related activity.

5.4.1 Event-related Functional Imaging

Statistical analysis of fMRI time series on a trial-by-trial basis was introduced by Boynton et al. (1996) and Buckner et al. (1996, see also Zarahn, Aguirre, and D’Esposito, 1997). Buckner and colleagues (1996) showed that activation maps derived from a single trial analysis resembled those derived from blocked runs. A prerequisite for modeling the hemodynamic response in event-related designs is the assumption that the BOLD response is the output of a linear time-invariant system, i.e. it is linearly related to the underlying neuronal activity (Boynton et al., 1996; Dale & Buckner, 1997). However, this assumption is violated when stimuli are delivered at high rate ($ISI < 2$ s; Friston et al., 1998b) or stimuli are presented at short durations (Vazquez & Noll, 1998).¹⁵

In their seminal study, Dale and Buckner (1997, cf. Buckner, 1998; Rosen, Buckner & Dale, 1998), showed that the BOLD response for multiple rapidly presented trials ($ISI \geq 2$ s) adds in a roughly linear fashion. The authors presented one-trial, two-trial, and three-trial clusters. For example, by subtracting the two-trial cluster from the one-trial cluster the estimated BOLD signal to the second trial could be derived. As apparent from Figure 5.6, the estimated signals for all three trials are almost identical, irrespective of the trials presented before (see Friston et al., 1998a; Josephs et al., 1997, for mathematical formulations and Josephs and Henson, 1999, for an excellent overview of event-related fMRI). Event-related designs have several advantages over blocked designs (see Chein & Schneider, in press; Josephs & Henson, 1999):

- Trials pertaining to different conditions can be presented in randomized order, e.g. maintenance vs manipulation trials in a working memory paradigm (e.g.

¹⁵Moreover, the nonlinearities of the BOLD response have a spatial dependence (Birn, Saad & Bandettini, 2001; Pfeuffer, McCullough, Van de Moortele, Ugurbil & Hu, 2003).

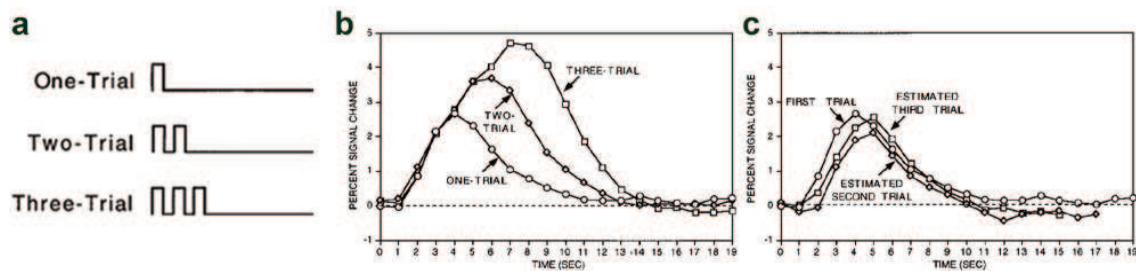


Figure 5.6. Results of the Dale and Buckner (1997) Study. (a) The three different types of trial clusters used in the study of Dale and Buckner (1997). (b) BOLD signals for one trial, two trial, and three trial clusters. (c) Hemodynamic responses for one trial clusters and estimated BOLD signals for two trial and three trial clusters. [Adapted from Dale and Buckner (1997). Copyright (1997) National Academy of Sciences, USA (Proceedings of the National Academy of Sciences of the United States of America).]

Postle, Berger & D'Esposito, 1999).

- It is possible to categorize conditions post hoc, e.g. in the subsequent memory paradigm, where the fMRI signal is measured during an encoding task and then classified based on subjects' memory performance during a later recognition phase (e.g. Wagner et al., 1998).
- In some experimental procedures, it is convenient to use event related designs, e.g. the oddball paradigm, where infrequent, deviant stimuli are presented among frequent, standard stimuli (Doeller et al., 2003; Opitz et al., 2002).
- In event-related designs, it is possible to model properties of the BOLD signal explicitly, e.g. the onset latency (Liao et al., 2002, see Section 5.2.3).
- Event-related fMRI allows the separation of subcomponents of trials, e.g. encoding, delay, and retrieval-related activity in complex working memory studies (Aguirre & D'Esposito, 1999; D'Esposito, Postle, Jonides & Smith, 1999, see Ollinger et al., 2001a, 2001b, for an estimation method).
- Finally, event-related designs allow for a direct comparison (and integration) with other trial-based methods (e.g. single cell recordings, MEG, ERP, behavioral measures; Dale & Halgren, 1999; Luck, 1999, see Doeller et al., 2003, and Opitz et al., 2002, for examples of integrating ERP and fMRI measures by means of dipole analysis).

5.4.2 Design Efficiency

The inherent temporal limitations of fMRI (with T_R normally ranging from 1-4 s, see Section 5.1) result in a low sampling rate. However, the effective sampling rate can be simply improved by using an interleaved procedure, i.e. by varying the relationship between the onset of stimuli and image acquisition (Miezin et al., 2000). Furthermore, given the temporal characteristics of the BOLD signal (i.e. it's extended time course; see Section 5.2.3), the HRF can be considered as a low-pass filter (Chein & Schneider, in press; Friston et al., 2000a; Josephs & Henson, 1999). Therefore, it is important to determine timing parameters appropriately in event-related designs (namely the distribution of the stimulus onset asynchrony, SOA) to maximize the variance of the experimental task passing the HRF. By this, it is possible to increase the signal-to-noise ratio. For instance, it has been shown that the efficiency of an event-related design (see Chein & Schneider, in press; Donaldson & Buckner, 2001, for an overview of design efficiency) is higher if the SOA is temporally jittered or randomized during the time course of the experiment as compared to a fixed SOA (Dale, 1999). Josephs and Henson (1999) formally defined an event-related design by the SOA and a $N^m \times N$ transition matrix T , describing the probability of an event given m previous events. N denotes the number of different event types. Based on this definition, several design types can be distinguished (Friston, Zarahn, Josephs, Henson & Dale, 1999b). Designs with events occurring at a prespecified time are termed deterministic designs and the probability p_e of each event type e is given by $p_e \in [0, 1]$. Furthermore, in nonstationary stochastic designs p_e changes over time ($p_e < 1$), whereas in stationary stochastic designs p_e is given by $p_e = \frac{1}{N} \forall e, e = 1, \dots, N$.

Optimizing an experimental design is equivalent to minimizing the covariance of a contrast $\mathbf{c}$ of a parameter estimate $\mathbf{b}$ (Friston et al., 1999b, 2000a, see also Section 5.3.2, in particular Equation 5.12-5.14) which is given by

$$\text{cov}\{\mathbf{c}^T \mathbf{b}\} = \sigma^2 \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c} \quad (5.21)$$

Therefore, the efficiency E (Friston et al., 1999b) of a design matrix can be defined as

$$E(X) \propto \text{trace}\{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}\}^{-1} \quad (5.22)$$

As apparent from Equation 5.21 and 5.22, maximizing the variance of the design matrix yields increased efficiency. Is it also important to note that the efficiency of a design critically depends on the specified contrast, e.g. a main effect or a differential effect. Simulation studies (Josephs & Henson, 1999) suggest that stationary stochastic designs with short SOAs (2-4 s) result in high efficiency for differential effects, whereas for main effects the optimal SOA is roughly 18 s. As pointed out by Liu, Frank, Wong, and Buxton (2001), a distinction should be made between detection power, i.e. the ability to detect activated voxels and estimation efficiency, i.e. the ability to estimate the shape of the response. Recent simulation studies have shown that estimation efficiency is highest for rapidly alternating stimuli, whereas blocked stimulus presentation reveals the best detection power (Birn, Cox & Bandettini, 2002). Moreover, it has been shown that exponentially distributed SOAs seem to be a good compromise, leading to high estimation efficiency as well as high detection power (Hagberg, Zito, Patria & Sanes, 2001, see Experiment 1A, 1B, and 3).

Part IV

Experiments

Chapter 6

Rationale of the Present Experiments

Contents

6.1	The Aim of the Present Studies	100
6.2	The Paradigm	102
6.3	Hypotheses and Predictions	105
6.4	Methods	106

6.1 The Aim of the Present Studies

The distinction between two separable learning and memory systems provides a theoretical starting point for the present thesis. It appears that one system is specialized for rapid learning unique episodes (first system), whereas the other system is best suited for gradual learning regularities across multiple episodes (second system) (cf. Chapter 2 and 3). This thesis aimed at specifying the neural and cognitive processes underlying the second learning and memory system. The term ‘regularity learning’ was introduced to characterize the learning process of the second system. The following main questions were addressed: (1) Which brain structures subserve regularity learning (and also which brain structures mediate learning unique episodes)? (2) Are these regions preferentially recruited during separate phases of

the learning process, i.e. which regions support regularity learning during the initial and later phases of the learning process, respectively? (3) Finally, is regularity learning affected by specific stimulus and task parameters, e.g. is regularity learning a domain-specific or a domain-general mechanism? For that purpose, a task was developed which allowed to disentangle both learning systems.

To be more precisely, evidence coming from different fields of cognitive neuroscience indicates that the hippocampus seems to play a crucial role during episodic memory formation (see Chapter 2). Both the Eichenbaum (Eichenbaum, 2000) and the O'Reilly model (O'Reilly & Norman, 2002) propose that the hippocampus subserves episodic memory formation by means of a specific memory function, namely relational binding of different episodic features (cf. Zimmer et al., in press). However, only a few imaging studies directly tested this claim in humans (Davachi & Wagner, 2002; Giovanello et al., 2004; Preston et al., 2004). Thus, one aim of the present thesis - even though not the core aim - was to investigate the role of the hippocampus during relational memory binding in more detail by implementing a task which requires binding of several episodic features in each trial.

The main purpose of the present thesis was to elucidate the neural basis of regularity learning. As outlined in Chapter 3, in numerous studies using different learning tasks, e.g. reinforcement learning, category learning, artificial grammar learning, and sequence learning tasks, regularities across episodes are introduced and I proposed that the existence of these regularities is one crucial factor which allows learning to occur. However, none of these studies directly tested the influence of regularities across multiple complex episodes by means of selectively manipulating the binding requirements (in comparison to a condition which does not include any regularities and hence entails constant binding requirements; see below). Thus, the present experiments investigated regularity learning by means of a design including two conditions, one condition *with* regularities (and hence decreasing binding requirements) and one condition *without* any regularities (and hence constant binding requirements).

Another purpose of the present thesis was to examine the learning *process*, and especially brain regions involved during the time course of learning in more detail. As

mentioned in Chapter 3, in several imaging studies a learning condition was solely compared with a visuomotor control condition. In contrast, in the present fMRI experiments the neural correlates of regularity learning were examined by means of parametric fMRI analyses which allow to investigate learning-related activation pattern on a trial-by-trial level during the entire learning process (cf. Opitz & Friederici, 2003; Poldrack et al., 2001).

Finally, the present thesis aims at specifying whether regularity learning is a domain-general or a domain-specific process and most importantly, whether particular brain regions support learning in a domain-general or a domain-specific manner. For that purpose, regularities were introduced on two separable domains (see Section 6.2). Several recent imaging studies suggest a domain-specific lateralization of both medial temporal and lateral prefrontal regions during *memory* formation (cf. Burgess et al., 2002; Paller & Wagner, 2002; Wagner, 2002, see Chapter 2 for further details). The present experiments tried to elucidate, whether such a hemispheric specialization also holds during *learning*.

6.2 The Paradigm

Experimental Procedures In all experiments reported in this thesis, 16 different objects were used as stimulus material. Within one experimental trial, stimuli were presented within a 4×4 grid, i.e. there were sixteen possible positions. As apparent from Figure 6.1, in each trial several stimuli were presented sequentially during a sample phase with a constant interstimulus interval (ISI).¹ After a short delay, a probe stimulus was presented at a particular position within the grid. Participants had to indicate whether or not the probe stimulus was identical to one of the object-position combinations seen before during the sample phase. Responses were delivered by a button press with the right or left index finger. Response-to-hand mappings were counterbalanced across subjects. Probes in each experimental block (see below) comprised 50% old (old object at old position) and 50% new

¹In Experiment 2F (Chapter 9), objects were presented simultaneously in each trial.

object-position conjunctions (3 equally distributed categories: old object/new position, new object/old position, and new object/new position). In all experiments, visual feedback was given for 500 ms by means of the color of a fixation cross (green = correct, red = false, yellow = timeout). Stimulus type, number of stimuli presented in the sample phase, stimulus duration, ISI, delay duration, probe duration, intertrial interval (ITI), block length, and the number of blocks per condition were selectively manipulated in the experiments (see Chapters 7, 8, 9, and 10, for details). Each experimental session comprised several blocks. Each of these blocks included 16-36 trials. The design of a single experiment comprised two conditions (see below). In Experiment 1A (Chapter 7), Experiment 1B (Chapter 8), and Experiment 2A and 2B (Chapter 9), blocks of both conditions were presented in randomized order with the constraint that participants had to perform at most two blocks of the same condition in succession. The sequence of blocks was balanced across subjects. In Experiment 2C-2F (Chapter 9) and Experiment 3A and 3B (Chapter 10), trials were blocked by condition. Participants were informed about the beginning of a new block.

Experimental Conditions Two conditions, a context-specific condition (CS) and an invariant learning condition (IL) were used in the experiments (Figure 6.2). In each experiment, the CS condition was paired with one version of the IL condition, either a *spatial* (Experiment 1A and 3B) or an *object* IL condition (Experiment 1B, 2, and 3A).² Two versions of the IL condition were used to investigate whether regularity learning is differentially modulated by a specific stimulus domain (see below). The sequence of several object-position conjunctions in the sample phase of each trial was defined as an ‘episode’ in the experimental framework of the present thesis. This definition meets the aforementioned criteria of an episode (Eichenbaum, 2000, see Chapter 2), i.e. including several temporally disparate episodic features. This structure of an episode requires relational binding of the episodic features (Eichenbaum, 2000) to entail an exhaustive, pattern-separated representation of the episode (O’Reilly & Norman, 2002). In the CS condition, object-position conjunc-

²In Experiment 2C-2F (Chapter 9), subjects solely performed blocks in the IL condition.

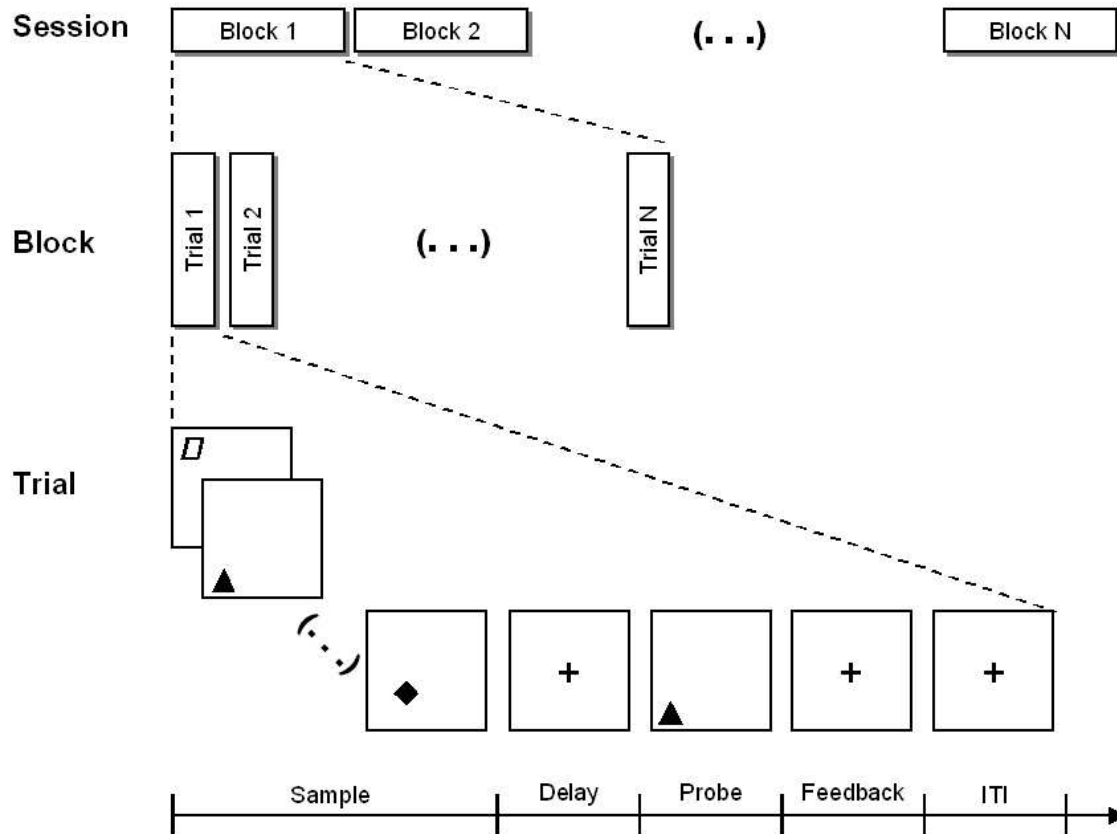


Figure 6.1. Session, Block, and Trial Structure of the Experiments. One session comprised several blocks which in turn comprised several trials. In each trial, different objects were presented at different positions of a 4×4 grid during the sample phase (grid not shown in the Figure). After a short delay, a probe stimulus was presented and subjects were required to indicate whether or not the probe is identical to one of the object-position conjunctions presented during the sample phase. Feedback was given by means of the color of a fixation cross.

tions were unique in each trial, i.e. objects were variably mapped to positions. Here, each object-position conjunction in each trial and each block was selected randomly from the set of objects and positions. In this condition, processing of unique episodes was required. Thus, an involvement of the first learning system was expected. In contrast, in the IL condition, either positions or objects were held constant across trials within blocks. In the *spatial* IL condition, variable objects were presented at constant positions. Accordingly, in the *object* IL condition, constant objects were presented at variable positions. By this, processing of regularities across multiple episodes was induced in this condition. Thus, an involvement of the second learning system was expected. The number of invariant positions and objects, respectively, resembled the

number of object-position conjunctions presented during the sample phase. Objects or positions were held constant across all trials of one experimental block. The fixed configuration of invariant objects and positions held constant across all trials of one experimental block changed from block to block, i.e. when a new block starts, a new set of positions or objects in object-position conjunctions was introduced.

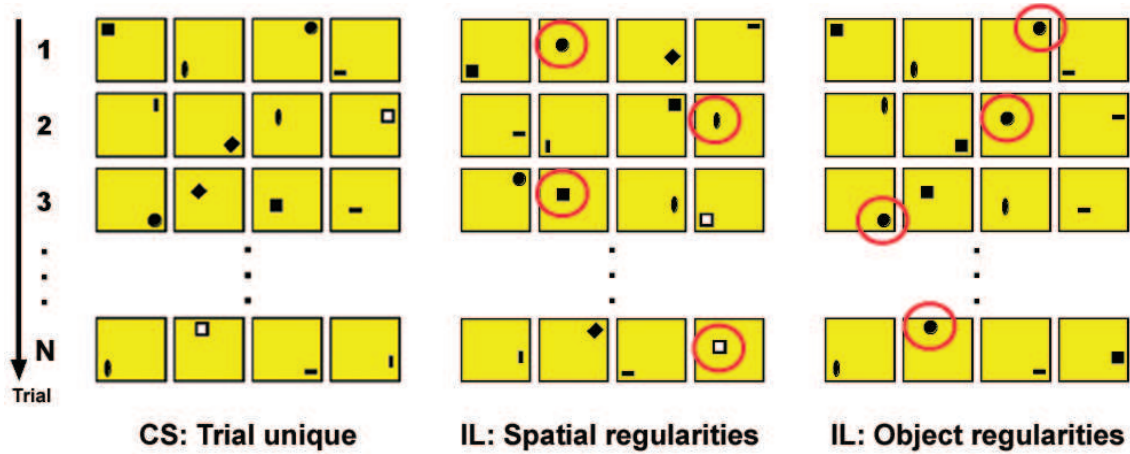


Figure 6.2. Experimental Conditions. Examples of the block structure, separately for the CS condition (left panel), the spatial IL condition (middle panel), and the object IL condition (right panel). In this example one trial comprises four sequential object-position conjunctions (from left to right in each panel). In the CS condition, each trial comprised unique object-position conjunctions. Importantly, in the spatial and object IL condition, positions and objects, respectively were held constant across all trials of one block. For illustration, one invariant position and one invariant object is highlighted by circles in the blocks of the spatial and object IL condition, respectively.

6.3 Hypotheses and Predictions

Behavioral Predictions On a behavioral level, a performance increase across trials within blocks was expected in the IL conditions. Here, subjects should benefit from object and spatial regularities in object-position conjunctions during the time-course of the blocks. In contrast, in the CS condition constant performance was expected within blocks, since variable objects at variable positions have to be encoded and retrieved in each trial of this condition. Furthermore, a modulation of the within-block learning effects across experimental blocks was expected in the IL condition (cf. Fletcher et al., 1999).

Imaging Predictions All fMRI analyses were restricted to the sample phase. It was assumed that learning-related neural activity should be optimally reflected in the sample phase, when subjects encode and maintain four object-position conjunctions. In agreement with the O'Reilly model (O'Reilly & Norman, 2002) and the Eichenbaum model (Eichenbaum, 2000), constant hippocampal activity was expected in the CS condition. Here, each trial comprised a unique episode, since variable objects have to be bound to variable positions in each trial. In accordance with both models, it is assumed that the hippocampus is crucial for building pattern-separated representations of episodes by means of relational feature binding. A corollary of this view is that the relational binding requirements are approximately constant across trials in the CS condition, since object-position conjunctions are trial-unique. In the IL condition, reduced hippocampal activation was expected during the time-course of experimental blocks. Here, the relational binding requirements of each episode should decrease across trials, since the probability for specific object-position conjunctions is substantially increased, whereas the probability for other object-position conjunctions is reduced to zero.

Importantly, a learning-related modulation of activity in the lateral prefrontal cortex was predicted in the IL condition. As mentioned in Chapter 3, there is converging evidence from studies using artificial grammar learning (Fletcher et al., 1999; Opitz & Friederici, 2003), sequence learning (Aizenstein et al., 2004; Schendan et al., 2003) and category learning tasks (Freedman et al., 2001; Reber et al., 1998b; Strange et al., 2001) that the lateral prefrontal cortex is involved in learning task-relevant regularities. Based on these results it is proposed that the lateral PFC - beyond the hippocampus - supports regularity learning in the IL condition.

6.4 Methods

The functional MRI methods have been described in Chapter 5. In the following, a brief overview of the behavioral methods will be given. In all experiments, Pr values were used as the main behavioral measure. The discrimination index Pr is a non-parametric equivalent of d' and refers to the proportion of hits (correct responses to

old probes) minus the proportion of false alarms (false responses to new probes). In contrast to d' which is defined by means of the signal detection theory, Pr is based on a two-high threshold model (see Feenan & Snodgrass, 1990; Snodgrass & Corwin, 1988). To evaluate learning/performance across trials, Pr values were calculated for different time steps within experimental blocks by averaging Pr values across several consecutive trials. Furthermore, to investigate performance modulations across trials in more detail, different types of false alarms were analyzed, i.e. false alarms to probes comprising an old object at a new position, a new object at an old position, and a new object at a new position.

Chapter 7

Experiment 1A:

Prefrontal-Hippocampal Dynamics Involved in Learning Regularities across Episodes¹

Contents

7.1	Introduction	109
7.2	Materials and Methods	112
7.3	Results	119
7.4	Discussion	125

In the present Experiment 1A, fMRI was used to investigate the neural basis of regularity learning in the spatial domain. Subjects performed several blocks in the

¹This Chapter contains the following article: **Doeller C. F., Opitz B., Krick, C. M., Mecklinger, A., & Reith, W. (in press). Prefrontal-hippocampal dynamics involved in learning regularities across episodes. *Cerebral Cortex*.** For stylistic reasons the format was adapted. Thus, this version differs in minor ways from the final published version. In the following, this study will be referred to as Experiment 1A. This work was supported by grants from the German Research Foundation (DFG; Research Group FOR-448). The authors wish to thank Ben Eppinger, Patric Meyer, and Markus Werkle for helpful comments and stimulating discussions.

CS condition and in the spatial version of the IL condition as outlined in Chapter 6. A performance increase within blocks of the IL condition was predicted, whereas performance should be constant across trials in the CS condition. Moreover, constant hippocampal activity was expected within blocks of the CS condition. In contrast, a learning-related decrease of hippocampal activity and a learning-related increase of lateral prefrontal activity was predicted in the IL condition.

Abstract Using fMRI the neural correlates of context-specific memories and invariant memories about regularities across episodes were investigated. Volunteers had to learn conjunctions between objects and positions. In an invariant learning condition, positions were held constant, enabling subjects to learn regularities across trials. In contrast, in a context-specific condition object-position conjunctions were trial unique. Performance increase in the invariant learning condition was paralleled by a learning-related increase of inferior frontal gyrus activation and ventral striatal activation and a decrease of hippocampus activation. Conversely, in the context-specific condition hippocampal activation was constant across trials. We argue that the learning-related hippocampal activation pattern might be due to reduced relational binding requirements once regularities are extracted. Furthermore, we propose that the learning-related prefrontal modulation reflects the requirement to extract and maintain regularities across trials and the adjustment of object-position conjunctions on the basis of the extracted knowledge. Finally, our data suggest that the ventral striatum encodes the increased predictability of spatial features as a function of learning. Taken together, these results indicate a transition of the relative roles of distinct brain regions during learning regularities across multiple episodes: regularity learning is characterized by a shift from a hippocampal to a prefrontal-striatal brain system.

7.1 Introduction

Episodic memory refers to the ability to remember specific events set in a spatio-temporal context (Tulving, 1983). A large number of lesion studies (Zola et al., 2000) and single cell recordings in animals (Fortin et al., 2002), as well as neuropsycholog-

ical investigations (Yonelinas et al., 2002), intracranial EEG data (Fernández et al., 1999) and functional imaging data in humans (Davachi et al., 2003) suggest that the hippocampus as part of the medial temporal lobe memory system plays a crucial role in encoding and retrieval of episodic memories. One influential view on hippocampal function suggests that this brain region is critically involved in binding of different features and events that compose episodic memories (Eichenbaum, 2000). For example, it has been shown that hippocampal cells in the rat bind singular conjunctions of episodic features, such as a specific odor occurring at a particular position (Wood et al., 1999). Furthermore, the hippocampal binding mechanism has to prevent interference between distinct episodes. One seminal framework (Norman & O'Reilly, 2003; O'Reilly & Norman, 2002) proposes that the hippocampus assigns distinct (pattern-separated) representations to different episodes to minimize interference. In a similar vein, Shastri (2002) proposes two levels of hippocampal binding: (1) entities occurring in the event are bound to the roles they fill in the event, and (2) all role-entity bindings defining an event are grouped together in order to separate them from role-entity bindings of other events. To exemplify the first binding type imagine the situation when a person paints a picture in a studio. In the representation of this specific episode the role 'painter' is bound to this particular person, the role 'location' is bound to studio, and the role 'object' is bound to picture. Given the episode that this person watches television in his apartment, the second level of binding is required to distinguish both episodes.

How does the brain represent repeated overlapping features of episodes? In addition to context-specific memories about particular episodes, individuals are able to acquire knowledge about regularities across such specific episodes (Shanks & St. John, 1994). Several recent models have addressed this issue. For instance, it has been assumed that this knowledge about regularities comprises a measure of the probability that certain types of entities are bound to a certain role (Shastri, 2002). Moreover, Eichenbaum (2000) assumes that particular hippocampal neurons are specialized to represent features that are common across many episodes. In contrast, O'Reilly and Norman (2002) suggest that the rhinal and parahippocampal cortices represent regularities in the environment by assigning overlapping distributed representations

to similar stimuli.

Recent neuroscientific research has begun to study the neural correlates of learning regularities across specific episodes. For instance, studies using category learning tasks in humans (Reber et al., 1998b; Strange et al., 2001) and monkeys (Freedman et al., 2001) have identified the lateral prefrontal cortex (PFC) as a core structure in mediating this kind of learning. In a study by Strange et al. (2001) subjects had to learn abstract rules that define category membership of four-letter strings (e.g. ‘The first two letters are always identical’). The lateral PFC was selectively engaged following rule change. Furthermore, Freedman et al. (2001) found category-sensitive neurons in the monkey’s lateral PFC. After defining new categories based on the same stimulus set, the same neurons then adaptively represented the new categories, indicating that the lateral PFC is involved in the flexible detection of regularities in this task.

Another line of evidence for prefrontal involvement in learning regularities comes from artificial grammar learning research, where subjects acquire abstract knowledge about the rules of a finite-state grammar (Cleeremans et al., 1998). Recent functional imaging studies have implicated the lateral PFC in artificial grammar learning (Fletcher et al., 1999; Opitz & Friederici, 2003). The lateral PFC also supports sequence learning, as indicated by findings from several fMRI studies (Aizenstein et al., 2004; Schendan et al., 2003). In these studies, subjects acquire knowledge about a regular sequence of stimulus events. Moreover, the lateral PFC is associated with causal associative learning (Fletcher et al., 2001) and with the detection of abstract sequence violations (Huettel, Mack & McCarthy, 2002). It also shows learning-related activity during arbitrary rule learning (Toni et al., 2001; Wallis et al., 2001). Despite the different tasks used in the above mentioned studies, taken as a whole, these latter findings underscore the importance of the lateral PFC in learning task-relevant regularities across different episodes, leading to the acquisition of abstract knowledge structures. However, it should be noted that the exact location of lateral PFC activation varies between studies, depending on stimulus properties and task requirements.

Using fMRI we investigated the neural correlates of context-specific and invariant

memories, i.e. memory for invariant features of episodes, thereby bringing together two research fields in cognitive neuroscience, episodic memory and rule learning. More specifically, we were interested in brain areas responsible for the acquisition of invariant memories by regularity learning. In each trial subjects were required to learn four sequential object-position conjunctions (see Methods section and Figure 7.1). Subsequently, they had to indicate whether or not a probe stimulus was identical to one of the four object-position conjunctions (i.e. whether or not there is an exact match of both stimulus features, the object and the position). In an invariant learning condition, the positions were invariant within a block. That is, different objects were presented at the same four positions in each trial of an experimental block, thereby enabling the extraction of spatial regularities and the formation of invariant memories. In a context-specific condition, in contrast, objects and positions were variable within experimental blocks, i.e. there were no spatial regularities across trials, requiring the processing of unique object-position combinations from trial to trial.

We expected enduring hippocampal activity in the context-specific condition, since variable objects have to be bound to variable positions in each trial of the blocks (constant binding requirements). In the invariant learning condition, in contrast, we hypothesized reduced hippocampal activation as a function of time within blocks. Here, variable objects have to be bound to invariant positions (reduced binding requirements). Moreover, we expected a hippocampal-prefrontal shift of activity in the invariant learning condition. More precisely, we assumed that the lateral PFC is involved in the extraction of invariant spatial features and the flexible adjustment of the extracted knowledge to the requirements of the ongoing task block. Thus, we expected increased PFC activity in the invariant learning condition.

7.2 Materials and Methods

Subjects Ten subjects participated in the study (aged 21-32, mean age 27 years, 6 males). All subjects were right-handed with normal or corrected-to-normal vision and were paid for participating. Informed consent was obtained before scanning. All

participants reported to be in good health with no history of neurological disease. One subject had to be excluded from further analysis due to technical problems during fMRI acquisition.

Stimuli, Task, and Design Sixteen stimuli which consisted of basic shapes (e.g. square, circle) were used as stimulus materials and were presented within a 4×4 grid, i.e. there were sixteen possible positions. Stimuli were back projected onto a translucent screen that participants viewed through a mirror during fMRI acquisition.

Within one experimental trial four different stimuli were presented sequentially at four different positions. Each stimulus was delivered for 800 ms with a 500 ms interstimulus interval (Figure 7.1a). After a short delay of 1300 ms a probe stimulus was presented for 2000 ms at a particular position within the grid. Participants had to indicate whether or not the current object-position conjunction (the probe stimulus) was identical to one of the four object-position combinations seen before. Responses were delivered by a button press with the right or left index finger (2000 ms response window) by means of MR-compatible response keys. Response-to-hand mappings were counterbalanced across subjects. Visual feedback (500 ms) was given by means of the color of a fixation cross (green = correct, red = false, yellow = time-out), immediately after probe offset. Probes in each block (see below) comprised 50% old (old object at old position) and 50% new object-position conjunctions (3 equally distributed categories: old object/new position, new object/old position, and new object/new position). An exponentially distributed intertrial interval (ITI) of 4-9 s (mean: 5 s) which was varied in steps of 1 s was used in order to get an optimal tradeoff between detectability and estimation efficiency of the BOLD response (Birn et al., 2002; Hagberg et al., 2001).

After subjects had practiced 20 trials outside and ten trials in the scanner, they performed 224 trials during the whole experiment. One experimental session comprised seven blocks (16 trials each) in the context-specific condition and seven blocks in the invariant learning condition. Blocks of both conditions were presented in randomized order with the constraint that participants had to perform at most two blocks

of the same condition in succession. Subjects were informed about the beginning of a new block. The sequence of blocks was balanced across subjects. In the invariant learning condition, four positions were held constant within one experimental block whereas in the context-specific condition, objects and positions were variable across trials (Figure 7.1b). Within one invariant learning block the temporal order of the four positions was randomized across trials. The fixed configuration of four positions changed from block to block in the invariant learning condition. All other parameters were held constant across both conditions. Before scanning, participants were instructed that the detection and application of a rule would make the task more easy and that this rule could change when a new block starts. Participants were unaware of the experimental manipulation, i.e. the existence of two different conditions.

Behavioral Analysis To get a first estimate for learning, mean Pr values (proportion hits - proportion false alarms) (Feenan & Snodgrass, 1990) were computed. Additional analyses were conducted for hits and false alarms. Furthermore, the false alarms were broken down by the type of false alarm committed. In all behavioral analyses, mean performance measures in the first and the last eight trials of each block (i.e. the two halves of the blocks) were averaged across all blocks separately for both conditions.

Imaging Parameters A Siemens SONATA MR scanner (Erlangen, Germany) operating at 1.5 Tesla with a standard circularly polarized whole head coil was employed to acquire both T_1 -weighted structural images and T_2^* -weighted BOLD-sensitive functional images. High-resolution (1 mm^3 voxel size) structural images were acquired using a 3-D MP RAGE sequence. Functional data were acquired using a gradient-echo EPI pulse sequence, with the following parameters: $T_R = 1.8 \text{ s}$, $T_E = 50 \text{ ms}$, flip angle = 85° , slice thickness = 4 mm , interslice gap = 1 mm , in-plane resolution = $3.5 \times 3.5 \text{ mm}$, FoV = 224 mm , 20 axial slices parallel to AC-PC plane. The first four volumes were discarded to allow for T_1 equilibration.

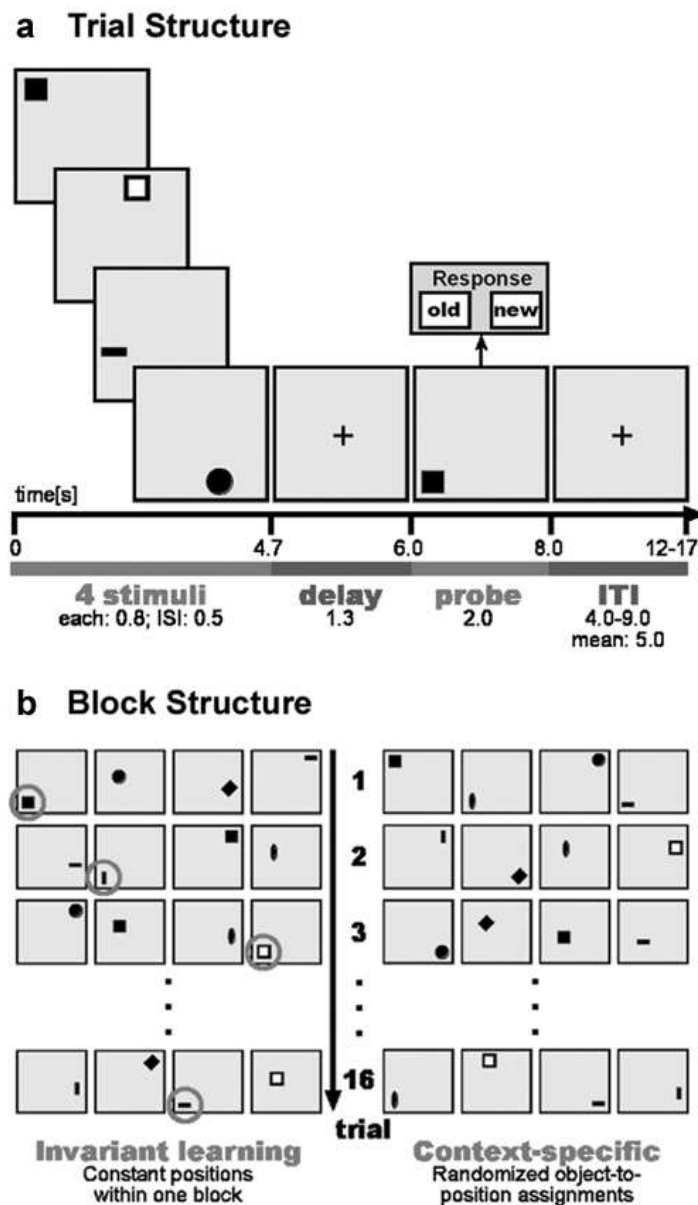


Figure 7.1. Trial and Block Structure of the Experiment. (a) Trial structure of the experiment. Subjects learned four sequential relations among objects and positions within a 4×4 matrix (grid not shown in the figure). Each stimulus was presented for 800 ms at a particular position with an interstimulus interval (ISI) of 500 ms. Subsequently, subjects were probed and had 2000 ms to determine whether the probe stimulus was identical to one of the four object-position combinations (i.e. an exact match of both stimulus features, the object and the position). Visual feedback was given for 500 ms. In the present example, the probe comprised an old object at a new position, therefore requiring a ‘new’ response. (b) Block structure of the experiment. In the invariant learning condition, positions were invariant in each trial of an experimental block, whereas in the context-specific condition, objects and positions were variable in each trial. For illustration, one invariant position is highlighted by circles. One experimental block comprised 16 trials. [Adapted from Doeller et al. (in press). Copyright Oxford University Press (Cerebral Cortex).]

Data Preprocessing FMRI time-series analysis was performed with Statistical Parametric Mapping (SPM2; Wellcome Department of Imaging Neuroscience, London, UK; <http://www.fil.ion.ucl.ac.uk/spm/spm2.html>). To account for the different sampling time of the slices, voxel time series were corrected using sinc interpolation and resampled using the middle slice as a reference point. All functional volumes were motion corrected. Using a least squares approach and a 6-parameter rigid body spatial transformation, realignment parameters were estimated. According to these determined parameters all volumes were resliced to the first volume using B-spline interpolation. After coregistering anatomical images to the mean functional image, the resulting images were normalized (Ashburner & Friston, 1999) to the standard T_1 template based on the MNI reference brain (Cocosco et al., 1997), using a 12-parameter affine transformation along with a nonlinear transformation (cosine basis functions). Based on the determined parameters the normalization algorithm was then applied to the functional volumes. Finally, the normalized images were resampled into 2 mm isotropic voxels and spatially smoothed with an isotropic 8-mm FWHM Gaussian kernel.

Basic Statistical Analysis Statistical analysis was performed in two stages in a mixed-effects model. For each subject, neural activity in each trial for both conditions was modeled by convolving a stimulus function, representing the onset of each sample phase with a canonical hemodynamic response function (HRF) (Friston et al., 1998a) (duration = 4.7 s) to cover the whole sample phase (epoch-related). The sample phase comprised the four sequential object-position conjunctions in each trial, including the interstimulus interval. To increase the power of our model, we additionally modeled neural activity for the probe stimulus in each trial for both conditions, using an event-related canonical HRF. For this purpose, an event train of δ -functions, time-locked to the onset of each probe stimulus was convolved with the canonical HRF. We assumed that learning-related neural activity should be optimally reflected in the sample phase, when subjects encode and maintain four object-position conjunctions. Thus, all analyses were restricted to the sample phase. The time series in each voxel were high-pass filtered to $\frac{1}{128}$ Hz to remove

low-frequency noise and proportionally scaled to a grand mean of 100 over voxels to remove effects due to global intensity fluctuations. Parameters for each covariate were estimated by a least-mean-squares fit of the model to the time series using a subject-specific fixed-effects model within the general linear model. During the estimation procedure serial correlations were estimated with a restricted maximum likelihood (ReML) algorithm using an AR(1) plus white noise model. In SPM2, the ReML estimates (hyperparameters) are then used to correct for non-sphericity (Friston et al., 2002b). Parameter estimates for the linear contrasts of interest (i.e. the direct contrasts of the invariant learning condition vs the context-specific condition and vice versa) entered into a second-level analysis treating subject as a random effect, using a one-sample t -test against a contrast value of zero at each voxel (Holmes & Friston, 1998). MNI-coordinates (Cocosco et al., 1997) of all reported activations have been transformed to the canonical Talairach space (Talairach & Tournoux, 1988) (see <http://www.mrc-cbu.cam.ac.uk/imaging/mnispace.html>). In the direct contrasts between both conditions, statistical parametric maps (SPMs) were thresholded at $p < 0.0005$, uncorrected for multiple comparisons. In addition, an extent threshold of 5 voxels was used to emphasize coherent activation.

Learning-Related Modeling In a second analysis step, we conducted a condition $\times$ time interaction analysis and two parametric fMRI analyses to directly investigate learning-related modulations of brain activity. To get a first estimate of a differential learning-related activation pattern in both conditions, we conducted a 2×2 ANOVA with the factors condition and time (mean BOLD activity for trials 1-8 and 9-16 of the experimental blocks, separately for both conditions). This analysis was restricted to the brain regions showing a main effect of condition in the basic statistical analysis ($p < 0.05$, small volume corrected; see below). Given our a priori hypotheses with respect to the hippocampus, this analysis was also conducted for both hippocampi ($p < 0.05$, small volume corrected; see below).

In two subsequent parametric fMRI analyses, we examined learning-related activation pattern in more detail by incorporating subjects' individual performance and learning functions on a trial-by-trial level into the imaging analysis. In the

first parametric analysis, we tested our a priori hypotheses that (a) the hippocampus shows a learning-related decrease in activity in the invariant learning condition *and* (b) constant activity in the context-specific condition. In this analysis subject-specific time series of each experimental block were weighted with subjects' behavioral performance measure. To increase the sampling rate of the learning functions, mean Pr values for four consecutive trials in each block (i.e. trials 1-4, 5-8, 9-12, and 13-16, respectively; see Figure 7.4d, upper panel for examples) were averaged across all blocks, separately for both conditions, resulting in four time bins for each subject and condition. The model functions were derived from each individual's learning curve by fitting a logarithmic function $y = a \cdot \ln(t_i) + b$ separately for both conditions (see Figure 7.4d, lower panel for examples). The two original condition-specific stimulus functions were multiplied by these parametric modulation functions, leading to additional regressors (Büchel, Holmes, Rees & Friston, 1998), either reflecting learning-related increase or decrease in the invariant learning condition or continuously enduring hemodynamic activity in the context-specific condition. Linear contrasts of the parameter estimates for each regressor were calculated for each subject and brought to the second level random effects analysis. To test our specific hypothesis about the differential hippocampal involvement in both conditions, we performed a conjunction analysis (Friston, Holmes, Price, Büchel & Worsley, 1999a) across subjects. Here, we used linear contrasts of the parameter estimates for the respective regressors in both conditions, i.e. the regressor including a decreasing model function in the invariant-learning condition and the regressor including a constant model function in the context-specific condition. This analysis was restricted to the hippocampus proper. Hippocampal regions of interest were determined a priori according to a detailed neuroanatomy atlas (Warner, 2001). An appropriate mask image was generated using the software package MRIcro (<http://www.psychology.nottingham.ac.uk/staff/cr1/micro.html>). Statistical tests were corrected for multiple comparisons ($p < 0.05$, small volume corrected) (Worsley et al., 1996). Statistical tests were corrected for multiple comparisons ($p < 0.05$), using appropriate small volume corrections (Worsley et al., 1996). In a second parametric analysis, we examined whether brain regions showing a main effect of con-

dition in the direct comparisons also exhibit learning-related changes of activation in the invariant learning condition. Thus, learning-related increase and decrease of activity in this condition was modeled. In this analysis, SPMs from the invariant-learning > context-specific contrast and the context-specific > invariant-learning contrast served as mask images, respectively ($p < 0.05$, small volume corrected).

7.3 Results

Behavioral Results Figure 7.2a shows mean Pr values (proportion hits - proportion false alarms) (Feenan & Snodgrass, 1990), separately for the context-specific and invariant learning condition, collapsed across experimental blocks and subjects. In the invariant learning condition, Pr values increased within task blocks. This was not the case in the context-specific condition. This observation was confirmed by a two-way repeated-measure ANOVA with the factors condition (context-specific vs invariant learning) and time (2 levels: mean Pr values for eight consecutive trials each). This analysis revealed a main effect of time, $F(1, 8) = 6.88$, $p < 0.05$, and a condition $\times$ time window interaction, $F(1, 8) = 24.66$, $p < 0.005$. Additional one-way ANOVAs separately for both conditions showed a significant effect of time in the invariant learning condition, $F(1, 8) = 20.58$, $p < 0.005$, but not in the context-specific condition, $F(1, 8) = 5.27$, $p < 0.1$.

To elucidate in more detail the learning mechanisms and performance pattern in both condition, additional analyses were conducted separately for the mean hit and false alarm rates (Figure 7.2b-c). There was a condition $\times$ time window interaction for hits, $F(1, 8) = 7.04$, $p < 0.05$, and false alarms, $F(1, 8) = 5.85$, $p < 0.05$. Furthermore, a marginally significant increase of hits, $F(1, 8) = 3.48$, $p < 0.1$, and a decrease of false alarms, $F(1, 8) = 6.83$, $p < 0.05$, within blocks could be observed in the invariant learning condition. In contrast, in the context-specific condition, mean hit rate, $F(1, 8) = 1.27$, $p > 0.2$, and mean false alarm rate, $F(1, 8) < 1$, remained constant within blocks.

In our task, subjects can commit a false positive response by classifying as old an old object at a new position, a new object at an old position, and a new object at

a new position. Given this, we assumed that learning, i.e. the extraction of spatial regularities, should be most clearly revealed by a reduction of false alarms to trials that include new positions, since subjects could reject these probes solely on the basis of their knowledge about the invariant positions. To examine this, the false alarms in both conditions were broken down by the three types of errors described above (Figure 7.2d-f). These supplementary analyses revealed a main effect of condition, $F(1, 8) = 38.95$, $p < 0.001$, and a condition $\times$ time window interaction, $F(1, 8) = 7.69$, $p < 0.05$, for false alarms to old objects at new positions (Figure 7.2d), but neither for false alarms to new objects at old positions (main effect: $F(1, 8) = 1.31$, $p > 0.2$; interaction: $F(1, 8) < 1$; Figure 7.2e) nor for false alarms to new objects at new positions (main effect and interaction: $F(1, 8) = 2.29$, $p > 0.1$; Figure 7.2f). Note that the latter type of errors was rarely committed, causing this non-significant effect (floor effect). Moreover, an analysis separately for both conditions revealed that the false alarm rate to old objects at new positions decreased in the invariant learning condition, $F(1, 8) = 10.36$, $p < 0.05$, but not in the context-specific condition, $F(1, 8) = 2.72$, $p > 0.1$. This differential modulation of the false alarm rates within the invariant learning condition confirms our initial hypothesis that learning takes place in the form of strengthening the representation of invariant positions within blocks.

Imaging Results: Basic Contrasts In a first step, we calculated direct contrasts between the context-specific and the invariant learning condition (Figure 7.3; Table 7.1). In the direct comparisons, only regions that survived a statistical threshold of $p < 0.0005$ (uncorrected) were considered significant. Regions that exhibited greater activation for invariant learning trials than for context-specific trials were considered to be sensitive for the formation of invariant memories. Several prefrontal regions, including the inferior portion of the left middle frontal gyrus (Brodmann Area [BA] 46; peak Talairach coordinates x, y, z : $-40, 30, 24$), the opercular part of the right inferior frontal gyrus (BA 45; $40, 28, 19$), the triangular part of the left inferior frontal gyrus (BA 45; $-57, 20, 17$), and the right inferior frontal sulcus (BA 9; $36, 21, 25$) were activated in the invariant learning relative to the context-specific

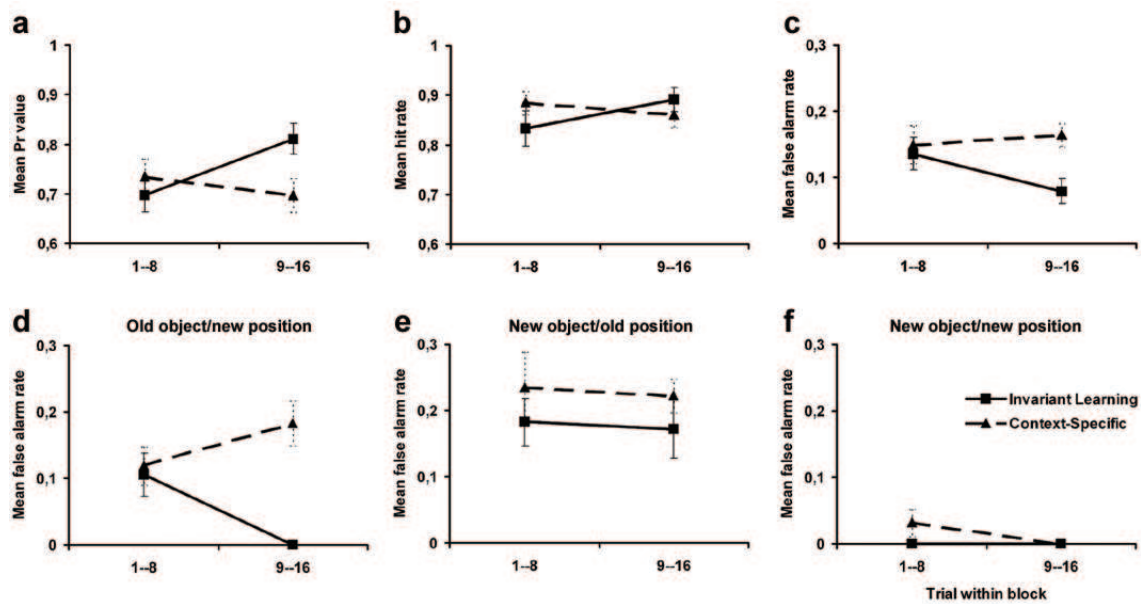


Figure 7.2. Behavioral Results. Performance measures ($\pm$ s.e.m.) are plotted as a function of time within experimental blocks for the invariant learning (solid) and context-specific condition (dashed), collapsed across all blocks and subjects. (a) Mean Pr values, (b) mean hit rate, (c) mean false alarm rate. (d)-(f) The false alarms are further divided in three categories, depending on the specific probe type, i.e. the mean false alarm rate for probes including (d) an old object at a new position, (e) a new object at an old position, and (f) a new object at a new position, respectively. The performance measure was averaged across trials 1-8 and 9-16, respectively. [Adapted from Doeller et al. (in press). Copyright Oxford University Press (Cerebral Cortex).]

condition. In addition to this lateral prefrontal pattern this contrast also revealed activation in the right ventral striatum (26, 2, -7), the right inferior parietal lobule (BA 40; 51, -42, 44 / 42, -56, 45), and the right lingual gyrus (22, -47, 2).

The comparison between the context-specific and the invariant learning condition revealed enhanced activity in the left posterior cingulate cortex (BA 31; -20, -33, 48) and along the right superior occipital sulcus (BA 19; center activity: 34, -75, 28). All experimental trials were included in this analysis. When solely considering correct responses in the analysis, a highly similar pattern for both contrasts could be observed, with the exception that the left inferior parietal lobule additionally showed a significant activation in the contrast between the invariant learning condition and the context-specific condition (see Table 7.2 for details).

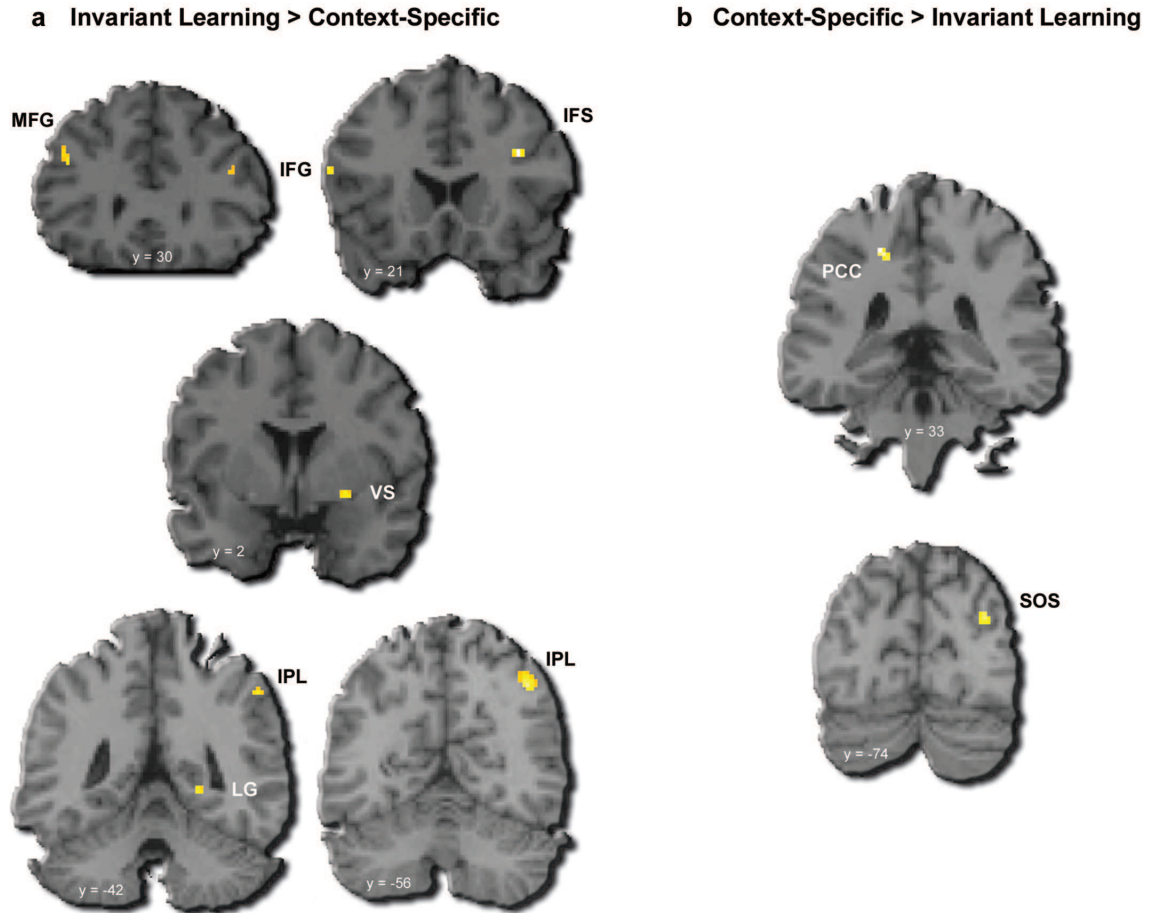


Figure 7.3. Direct Comparisons between Conditions. SPMs superimposed on coronal sections of the MNI T_1 -weighted MRI template showing activated regions for the contrasts comparing the invariant learning and the context-specific condition. Talairach y-coordinate is given below each image, respectively. (a) Trials in the invariant learning condition revealed greater hemodynamic activity than trials in the context-specific condition in the left middle frontal gyrus (MFG), the opercular part of the right inferior frontal gyrus (IFG), the triangular part of the left IFG, the right inferior frontal sulcus (IFS), the right ventral striatum (VS), the right inferior parietal lobule (IPL), and the right lingual gyrus (LG). (b) The left posterior cingulate cortex (PCC) and the right superior occipital sulcus (SOS) showed stronger BOLD responses in the context-specific condition as compared to the invariant learning condition (for further details see Table 7.1). [Adapted from Doeller et al. (in press). Copyright Oxford University Press (Cerebral Cortex).]

Imaging Results: Learning-Related Activity In accordance with the behavioral analysis, we conducted a 2×2 ANOVA with the factors condition and time (BOLD activity for trials 1-8 and trials 9-16 of the experimental blocks, separately for both conditions) to get a first estimate of differential learning-related activation pattern in both conditions. This analysis was restricted to the hippocampus

Table 7.1

fMRI Activation Foci: Direct Contrasts (All Trials). Regions activated in the direct contrasts between the context-specific and invariant learning condition (from anterior to posterior), described in terms of Brodmann areas (BA), hemisphere (L, left; R, right), Talairach coordinates (mm; transformed from the MNI-space), and peak Z-score. All regions pass the statistical threshold of $p < 0.0005$ (uncorrected). All trials were included in this analysis.

Region	BA	Hemisphere	Talairach			Z-score
			x	y	z	
(1) Invariant Learning > Context-Specific						
Inferior middle frontal gyrus	46	L	-40	30	24	3.66
Inferior frontal gyrus (opercular part)	45	R	40	28	19	3.50
Inferior frontal gyrus (triangular part)	45	L	-57	20	17	3.86
Inferior frontal sulcus	9	R	36	21	25	4.30
Ventral striatum	–	R	26	2	-7	3.67
Inferior parietal lobule	40	R	51	-42	44	3.58
	40	R	42	-56	45	4.02
Lingual gyrus	30	R	22	-47	2	3.81
(2) Context-Specific > Invariant Learning						
Posterior cingulate cortex	31	L	-20	-33	48	4.11
Superior occipital sulcus	19	R	34	-74	28	3.76

and to the activation foci found in the direct comparisons between conditions (see Methods). A subset of the regions showing a main effect of condition in the direct comparisons also exhibited a significant condition $\times$ time interaction, including the triangular part of the left inferior frontal gyrus ($-57, 20, 17; z_{max} = 1.75$), the left posterior cingulate cortex ($-16, -29, 47; z_{max} = 2.70$), the right inferior parietal lobule ($48, -46, 48; z_{max} = 3.12$), and the right lingual gyrus ($24, -47, 1; z_{max} = 2.48$). Moreover, this analysis revealed a significant interaction between condition and time in the right ($24, -8, -10; z_{max} = 2.70$ / $38, -20, -14; z_{max} = 2.53$) and left ($-18, -10, -13; z_{max} = 3.41$ / $-26, -18, -16; z_{max} = 4.46$) hippocampus.

On the basis of the interaction between condition and time, suggesting a differential involvement of several brain regions in both conditions as a function of time, we

Table 7.2

fMRI Activation Foci: Direct Contrasts (Correct Trials). Regions activated in the direct contrasts between the context-specific and invariant learning condition. Only correct trials were included in this analysis (for further details see Table 7.1 legend).

Region	BA	Hemisphere	Talairach			Z-score
			x	y	z	
(1) Invariant Learning > Context-Specific						
Inferior middle frontal gyrus	46	L	-40	28	24	3.75
Inferior frontal gyrus (opercular part)	45	R	40	28	19	3.66
Inferior frontal gyrus (triangular part)	45	L	-55	26	19	3.88
Inferior frontal sulcus	9	R	36	21	25	4.22
Ventral striatum	–	R	28	4	-5	3.83
Inferior parietal lobule	40	R	54	-41	44	3.45
	40	L	-42	-45	41	3.77
	40	R	40	-56	45	4.10
Lingual gyrus	30	R	22	-47	2	3.82
(2) Context-Specific > Invariant Learning						
Posterior cingulate cortex	31	L	-18	-35	46	4.06
Superior occipital sulcus	19	R	34	-74	28	3.70

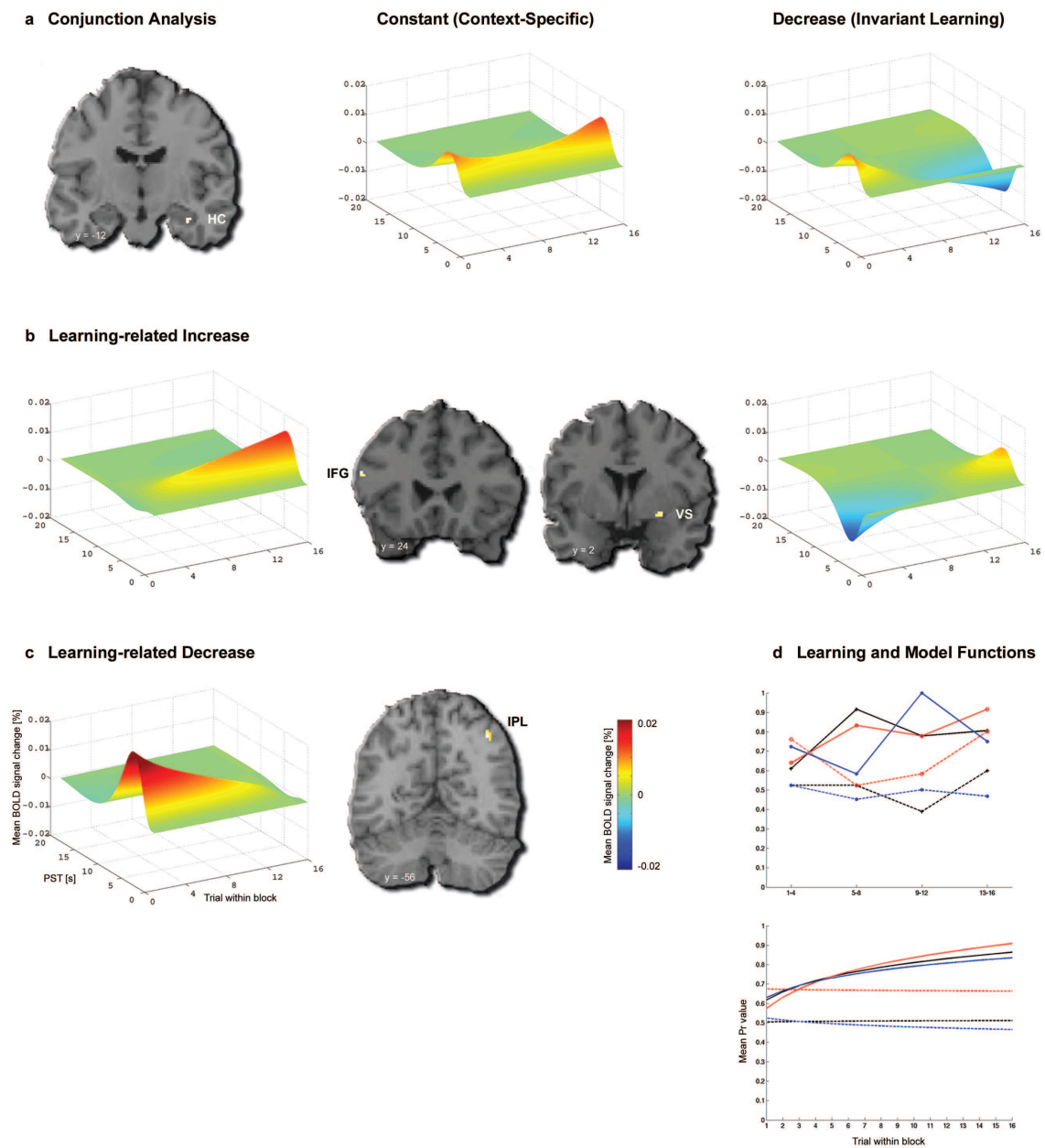
further examined learning-related activation pattern in more detail by means of parametric fMRI analyses. For this purpose we weighted the fMRI time series separately for both conditions with individually fitted logarithmic learning functions. Thereby learning-related changes in hemodynamic activation could be modeled explicitly (see Methods). All individual learning functions were well approximated by a logarithmic function (mean functions and goodness of fit: $y = 0.10 \cdot \ln(t_i) + 0.67$; $R^2 = 0.72$ [invariant learning condition]; $y = -0.07 \cdot \ln(t_i) + 0.72$; $R^2 = 0.71$ [context-specific condition]). In a first parametric analysis, we tested our a priori hypotheses about a differential learning-related hippocampal activation pattern for both conditions. For this purpose, we conducted the following conjunction analysis (Friston et al., 1999a), restricted to both hippocampi ($p < 0.05$, small volume corrected): learning-related decrease in the invariant learning condition *and* constant activity in the context-

specific condition. This conjunction analysis revealed an activation of the right hippocampus (34, -12, -16; Figure 7.4a; Table 3). As apparent from Figure 7.4a (middle panel), the hippocampus showed continuously enduring activation within blocks of the context-specific condition. To further control whether the hippocampus was responsive across all trials in the context-specific condition, we conducted a *t*-test against the null-hypothesis of no activation for the peak BOLD response of the hippocampus. For each subject, the peak BOLD response was determined for each trial of the blocks and subsequently averaged across all trials. This analysis revealed a significant activation of the hippocampus, $t(8) = 3.84$, $p < 0.01$.

A second parametric analysis tested whether brain regions showing a main effect of condition in the direct comparisons also exhibit a learning-related activation pattern in the invariant learning condition. Thus, this analysis was restricted to the activation foci found in the direct comparisons between conditions ($p < 0.05$, small volume corrected; see Methods). The triangular part of the left inferior frontal gyrus (BA 45; -53, 24, 19) and the right ventral striatum (28, 2, -7) showed learning-related increases of activity within the experimental blocks (Figure 7.4b). In contrast, right inferior parietal lobule (BA 40; 38, -56, 47) activity decreased as a function of learning within blocks (Figure 7.4c). Moreover, a learning-related decrease of activation could be observed in the superior occipital sulcus (BA 19; 34, -76, 26). None of the additional regions found in the direct comparison between both conditions exhibited a learning-related activation pattern in this analysis.

7.4 Discussion

Our behavioral data show that subjects learned spatial regularities across trials and did benefit from this knowledge in the invariant learning condition. This was reflected in increased *Pr* values, due to increasing correct detection and decreasing false alarm behavior relative to the context-specific condition during the time-course of experimental blocks. The data also indicate that learning is based on reducing false alarm responses to those probes that include new positions. Here, subjects benefit most of all from learning spatial regularities, probably based on an enhanced



Please see next page for Figure legend.

Figure 7.4. Brain Regions Showing a Learning-related Activation Pattern. (a) Results of the conjunction analysis. The right hippocampus (HC) showed constant activity in the context-specific condition *and* decreased activity during the time course of learning in the invariant learning condition. (b) Regions demonstrating increased learning-related activity in the invariant learning condition included the opercular part of the left IFG and the right VS. (c) Furthermore, the parametric analysis revealed a learning-related activation decrease in the right IPL. Time courses of the best-fitting parametric BOLD response relative to grand mean over voxels, time-locked to the sample phase onset against post-stimulus time (PST) are plotted across trials of the experimental blocks. The response is collapsed across experimental blocks and averaged across participants. For the conjunction analysis (a), parametric BOLD responses are plotted for the hippocampus separately for the context-specific (middle panel) and invariant-learning condition (right panel). For (b) and (c), parametric responses are depicted on the left and right side of the structural scan, respectively (for further details see Table 7.3). (d) Examples of learning functions (upper panel) and model functions (lower panel) from three selected subjects (black, red, blue), separately for the invariant learning (solid) and the context-specific condition (dashed). Learning functions were derived by averaging mean Pr values for four consecutive trials across all blocks separately for both conditions. The model functions were derived from the individual learning functions by fitting a logarithmic function $y = a \cdot \ln(t_i) + b$. [Adapted from Doeller et al. (in press). Copyright Oxford University Press (Cerebral Cortex).]

Table 7.3

fMRI Activation Foci: Learning-related Activity. Regions showing a learning-related activation pattern. (1) Conjunction analysis: learning-related decrease (invariant learning condition) and constant activity (context-specific condition). (2-3) Regions with a learning-related (2) increase and (3) decrease in the invariant learning condition. SPMs were thresholded at $p < 0.05$ (corrected), using small volume corrections (for further details see Table 7.1 legend).

Region	BA	Hemisphere	Talairach			Z-score
			x	y	z	
(1) Conjunction Analysis						
Hippocampus	—	R	34	-12	-16	1.73
(2) Learning-Related Increase						
Inferior frontal gyrus (triangular part)	45	L	-53	24	19	2.06
Ventral striatum	—	R	28	2	-7	1.88
(3) Learning-Related Decrease						
Inferior parietal lobule	40	R	38	-56	47	2.37
Superior occipital sulcus	19	R	34	-76	26	2.34

selectivity of the spatial representations (Yeshurun & Carrasco, 1998).

The fMRI results point to dissociable neural correlates for context-specific and invariant memories. In the parametric analysis the right hippocampus showed continuously enduring activity in the context-specific condition, in which subjects encode variable objects at variable positions. Importantly, when subjects benefit from learning invariant spatial features across trials right hippocampal activity decreased in the invariant learning condition. Furthermore, we could dissociate a prefrontal-striatal-parietal network supporting learning regularities. The lateral prefrontal cortex and the ventral striatum showed a learning related increase of activity, whereas activity in the inferior parietal lobule decreased as a function of learning.

Hippocampus and pattern-separated, relational representation It has been proposed that the hippocampus uses sparse, pattern-separated representations to encode arbitrary conjunctions or bindings of features defining an episode (O'Reilly & Norman, 2002). This mechanism results in highly distinct, non-overlapping representations of episodes and minimizes interference between episodes. In a similar vein, Eichenbaum (2000) argues that the hippocampus is responsible for binding multiple inputs and by this optimally represents the relations between temporally and spatially disparate features comprising a complex episode (relational memory framework) (Cohen & Eichenbaum, 1993; Cohen et al., 1999). Accordingly, in our task four different objects have to be bound to four different positions and these four object-positions bindings have further to be bound together in each trial to provide a complete, relational representation of each episode. In accordance with both frameworks (Eichenbaum, 2000; O'Reilly & Norman, 2002), in the present experiment the hippocampus was activated continuously across experimental blocks in the context-specific condition. Here, each trial comprised a unique episode, i.e. the probability of each possible object-position conjunction was constant across trials, resulting in highly variable feature conjunctions in each trial. Consequently, constant hippocampal relational binding operations were required to encode unique episodes into separated memory representations.

In the invariant learning condition, in contrast, there were four invariant positions in

each block. Here, the relational binding requirements of each episode decreased during the time-course of each block, since the probability for specific object-position conjunctions was substantially increased (e.g. a specific object presented at one of the four positions held constant in a block), whereas the probability for other object-position conjunctions was reduced to zero (e.g. an object presented at one of the remaining 12 positions never occurring in one block). Thus, invariant positions in object-position conjunctions might cause a reduced relational representation of the current episode. As the objects can be bound to the same positions within a block, learning presumably results from facilitated object-position binding. The decreasing hippocampal activation in the invariant learning condition may thus reflect these lower relational binding requirements. It is important to note that in the invariant learning condition a singular object-position conjunction is presented only once in each block, i.e. is not repeated within one block. By this, our experimental manipulation did not entail the repetition of a specific object-position conjunction. Thus, the decrease of hippocampal activation cannot be attributed to a higher amount of repetition of object-position conjunctions in the invariant learning condition.

This line of arguments is confirmed by previously observed decreased hippocampal activity during the time-course of probabilistic cue-outcome learning (Poldrack et al., 2001), associative learning (Zeineh et al., 2003), artificial grammar learning (Opitz & Friederici, 2003), and sequence learning (Grafton et al., 1995; Schendan et al., 2003). For instance, Schendan et al. (2003) could show a learning-related modulation of hippocampal activity using a serial reaction time task. Subjects learned complex sequential structures of stimuli and responses. The authors observed activation of the hippocampus that was more pronounced during initial phases of learning than in the final phase of learning. In line with the relational memory account (Cohen & Eichenbaum, 1993; Eichenbaum, 2000), Schendan and colleagues argue that the hippocampus is involved in the acquisition of higher-order associations, i.e. relations among temporally discontinuous events, most pronounced at the beginning of learning. It is conceivable that in the Schendan et al. study as well as in our study, relational representations of episodes get reduced due to overlapping episodic features and by this less hippocampal relational processing is required.

Moreover, the hippocampal activation was right-lateralized in the present study. Numerous studies have demonstrated that the hippocampus mediates the processing of spatial relationships (Burgess et al., 2002; Eichenbaum et al., 1999). Neuroimaging and neuropsychological studies suggest a preferential role of the right hippocampus in spatial memory (Maguire, Valentine, Wilding & Kapur, 2003; Smith & Milner, 1981) and spatial navigation (Burgess et al., 2001; Grön et al., 2000; Maguire et al., 2000). Consistent with these findings, the present right hippocampal activation decrease in the invariant learning condition may reflect the lower spatial processing requirements as compared to the context-specific condition, i.e. the processing of invariant spatial features from trial to trial.

Given the small set of 16 objects and 16 positions used in the present study, it could be argued that learning was not restricted to the invariant learning condition but also took place in the context-specific condition. To test this hypothesis, we conducted a post-hoc analysis, contrasting mean performance (Pr values) in blocks 1-3 with mean performance in blocks 5-7 for the context-specific condition. This analysis revealed no differences between both blocks ($t(8) < 1$; mean Pr: 0.65 ± 0.05 [block 1-3]; 0.66 ± 0.06 [block 5-7]). Furthermore, in an additional fMRI analysis we contrasted hemodynamic responses in block 1-3 with block 5-7 (and vice versa) in the context-specific condition. Hippocampal activation did not differ significantly between the first and the last blocks ($p < 0.05$, small-volume corrected). These supplementary analyses suggest that learning was absent or negligible and hippocampal activation did not change across blocks in the context-specific condition.

Brain regions exhibiting greater activity for context-specific as compared to invariant learning trials include the posterior cingulate cortex (PCC) and the superior occipital sulcus (SOS). The PCC is part of the extended hippocampal diencephalic system (Aggleton & Brown, 1999). The integrity of this system is a prerequisite for successful memory operations. However, the precise function of the PCC is an issue of current debate. Furthermore, the SOS as part of the dorsal visual processing stream (Ungerleider & Mishkin, 1982) might be recruited due to processing variable spatial features from trial to trial in the context-specific relative to the invariant learning condition. In accordance with this view, SOS activity decreased during the

time course of experimental blocks in the invariant learning condition, since spatial processing requirements get reduced.

Lateral prefrontal cortex and learning regularities As predicted, lateral prefrontal regions were activated in the invariant learning relative to the context-specific condition. Moreover, the triangular part of the left inferior frontal gyrus showed an activation increase as a function of learning in the invariant learning condition. Based on these results, we propose that the increase of lateral PFC activity in the invariant learning condition reflects (1) the extraction and the maintenance of the invariant positions *across* trials and (2) the adjustment of object-position conjunctions in each trial on the basis of the extracted information, resulting in an efficient encoding strategy.

First, we argue that the lateral PFC is involved in extracting task-relevant invariant spatial features during the initial trials. After regularities are successfully extracted, these regularities have to be actively maintained during the whole experimental block. This view receives support from numerous studies demonstrating that the lateral PFC plays a crucial role in active maintenance of information against distraction (Miller, Erickson & Desimone, 1996; Sakai, Rowe & Passingham, 2002) and in suppressing interference from varying sources (Bunge, Ochsner, Desmond, Glover & Gabrieli, 2001; Mecklinger, Weber, Gunter & Engle, 2003). These processes are an important prerequisite for rule extraction and formation and by this might mediate rule-based learning (Ashby & Ell, 2001).

Second, during the time-course of the experimental blocks object-position conjunctions could be reorganized in light of the representations of the extracted regularities. Supporting evidence is provided by a recent fMRI study (Bor et al., 2003). When subjects could integrate items into higher-level chunks, performance as well as lateral prefrontal activity increased. In the same vein, lateral prefrontal regions showed increased activity, when information is stored in a bound, rather than in a separated representation (Prabhakaran, Narayanan & Gabrieli, 2000). In both studies, subjects benefit from a reorganization of items during encoding. Lateral PFC involvement has been consistently found during episodic memory tasks (Wagner, 2002). The lat-

eral PFC has been posited to subserve specific control operations during memory encoding, like the reorganization, evaluation, and manipulation of the items to be memorized (Fletcher & Henson, 2001; Simons & Spiers, 2003). In a recent computational model of prefrontal memory control, Becker and Lim (2003) could show that the PFC represents internal mnemonic codes, rapidly adjusted to current task requirements. Furthermore, Becker and Lim's simulations reveal that these mnemonic codes evolve via reinforcement mechanisms during the time-course of the experiment. In accordance with these accounts, we propose that the PFC is involved in the adjustment of representations of incoming object-position conjunctions on the basis of the extracted regularities, i.e. the knowledge of four invariant positions in one block. It could be argued that the extracted regularities are represented as an internal mnemonic code (Becker & Lim, 2003). In light of this mnemonic code, object-position conjunctions are encoded more efficiently, as reflected in increased PFC activity in the invariant learning as compared to the context-specific condition. Moreover, the learning-related increase of lateral PFC activity indicates that this process evolves across trials and by this entails increased task performance in the invariant learning condition. Taken together, the dynamic PFC activation pattern seems to reflect the implementation of an efficient encoding strategy in the face of task-relevant regularities. These data underscore the special importance of the lateral prefrontal cortex during regularity learning.

Beyond the lateral prefrontal cortex right ventral striatum activity increased as a function of learning. The ventral striatum as part of the basal ganglia has been associated with habit learning, the gradual acquisition of stimulus-stimulus and stimulus-response associations (Jog et al., 1999; Packard & Knowlton, 2002), e.g. during probabilistic classification learning (Knowlton et al., 1996; Poldrack et al., 2001) and sequence learning (Schendan et al., 2003; Willingham et al., 2002). It has been proposed that the striatum and the hippocampus comprise interactive memory systems, specialized for distinct memory processes, namely relational binding of features defining episodes (hippocampus) and the acquisition of stimulus-stimulus associations (striatum), respectively (Packard & Knowlton, 2002; Poldrack et al., 2001; Poldrack & Packard, 2003). This is in accordance with the complementary

learning-related activation pattern observed in the hippocampus and the ventral striatum, respectively. Hippocampal activity decreased, whereas ventral striatal activity increased as a function of learning.

Another line of neurophysiological (Lauwereyns et al., 2002a; Waelti et al., 2001) and neuroimaging research (McClure et al., 2003; O'Doherty et al., 2003) suggests that the mesolimbic dopamine system and its main target areas, especially the striatum, play a pivotal role during classical conditioning. Based on assumptions of formal reinforcement learning theories (Rescorla & Wagner, 1972; Sutton & Barto, 1990), it has been proposed that these brain regions represent predictions of future reward delivery and by this drive learning of associations between stimuli (Schultz, 2002). In accordance with these views, it is conceivable that the observed learning-related activation of the right ventral striatum might reflect the processing of increased predictability of spatial features during the time-course of learning. It could further be argued that these modified expectations of positions serve as internal reinforcement signals supporting regularity learning (Koechlin et al., 2002; Schultz, 2002). Moreover, it has been proposed that mesolimbic reinforcement signals provided to the PFC play a crucial role in maintaining and updating of prefrontal memory representations (Cohen, Braver & Brown, 2002; Miller, 2000). Thus, it could be speculated that bottom-up reinforcement signals from mesolimbic and striatal dopamine neurons, sensitive to the predictability of spatial features might modulate processing in higher-level brain structures like the PFC (Schultz, 2002). On the basis of these reinforcement signals the PFC might implement an efficient encoding strategy (Becker & Lim, 2003).

The right inferior parietal lobule (IPL) showed stronger BOLD responses in the invariant learning as compared to the context-specific condition. However, in contrast to the PFC, the IPL activation decreased as a function of learning. The parietal cortex is assumed to be a core structure for attentional-based selection and representation of spatial features (Gottlieb, 2002). In addition, the right lingual gyrus was activated in the invariant learning condition. Similar to the IPL, this region has been associated with attentional feature processing (Hopfinger, Buonocore & Mangun, 2000). In light of these findings, we assume that the IPL is involved in

the allocation of attention to upcoming positions and in maintaining an attentional set (Corbetta & Shulman, 2002) across trials and by this supports the extraction of spatial regularities. During the time-course of experimental blocks, attentional maintenance requirements get reduced as reflected in a decrease of IPL activity.

To summarize, the present results point to a dynamic interplay between medial temporal, striatal, and lateral prefrontal brain regions during the formation of invariant memories. During initial trials, the right hippocampus is involved, since unique object-position conjunctions require hippocampal relational binding processes. After regularities are extracted, hippocampal activation decreases as a function of learning due to the reduced relational binding requirements. In contrast, both the lateral PFC and the ventral striatum showed a learning-related increase of activity. However, further studies will be required in order to examine the differential involvement of the hippocampus, the striatum, and the lateral PFC in other forms of regularity learning. In conclusion, these data indicate a transition of the relative roles of distinct neural systems during the time-course of learning, i.e. learning is accompanied by a shift from a hippocampal to a prefrontal-striatal brain system (Opitz & Friederici, 2003; Poldrack et al., 2001).

Chapter 8

Experiment 1B: Domain-specific Hemispheric Specialization of the Hippocampus during Learning Regularities across Episodes¹

Contents

8.1	Introduction	136
8.2	Materials and Methods	138
8.3	Results	139
8.4	Discussion	141

In Experiment 1A, regularity learning in the spatial domain was accompanied by a transition from a hippocampal to a prefrontal striatal brain system. In the present Experiment 1B, the question was addressed whether the observed behavioral and neural pattern could be replicated for regularity learning in another domain. More precisely, is regularity learning a domain-general or a domain-specific mechanism?

¹This Chapter contains the following manuscript: **Doeller C. F., Opitz B., Krick, C. M., Mecklinger, A., & Reith, W. (in preparation). Domain-specific hemispheric specialization of the hippocampus during learning regularities across episodes.** For stylistic reasons the format was adapted. In the following, this study will be referred to as Experiment 1B. This work was supported by grants from the German Research Foundation (DFG; Research Group FOR-448). The authors wish to thank Patric Meyer for comments on an earlier version of this manuscript.

Therefore, invariant objects in object-position conjunctions were introduced in the IL condition. All other task and design parameters were held constant as compared to Experiment 1A.

Abstract It is a widely held view that the hippocampus plays a crucial role during building episodic memories. However, it is still controversial which brain regions subserve the representation of regularities across episodes. In a recent study (Doeller et al., in press²), in which subjects had to learn object-position conjunctions in several trials, learning spatial regularities across episodes was accompanied by a shift from a hippocampal to a prefrontal-striatal brain system. Based on the evidence for a domain-specific lateralization of the hippocampus during memory encoding, the present fMRI study aimed at specifying whether a hemispheric specialization also holds for regularity learning. Thus, in contrast to invariant positions in our previous study we introduced invariant objects in object-position conjunctions across trials in the present study. In showing a learning-related increase of prefrontal-striatal activity and a decrease of hippocampal activity, the present findings resembled the results of our previous study. Most notably, relative to a right-lateralized activation decrease during learning invariant spatial features, learning invariant objects was associated with a decrease of bilateral hippocampal activation. These data indicate a domain-specific hemispheric specialization of the hippocampus during regularity learning, possibly reflecting hippocampal sensitivity to perceptual attributes of invariant episodic features.

8.1 Introduction

The formation of episodic memory critically depends on the integrity of the hippocampus within the medial temporal lobe, MTL (Squire, 1992). The importance of this brain region for memory processes has been highlighted by studies using a wide variety of neuroscientific methods in animals (Fortin et al., 2002, 2004; Wirth et al., 2003) as well as in humans (Davachi et al., 2003; Fernández et al., 1999). Recent neuroimaging research provides evidence for a domain-specific lateralization within the

²Experiment 1A

human MTL during memory encoding (cf. Paller & Wagner, 2002). Later successful memory performance has been associated with left hippocampus/MTL activation during encoding in the verbal domain (Golby et al., 2001; Kirchhoff et al., 2000; Otten et al., 2001), with bilateral hippocampus/MTL activation in the figural domain (Golby et al., 2001; Kelley et al., 1998, see also Stern et al., 1996), and with right (and bilateral) activation in the spatial domain (Golby et al., 2001; Kelley et al., 1998; Kirchhoff et al., 2000). Furthermore, recent imaging studies reported right hippocampal involvement during spatial navigation and spatial memory (Maguire et al., 1998, 2003, cf. Burgess et al., 2002).

In his relational memory account, Eichenbaum (2000) proposes that the hippocampus is involved in binding different features defining episodes (see Giovanello et al., 2004; Preston et al., 2004, for recent fMRI evidence for hippocampal involvement in relational memory binding). However, less is known about the role of MTL structures in learning invariant episodic features, i.e. regularities across multiple episodes. In a recent fMRI study, we investigated the neural correlates of learning spatial regularities across episodes (Doeller et al., in press). In each trial of several experimental blocks, subjects had to learn object-position conjunctions. The design of the experiment comprised two conditions, a context-specific (CS) and an invariant learning (IL) condition. In the CS condition, objects and positions were variably mapped across trials. In contrast, in the IL condition positions were held constant in each trial of the experimental blocks, enabling subjects to extract spatial regularities across trials, i.e. invariant positions in object-position conjunctions. The introduction of invariant spatial features entailed increased performance across trials. Our imaging results indicate that this performance increase is mediated by a shift from a hippocampal to a prefrontal-striatal brain system. Left prefrontal and right striatal (ventral putamen) brain regions showed a learning-related activation increase across trials, whereas right hippocampal activation decreased as a function of learning. We argued that the learning-related right hippocampal activation pattern reflects reduced relational binding requirements once spatial regularities are extracted. Given the evidence for right hippocampus' preferential role in spatial memory, these data further indicate that the right-lateralized hippocampal activation decrease is pre-

sumably due to reduced binding of spatial features in object-position conjunctions. In contrast to invariant *positions* (Doeller et al., in press), in the present study we examined the role of the hippocampus in learning invariant *objects* in object-position conjunctions in the same task. Given the aforementioned domain-specific hemispheric asymmetry of the hippocampus during *memory* formation, we addressed the question, whether this hemispheric specialization of the hippocampus also holds for *learning* regularities across episodes. Stimuli, task, and design parameters were identical as in the previous experiment, with the following critical exception: In contrast to Doeller et al., in press, objects, not positions were held constant within experimental blocks in the IL condition. No changes were made in the CS condition (see Figure 8.1a and Methods).

8.2 Materials and Methods

Eight right-handed subjects participated in the study (aged 22-29, mean age 25 years, 4 females). Informed consent was obtained before the experiment. Sixteen objects (basic shapes, e.g. square, circle, trapeze; black and white, filled and unfilled) were used as stimulus material. As shown in Figure 8.1a, in the sample phase of each trial four different objects were presented sequentially at four different positions within a 4×4 grid (grid not shown in the figure). Stimuli were presented for 800 ms with a 500 ms interstimulus interval. After a brief delay (1300 ms), subjects were probed for 2000 ms and had to indicate whether or not the probe stimulus was identical to one of the four object-position conjunctions presented in the sample phase, followed by a variable intertrial interval (4-9 s; mean = 5 s; including visual feedback for 500 ms). The whole experiment included seven blocks in the CS and seven blocks in the IL condition, which in turn comprised 16 trials. In the IL condition, the four objects presented in each trial were held constant across all trials of one block, i.e. the same four objects were presented in each trial of one block. In contrast, in the CS condition objects and positions were variably mapped across trials. Blocks of both condition were presented in randomized order.

BOLD-sensitive, T_2^* -weighted functional images and T_1 -weighted structural images

(MPRAGE) were acquired at 1.5 T (Siemens Sonata; $T_R = 1.8$ s, $T_E = 50$ ms, flip angle = 85° ; 20 slices; voxel size: $3.5 \times 3.5 \times 4$ mm). Analysis was performed with SPM2 (<http://www.fil.ion.ucl.ac.uk/spm/spm2.html>). Standard preprocessing procedures were applied, including correction for slice acquisition time and motion, coregistration of anatomical to functional images, normalization to the MNI T_1 standard template, resampling into 2 mm isotropic voxels, and spatial smoothing (FWHM = 8 mm). Time-series were analyzed with a two-stage random effects analysis. At a first level, for each trial of both conditions trial-specific effects were modeled using a canonical HRF for the sample phase (epoch-related; duration = 4.7 s) and the probe phase (event-related), respectively. Additional parametric analyses were conducted. Here, subject-specific time series of each experimental trial were weighted with logarithmic fitted individual learning functions (Pr values for four consecutive trials, averaged across all blocks), separately for both conditions. Linear contrasts of the parameter estimates for each regressor were calculated for each subject and brought to the second level random effects analysis. Data were high-pass filtered to $\frac{1}{128}$ and scaled for global activity. All analyses were restricted to the sample phase. MNI-coordinates of all reported activations have been transformed to the canonical Talairach space (<http://www.mrc-cbu.cam.ac.uk/imaging/mnispace.html>). See Doeller et al., 2004, for a detailed description of the methods.

8.3 Results

In the behavioral analysis, a repeated-measures ANOVA with the factors condition (IL vs CS) and time (trials 1-8 vs 9-16 within blocks) revealed constant Pr values (hits - false alarms) in the CS condition across trials within blocks, $F(1, 7) < 1$, whereas Pr values tended to increase across trials in the IL condition, $F(1, 7) = 4.25$, $p < 0.08$ (Figure 8.1b). Furthermore, the mean hit rate increased in the IL condition, $F(1, 7) = 7.02$, $p < 0.05$, but not in the CS condition, $F(1, 7) = 2.67$, $p > 0.1$. Finally, separate analyses of false alarms for probes including (a) a new object at an old position and (b) an old object at a new position revealed a marginally significant condition $\times$ time interaction for (a), $F(1, 7) = 5.17$, $p < 0.06$, but not

for (b), $F(1, 7) < 1$. Thus, relative to the CS condition the false alarm rate in the IL condition selectively diminished for probes including a new object at an old position, since subjects could reject these specific probes solely on the basis of the knowledge of invariant objects. In summary, these data indicate that learning occurred only in the IL condition.



Figure 8.1. Experimental Procedure and Behavioral/Imaging Results. (a) Trial and block structure of the experiment. In each trial (horizontal), four objects were presented sequentially at four positions during the sample phase. Upon probe presentation, subjects had to indicate whether or not the probe stimulus was identical to one of the four object-position conjunctions seen before. Each block comprised 16 trials (vertical). In the blocks of the IL condition (shown in the Figure), objects were held constant in each trial of one block. For illustration, one invariant object is highlighted by circles. In contrast, in the CS condition object-position conjunctions were trial-unique. (b) Mean Pr values (hits - false alarms; $\pm$ s.e.m.) are plotted as a function of trial within experimental blocks for the IL condition (solid) and the CS condition (dashed), collapsed across all blocks. Pr values were averaged across trials 1-8 and 9-16, respectively. (c) The left hippocampus ($-24, -29, -2$) and the right hippocampus ($30, -31, -7$) showed weaker BOLD responses in the IL relative to the CS condition. The SPM is superimposed on a coronal section of the MNI T_1 -weighted MRI template.

On the basis of our a priori hypothesis and the results of our previous study (Doeller et al., in press), in the imaging analysis specific regions of interest sensitive to regularity learning were selected a priori: the hippocampus, the putamen, and the lateral inferior/middle frontal gyrus. Statistical tests were corrected for multiple comparisons, using appropriate small volume corrections (SVC; $p < 0.05$; Worsley et al., 1996). To get a first estimate of differential hippocampal involvement in both conditions, we contrasted activity in the CS with activity in the IL condition (related to the sample phase, see Methods). As apparent from Figure 8.1c, the left hippocampus ($-24, -29, -2$; $z_{max} = 2.84$) and the right hippocampus ($30, -31, -7$; $z_{max} = 2.03$)

showed a deactivation in the IL relative to the CS condition. In addition, in this contrast a higher amount of deactivated voxels could be observed in the left as compared to the right hippocampus (29 vs 10).

In a next analysis step, regions showing a learning-related activation pattern in the IL condition were identified using a parametric analysis. For this purpose, we weighted the fMRI time series with individual learning functions (see Methods). This analysis revealed a learning-related activation decrease in the left hippocampus ($-24, -30, -9$; $z_{max} = 1.95$) and in the right hippocampus ($38, -24, -6$; $z_{max} = 2.09$). As apparent from Figure 8.2a, the BOLD response in both hippocampi decreased across trials. This observation was confirmed by a t -test contrasting the mean peak BOLD response elicited by the first trial of the blocks with the hemodynamic signal associated with the last trial of the blocks, separately for both hippocampi (left: $t(7) = 4.31$, $p < 0.005$; right: $t(7) = 5.54$, $p < 0.001$). It is noteworthy, that the bilateral hippocampal activation decrease could be observed in every subject (Figure 8.2b), emphasizing the robustness of this effect. To investigate differential hemispheric involvement in more detail, we compared the effect size (parameter estimate of the above parametric analysis) observed for the left and right hippocampus, separately for the present experiment (object regularities) and our previous study (spatial regularities; Doeller et al., in press). This analysis revealed no differences between hemispheres during learning object regularities ($p > 0.9$), however we found higher effect sizes for the right as compared to the left hippocampus during learning spatial regularities ($p < 0.05$; Figure 8.3).

In accordance with the results of our previous study (Doeller et al., in press), in the parametric analysis the right putamen ($-28, -1, 13$; $z_{max} = 3.11$) and the left middle frontal gyrus (BA 8; $-46, 12, 38$; $z_{max} = 3.01$) showed a learning-related activation increase across trials in the IL condition.

8.4 Discussion

In this study, we investigated the neural correlates of learning invariant objects in object-position conjunctions across multiple trials. In showing a learning-related

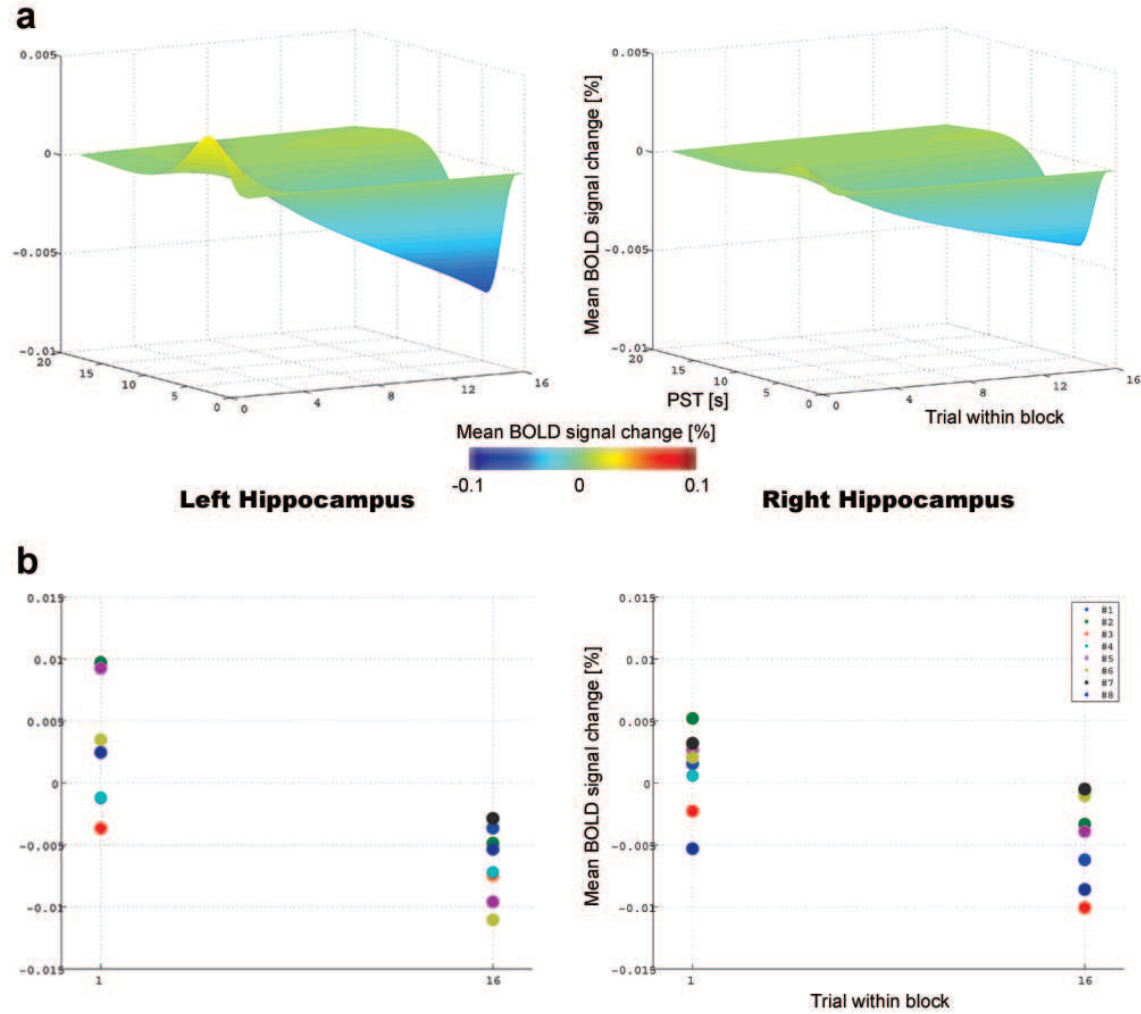


Figure 8.2. Learning-related Activation Pattern of the Hippocampus. (a) For the left ($-24, -30, -9$) and right ($38, -24, -6$) hippocampus, time courses of the best-fitting parametric BOLD response are plotted across trials of the experimental blocks in the IL condition, relative to grand mean over voxels and time-locked to the sample phase onset against post-stimulus time (PST). (b) Single subject data showing learning-related temporal changes of left and right hippocampal activity. Peak signals of individual parametric BOLD responses are plotted across the first and the last trials of the experimental blocks in the IL condition, separately for subjects #1-#8.

increase of activity in lateral prefrontal and striatal regions and a learning-related decrease of hippocampal activity, we could replicate the results of our previous study (Doeller et al., in press). Whereas the left prefrontal and right striatal activity resembled the activation pattern found during learning spatial regularities (Doeller et al., in press), a hemispheric asymmetry could be observed in the hippocampus for learning spatial relative to object regularities. In contrast to the right-lateralized ac-

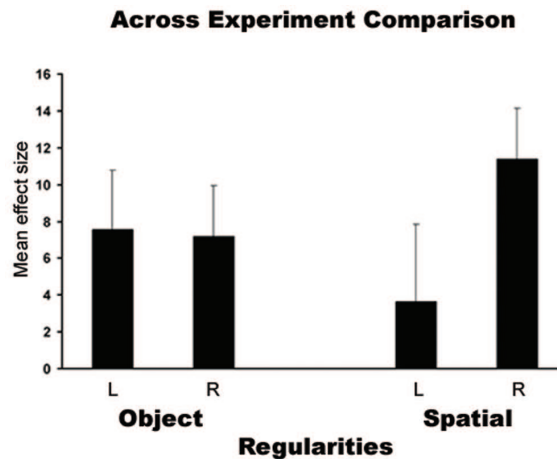


Figure 8.3. Across Experiment Comparison. Mean effect size (parameter estimate) of the parametric analysis ($\pm$ s.e.m.), separately for both hippocampi (L vs R) and experiments (left: object regularities [the present study]; right: spatial regularities [Doeller et al., in press]).

tivation in our previous study, the hippocampus showed a bilateral learning-related activation pattern in the present experiment.

In terms of the relational memory framework (Eichenbaum, 2000), in each trial of the present task four objects have to be bound sequentially to four positions. Furthermore, an exhaustive representation of the whole episode (i.e. the four object-position conjunctions in one trial) requires pattern separation of the specific conjunctions defining this episode (O'Reilly & Norman, 2002). We argue that the learning-related decrease of hippocampal activation reflects lower binding requirements in the IL condition, since invariant objects could be bound to variable positions. It might be possible, that long-term potentiation in the hippocampus (cf. Malenka & Nicoll, 1999) could serve as the molecular mechanism underlying the reduction of relational binding and by this might support regularity learning. Conversely, in the CS condition, trial-unique object-position bindings have to be encoded. Thus, relational binding requirements remain constant across episodes, as indicated by enhanced hippocampal activity in the CS relative to the IL condition.

In addition to the aforementioned fMRI studies, supporting evidence for a domain-specific lateralization is provided by patient studies, suggesting a preferential role of the left hippocampus for verbal memory (Frisk & Milner, 1990) and of the right

hippocampus for spatial memory (Smith & Milner, 1981). Furthermore, in patients with Alzheimer's disease left hippocampal volume is correlated with verbal memory recall, whereas right hippocampal volume is correlated with spatial memory recall (de Toledo-Morrell et al., 2000, but see Helmstaedter, Brosch, Kurthen, and Elger, 2004, and Richardson, Strange, Duncan, and Dolan, 2003, for recent evidence for a hemispheric reorganization of verbal memory function following left hippocampal pathology). One possible explanation for the hippocampal lateralization during memory formation was provided by Golby et al., 2001. The authors argue that the lateralization depends on the availability of verbal codes. Moreover, an influential theory proposes that memory encoding is subserved by two independent subsystems, a verbal and a non-verbal processing system (Paivio, 1986). In accordance with these views, it could be speculated that objects, in contrast to positions, could be more easily transformed into verbal codes and that these verbal codes are more easily accessed for objects than for positions. Consequently, objects might induce both verbal and non-verbal codes, whereas the representation of spatial features might be predominantly based on non-verbal codes. Thus, the observed bilateral hippocampal activation decrease during learning object relative to spatial regularities might be due to reduced binding of verbally and non-verbally coded object features in object-position conjunctions. However, future work is required to address this issue in more detail.

The domain-specific hippocampal specialization during regularity learning is consistent with the hemispheric asymmetry during memory encoding (e.g. Golby et al., 2001), thus indicating that the hippocampal contribution to regularity learning and memory encoding might be based on highly similar mechanisms. Furthermore, this hemispheric asymmetry might suggest that the hippocampus operates in a stimulus-driven processing mode, sensitive to perceptual attributes of invariant episodic features. This stimulus-driven processing mode is possibly mediated by projections from distinct visual brain regions via the rhinal cortex, providing information about various stimulus attributes to the hippocampus (cf. Murray & Bussey, 1999).

Similar to the hippocampus, there is also evidence for a domain-specific lateralization in the prefrontal cortex (PFC) during memory formation (Golby et al., 2001; Kelley

et al., 1998), i.e. left PFC regions have been associated with phonological memory operations, whereas right PFC areas have been implicated in visuo-spatial memory processes (cf. Wagner, 2002). Interestingly, we could not observe any hemispheric differences between our experiments with respect to the PFC. During learning object regularities as well as during learning spatial regularities (Doeller et al., in press), prefrontal activation foci were left-lateralized. Thus, prefrontal operations involved during regularity learning seem to be, at least partially different from those supporting memory formation. In contrast to the bottom-up, stimulus-driven hippocampal operations, the prefrontal cortex might mediate regularity learning in a top-down manner by means of more abstract, rule-driven operations (Opitz & Friederici, 2003, 2004; Strange et al., 2001; Wallis et al., 2001). However, it is noteworthy that two recent fMRI studies (Dobbins, Simons & Schacter, 2004; Mitchell, Johnson, Raye & Greene, 2004) indicate that PFC laterality during memory retrieval rather depends on the type of retrieval process than on the domain of the material being retrieved. In both studies, the left lateral PFC was recruited during monitoring specific memory contents, whereas the right lateral PFC was preferentially engaged during heuristic memory retrieval operations. However, beyond the area of memory further research is required to investigate whether distinct learning operations might differentially rely on the left and the right PFC, respectively.

In summary, the present findings suggest a domain-specific hemispheric specialization of the hippocampus during regularity learning, thus indicating that the hippocampus is sensitive to invariant perceptual stimulus features defining regularities across multiple episodes.

Chapter 9

Experiment 2: Cognitive Mechanisms Subserving Regularity Learning¹

Contents

9.1	Introduction	148
9.2	General Methods	150
9.3	Experiment 2A: Distinctiveness of Objects I: Object Features	152
9.3.1	Introduction	152
9.3.2	Experimental Procedures	153
9.3.3	Results	153

¹This Chapter contains main parts of the following article: **Doeller, C. F. & Opitz, B. (2004). Cognitive and neural mechanisms subserving regularity learning. In A. Mecklinger, H. Zimmer, & U. Lindenberger (Eds.), *Bound in memory: insights from behavioral and neuropsychological studies* (pp. 129-157). Aachen: Shaker Verlag.** For stylistic reasons the format was adapted. Thus, this version differs in minor ways from the final published version. In the following, this study will be referred to as Experiment 2. The six experiments of this study will be referred to as Experiment 2A-2F, respectively. This work was supported by grants from the German Research Foundation (DFG; Research Group FOR-448). The authors wish to thank Lea Meyer and Henning Loebbecke for their assistance during data acquisition. The authors are also grateful to Hubert Zimmer who provided helpful comments.

9.3.4	Discussion	154
9.4	Experiment 2B: Distinctiveness of Objects II: Object Types	155
9.4.1	Introduction	155
9.4.2	Experimental Procedures	155
9.4.3	Results	156
9.4.4	Discussion	157
9.5	Experiment 2C: Memory Load and Learning Phase I . .	157
9.5.1	Introduction	157
9.5.2	Experimental Procedures	158
9.5.3	Results	159
9.5.4	Discussion	159
9.6	Experiment 2D: Memory Load and Learning Phase II .	160
9.6.1	Introduction	160
9.6.2	Experimental Procedures	161
9.6.3	Results	161
9.6.4	Discussion	162
9.7	Experiment 2E: Distinctiveness of Objects Revisited . .	162
9.7.1	Introduction	162
9.7.2	Experimental Procedures	163
9.7.3	Results	163
9.7.4	Discussion	164
9.8	Experiment 2F: Simultaneous Object-Position Bindings	164
9.8.1	Introduction	164
9.8.2	Experimental Procedures	165
9.8.3	Results	165
9.8.4	Discussion	166
9.9	General Discussion	166

In Experiment 1A and 1B, regularity learning was reflected in an increase of performance across trials within blocks in the invariant learning condition. On a behavioral

level, differences could be observed between regularity learning in the spatial and object domain, especially with respect to a selective decrease of false alarms for probes including new positions and objects, respectively. The aim of the present series of behavioral experiments was to investigate the cognitive processes involved in regularity learning in more detail.

Abstract It is an issue of current debate which brain structures support regularity learning. In a previous fMRI study² we could show that prefrontal-striatal brain regions are involved in learning spatial regularities across episodes, whereas hippocampal activity decreased as a function of learning. In a follow-up study,³ we examined the neural correlates of learning object regularities. In comparison to our previous study we found neural and behavioral differences between both stimulus domains. Based on these results, we conducted six behavioral experiments investigating the cognitive processes underlying object regularity learning in more detail which will be described in this chapter. In each trial of the experiments, subjects had to learn object-position conjunctions. Crucially, objects in object-position conjunctions were held constant across trials of experimental blocks, enabling subjects to learn regularities across episodes. More precisely, we investigated the influence of object distinctiveness (Experiment 2A-2B), memory load and learning duration (Experiment 2C-2E), and the effect of simultaneous object-position bindings (Experiment 2F) on learning object regularities in object-position conjunctions. Our data indicate that the increase of object distinctiveness mainly affects overall memory performance, whereas manipulation of the learning duration and the memory load affects the learning process.

9.1 Introduction

In Experiment 1A (Chapter 7), subjects showed increased performance across trials in the invariant learning condition, but not in the context-specific condition. During learning spatial regularities, lateral prefrontal and striatal activity increased, whereas hippocampal activity decreased. In Experiment 1B (Chapter 7), the behav-

²Experiment 1A

³Experiment 1B

ioral and imaging results of Experiment 1A could be replicated. In contrast to the context-specific condition, increased performance could be observed in the invariant learning condition. Moreover, learning object regularities was accompanied by a learning-related activation increase of prefrontal-striatal brain regions and by a learning-related decrease of hippocampal activity. In contrast to a right-lateralized hippocampal dynamic during learning spatial regularities, learning object regularities was associated with a more bilateral (left-pronounced) hippocampal activation decrease. However, learning effects were slightly more pronounced in Experiment 1A as compared to Experiment 1B. When contrasting performance in the invariant learning condition of both experiments, separately for the first and the last half of the experimental blocks, performance did not differ in the first half ($t(15) = 1.01$, $p > 0.3$; mean Pr values: 0.70 ± 0.03 [Experiment 1A], 0.64 ± 0.04 [Experiment 1B]), but in the second half ($t(15) = 2.12$, $p < 0.06$; mean Pr values: 0.81 ± 0.03 [Experiment 1A], 0.72 ± 0.03 [Experiment 1B]). Several explanations of these findings remain possible. For instance, these data might suggest that learning object regularities evolves slower as compared to learning spatial invariances. It might also be possible that objects and by this object regularities are less distinctive as compared to positions. Consequently, the behavioral experiments described in this chapter aimed at specifying in more detail the cognitive processes underlying learning object regularities.

In a series of six behavioral studies the effect of object distinctiveness on learning effects were investigated by means of using different object features (Experiment 2A) and different object types (Experiment 2B). By this, the detection of regularities should be facilitated. Furthermore, several critical changes of the core paradigm were made, especially increased memory load (increased number of object-position bindings) in each trial and increased block length to cover the learning phase more widely (Experiment 2C and 2D). Finally, the interaction of object distinctiveness and learning phase (Experiment 2E) and the effect of simultaneous object-position bindings (Experiment 2F) on regularity extraction was investigated in more detail.

9.2 General Methods

Stimuli, Task, and Design In all six behavioral experiments described in this chapter, 16 different objects were used as stimulus material. Stimuli in Experiment 2A consisted of basic shapes, presented in four different colors. In Experiment 2B-2D, pictures denoting real-life objects were used as stimulus material. Abstract objects were used in Experiments 2E and 2F. Within one experimental trial, stimuli were presented within a 4×4 grid, i.e. there were sixteen possible positions. In each trial, four (Experiment 2A-2B) or six stimuli (Experiment 2C-2F) were presented sequentially (Experiment 2A-2E) or simultaneously (Experiment 2F) during a sample phase (cf. Figure 6.1). After a short delay a probe stimulus was presented at a particular position within the grid. Participants had to indicate whether or not the probe stimulus was identical to one of the object-position combinations seen before during the sample phase. Responses were delivered by a button press with the right or left index finger. In all experiments, visual feedback was given for 500 ms. Stimulus type, number of stimuli presented in the sample phase, stimulus duration, interstimulus (ISI) interval, delay duration, and probe duration were selectively manipulated in the experiments (see Table 9.1). Each experimental session comprised several blocks in the invariant learning condition. Each of these blocks comprised 16-36 trials. In these blocks, objects were held constant across trials, i.e. in each trial of one block, the same objects were presented. In contrast, positions were selected randomly from trial to trial. The temporal order of the invariant objects was randomized across trials of one block. The configuration of objects held constant across trials of one block changed from block to block. Participants were informed about the beginning of a new block. In Experiment 2A and 2B, in addition to the invariant learning condition, the design comprised a second condition, the context-specific condition. Here, object-position conjunctions were unique in each trial, i.e. there were no object regularities across trials. Blocks of both conditions were presented in randomized order in Experiment 2A and 2B. The sequence of blocks was balanced across subjects. Volunteers were paid or were given course credit for participating. Informed consent was obtained before the measurements (see also Methods sections

for each experiment).

Table 9.1
Task and Design Parameters of Experiment 2A-2F

Exp.	Stimuli	Stimuli/ sample phase	Stimulus duration [ms]	Presen- tation format	ISI [ms]	Delay [ms]	Probe duration [ms]	Trials/ blocks	Blocks	Subjects	Age
2A	basic shapes	4	800	sequ.	500	1300	2000	16	7 (per cond.)	16 (11 female)	17-35 (24)
2B	real-life objects	4	800	sequ.	500	1300	2000	16	7 (per cond.)	16 (10 female)	19-33 (23)
2C	real-life objects	6	800	sequ.	500	1300	2000	36	6	16 (8 female)	19-27 (22)
2D	real-life objects	6	500	sequ.	0	2000	1000	36	6	16 (13 female)	19-28 (24)
2E	abstract objects	6	500	sequ.	0	2000	1000	36	6	16 (13 female)	19-33 (22)
2F	abstract objects	6	2000	simult.	0	1000	1000	36	8	16 (10 female)	18-26 (22)

Note. sequ. = sequential presentation, simult. = simultaneous presentation of object-position conjunctions in the sample phase.

Data Analysis To have a more sensitive index for object regularity extraction, Pr values (proportion hits - proportion false alarms; Feenan & Snodgrass, 1990) were used as performance measure in all behavioral experiments. Instead of the false alarm rate to all new probes only false alarms to probes comprising new objects were included, i.e. probes comprising a new object at an old or a new position, respectively. Based on results of Experiment 1A (Chapter 7), we assumed that learning should be most clearly revealed by a reduction of false alarms for these probe types, since subjects could reject these probes solely on the basis of their knowledge of the invariant objects. In the following, this performance measure will be referred to as ‘corrected Pr’.

In Experiment 2A and 2B, in a first analysis step two-way repeated-measure ANOVAs with the factors condition (context-specific vs invariant learning) and time (four levels: mean corrected Pr values for four consecutive trials each, i.e. mean performance in the four quarters of the blocks) were conducted. In a subsequent step, separately for each condition one-way repeated-measure ANOVAs with the factor time were conducted.

In Experiment 2C-2F, in a first analysis step one-way repeated-measure ANOVAs with the factor time (six levels: mean corrected Pr values for six consecutive trials

each, i.e. mean performance in the six sixth parts of the blocks) were conducted. In a subsequent step, one-way repeated-measure ANOVAs with the factor time were conducted separately for block 1-3 (block 1-4, Experiment 2F) and block 4-6 (block 5-8, Experiment 2F). Finally, two-way repeated-measure ANOVAs with the factor block (two levels: block 1-3 vs block 4-6 [block 1-4 vs block 5-8 in Experiment 2F]) and time (six levels: mean Pr values for six consecutive trials each) were conducted. Corrections for multiple contrasts were used where appropriate. Uncorrected degrees of freedom and corrected p values are reported. In all experiments, post-hoc contrasts between specific levels of the factor time and tests for linear trends were calculated.

9.3 Experiment 2A: Distinctiveness of Objects I: Object Features

9.3.1 Introduction

The stimuli used in Experiment 1A and 1B comprised six basic shapes. By additionally presenting some of these stimuli rotated and in a filled and an unfilled version, the study comprised 16 stimuli. A possible explanation for the slightly less pronounced learning effects for object regularities in Experiment 1B could be that the 16 stimuli used in this study were not distinctive enough and by this object regularities were harder to detect as compared to spatial regularities. Accordingly, Experiment 2A addresses the question whether more distinctive objects and by this more distinctive object regularities would lead to increased performance within experimental blocks in the invariant learning condition. For this purpose, the same stimuli used in Experiment 1B were presented in four different colors. Importantly, the subsets of objects which were held constant across trials of experimental blocks in the invariant learning condition were made more distinctive (see below). By this, the extraction of object regularities should be facilitated. A performance increase within task blocks in the invariant learning condition was predicted, whereas performance in the context-specific condition should be constant across trials.

9.3.2 Experimental Procedures

Stimuli, Task, and Design Sixteen stimuli which consisted of basic shapes (e.g. square, circle) were used as stimulus material (the same stimuli were used in Experiment 1A and 1B), with the exception that an additional object feature, color, was introduced (see Figure 9.1a). Four colors were used, blue, green, red, and yellow, i.e. four objects were presented in the same color, respectively. The four colors were systematically assigned to objects of the same shape (four sets consisting of four triangles, circles/ellipses, trapezes, squares/rhomboids, respectively), e.g. all four triangles were blue.

Each experimental block comprised 16 trials. One experimental session comprised seven blocks in the context-specific condition and seven blocks in the invariant learning condition. Blocks of both conditions were presented in randomized order with the constraint that participants had to perform at most two blocks of the same condition in succession. Subjects were informed about the beginning of a new block. The sequence of blocks was balanced across subjects.

In the invariant learning condition, four objects were held constant within one block, whereas in the context-specific condition, object-to-position assignment was randomized. In contrast to Experiment 1A and 1B, the seven subsets of the four stimuli which were held constant in the seven experimental blocks in the invariant learning condition, were not selected randomly. In fact, to make the regularities more distinctive each of the seven subsets contained each color, each basic shape, and consisted of either four filled or four unfilled objects. Moreover, an intertrial interval (ITI) of 2000 ms was used. All other task and design parameters were held constant as compared to Experiment 1A and 1B.

9.3.3 Results

Figure 9.1b shows mean corrected Pr values for both conditions. Subjects showed better performance in the invariant learning as compared to the context-specific condition. This observation was confirmed by a main effect of condition, $F(1, 15) = 21.36$, $p < 0.001$. Furthermore, the ANOVA revealed a main effect of time, $F(3, 45) =$

7.80, $p < 0.001$. However, in contrast to our predictions no condition $\times$ time interaction was observed, $F(3, 45) < 1$. A main effect of time in the invariant learning condition was marginally significant, $F(3, 45) = 2.75$, $p < 0.07$. This was however partially due to a performance increase at the end of the blocks, as reflected in a significant performance increase between the third and the fourth quarter of the blocks, $F(1, 15) = 8.23$, $p < 0.05$. Critically, a contrast between the first and last quarter revealed no performance difference in the invariant learning condition, $F(1, 15) < 1$. Contrary to the predictions, performance in the context-specific condition was not constant across trials (main effect of time: $F(3, 45) = 3.70$, $p < 0.05$), mainly caused by a decrease of performance between the first and second quarter of the blocks, $F(1, 15) = 7.58$, $p < 0.05$. A comparison of the first and the last quarter revealed a significant decrease of performance within blocks in the context-specific condition, $F(1, 15) = 4.83$, $p < 0.05$.

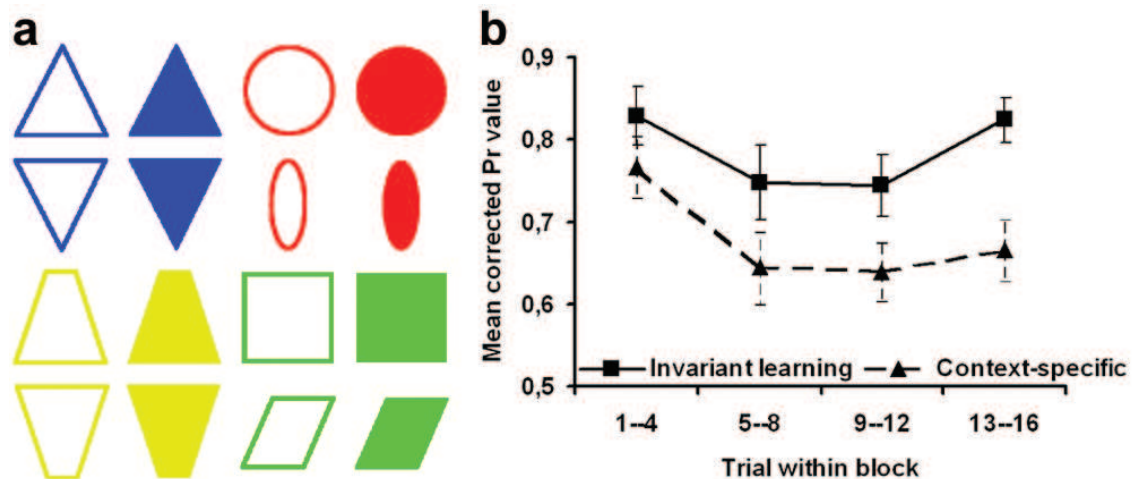


Figure 9.1. Stimuli and Results in Experiment 2A. (a) The modified 16 stimuli used in the study. (b) Mean corrected Pr values (proportion hits - proportion false alarms to new objects; $\pm$ s.e.m.) plotted as a function of trial number within experimental blocks for the invariant object learning (solid) and context-specific (dashed) condition collapsed across all blocks. The performance measure was averaged across four consecutive trials, respectively. [Adapted from Doeller and Opitz, 2004. Copyright (2004) Shaker Verlag.]

9.3.4 Discussion

Subjects showed better task performance in the invariant learning condition than in the context-specific condition. However, we could not observe any performance

increase in the invariant learning condition, when comparing performance at the beginning and the end of the blocks (a performance increase was observed only at the end of the blocks). Moreover, contrary to our predictions, performance decreased in the context-specific condition. One possible explanation of the present results could be that the color stimuli are still not distinctive enough to facilitate the proposed extraction process. The following Experiment 2B will test this hypothesis by using more distinctive objects.

9.4 Experiment 2B: Distinctiveness of Objects II: Object Types

9.4.1 Introduction

The aim of the present study was to test the hypothesis that object distinctiveness would facilitate detecting object regularities and by this would lead to a performance increase within task blocks in the invariant learning condition. It could be argued that objects used in Experiment 2A per se are too similar to each other. For this purpose highly distinctive line-drawings denoting real-life objects were used as stimulus materials. We expected increasing performance in the invariant learning condition and constant performance across trials in the context-specific condition.

9.4.2 Experimental Procedures

Stimuli, Task, and Design In this experiment, the stimulus set comprised 16 real-life objects (see Figure 9.2a). The stimulus set contained eight living and eight non-living objects. The subsets of objects which were held constant within the invariant object learning blocks were randomly selected from the stimulus set with the constraint that each subset comprised two living and two non-living objects. All other task and design parameters were held constant as compared to Experiment 2A.

9.4.3 Results

Figure 9.2b shows mean corrected Pr values for both conditions. The two-way repeated-measure ANOVA revealed a condition $\times$ time window interaction, $F(3, 45) = 5.21$, $p < 0.01$. Moreover, a marginally significant main effect of condition, $F(1, 15) = 4.17$, $p < 0.06$, and time, $F(3, 45) = 2.57$, $p < 0.08$, was observed. Contrary to the predictions, ANOVAs performed separately for both conditions revealed a significant effect of time in the context-specific condition, $F(3, 45) = 6.40$, $p < 0.01$, but not in the invariant learning condition, $F(3, 45) = 1.88$, $p > 0.1$. A (marginally) significant linear trend was observed in the context-specific condition, $F(1, 15) = 7.12$, $p < 0.05$, and in the invariant learning condition, $F(1, 15) = 4.32$, $p < 0.06$. Corrected Pr values in the first and fourth quarter of the invariant learning condition differed at a more liberal statistical threshold, $F(1, 15) = 3.42$, $p < 0.09$. Finally, an additional ANOVA contrasting the first and fourth quarter in both conditions revealed a significant condition $\times$ time interaction, $F(1, 15) = 8.25$, $p < 0.05$.

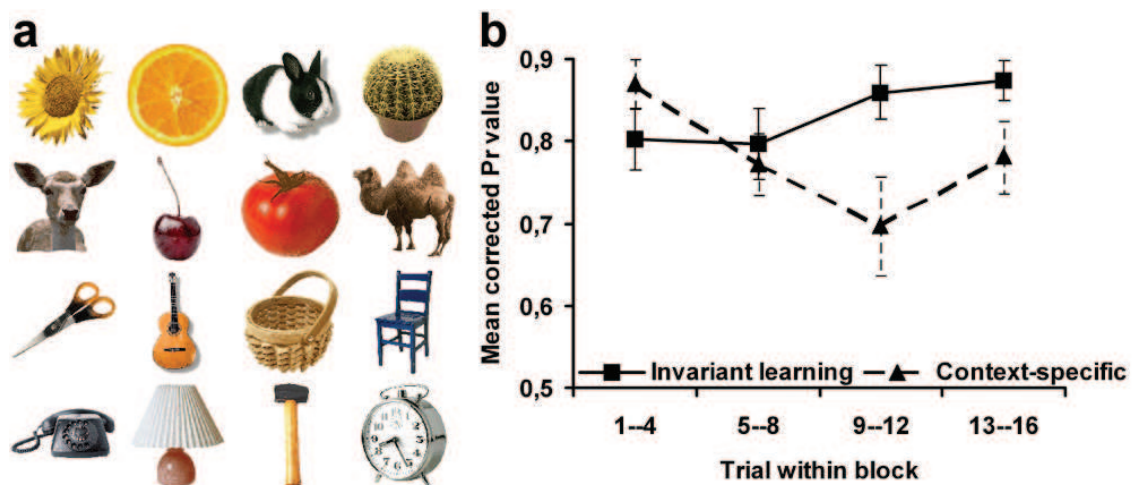


Figure 9.2. Stimuli and Results in Experiment 2B. (a) The 16 real-life stimuli used in the study. (b) See Figure 9.1 legend for more details. [Adapted from Doeller and Opitz, 2004. Copyright (2004) Shaker Verlag.]

9.4.4 Discussion

Similar to Experiment 2A, subjects were better in the invariant learning as compared to the context-specific condition and a marginally significant linear trend was observed in the invariant learning condition, indicating that participants did benefit from constant objects across trials. Moreover, a significant condition x time window interaction was observed when considering the first and last quarter of the blocks in both conditions. These data indicate that increased distinctiveness facilitated learning regularities across trials. However, these effects were small. Moreover, no significant increase in performance during the time-course of the invariant learning blocks was observed. Three explanations for these results remain possible. First, as compared to Experiment 1B overall performance was extremely high in Experiment 2A and 2B. Given this high performance level, even at the beginning of the blocks, it could be argued that a performance increase across trials is hardly detectable (ceiling effect). Second, especially in Experiment 2A a slight increase of performance was observed at the end of the blocks, indicating that extracting object regularities might be a slower mechanism as compared to the extraction of spatial regularities. Finally, the pseudorandom sequence of blocks in the invariant learning and the context-specific condition, respectively could have entailed that subjects performed the task by a common strategy for both conditions. More precisely, the task allowed subjects to succeed with a solely ‘context-specific strategy’ in both conditions, i.e. not to extract regularities. The following experiments aim at specifying these hypotheses.

9.5 Experiment 2C: Memory Load and Learning Phase I

9.5.1 Introduction

The goal of the present study was to investigate the influence of memory load and the duration of the learning phase on performance measures. Given the high overall performance in the previous experiments, even at the beginning of the blocks, mem-

ory load was increased by increasing the number of object-position bindings per trial from four to six. By this manipulation, a lower performance baseline in the first trials of the experimental blocks was expected. In addition, block length was substantially increased in Experiment 2C from 16 to 36 trials. By increasing the learning duration, we allowed subjects to reach a maximum performance level. Moreover, based on the observation that performance increase in Experiment 2A occurred at the end of the block, it might be possible that the process of extracting object regularities is slower than spatial regularity extraction. To minimize the probability that subjects based their judgment on a common strategy for both conditions, trials were blocked by condition.

9.5.2 Experimental Procedures

Stimuli, Task, and Design The same stimuli as in Experiment 2B were used. As compared to the previous experiments we modified the following trial and block parameters. In each trial six different stimuli were presented sequentially at six different positions within the grid. Moreover, one experimental block comprised 36 instead of 16 trials. Each subject performed a total of six blocks in the invariant learning condition, including a 10 min brake after 3 consecutive blocks. No context-specific condition was measured.⁴ In contrast to Experiment 2A and 2B, an exponential distributed ITI was used (mean ITI: 5 s).⁵ In order to increase the reliability of the Pr values, the performance measure was averaged across six consecutive trials which comprised three old and three new responses. Finally, the subsets of objects which were held constant within the invariant learning blocks were randomly selected from

⁴The focus of the present and the following behavioral experiments was to investigate regularity extraction in more detail. Therefore the number of blocks in the invariant learning condition was maximized, however at the cost of excluding the context-specific condition.

⁵In this and the following experiments, the same ITI was used as in the fMRI experiments to minimize the task differences between behavioral and imaging experiments. In a similar vein, a 10 min brake after three experimental blocks was introduced since in the following fMRI studies (Experiments 3A and 3B, Chapter 10), in the middle of the experiment the structural MRI measurements were conducted (approx. 10 min).

the stimulus set with the constraint that each subset comprised three living and three non-living objects. In contrast to the previous experiments, the subsets were variable across subjects.

9.5.3 Results

When analyzing mean corrected Pr values collapsed across all six blocks, neither a main effect of time, $F(5, 75) = 1.74$, $p > 0.1$, nor a significant difference between the first and last sixth part, $F(1, 15) = 3.07$, $p > 0.1$, was observed (see Figure 9.3a). However, ANOVAs separately for the first three and the last three blocks of the experiment (see Figure 9.3b) revealed a main effect of time in blocks 1-3, $F(5, 75) = 2.38$, $p < 0.05$, but not in blocks 4-6, $F(5, 75) = 1.66$, $p > 0.1$. This was confirmed by a block $\times$ time interaction, $F(5, 75) = 2.36$, $p < 0.05$. Moreover, a significant difference between the first and last sixth part was observed in blocks 1-3, $F(1, 15) = 8.25$, $p < 0.05$, but not in blocks 4-6, $F(1, 15) < 1$. Similarly, the trend analysis revealed a significant linear trend in blocks 1-3, $F(1, 15) = 10.36$, $p < 0.01$, but not in blocks 4-6, $F(1, 15) < 1$. Finally, t -tests were conducted to compare performance in Experiment 2A and 2C. Performance was better in the first quarter of Experiment 2A (first 4 trials across blocks) as compared to the first sixth part of Experiment 2C (first 6 trials across blocks), $t(30) = 2.12$, $p < 0.05$. Performance at the end of the blocks of both experiments did not differ significantly (last quarter in Experiment 2A and last sixth part of Experiment 2C), $t(30) < 1$.

9.5.4 Discussion

As predicted, the increased memory load in each trial entailed reduced Pr values at the beginning of experimental blocks as compared to Experiment 2A. Furthermore, a performance increase and a linear trend across trials was observed in the first three blocks, but not in blocks 4-6. The main experimental manipulations were successful. The increased memory load and the prolonged learning phase entailed learning effects, however only in the first three blocks. These data indicate that the learning maximum is reached earlier in the last as compared to the first blocks. This

dissociation between both halves of the experiment might indicate that regularity extraction across episodes is based on two distinct mechanisms: a process operating within blocks and a process operating across blocks. It could be speculated that the former extraction process might be involved in the detection of feature invariances and the latter, higher-order process might be involved in transferring the extraction mechanism to new exemplars.

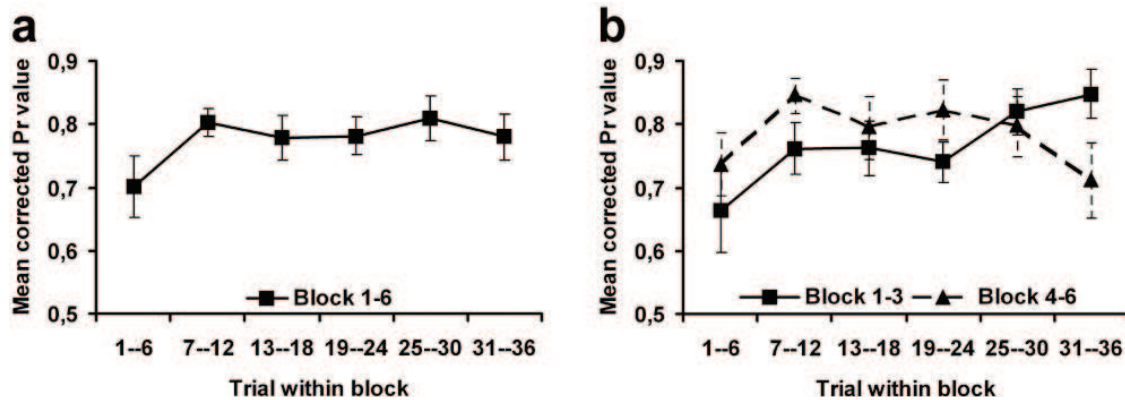


Figure 9.3. Results in Experiment 2C. Mean corrected Pr values ($\pm$ s.e.m.) plotted as a function of trial number within experimental blocks for the invariant object learning condition collapsed across (a) all blocks (solid) and (b) separately across blocks 1-3 (solid) and blocks 4-6 (dashed). The performance measure was averaged across six consecutive trials, respectively. [Adapted from Doeller and Opitz, 2004. Copyright (2004) Shaker Verlag.]

9.6 Experiment 2D: Memory Load and Learning Phase II

9.6.1 Introduction

Experiment 2C revealed a learning effect across trials in the first half of the experiment. Overall performance at the beginning of the blocks was reduced as compared to Experiment 2A, however it could be hypothesized that the learning effect might be further enhanced if the task would be more difficult, driving the need for learning to succeed in the task. By this, reduced performance in the first phase of the experimental blocks was expected. Based on this manipulation subjects may benefit to a higher degree from extracting regularities in a late phase of the blocks. A final

aim of the following experiment was to replicate the dissociation between blocks 1-3 and blocks 4-6 observed in Experiment 2C.

9.6.2 Experimental Procedures

Stimuli, Task, and Design In contrast to Experiment 2C, we increased task difficulty by modifying the timing parameters in the experimental trials. The stimulus duration was reduced to 500 ms and the interstimulus interval to 0 ms. The probe was presented for 1000 ms instead of 2000 ms. Accordingly, the response window was reduced to 1000 ms. Finally, the delay between sample and probe presentation was increased from 1300 to 2000 ms. All other task and design parameters were held constant as compared to Experiment 2C.

9.6.3 Results

As apparent from Figure 9.4a, performance increased across trials when collapsing performance across all six blocks. This impression was confirmed by a main effect of time, $F(5, 75) = 5.00$, $p < 0.001$, and a significant performance difference between the first and the last sixth part of the blocks, $F(1, 15) = 22.71$, $p < 0.001$. Moreover, a significant linear trend was observed when considering all blocks, $F(1, 15) = 17.98$, $p < 0.001$. Figure 9.4b shows mean corrected Pr values across trials separately for the first and last three blocks. Performance increased in blocks 1-3, $F(5, 75) = 2.82$, $p < 0.05$, but not in blocks 4-6, $F(5, 75) < 1$. In accordance with this observation, performance in the first and last sixth part differed significantly in blocks 1-3, $F(1, 15) = 4.99$, $p < 0.05$, but not in blocks 4-6, $F(1, 15) = 1.40$, $p > 0.2$. In addition, corrected Pr values in both halves of the experiment differed significantly, $F(1, 15) = 22.13$, $p < 0.001$. The trend analysis separately for both halves of the experiment revealed a marginally significant linear trend in blocks 1-3, $F(1, 15) = 4.45$, $p < 0.06$, but not in blocks 4-6, $F(1, 15) < 1$. When comparing the first and last sixth part of Experiment 2C and 2D, a marginally significant difference was observed at the beginning of experimental blocks, $t(30) = 1.78$, $p < 0.09$, but not at the end of the blocks, $t(30) < 1$.

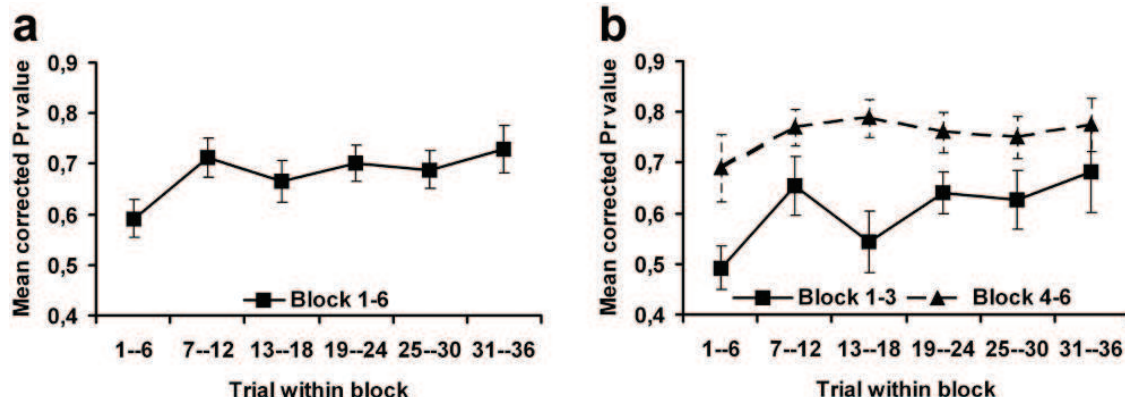


Figure 9.4. Results in Experiment 2D. See Figure 9.3 legend for more details. [Adapted from Doeller and Opitz, 2004. Copyright (2004) Shaker Verlag.]

9.6.4 Discussion

Similar to Experiment 2C, the analyses revealed a dissociation between the first and second half of the experiment, i.e. increased performance across trials in blocks 1-3 and constant performance across trials in the last three blocks. This pattern of results underscores the existence of two distinct mechanisms involved in regularity extraction. Furthermore, the task difficulty manipulation was successful. Performance was reduced at the beginning but not at the end of the blocks and by this the learning effect was slightly more pronounced as compared to Experiment 2C.

9.7 Experiment 2E: Distinctiveness of Objects Revisited

9.7.1 Introduction

The main design modifications in Experiments 2C and 2D, i.e. increased memory load and increased block length, were successful. Learning-effects were found, most pronounced in the first three blocks. However, based on the previous findings it could not be ruled out that the learning effects were, at least partially due to the distinctiveness of the real-life objects. Thus, the present experiment was designed to investigate the influence of less distinctive objects on performance by using six ab-

stract geometrical stimuli instead of real-life objects (see Figure 9.5a). If the learning effect solely depends on the modified task parameters (increased memory load and prolonged learning phase), similar effects as in the two previous experiments were expected. If regularity extraction and by this increased performance across trials depends on the distinctiveness of the objects, diminished learning effects for abstract geometrical stimuli were expected.

9.7.2 Experimental Procedures

Stimuli, Task, and Design In this study, 16 abstract geometrical objects derived from the Microsoft Windows font Klinzhai ('Klingon letters') were used as stimulus material (see Figure 9.5a; cf. Mecklinger, Bosch, Gruenewald, Bentin & von Cramon, 2000). All other task and design parameters were held constant as compared to Experiment 2D.

9.7.3 Results

In Figure 9.5b and 9.5c, subjects performance is plotted as a function of time within experimental blocks collapsed across all blocks and separately for blocks 1-3 and blocks 4-6, respectively. The statistical analysis of all blocks revealed a marginally significant effect of time, $F(5, 75) = 2.17$, $p < 0.07$. However, neither the first and last sixth part differed significantly, $F(1, 15) < 1$, nor a significant linear trend was observed, $F(1, 15) = 1.12$, $p > 0.3$. Similar to Experiment 2C and 2D, a significant effect of time was found in the first three blocks, $F(5, 75) = 3.63$, $p < 0.05$, but not in blocks 4-6, $F(5, 75) = 1.34$, $p > 0.2$. When comparing performance in the first and last sixth part of the blocks, neither in blocks 1-3, $F(1, 15) = 1.66$, $p > 0.2$, nor in blocks 4-6, $F(1, 15) < 1$, performance differed significantly. Finally, the trend analysis separately for both halves of the experiment revealed a significant linear trend in blocks 1-3, $F(1, 15) = 9.00$, $p < 0.01$, but not in blocks 4-6, $F(1, 15) < 1$.

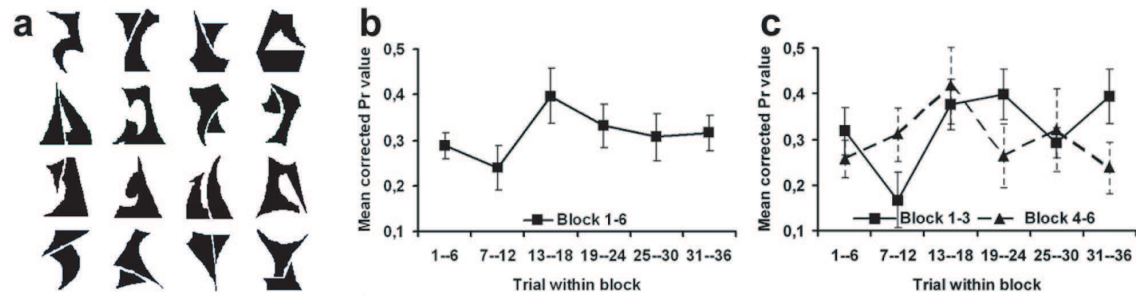


Figure 9.5. Stimuli and Results in Experiment 2E. (a) The 16 abstract geometrical objects used in the study. (b,c) See Figure 9.3 legend for more details. [Adapted from Doeller and Opitz, 2004. Copyright (2004) Shaker Verlag.]

9.7.4 Discussion

The observed results are consistent with the learning pattern of Experiment 2C and 2D. A dissociation between both halves of the experiment was found, i.e. increased performance across trials in the first blocks and constant performance across trials in the last blocks. Thus, these findings are compatible with the interpretation that regularity extraction is based on two distinct mechanisms, one operating within and one operating across blocks. Furthermore these data suggest that mainly the modifications of the task parameters (in particular increased memory load and block length) entailed the observed learning effects and not object distinctiveness per se.

9.8 Experiment 2F: Simultaneous Object-Position Bindings

9.8.1 Introduction

The present experiment was designed to investigate the influence of simultaneously presenting the six object-position conjunctions in each trial in contrast to the successive presentation in all previous experiments. If regularity extraction is independent of the presentation form and the presentation duration, similar learning effects were expected, most pronounced in the first blocks comparable to Experiments 2C, 2D, and 2E. An additional motivation of this study was an efficiency consideration. By simultaneously presenting all objects within one trial, trial duration could be re-

duced significantly. This gives rise to an increased trial number and by this to more reliable sampling of the learning functions.

9.8.2 Experimental Procedures

Stimuli, Task, and Design In this study, the same 16 abstract geometrical objects as in Experiment 2E were used as stimulus material. In contrast to the previous studies, we modified the trial structure of the experiment. In each trial, six different stimuli were presented simultaneously at six different positions within the grid for 2000 ms. By this the total presentation time for all six stimuli within one trial was reduced from 3000 ms to 2000 ms as compared to Experiment 2D and 2E. After the presentation of a mask for 500 ms and a fixation cross for 500 ms the probe stimulus was presented for 1000 ms. Thus, the delay between sample and probe presentation was reduced from 2000 ms to 1000 ms (mask and fixation presentation) as compared to Experiments 2C-2E. The experiment comprised eight experimental blocks in the invariant learning condition, including a 10 min brake after four blocks. All other task and design parameters were held constant as compared to Experiment 2E.

9.8.3 Results

A marginally significant main effect of time was observed when considering corrected Pr values for all eight blocks, $F(5, 75) = 2.01$, $p < 0.1$ (Figure 9.6a), but neither in blocks 1-4, $F(1, 15) = 1.40$, $p > 0.2$, nor in blocks 5-8, $F(1, 15) < 1$, performance across trials increased significantly (Figure 9.6b). Similarly, the contrast between the first and the last sixth part revealed no significant differences when considering all blocks, $F(1, 15) = 3.33$, $p < 0.09$, and separately blocks 1-4, $F(1, 15) = 2.05$, $p < 0.1$, and blocks 5-8, $F(1, 15) < 1$. The trend analysis revealed a significant linear trend when considering all blocks, $F(1, 15) = 5.06$, $p < 0.05$. However the trend analysis separately for both halves of the experiment did not reach significance in blocks 1-4, $F(1, 15) = 2.43$, $p > 0.1$, and blocks 5-8, $F(1, 15) = 3.70$, $p < 0.08$.

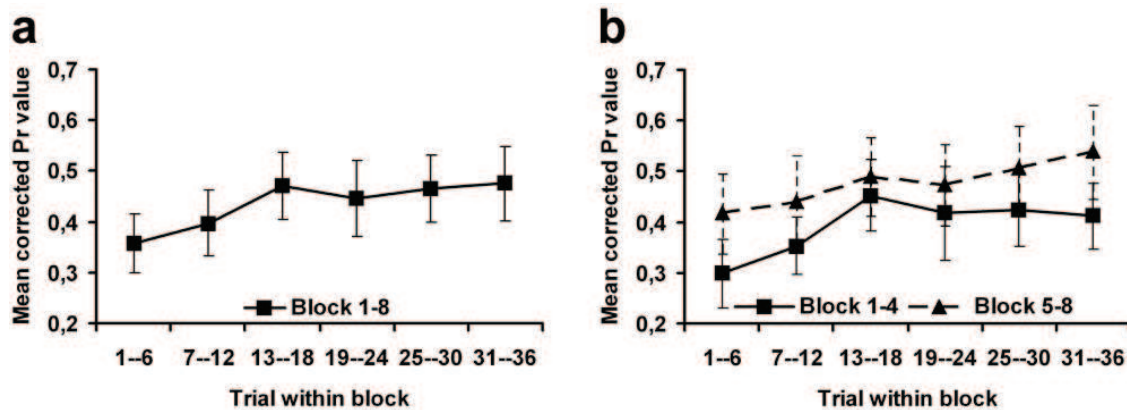


Figure 9.6. Results in Experiment 2F. See Figure 9.3 legend for more details. Note, this experiment comprised eight blocks. [Adapted from Doeller and Opitz, 2004. Copyright (2004) Shaker Verlag.]

9.8.4 Discussion

Contrary to Experiment 2C-2E, the learning effects did not reach significance neither in the first half nor in the second half of the experiment. However, a linear trend was observed within blocks. It could be speculated that the evolvement of the extraction process might be hampered when stimulus presentation time falls below a critical threshold. An alternative explanation remains possible. It could be argued that the detection of invariant objects initially requires a clear separated representation of the single objects defining the critical regularity. The strength of such ‘entity representations’ might be diminished when objects are presented simultaneously for a brief period.

9.9 General Discussion

In six behavioral experiments the cognitive mechanisms underlying the extraction of object regularities were investigated. When positions were held constant within learning blocks, a performance increase across trials was observed (Experiment 1A). Using similar design and task parameters, performance increase was slightly less pronounced during learning object regularities (Experiment 1B). In the present series of behavioral experiments, several hypotheses for these findings were tested. In Experiment 2A and 2B, distinctiveness of objects was increased. However, we could

not find pronounced learning effects in both experiments. Neither the introduction of additional, distinctive but redundant object features (color) in Experiment 2A, nor the application of distinct, real-life objects in Experiment 2B entailed substantial learning effects across trials. One explanation for these results could be that learning was hardly detectable in these experiments, since overall performance was on a high level. Therefore we increased task difficulty in the following experiments by increasing memory load and by decreasing stimulus duration. Furthermore, we found a slight learning effect at the end of the learning blocks in Experiment 2A. Based on this observation, it could be further hypothesized that object regularity extraction is a slow process as compared to spatial regularity extraction. Therefore, the learning phase was substantially augmented in Experiment 2C-2F. These experimental manipulations succeeded. In Experiment 2C and 2D we observed pronounced learning effects. Furthermore, the sustained learning increase, at least in the first blocks of Experiment 2C and 2D indicates that regularity extraction seems to be a gradual process evolving across multiple experiences. Furthermore, Experiment 2E, also employing an augmented learning phase similar to Experiment 2C and 2D revealed a small learning effect even though abstract, less distinctive objects were used as compared to Experiments 2C and 2D. These data provide supporting evidence for the view that an extended learning phase is a necessary and sufficient prerequisite for regularity extraction (if performance is on a marginal level at the beginning of learning, as in Experiment 2C-2E where memory load was increased as compared to Experiment 2A and 2B). Given the more robust learning effects in Experiments 2C and 2D, where salient real-life objects were used as stimulus material it could be argued that object distinctiveness might be a factor that additionally enhances the proposed extraction process. In line with this, the diminished learning effects in Experiment 2F indicate that the extraction process might further depend on separated, distinct instead of simultaneous representations of the object-position conjunctions. Finally, the analyses of Experiments 2C and 2D revealed a clear time-sensitive dissociation of the learning effects. Learning effects were pronounced in the first half of the experiments, but diminished in the second half, indicating that learning was faster in the second as compared to the first half of the experiments.

Taken together, the results of the present behavioral experiments indicate that in comparison to Experiment 1B the increase of object distinctiveness mainly affects overall memory performance, whereas manipulation of the learning duration and the memory load affects the learning process. The results of the present six behavioral experiments will be discussed in more detail in the General Discussion of the present thesis (Chapter 11).

Chapter 10

Experiment 3: Differential Hippocampal and Prefrontal-striatal Contributions to Instance-based and Rule-based Learning¹

Contents

10.1 Introduction	171
10.2 Materials and Methods	173
10.3 Results	178
10.4 Discussion	187

¹This Chapter contains the following manuscript: **Doeller C. F., Opitz B., Krick, C. M., Mecklinger, A., & Reith, W. (in preparation). Differential hippocampal and prefrontal-striatal contributions to instance-based and rule-based learning.** For stylistic reasons the format was adapted. In the following, Experiment 1 and Experiment 2 of this study will be referred to as Experiment 3A and Experiment 3B, respectively. This work was supported by grants from the German Research Foundation (DFG; Research Group FOR-448). The authors wish to thank Lea Meyer and Henning Loebbecke for their assistance during data acquisition.

Experiment 2 revealed a dissociation of two learning mechanisms, namely within-block learning and across-block learning. The learning asymptote within blocks was reached earlier in the final blocks as compared to the initial blocks. The present Experiment 3 aimed at specifying whether separable brain regions are specialized for either mechanism. In two experiments either an object IL condition (Experiment 3A) or a spatial IL condition (Experiment 3B) was paired with a respective CS condition. In both experiments, blocks of both conditions were presented in separate sessions to minimize the probability that subjects based their judgment on a common strategy for both conditions.

Abstract It is a topic of current interest whether learning in humans relies on the acquisition of abstract rule knowledge (rule-based learning) or whether it depends on superficial item-specific information (instance-based learning). Here we identified brain regions that mediate either of the two learning mechanisms by combining fMRI with an experimental protocol shown to be able to dissociate both learning mechanisms. Subjects had to learn object-position conjunctions in several trials and blocks. In a learning condition, either objects (Experiment 1) or positions (Experiment 2) were held constant within blocks. In contrast to a control condition in which object-position conjunctions were trial-unique, a performance increase within and across blocks was observed in the learning condition of both experiments. We argue that within-block learning mainly relies on instance-based processes, whereas across-block learning might depend on rule-based mechanisms. A within-block parametric fMRI analysis revealed a learning-related increase of lateral-prefrontal and striatal activity and a learning-related decrease of hippocampal activity in both experiments. Across-block learning was associated with an activation increase of distinct lateral prefrontal brain regions, whereas striatal activation decreased as a function of learning across blocks. Finally, a modulation of within-block learning-related activation pattern across blocks was observed in the anterior prefrontal cortex and the striatum. These data indicate that hippocampal and prefrontal-striatal brain regions differentially contribute to instance-based and rule-based learning.

10.1 Introduction

Learning regularities across multiple episodes is a core cognitive ability. In the last decades, controversy currently surrounds whether humans learn the surface structure of regular input pattern based on the superficial similarity between learning instances or whether humans acquire abstract rule knowledge. Three main learning tasks have been used extensively in experimental psychology: artificial grammar learning, category learning, and sequence learning tasks. Results of several studies provide evidence for the notion that learning in these tasks is mainly based on the knowledge of (fragments of) learning instances (Nosofsky, 1986; Perruchet, 1994). In contrast, rule-based accounts assume that subjects acquire a set of abstract rules, defining an input pattern as grammatical, as a category member, and as a regular sequence, respectively (Ashby & Perrin, 1988; Reber, 1989). Alternative views posit that learning is subserved by both instance-based and rule-based mechanisms (Dominey et al., 1998; Erickson & Kruschke, 1998; Knowlton & Squire, 1996; Meulemans & Van der Linden, 1997; Shanks & St. John, 1994; Shanks, 1995). However, it is still not clear whether instance-based and rule-based processes could be experimentally separated (cf. Pothos, in press).

Tightly coupled with the debate about instance-based vs rule-based learning is the question which brain structures might subserve either mechanism. However, only a few studies examined the neural basis of instance-based vs rule-based learning. In an artificial grammar learning study, Fletcher et al. (1999) demonstrated that learning within experimental blocks is mediated by the right lateral prefrontal cortex (PFC), whereas the left lateral PFC subserves learning across the entire experiment. The authors argue that within-block learning effects mainly rely on the explicit retrieval of individual items based on the surface structure of items. In contrast, the authors propose that across-block learning is based on the acquisition of abstract rule knowledge. In another study (Strange et al., 2001), subjects were required to learn rules which define the category membership of four-letter strings. Changes in abstract rules were associated with an increase of anterior PFC activity, whereas hippocampal activation was modulated by the introduction of new instances. Results of patient

studies and computational models have implicated the basal ganglia in instance-based learning, whereas lateral prefrontal regions seem to be involved in rule-based learning (Dominey, Lelekov, Ventre-Dominey & Jaennerod, 1998). Contrary, a recent fMRI study of artificial grammar learning revealed that the basal ganglia subserve rule-based learning and that the hippocampus is associated with instance-based learning, respectively (Lieberman, Chang, Chiao, Bookheimer & Knowlton, 2004). In the present fMRI study, we sought to distinguish brain regions subserving instance-based and rule-based learning. We adopted the experimental logic of the Fletcher et al. (1999) study, however using a more conventional learning paradigm (cf. Doeller et al., in press). In two experiments, subjects were required to encode and recognize six object-position conjunctions in each trial of several experimental blocks. Both experiments included two conditions, a control condition (context-specific condition, CS) and a learning condition (invariant learning condition, IL). In the CS condition, object-position conjunctions were trial-unique, whereas in the IL condition either objects (Experiment 1) or positions (Experiment 2) were held constant in each trial of the experimental blocks, by this introducing regularities across episodes. We expected that the introduction of invariant objects and positions in object-position conjunctions entails increased task performance across trials within blocks of the IL condition. These *within*-block learning effects were supposed to reflect mainly instance-based learning, since subjects' judgments could rely solely on similarity processes. Critically, the set of invariant objects and positions changed from block to block in the IL condition. By this, subjects were able to transfer their knowledge about regularities to new instances when a new block starts. This transfer has been implicated as a possible experimental test to dissociate instance-based and rule-based learning (e.g. Gomez & Schvaneveldt, 1994; Mathews et al., 1989). If subjects acquire abstract rule knowledge, they should be able to transfer this knowledge to new instances. Thus, a performance modulation *across* learning blocks was assumed to be a main index of rule-based learning. Based on previous results (Doeller et al., in press), we predicted a learning-related decrease of hippocampal and an increase of prefrontal-striatal activation as a function of learning within blocks. In contrast, we expected a modulation of distinct prefrontal-striatal regions during across-block

learning (cf. Fletcher et al., 1999; Lieberman et al., 2004).

10.2 Materials and Methods

Subjects Twenty-four subjects participated in the study, 12 subjects in Experiment 1 (aged 22-33, mean age 24.6 years, 6 females) and 12 subjects in Experiment 2 (aged 22-29, mean age 24.3 years, 5 females). All subjects were right-handed with normal or corrected-to-normal vision and were paid for participating. Informed consent was obtained before scanning. All participants reported to be in good health with no history of neurological disease.

Stimuli, Task, and Design Sixteen line drawings denoting real-life objects were used as stimulus material in both experiments. The stimulus set contained eight living and eight non-living objects. Stimuli were presented within a 4×4 grid. In each experimental trial (Figure 10.1), six different objects were presented sequentially at six different positions (sample phase). Each object was presented for 600 ms, followed by a 100 ms interstimulus interval. After a 2000 ms delay (fixation cross), subjects were shown a probe stimulus for 1000 ms. Participants were required to indicate whether or not the current object-position conjunction (the probe stimulus) was identical to one of the six object-position conjunctions presented during the sample phase. Responses were delivered by a button press with the right or left index finger (1000 ms response window). Response-to-hand mappings were counterbalanced across subjects. Probes in each block (see below) comprised 50% old (old object at old position) and 50% new object-position conjunctions (3 equally distributed categories: old object/new position, new object/old position, and new object/new position). Visual feedback was provided for 500 ms immediately after probe offset. An exponentially distributed intertrial interval (ITI) of 2.5-7.5 s (mean: 3.5 s) was used (cf. Doeller et al., in press, for a detailed description of the experimental procedures).

The design of both experiments included two conditions, a control condition (context-specific condition, CS) and a learning condition (invariant learning condition, IL).

To minimize the probability that subjects based their judgment on a common strategy for both conditions, trials were blocked by condition. In contrast to our previous study (Doeller et al., in press), in which blocks of both conditions were presented in randomized order, subjects performed four blocks in sequence in each condition in the present study. In both conditions, participants were informed about the beginning of a new block. Each experimental block comprised 36 trials. The temporal order of conditions was balanced across subjects. The 3D structural MR sequence (see below) was measured in the middle of the experiments to separate the sessions of both conditions. Participants were unaware of the experimental manipulation, i.e. the existence of two different conditions. In the CS condition, object-position conjunctions were unique in each trial. Here, each object-position conjunction in each trial and each block was selected randomly from the set of objects and positions. In contrast, in the IL condition either objects (Experiment 1) or positions (Experiment 2) were held constant in each trial of the experimental blocks. In the IL condition of Experiment 1, invariant objects were presented at variable positions (object regularities). Accordingly, in the IL condition of Experiment 2, variable objects were presented at invariant positions (spatial regularities). The number of invariant objects and positions, respectively resembled the number of object-position conjunctions presented during the sample phase, i.e. the same six objects or positions were presented in each trial of one block. Objects or positions were held constant across all trials of one experimental block. The fixed configuration of invariant objects and positions changed from block to block, i.e. when a new block starts, a new set of objects or positions in object-position conjunctions were replicated. All other task parameters were held constant across conditions and experiments. In particular, the CS condition was identical in Experiment 1 and 2. Subjects were given a training session immediately before the experiment, including 36 training trials.

Behavioral Analysis For each subject, Pr values (proportion hits - proportion false alarms; Feenan & Snodgrass, 1990) were used as performance measure in both experiments. Instead of the false alarm rate to all new probes only false alarms to probes comprising new objects (Experiment 1) and new positions (Experiment

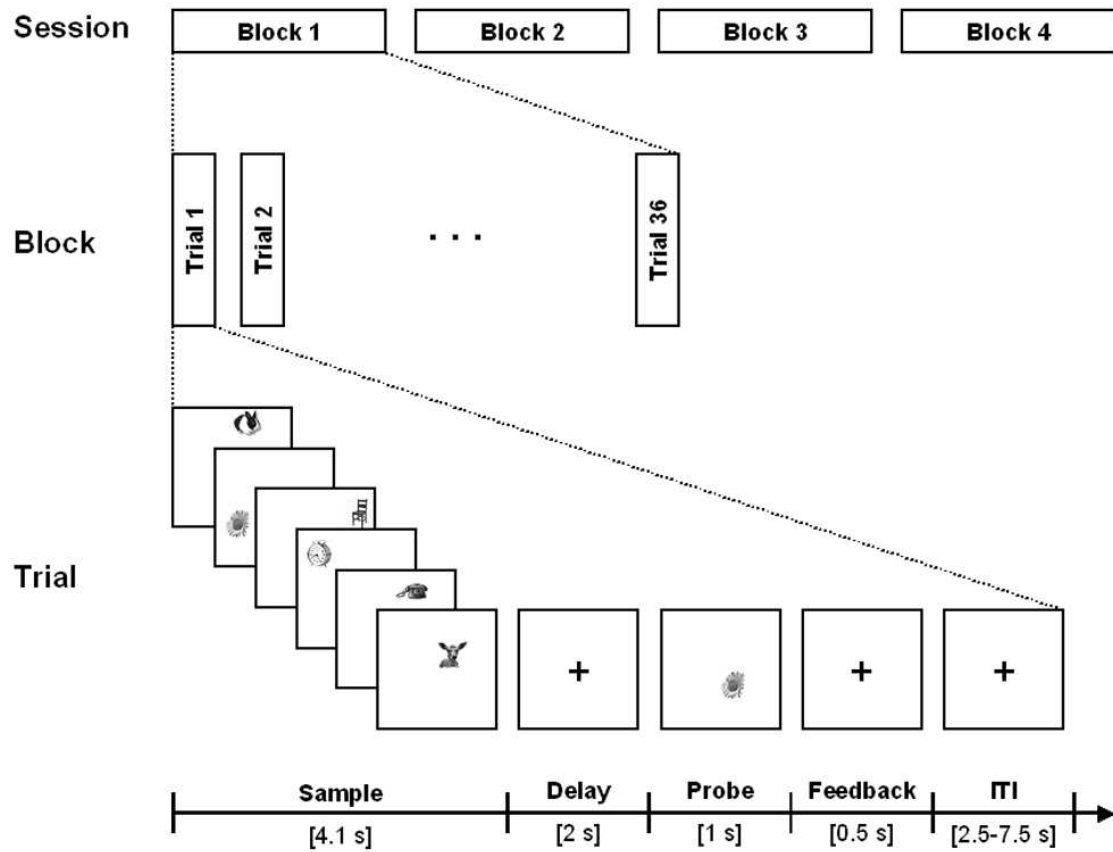


Figure 10.1. Session, Block, and Trial Structure. Subjects performed four blocks in each experimental condition. Each block comprised 36 trials. During the sample phase of the trial, six different objects were presented sequentially at six different positions within a 4×4 grid (grid not shown in the Figure). Upon presentation of a probe stimulus, participants were required to indicate whether the probe is identical to one of the six object-position conjunctions presented during the sample phase. In the present example, the probe comprised an old object at a new position, therefore requiring a “new” response. Duration (in seconds) of the respective trial phases are depicted at the bottom of the figure. ITI, intertrial interval.

2), respectively were included, i.e. probes comprising a new object at an old or a new position (Experiment 1) and probes comprising a new or an old object at a new position (Experiment 2). Based on previous results (Doeller & Opitz, 2004; Doeller et al., in press), we assumed that learning should be most clearly revealed by a reduction of false alarms to these specific probe types, since subjects could reject these probes solely on the basis of the knowledge of invariant objects and positions, respectively. In the following, this performance measure will be referred to as ‘corrected Pr’. In the CS condition of Experiment 1 and 2, corrected Pr values were computed in the same manner as in the corresponding IL condition. To analyze

within-block learning effects, mean performance measures in nine consecutive trials (i.e. trials 1-9, 10-18, 19-27, and 28-36) were averaged across all blocks separately for both conditions. To evaluate across-block learning effects, mean corrected Pr values for each entire block of both conditions were averaged separately for blocks 1-2 and 3-4, respectively.

Imaging Parameters BOLD-sensitive, T_2^* -weighted functional images and T_1 -weighted structural images (3D MP-RAGE) were acquired at 1.5 T (Siemens Sonata, Siemens AG, Erlangen, Germany). Functional data were acquired using a gradient-echo EPI pulse sequence, with the following parameters: $T_R = 1.8$ s, $T_E = 50$ ms, flip angle = 85° , slice thickness = 4 mm, interslice gap = 1 mm, in-plane resolution = 3.5×3.5 mm, FoV = 224 mm^2 , 20 axial slices parallel to AC-PC plane. The first four volumes were discarded to allow for T_1 equilibration. Both conditions were conducted in separate functional sessions. An additional 2D T_1 -weighted structural sequence ($T_R = 600$ ms, $T_E = 13$ ms, flip angle = 80° , slice thickness = 4 mm, interslice gap = 1 mm, in-plane resolution = 0.9×0.9 mm, FoV = 224 mm^2 , 20 axial slices parallel to AC-PC plane) was measured in-plane with respect to the functional sequence and applied during the coregistration procedure (see below).

Imaging Preprocessing Imaging analysis was performed with SPM2 (<http://www.fil.ion.ucl.ac.uk/spm/spm2.html>). FMRI time series were sinc-interpolated in time to correct for differences in slice acquisition time and motion-corrected, using a 6-parameter rigid body spatial transformation and a B-spline interpolation. An additional unwarping procedure was applied to account for movement-by-magnetization inhomogeneity interactions and by this to minimize the movement-related residual variance (Andersson et al., 2001). Coregistration included two steps. First, images of the 2D T_1 -weighted sequence, which was measured in-plane with respect to the EPI sequence were coregistered to the mean functional image. Second, images of the 3D MP-RAGE sequence were coregistered to the images of the resliced 2D T_1 -weighted sequence. Subsequently, resliced 3D anatomical images were normalized to the standard T_1 template (MNI reference brain). Based on the determined pa-

rameters, the normalization algorithm was then applied to the functional volumes. Finally, the normalized functional images were resampled into 2 mm isotropic voxels and spatially smoothed with an isotropic 7-mm FWHM Gaussian kernel.

Statistical Analysis of fMRI Time Series Time series were analyzed using a two-stage random effects analysis. At a first level, for each trial of both conditions trial-specific effects were modeled using the canonical HRF (Friston et al., 1998a) separately for the sample phase (epoch-related; duration = 4.1 s) and the probe phase (event-related). Additional parametric analyses were conducted (see below). Data were high-pass filtered to $\frac{1}{128}$ and scaled for global activity. Parameters for each covariate were estimated by a least-mean-squares fit of the model to the time series using a subject-specific fixed-effects model within the general linear model. During the estimation procedure, serial correlations were estimated with a restricted maximum likelihood (ReML) algorithm using an AR(1) plus white noise model. In SPM2, the ReML estimates (hyperparameters) are then used to correct for non-sphericity (Friston et al., 2002b). Linear contrasts of the parameter estimates for each regressor were calculated for each subject and brought to the second level random effects analysis. All analyses were restricted to the sample phase. MNI-coordinates of all reported activations have been transformed to the canonical Talairach space (<http://www.mrc-cbu.cam.ac.uk/imaging/mnispac.html>).

Given our a priori hypotheses and our previous results (Doeller et al., in press), all fMRI analyses were restricted to the lateral prefrontal cortex, the striatum, and the hippocampus. To investigate brain regions involved during regularity learning, we conducted parametric fMRI analyses to explicitly model learning-related activation pattern (see Doeller et al., in press, for details). For this purpose, individual fMRI time series (associated with each sample phase) were weighted with parametric modulation functions, separately modeling (1) within-block learning (Figure 10.4), (2) across-block learning (Figure 10.5), and (3) the modulation of within-block learning effects across blocks (Figure 10.6). The first set of model functions was derived by averaging the individual within-block learning functions (mean corrected Pr values for nine consecutive trials, collapsed across all blocks; see above) across subjects,

separately for both experiments. A second set of model functions was used to analyze across-block learning effects. Here, performance measures (corrected Pr values) were collapsed across all trials, separately for each block and averaged across subjects (separately for each experiment). Finally, modulations of within-block learning effects across blocks were evaluated in a separate model, using the slope of the mean fitted learning functions separately for each entire block as model function. The model function was derived from the mean learning functions (corrected Pr values) for *each* block by fitting a logarithmic function $y = a \cdot \ln(t_i) + b$. To increase the model fit, mean learning functions were derived by averaging individual learning functions for each block collapsed across all 24 subjects of both experiments.

Significant activated regions were identified using a statistical threshold of $p < 0.0005$ (uncorrected; cluster size: 5 contiguous voxels). Given our previous results with respect to the within-block learning effects (Doeller et al., in press), additional region-of-interest (ROI) analyses were conducted for the striatum and the hippocampus in the within-block fMRI analysis ($p < 0.01$, small volume corrected, SVC; Worsley et al., 1996). Striatal and hippocampal ROIs were determined according to a detailed neuroanatomy atlas (Warner, 2001). Appropriate mask image was generated using MRICro (<http://www.psychology.nottingham.ac.uk/staff/cr1/mricro.html>).

10.3 Results

Behavioral Results: Within-block Learning In a first step, within-block learning effects were analyzed (Figure 10.2). As in our previous behavioral study (Doeller & Opitz, 2004), we used corrected Pr values as behavioral measures (see Methods). A two-way repeated measures ANOVA with the factors condition (CS vs IL) and trial (mean corrected Pr values for trials 1-9, 10-18, 19-27, and 28-36 within blocks, collapsed across all blocks) revealed a main effect of condition, $F(1, 11) = 21.84$, $p < 0.001$, in Experiment 1 (Figure 10.2a, left panel). Further comparisons between the first (trials 1-9) and last (trials 28-36) quarter of trials within blocks were conducted to examine changes of performance during the time course of the blocks. Mean corrected Pr values increased from the first to the last quarter of trials within

blocks in the IL condition, $F(1, 11) = 4.86$, $p < 0.05$, but not in the CS condition, $F(1, 11) < 1$. Accordingly, performance was better in the IL condition as compared to the CS condition in the last quarter of trials, $F(1, 11) = 16.12$, $p < 0.005$, but not in the first quarter of trials, $F(1, 11) = 2.95$, $p > 0.1$. A marginally significant linear trend across trials was restricted to the IL condition, $F(1, 11) = 4.32$, $p < 0.07$ (CS condition: $F(1, 11) = 1.50$, $p > 0.2$).

The analyses in Experiment 2 revealed similar results (Figure 10.2a, right panel). Here, an increase of corrected Pr values between the first and last quarter of trials was observed in the IL condition, $F(1, 11) = 5.40$, $p < 0.05$, but not in the CS condition, $F(1, 11) = 1.29$, $p > 0.2$. In contrast to the first quarter of trials, $F(1, 11) = 1.28$, $p > 0.2$, performance between conditions differed at a marginally significant level in the last quarter of trial, $F(1, 11) = 4.05$, $p < 0.07$. Moreover, mean corrected Pr values increased linearly in the IL condition, $F(1, 11) = 5.70$, $p < 0.05$, but not in the CS condition, $F(1, 11) = 1.38$, $p > 0.2$. Thus, these data indicate that an increase of performance across trials within blocks was restricted to the IL condition in both experiments.

For both experiments, additional statistical analyses were conducted separately for false alarms to probes including new objects and new positions, respectively. A dissociation between both experiments was expected with respect to the two types of false alarms. In Experiment 1 (object regularities), a reduction of false alarms to probes including new objects was predicted, whereas in Experiment 2 (spatial regularities), we hypothesized a selective reduction of false alarms to probes including new positions, since subjects could reject these specific probes solely on the basis of their knowledge of invariant objects and positions, respectively (denoted as “critical probe category” in Figure 10.2). In Experiment 1, the false alarm rate to probes including new objects (Figure 10.2b, left) was significantly reduced in the IL condition relative to the CS condition, $F(1, 11) = 22.71$, $p < 0.001$. Additional comparisons between both conditions of Experiment 1 revealed no differences with respect to this false alarm type in the first quarter of trials, $F(1, 11) < 1$. However, both conditions differed significantly in the last quarter of trials, $F(1, 11) = 9.12$, $p < 0.05$. By contrast, no differences between conditions were observed for false alarms to probes including

new positions (Figure 10.2c, left), neither for all trials, $F(1, 11) = 1.25$, $p > 0.2$, nor for the first, $F(1, 11) = 3.53$, $p > 0.08$, and the last quarter of trials, $F(1, 11) < 1$. In Experiment 2, a reversed pattern was observed. Analyses of false alarms to probes including new objects (Figure 10.2b, right) revealed no differences between both conditions, neither for all trials, $F(1, 11) < 1$, nor for the first, $F(1, 11) = 1.36$, $p > 0.2$, and the last quarter of trials, $F(1, 11) < 1$. Contrary, the false alarm rate to probes including new positions (Figure 10.2c, right) tended to be reduced in the last quarter of trials in the IL as compared to the CS condition, $F(1, 11) = 4.49$, $p < 0.06$. No differences between conditions were observed for the first quarter of trials, $F(1, 11) < 1$. Accordingly, false alarms of this type significantly decreased between the first and last quarter of trials selectively in the IL condition, $F(1, 11) = 5.71$, $p < 0.05$ (CS condition: $F(1, 11) < 1$). Furthermore, a significant linear trend across trials was restricted to the IL condition, $F(1, 11) = 5.76$, $p < 0.05$ (CS condition: $F(1, 11) < 1$).

Behavioral Results: Across-block Learning In a second analysis step, across-block learning effects were examined (Figure 10.3). A two-way repeated measures ANOVA with the factors condition (CS vs IL) and block (mean corrected Pr value for blocks 1-2 vs 3-4) revealed a main effect of condition in Experiment 1, $F(1, 11) = 21.83$, $p < 0.001$ (Figure 10.3a, left), but not in Experiment 2, $F(1, 11) < 1$ (Figure 10.3a, right). Moreover, mean corrected Pr values increased across blocks in both experiments in the IL condition (Experiment 1: $F(1, 11) = 13.90$, $p < 0.005$; Experiment 2: $F(1, 11) = 6.50$, $p < 0.05$), but not in the CS condition ($F(1, 11) < 1$ in both experiments).

Similar to the within-block learning effects, we conducted additional analyses of false alarms to probes including new objects and new positions, respectively (Figure 10.3). In Experiment 1, the ANOVA revealed a main effect of condition for probes including new objects, $F(1, 11) = 22.70$, $p < 0.001$ (Figure 10.3b, left). Accordingly, false alarms of this type were reduced in blocks 1-2, $F(1, 11) = 10.03$, $p < 0.01$, and blocks 3-4, $F(1, 11) = 14.43$, $p < 0.005$, in the IL relative to the CS condition. The false alarm rate to probes including new positions (Figure 10.3c, left) did not differ between conditions, $F(1, 11) = 1.25$, $p > 0.2$, neither in blocks 1-2, $F(1, 11) < 1$,

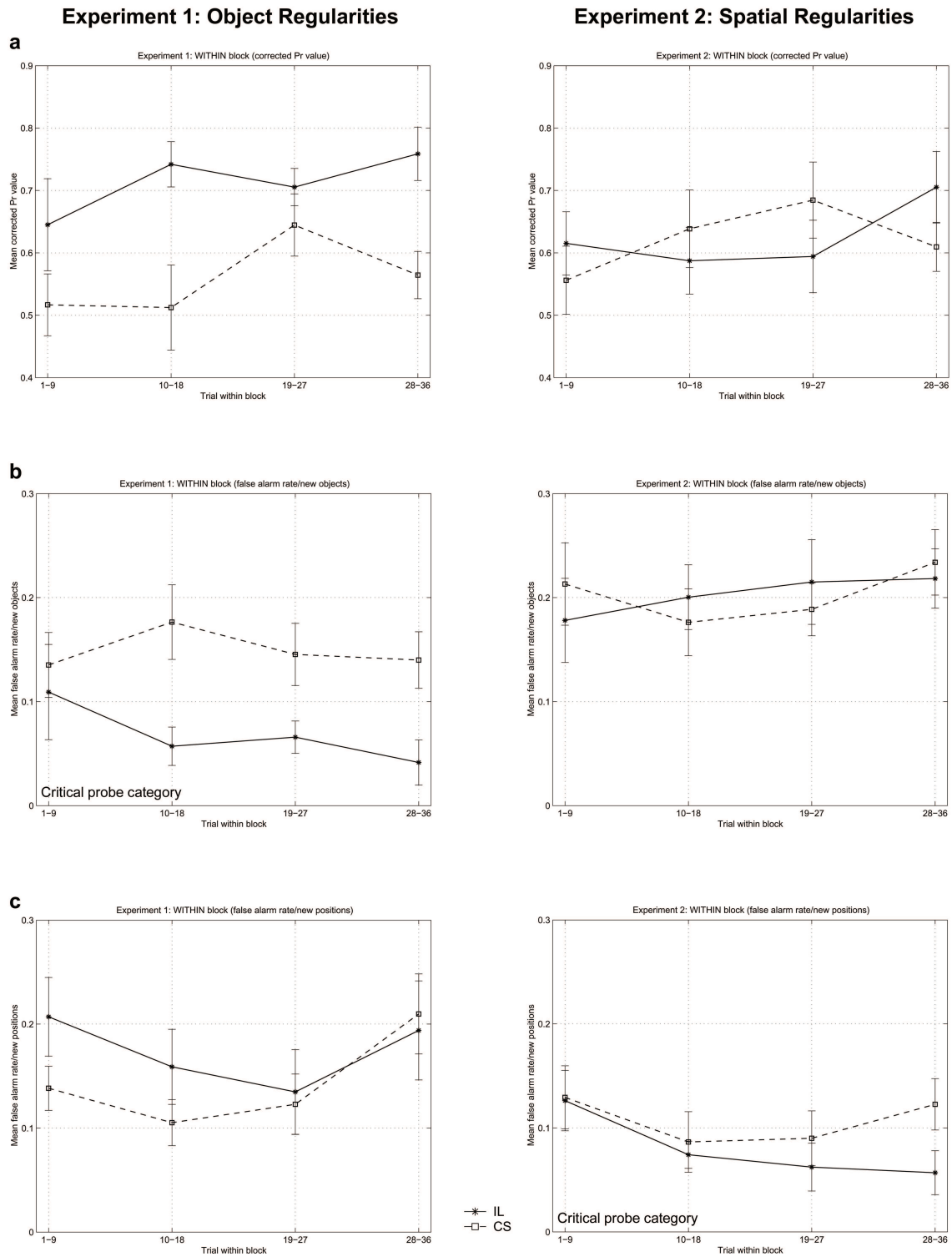


Figure 10.2. Behavioral Results: Within-block Learning. (a) Mean corrected Pr values, (b) mean false alarm rate to probes including new objects, and (c) mean false alarm rate to probes including new positions. The performance measures are averaged across trials 1-9, 10-18, 19-27, and 28-36, respectively and collapsed across all blocks, separately for the invariant learning condition (solid) and the context-specific condition (dashed) in Experiment 1 (object regularities; left panels) and Experiment 2 (spatial regularities; right panels).

nor in blocks 3-4, $F(1, 11) < 1$. In accordance with the results of the within-block learning effects, a dissociation between experiments was observed. In Experiment 2, the false alarm rate to probes including new objects (Figure 10.3b, right) was similar in the IL and CS condition ($F(1, 11) < 1$, for the analysis of all blocks and separate analyses for blocks 1-2 and 3-4, respectively). By contrast, subjects showed a reduced false alarm rate to probes including new positions (Figure 10.3c, right) in the IL relative to the CS condition (all blocks: $F(1, 11) = 7.85$, $p < 0.05$; blocks 1-2: $F(1, 11) = 1.91$, $p > 0.1$; blocks 3-4: $F(1, 11) = 5.87$, $p < 0.05$).

To evaluate the across-block modulation of the within-block learning effects, we calculated performance differences (corrected Pr values) between the last quarter of blocks and the first quarter of blocks, separately for blocks 1-2 and blocks 3-4. These difference values were considered as a measure of learning within blocks. Difference values entered into an ANOVA with the factors condition (CS vs IL) and block (difference values for blocks 1-2 and blocks 3-4, respectively). Given the decreased number of datapoints in this analysis as compared to the above analyses, we increased the power of this analysis by including data from Experiment 1 and 2. This analysis revealed a marginally significant condition $\times$ block interaction, $F(1, 23) = 4.19$, $p < 0.06$. Moreover, difference values were higher in the IL relative to the CS condition in blocks 1-2, $F(1, 23) = 4.28$, $p < 0.05$, but not in blocks 3-4, $F(1, 23) < 1$, indicating that the performance increase within blocks in the IL condition was attenuated across blocks relative to the CS condition.

Imaging Results: Within-block Learning According to the behavioral analysis, in a first step within-block learning was evaluated in a parametric fMRI analysis (Figure 10.4; Table 10.1). For this purpose, fMRI time series were weighted with mean within-block learning functions, separately for both experiments (see Methods). All fMRI analyses were restricted to the IL condition, since we were interested in learning-related activation pattern. In Experiment 1, a learning-related increase of activity within blocks was observed in the left superior frontal gyrus (SFG; Brodmann Area [BA] 9; peak Talairach coordinates x, y, z : $-8, 54, 32$) and the left middle frontal gyrus (MFG; BA 8; $-24, 31, 33$). Furthermore, the left putamen ($-24, -1, 9$)

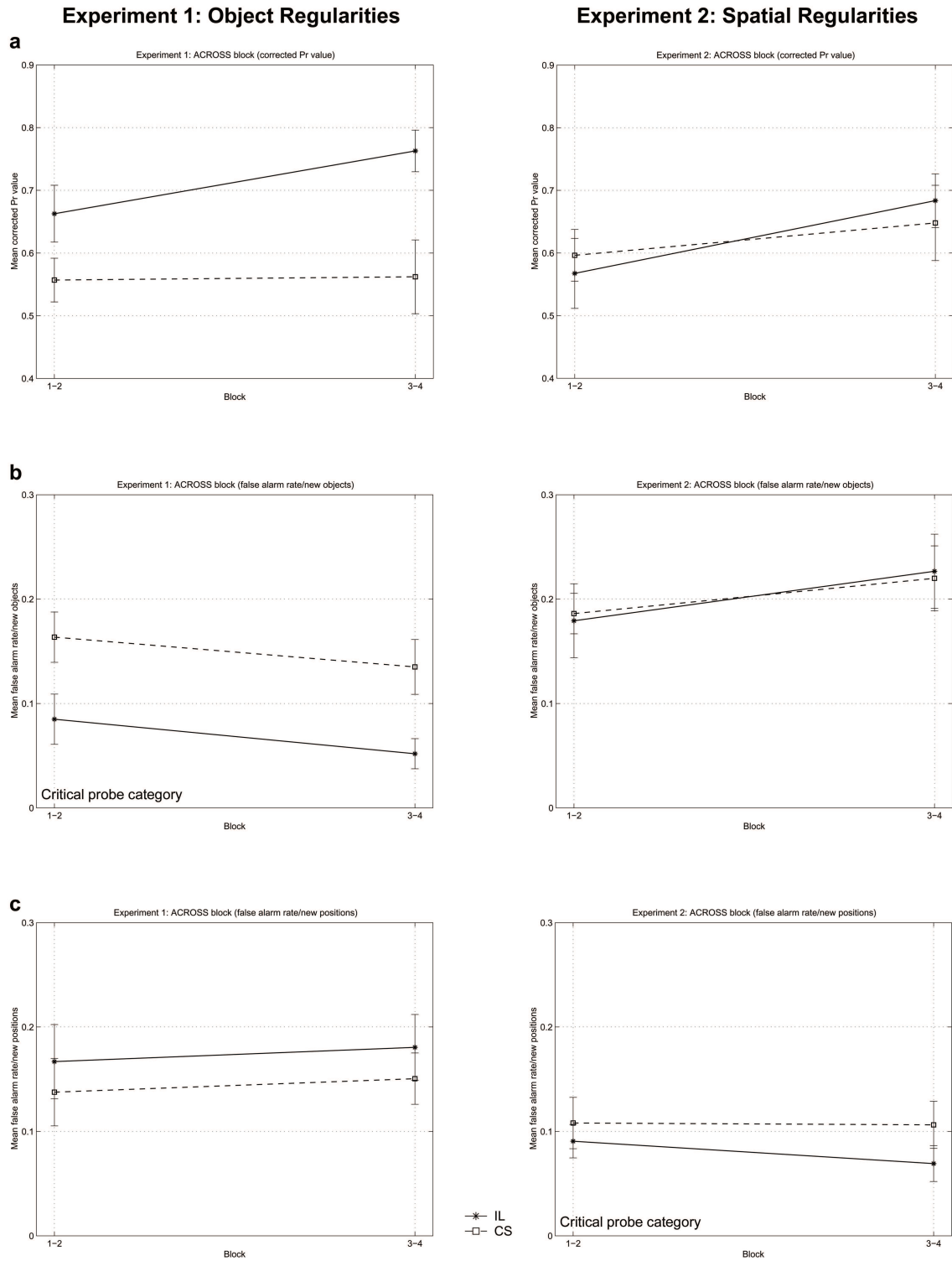


Figure 10.3. Behavioral Results: Across-block Learning. (a) Mean corrected Pr values, (b) mean false alarm rate to probes including new objects, and (c) mean false alarm rate to probes including new positions. The performance measures are averaged across blocks 1-2 and 3-4, respectively and collapsed across all trials, separately for the invariant learning condition (solid) and the context-specific condition (dashed) in Experiment 1 (object regularities; left panels) and Experiment 2 (spatial regularities; right panels).

showed a learning-related increase within blocks. A similar prefrontal-striatal pattern was observed in Experiment 2. Right MFG (BA 8; 24, 15, 32) and bilateral putamen (left: -26, 4, 3; right: 26, 1, 11) activity increased as a function of learning within blocks. By contrast, the right hippocampus showed a learning-related decrease of activation within blocks (Experiment 1: 26, -22, -6; Experiment 2: 28, -24, -9).²

Table 10.1

*Imaging Results: Within-block Learning. Brain regions showing a learning-related activation pattern within blocks, separately for Experiment 1 and 2 (from anterior to posterior), described in terms of Brodmann areas (BA), hemisphere (L, left; R, right), Talairach coordinates (mm; transformed from the MNI-space), and peak Z-score. SPMs were thresholded at $p < 0.0005$ (uncorrected), 5 voxel extent, except for $*p < 0.01$ (small volume corrected).*

Region	BA	Hemisphere	Talairach			Z-score
			x	y	z	
Experiment 1: Object Regularities						
(1) Learning-related increase						
Superior frontal gyrus (SFG)	9	L	-8	54	32	3.89
Middle frontal gyrus (MFG)	8	L	-24	31	33	4.36
Putamen (PUT)	—	L	-24	-1	9	*2.66
(2) Learning-related decrease						
Hippocampus (HC)	—	R	26	-22	-6	*3.87
Experiment 2: Spatial Regularities						
(1) Learning-related increase						
Middle frontal gyrus (MFG)	8	R	24	15	32	3.69
Putamen (PUT)	—	L	-26	4	3	*3.23
	—	R	26	1	11	*2.70
(2) Learning-related decrease						
Hippocampus (HC)	—	R	28	-24	-9	*2.71

Imaging Results: Across-block Learning In a second analysis step, neural correlates of across-block learning were investigated (Figure 10.5; Table 10.2). Mean

²When analyzing learning-related activation pattern in the IL condition relative to the CS condition, the bilateral hippocampus (Experiment 1) and the right hippocampus (Experiment 2), respectively showed a learning-related decrease of activation ($p < 0.05$, small volume corrected).

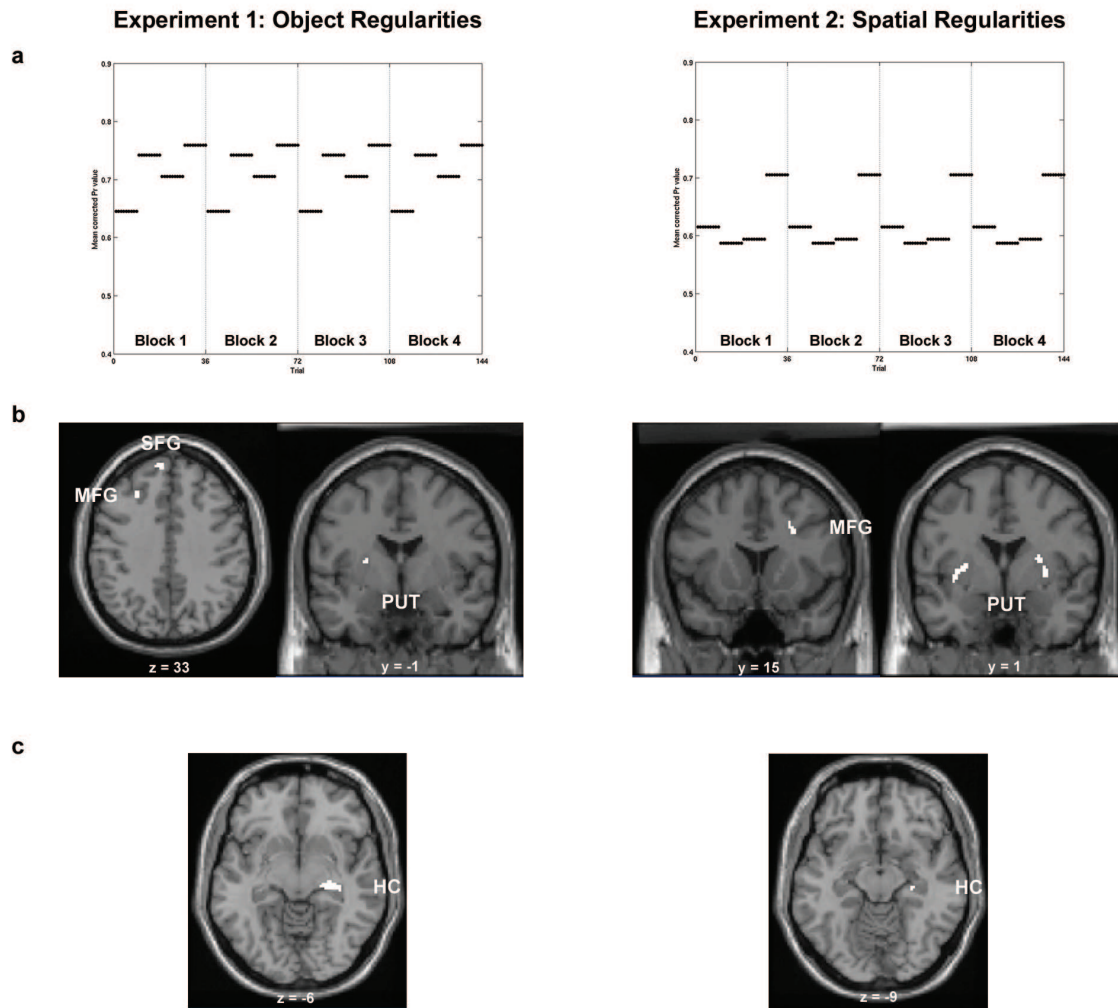


Figure 10.4. Imaging Results: Within-block Learning. Model functions and imaging results are depicted separately for Experiment 1 (object regularities; left panels) and Experiment 2 (spatial regularities; right panels). (a) Model functions for block 1 (trials 1-36), block 2 (trials 37-72), block 3 (trials 73-108), and block 4 (trial 109-144). (b) Learning-related increase of activity within blocks in the superior frontal gyrus (SFG), middle frontal gyrus (MFG), and putamen (PUT). (c) Learning-related decrease of activity within blocks in the hippocampus (HC). SPMs are superimposed on coronal or axial sections of the MNI T_1 -weighted MRI template. Talairach y -coordinate or z -coordinate is given below each image.

performance functions of each entire block were used as model functions in the fMRI analysis (see Methods). This analysis revealed a learning-related increase of activity across blocks in the left anterior MFG (BA 46; $-22, 43, 2$) in Experiment 1 and in the left posterior MFG (BA 8; $-30, 9, 35$) in Experiment 2. In contrast, the left putamen showed a learning-related decrease of activity in Experiment 1 ($-18, 13, -9$) and Experiment 2 ($-20, 12, -1$).

Finally, modulations of the within-block learning effects across blocks were analyzed in a separate model, using the slope of the mean fitted block-specific learning functions collapsed across both experiments as model function (see Methods; Figure 10.6; Table 10.2). The mean slope decreased across blocks and we suppose that this gradual decrease of the slope reflects the across-block modulation of the within-block learning effect. This analysis revealed activations in the left SFG (BA 10; $-28, 53, 14$), the bilateral putamen (left: $-22, 11, -4$; right: $24, 13, -6$), and the left caudate nucleus ($-14, 8, 5$).

Table 10.2

Imaging Results: Across-block Learning. Brain regions showing a learning-related activation pattern across blocks (for further details see Table 10.1 legend). SPMs were thresholded at $p < 0.0005$ (uncorrected), 5 voxel extent.

Region	BA	Hemisphere	Talairach			Z-score
			x	y	z	
Experiment 1: Object Regularities						
(1) Learning-related increase						
Middle frontal gyrus (MFG)	46	L	-22	43	2	4.46
(2) Learning-related decrease						
Putamen (PUT)	–	L	-18	13	-9	3.60
Experiment 2: Spatial Regularities						
(1) Learning-related increase						
Middle frontal gyrus (MFG)	8	L	-30	9	35	4.94
(2) Learning-related decrease						
Putamen (PUT)	–	L	-20	12	-1	3.53
Experiment 1/2: Slope Analysis						
Superior frontal gyrus (SFG)	10	L	-28	53	14	3.72
Putamen (PUT)	–	R	24	13	-6	4.87
	–	L	-22	11	-4	3.81
Caudate nucleus (CN)	–	L	-14	8	5	4.38

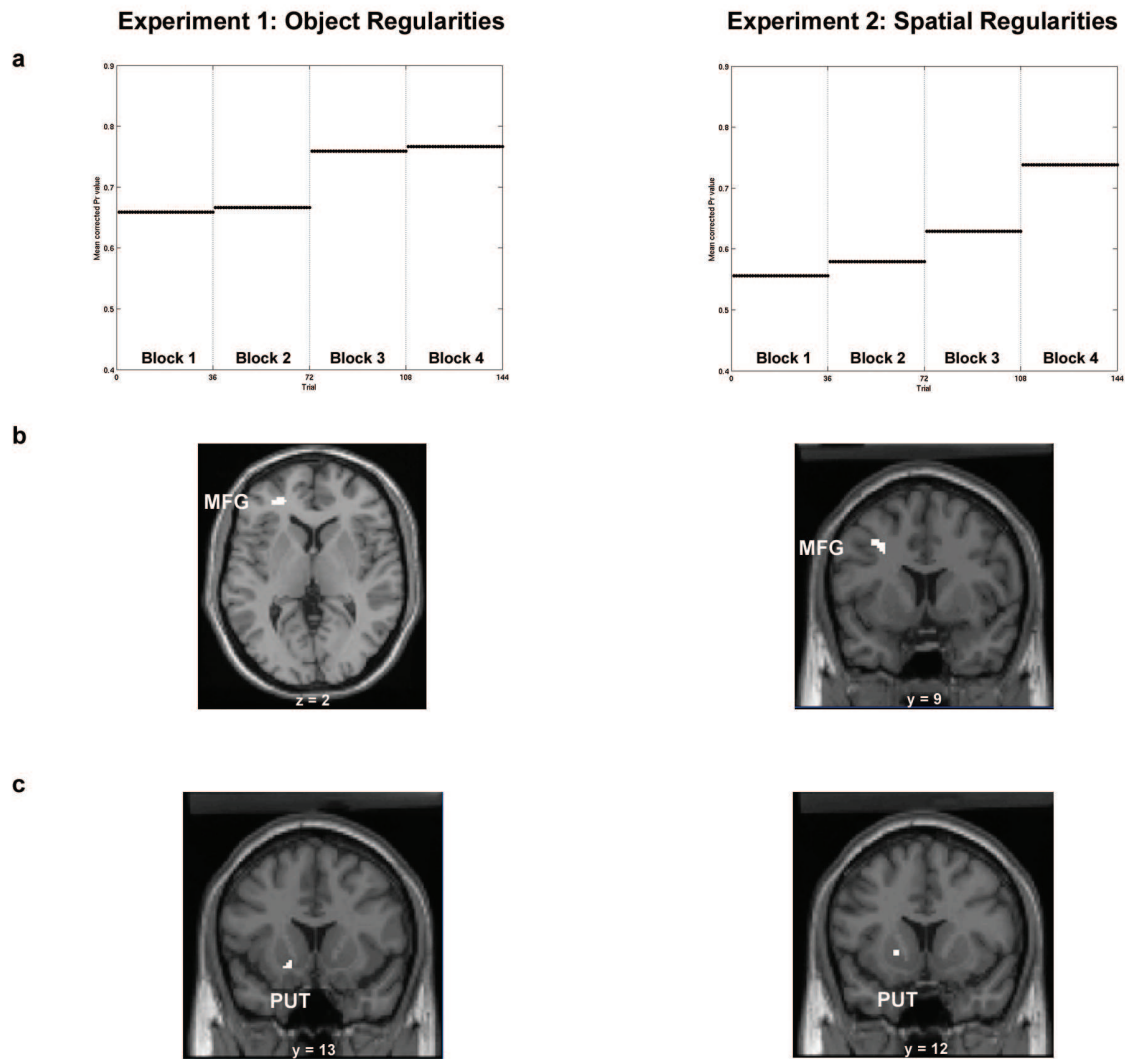


Figure 10.5. Imaging Results: Across-block Learning. Model functions and imaging results are depicted separately for Experiment 1 (object regularities; left panels) and Experiment 2 (spatial regularities; right panels). (a) Model functions separately for blocks 1-4. (b) Learning-related increase of activity across blocks in the MFG. (c) Learning-related decrease of activity across blocks in the putamen. See Figure 10.4 for details.

10.4 Discussion

The present study aimed at specifying the neural basis of instance-based and rule-based learning. The first learning-process should be reflected in a performance increase within experimental blocks (within-block learning), whereas the latter learning process should entail a performance increase across learning blocks (across-block learning). The introduction of object regularities (Experiment 1) and spatial regularities (Experiment 2) gave rise to both within-block and across-block learning.

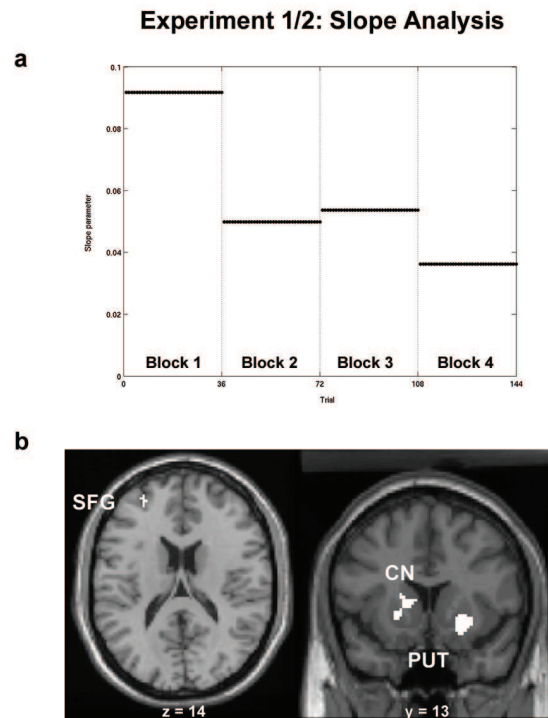


Figure 10.6. Imaging Results: Slope-related Activation Pattern. (a) Model functions separately for blocks 1-4. (b) Slope-related activity in the SFG, the putamen, and the caudate nucleus (CN). See Figure 10.4 for details.

In both experiments, within-block learning was associated with a decrease of hippocampal and an increase of middle frontal gyrus and putamen activity, whereas across-block learning was reflected in an increase of middle frontal gyrus activity and an attenuation of superior frontal gyrus and putamen/caudate nucleus activation. The behavioral data indicate that learning was restricted to the IL condition in both experiments. Here, performance increased within and across blocks. Contrary, in the CS condition performance remained constant within and across blocks. Furthermore, the analysis of false alarms revealed a dissociation between both experiments. During learning object regularities (Experiment 1), solely false alarms to probes including new objects decreased within and across blocks, whereas in Experiment 2 (spatial regularities), this selective decrease of false alarms was restricted to probes including new positions. Here, subjects benefit most of all from regularity learning since they can reject the probes solely on the basis of the knowledge of invariant objects and positions, respectively. A possible explanation for this differential domain-specific

reduction of false alarms might be an enhanced selectivity of object and spatial representations, respectively (cf. Yeshurun & Carrasco, 1998).

Hippocampal Contribution to Instance-based Learning The learning-related decrease of hippocampal activity within blocks in Experiment 1 and 2 is consistent with our previous study (Doeller et al., in press). In accordance with models of hippocampal function (Eichenbaum, 2000; Norman & O'Reilly, 2003), we argue that relational binding requirements gradually decrease in the IL condition, since invariant objects are bound to variable positions or vice versa. As a consequence, hippocampal involvement diminishes as a function of learning. This learning-related hippocampal activation pattern is in agreement with previous studies showing hippocampal decrease as a function of sequence learning (Schendan et al., 2003), probabilistic learning (Poldrack et al., 2001), and artificial grammar learning (Opitz & Friederici, 2003, 2004; Strange et al., 2001). Interestingly, the introduction of new exemplars (Strange et al., 2001) and changes of superficial features of sentences (Opitz & Friederici, 2004) entailed a phasic increase of hippocampal activation which rapidly attenuated as a function of learning (cf. Strange et al., 1999). In contrast, in both studies (Opitz & Friederici, 2004; Strange et al., 2001) hippocampal activation was not affected by the introduction of new grammatical rules. In line with these findings, hippocampal activity decreased as a function of learning within blocks. At the beginning of each new block, the set of invariant objects and positions was changed. Consequently, these changes of instances gave rise to hippocampal involvement at the beginning of the blocks with its activity gradually decreasing as a function of increasing performance within blocks.

Prefrontal-striatal Contributions to Rule-based Learning Distinct lateral prefrontal and striatal brain regions exhibited a learning-related modulation of activity within and across blocks in both experiments. In showing a prefrontal-striatal activation increase as a function of learning within blocks in both experiments, we could replicate our previous findings (Doeller et al., in press). Accordingly, we argue that the lateral PFC contributes to learning by means of mnemonic codes

(Becker & Lim, 2003), representing regularities of invariant episodic features. It is supposed that these rule-like aggregated representations facilitate encoding of new object-position conjunctions and give rise to increased performance across trials by providing a more efficient encoding strategy. The lateral PFC has been implicated in the acquisition and the implementation of task-relevant rules (Freedman et al., 2001; Miller, 2000; Wallis et al., 2001). Given the across-block modulation of middle frontal and superior frontal gyrus activity, it is conceivable that such rule-like representations are flexibly adapted to new task requirements, such as a new set of invariant episodic features at the beginning of a new block (cf. Duncan, 2001). Fletcher et al. (1999) observed an virtually identical increase of left middle frontal gyrus activity across blocks during artificial grammar learning. The authors argue that this incremental activation pattern reflects the “implementation of semantic knowledge” (Fletcher et al., 1999, p. 176) of abstract grammar rules. In accordance with this view, we argue that rule-like representations of task-relevant regularities in object-position conjunctions gradually evolve across blocks, since information is aggregated across multiple episodes (cf. O’Reilly & Norman, 2002).

The putamen exhibited a learning-related activation increase within blocks and a learning-related activation decrease across blocks. In addition, putamen and caudate nucleus activity correlated with the across-block attenuation of the slope of the within-block learning functions. The striatum has been implicated in stimulus-reward and stimulus-stimulus learning by representing predictions of future reward delivery (O’Doherty et al., 2003, 2004; Schultz, 2002; Schultz et al., 2003). Based on this evidence and assumptions of formal learning theories (Rescorla & Wagner, 1972; Sutton & Barto, 1990), the learning-related increase of striatal activity within blocks might indicate that the striatum codes the increased predictability (cf. Koechlin et al., 2002) of invariant episodic features and by this implements a statistical model of the task environment (Dayan, Kakade & Montague, 2000). The across-block attenuation of these striatal processes might reflect the facilitated update of this statistical model.

Finally, it could be speculated that there is a cooperative functional interaction between lateral prefrontal and striatal structures during learning within and across

blocks. Recent computational models (Braver & Cohen, 2000; Frank et al., 2001) support this view by showing a crucial role of the striatum in providing a gating mechanism which controls the flexible update of prefrontal representations by means of a reinforcement signal. Thus, striatal reinforcement signals, coding the predictive relationships among invariant episodic features might be involved in updating information about regularities maintained in the PFC, hence stabilizing prefrontal rule-like representations.

In summary, the present results suggest that instance-based and rule-based learning presumably depend on different brain regions. The hippocampus seems to be solely involved during instance-based learning, whereas distinct lateral prefrontal and striatal structures subserve both instance-based and rule-based learning.

Part V

General Discussion

Chapter 11

General Discussion

Contents

11.1 Summary	194
11.2 Two Separate Learning and Memory Systems: Fast Trial-unique vs Gradual Regularity Learning	198
11.3 A Definition of Regularity Learning	199
11.3.1 Regularity Extraction across Multiple Episodes	200
11.3.2 The Gradual Nature of the Extraction Process	201
11.3.3 Maintenance of Extracted Regularities	203
11.3.4 Abstract Knowledge Base	203
11.4 A Tentative Model of Regularity Learning	205
11.4.1 Hippocampus and Relational Binding	206
11.4.2 Striatum and Predictability Coding	209
11.4.3 Prefrontal Cortex and the Representation of Rule-like Information	211
11.4.4 Possible Functional Interactions between the Hippocampus, the Striatum, and the Prefrontal Cortex	212
11.5 Conclusions and Open Issues	217

In the last decades, the existence of separate learning and memory systems has been proposed in experimental psychology and cognitive neuroscience (McClelland et al., 1995; Mishkin et al., 1984; O’Keefe & Nadel, 1978; O’Reilly & Norman, 2002; Sherry & Schacter, 1987; Squire, 1992; Squire & Zola, 1996, cf. Packard &

Knowlton, 2002; Poldrack & Packard, 2003; Rolls, 2000; White & McDonald, 2002). In general, it has been proposed that two distinct systems are specialized for complementary functions, namely (1) fast learning unique episodes (first system) and (2) gradual learning regularities across multiple episodes (second system). In several fMRI and behavioral experiments, the present thesis aimed at specifying the neural and cognitive basis of the second system. In the present experimental framework, the general term ‘regularity learning’ was introduced to describe the learning process underlying the second system. Regularity learning was defined as the gradual extraction of overlapping, invariant features of multiple episodes. The main results of the present experiments indicate that regularity learning is accompanied by a shift from a hippocampal to a prefrontal-striatal brain system.

After summarizing the theoretical background and the results of the present thesis (11.1), the main characteristics of both learning and memory systems will be summarized (11.2). Furthermore, several criteria will be proposed to provide theoretical and empirical constraints for a definition of regularity learning in the scope of the present experimental framework (11.3). Moreover, a tentative model of regularity learning will be outlined, describing the functional roles of the hippocampus, the striatum, and the lateral prefrontal cortex and possible functional interactions between these structures during regularity learning (11.4). Finally, open issues and future directions will be described (11.5).

11.1 Summary

The formation of episodic memory is supposed to be a core function of the first proposed learning and memory system. In Chapter 2, an overview about the role of the hippocampus during the formation of episodic memories was given. Results from single cell recording and lesion studies in animals, neuropsychological studies with amnesic patients, single cell recordings, ERP studies and imaging studies in humans suggest that the hippocampus is critically involved during building episodic memories. Two influential models of hippocampal function, namely the Eichenbaum model (Eichenbaum, 2000) and the O’Reilly model (O’Reilly & Norman, 2002) were

described. The Eichenbaum model assumes that the hippocampus contributes to the formation of episodic memory by means of relational binding, i.e. binding of different features defining episodes. The O'Reilly model proposes that the hippocampus is characterized by a specific representational format, namely a sparse, pattern-separated, conjunctive representation, making the hippocampus best suited for encoding and retrieval of episodic memories. Both models make different predictions how regularities across episodes, i.e. overlapping features of multiple episodes are represented in the brain. Eichenbaum has pointed out that specific hippocampal neurons represent regularities across episodes, whereas O'Reilly assumes that the surrounding rhinal and parahippocampal cortex is specialized for this operation.

In Chapter 3, the cognitive and neural basis of the second learning and memory system was described. The term 'regularity learning' was introduced to describe the learning process associated with the second system. Regularity learning was defined as the gradual extraction of overlapping, invariant features of multiple episodes. In this chapter, results from studies investigating reinforcement learning mechanisms, e.g. classical and instrumental conditioning and probabilistic learning, as well as studies using artificial grammar learning, category learning, and sequence learning tasks were reported. On a cognitive level, the mechanisms underlying learning in these tasks have in common that subjects acquire a (fragmentary) knowledge of regularities across multiple episodes. On a neural level, reinforcement learning is mainly based on the mesolimbic dopamine system, especially the ventral tegmental area and the substantia nigra, and its main target areas, the striatum and the orbitofrontal cortex. The lateral prefrontal cortex has been identified as the core structure in mediating artificial grammar, category, and sequence learning. Evidence for medial temporal lobe and striatal involvement in these learning tasks is less consistent.

Two brain regions were of main interest in the scope of the present thesis, namely the hippocampus and the prefrontal cortex. The neuroanatomy of both brain structures was described in Chapter 4. In the main experiments of the present thesis, functional magnetic resonance imaging (fMRI) was used to investigate the neural correlates of regularity learning. An overview of the physical principals of MRI, the main characteristics of the BOLD signal, and important aspects of the analysis of

fMRI time series and the designing of fMRI experiments was given in Chapter 5. The topic of Chapter 6 was the rationale for the present experiments. In the experiments described in this thesis, subjects had to learn several object-position conjunctions during the sample phase of the experimental trials. Upon presentation of a probe stimulus, participants had to indicate whether or not the current object-position conjunction (the probe stimulus) was identical to one of the object-position conjunctions presented in the sample phase. In the current experimental framework, an episode was defined as the set of object-position conjunctions in the sample phase. The design of the experiments comprised two conditions, a context-specific (CS) and an invariant learning (IL) condition. The CS condition should tap processes associated with the first proposed learning and memory system. Therefore, episodes were defined trial-unique, i.e. objects and positions were variably mapped across trials. In contrast, the IL condition should elicit the involvement of the second learning and memory system. Here, either positions or objects in object-position conjunctions were held constant across trials of experimental blocks. By this, regularities across episodes were introduced. The IL condition was of main interest in the scope of the present thesis, thus the CS condition was designed as a control condition with respect to regularity learning. Each condition included several experimental blocks. On a behavioral level, an increase of performance was expected within blocks of the IL condition, since subjects should benefit from the introduction of invariant objects and positions, respectively. In contrast, in the CS condition a constant performance level was expected across trial. On a neural level, constant hippocampal activation was expected in the CS condition, since variable objects have to be bound to variable positions in each trial (constant binding requirements). In the IL condition, reduced binding requirements were expected during the time-course of learning. Thus, a learning-related decrease of hippocampal activity was predicted. Conversely, a learning-related increase of lateral prefrontal activity was hypothesized in the IL condition.

In Experiment 1, the introduction of invariant positions (Experiment 1A; Chapter 7) and invariant objects (Experiment 1B; Chapter 8) in object-position conjunctions entailed increased performance in the IL condition. As predicted, performance in the

CS condition was constant across trials. In both experiments, regularity learning in the IL condition was associated with a learning-related decrease of hippocampal activity. In contrast to a right hippocampal decrease during learning spatial regularities (Experiment 1A), bilateral hippocampal activation decreased as a function of learning object regularities (Experiment 1B). Lateral prefrontal and striatal activity increased as a function of learning in both experiments.

In Chapter 9, several behavioral experiments investigated learning object regularities in more detail (Experiment 2). In brief, the influence of object distinctiveness (Experiment 2A-2B), memory load and learning duration (Experiment 2C-2E), and the effect of simultaneous object-position bindings (Experiment 2F) on learning invariant objects in object-position conjunctions was examined. The behavioral data indicate that object distinctiveness mainly affects overall memory performance, whereas manipulation of the learning duration and the memory load affects the learning process. Moreover, the analyses revealed a dissociation of two learning mechanisms: a process operating within learning blocks and a process operating across blocks.

Finally, Experiment 3 (Chapter 10) aimed at specifying whether separable brain regions mediate these two learning mechanisms. In contrast to Experiment 1A and 1B, trials were blocked by condition to minimize the probability that subjects based their judgment on a common strategy for both conditions. On a behavioral level, subjects showed a performance increase within and across learning blocks in an object IL condition (Experiment 3A) and in a spatial IL condition (Experiment 3B), but not in the respective CS conditions. On a neural level, within-block learning was associated with a learning-related increase of lateral prefrontal and striatal activity and a learning-related decrease of hippocampal activity, by this replicating the results of Experiment 1A and 1B. In addition, distinct prefrontal-striatal regions were involved in across-block learning. Thus, these data might indicate that regularity learning in the present experimental framework is subserved by instance-based and rule-based learning mechanisms, as indexed by within-block and across-block learning, respectively. The hippocampus seems to be preferentially involved during instance-based learning, whereas distinct lateral prefrontal and striatal structures subserve both instance-based and rule-based learning.

Taken together, the results of the present experiments indicate that regularity learning is accompanied by a gradual transition from a hippocampal to a prefrontal-striatal brain system.

11.2 Two Separate Learning and Memory Systems: Fast Trial-unique vs Gradual Regularity Learning

The two separate learning and memory systems can be distinguished from each other along several characteristics which are summarized in Table 11.1. Each of these characteristics will be addressed in the following Section 11.3 and Section 11.4.¹ In summary, the first system is specialized for the rapid representation of unique episodes (O'Reilly & Norman, 2002; Squire, 1992) which comprise non-predictive relationships among stimuli. In contrast, the second system is best suited for learning regularities across multiple episodes, where predictive relationships among stimuli exist. Furthermore, the first system is characterized by fast encoding of episodes by means of pattern-separated representations (O'Reilly & Norman, 2002), whereas the second system is involved in gradual, incremental learning regularities across multiple episodes, i.e. this system integrates over many experiences, by this providing aggregated information (O'Reilly & Norman, 2002; Sherry & Schacter, 1987). The first system might store information by means of weight-based representations, whereas the storage principle of the second system might be best characterized by activation-based representations (cf. Frank et al., 2001). On a molecular level, information processing and storage might be mediated by LTP and dopamine in the first and second system, respectively. Finally, it is proposed that the hippocampus is the neural structure associated with the first system, whereas the striatum and the prefrontal cortex are best suited to meet the functional demands of the second system (Poldrack & Packard, 2003; Squire & Zola, 1996; White & McDermott, 2002).

¹The relevant characteristics will be highlighted in the text, e.g. [C1].

Table 11.1

Two Separate Learning and Memory Systems: Main Characteristics and Differences.

Characteristics	First System	Second System
[C1] Learned information	Unique episodes	Regularities across episodes
[C2] Relationship among Stimuli	Non-predictive	Predictive
[C3] Processing Speed	Fast	Slow
[C4] Representational Format	Pattern-separated	Aggregated
[C5] Storage Principle	Weight-based	Activation-based
[C6] Molecular Mechanism	LTP	Dopamine
[C7] Brain Structure	Hippocampus	Striatum, prefrontal cortex

See text for further details.

First, by providing four criteria I will describe the cognitive mechanisms underlying regularity learning (11.3). Second, a tentative model will be outlined, addressing the neural basis of regularity learning (11.4).

11.3 A Definition of Regularity Learning

I propose four criteria to provide constraints for a definition of regularity learning in the current experimental framework: (a) the extraction of regularities, i.e. the detection of overlapping features across multiple episodes, (b) the gradual nature of the learning process, (c) the maintenance of these regularities over time, and (d) the aggregated representational format of the extracted information. Each criterium is concerned with a different aspect of learning, namely the information which is learned (criterium a), temporal properties (criterium b) and ‘cognitive control’ properties (criterium c) of the learning process, and the representational format of the learned information (criterium d). In the following, the results of the present thesis will be discussed according to these criteria.

11.3.1 Regularity Extraction across Multiple Episodes

The first criterium describes the information which is learned and the properties of the learning environment. The essential prerequisite for learning is the existence of regularities across multiple episodes, i.e. overlapping features of episodes [C1]. As a consequence, structured, predictive relationships between episodic features gradually evolve, i.e. subjects gradually detect contingencies between episodic features [C2]. In the present experiments, either objects or positions in object-position conjunctions were held constant. Hence, regularities, i.e. overlapping features across multiple episodes were introduced. It is assumed that this feature overlap allows learning to occur. In the present experiments, subjects benefit from the presence of spatial or object regularities across trials. The decision upon the probe stimulus is facilitated in the presence of learned regularities across episodes as reflected in increased performance across trials in the IL condition. In contrast, in the CS condition each trial comprised a unique episode, i.e. no regularities were introduced and by this no predictive relationships among episodic features exist. Thus, learning was absent or negligible in the CS condition.

The existence of regularity learning is supported by analyses of false alarms committed by subjects in the present experiments. When spatial regularities are introduced (Experiment 1A and 3B), a selective reduction of false alarms to probes including new positions was observed within learning blocks. In contrast, object regularities in object-position conjunctions (Experiment 1B and 3A) entailed a selective decrease of false alarms for probes including new objects. Here, subjects benefit most of all from learning regularities, since they can reject the probes solely on the basis of the knowledge of invariant objects and positions, respectively. A possible explanation for this differential domain-specific reduction of false alarms might be an enhanced selectivity of object or spatial representations (cf. Yeshurun & Carrasco, 1998). In agreement with this view, the distinct hemispheric specialization during learning object and spatial regularities, respectively indicates that the hippocampus is sensitive to the specific domain of invariant episodic features. Taken together, these data suggest that in contrast to the CS condition the performance increase in the IL condition

is driven by the successful extraction of regularities across multiple episodes. In a similar vein, in artificial grammar learning tasks, subjects are exposed to repeating overlapping sequences of grammatical strings, in category learning tasks, subjects are exposed to repeating overlapping category-defining features (e.g. dot patterns), and in sequence learning tasks, subjects are exposed to repeating overlapping stimulus and/or response sequences. In all these tasks, regularities are extracted and entail better task performance.

The present fMRI experiments and especially the behavioral data in Experiment 2C and 2D (and partially Experiment 2E) further indicate that distinctiveness of invariant features facilitates the extraction of regularities, possibly by further strengthening the selectivity of object and spatial representations. In the spatial invariant learning conditions (Experiment 1A and 3B), invariant positions were highly distinctive. Stimuli were presented in a grid, by this positions were highly salient. Moreover, the grid was presented permanently for the whole trial phase. However, in Experiment 2A and 2B, although feature distinctiveness was increased, no learning effects were observed. When both feature distinctiveness and learning phase (together with higher memory load) were increased, learning occurred (Experiment 2C and 2D). Importantly, the observation of more pronounced learning effects in Experiment 2C and 2D (prolonged learning phase and highly distinctive features) than in Experiment 2E (prolonged learning phase and less distinctive features), provides evidence for the notion that feature distinctiveness is an additional factor which affects regularity extraction. Accordingly, the diminished learning effects in Experiment 2F suggest that regularity extraction might depend on separated, distinct representations of the features defining the critical regularities.

11.3.2 The Gradual Nature of the Extraction Process

The second criterion is concerned with temporal properties of the learning process. It is assumed that learning has to be gradual, since the extraction process integrates over many experiences (cf. O'Reilly & Norman, 2002) [C3]. Similar to the present experiments, in which performance increases across several trials, in all aforementioned learning tasks learning evolves across multiple experiences with the structured

stimulus material. Several presentations of artificial grammar strings, members of a specific category, or sequential structures are required to entail robust learning effects. The importance of this criterion is elucidated by the comparison between performance in Experiment 2A and 2B (short learning phase) with performance in Experiment 2C-2E (long learning phase). Learning occurred when learning blocks comprised 36 trials (Experiment 2C-2E), but less learning occurred when learning blocks comprised 16 trials (Experiment 2A and 2B). These results indicate that the extraction process is gradual in nature, similar to the mechanisms involved in learning artificial grammars, categories, and sequences.

With regard to this criterion, the time-sensitive dissociation of the learning effects in Experiments 2C-2E and Experiments 3A and 3B seems to be relevant. Here, learning effects were pronounced in the first half of the experiments, but diminished in the second half. These data indicate that the learning asymptote is reached earlier in the last as compared to the first blocks. Accordingly, the across-block distribution of the learning functions follows a step function, i.e. learning becomes faster in the final phase of the experiment (cf. Fletcher et al., 1999, 2001), where the predictability of invariant episodic features increases and by this uncertainty diminishes faster (cf. Rescorla & Wagner, 1972; Sutton & Barto, 1990). This dissociation between both halves of the experiment indicates that regularity extraction across episodes in the present experimental framework might be based on two distinct mechanisms: a process operating within blocks and a process operating across blocks. It could be speculated that the former extraction process might be responsible for detecting feature invariances (instance-based learning) and the latter, higher-order process might be involved in transferring the extraction mechanism to new exemplars (rule-based learning). Experiment 3 aimed at disentangling both mechanisms. The results of Experiment 3 indicate that distinct lateral prefrontal and striatal brain structures mediate both within-block and across-block learning, whereas the hippocampus is preferentially involved in within-block learning. Furthermore, an across-block modulation of within-block learning effects was observed. Most interestingly, the striatum (putamen and caudate nucleus) showed a striking correspondence with the interaction of both learning mechanisms as indexed by the gradual across-block decrease

of the slope of the within-block learning functions.

11.3.3 Maintenance of Extracted Regularities

After subjects have successfully extracted the set of invariant positions or objects in the present task, this set has to be maintained across the remaining block. This issue was not tested in the behavioral studies of Experiment 2. One possibility would be to concurrently present a second, interfering task. Learning should diminish when this second task hampers the maintenance processes. Interestingly, constraints of these processes could be provided by selectively manipulating the onset of the secondary task. More precisely, it could be speculated that the interfering influence of a secondary task is restricted to a critical time window. Supporting evidence is provided by sequence learning research, indicating that learning is preserved even under dual-task conditions when subjects have enough time to retain sequence elements in working memory (Frensch et al., 1994). Support for the relevance of this criterion might be provided by the imaging data (Experiment 1 and 3). During learning spatial and object regularities, the lateral prefrontal cortex showed a learning-related increase of activity across trials within experimental blocks in both experiments. One of the core cognitive control functions of the PFC is the active maintenance of task-relevant information and the protection of this information against distraction (Miller et al., 1996; Sakai et al., 2002; Smith & Jonides, 1999). Accordingly, the results of Experiment 1 and 3 highlight the importance of the lateral prefrontal cortex for the maintenance of extracted regularities (see below).

11.3.4 Abstract Knowledge Base

Which kind of knowledge is acquired in the present task? In the artificial grammar and category learning literature, one core controversy revolves around two contrary concepts: rule vs similarity (cf. Pothos, in press). Is learning mainly based on abstract rules or on the superficial similarity between items to be learned? For instance, several artificial grammar learning models posit that subjects learn the abstract rules of the artificial grammar system (Dulany et al., 1984, 1985; Reber, 1989), whereas others argue that learning is based on knowledge of fragments/chunks of gram-

mathematical strings, e.g. letter bigrams and trigrams and not on the acquisition of an abstract rule-system (Perruchet & Pacteau, 1990; Perruchet, 1994; Servan-Schreiber & Anderson, 1990). In a similar vein, a distinction could be made between exemplar-based (Kruschke, 1992; Nosofsky, 1986) and rule-based models (Ashby & Townsend, 1986; Ashby & Perrin, 1988) of category learning. According to exemplar-based models, subjects represent category exemplars and their categorization decision is based on a comparison between the current stimulus and the stored exemplars. In contrast, rule-based models assume that subjects respond to stimuli according to decision rules which partition the potential decision space.

It is not in the scope of the present framework to provide a solution for this controversy. However, it might be argued that both rule-based and similarity-based models (at least with regard to the chunk-based account in artificial grammar learning) are based on a common denominator, namely the idea that learning is characterized by the successful aggregation of information over time. For instance, chunk-based accounts posit that learning is driven by the aggregation or organization of the input pattern into compact chunks (e.g. Servan-Schreiber & Anderson, 1990). This form of knowledge is abstract in the sense that it is independent of a specific learning episode, in fact it evolves over multiple experiences. In line with this, in the present task subjects extract regularities over time, i.e. spatial or object invariances in object-position conjunctions. These invariances resemble the aggregated chunks in the model mentioned above (Servan-Schreiber & Anderson, 1990). The prefrontal cortex seems to be best suited for representing such aggregated, rule-like information (Bor et al., 2003; Strange et al., 2001; Wallis et al., 2001, see below) [C4].

A possible way to provide empirical constraints regarding the representational format of the knowledge acquired during learning regularities is to investigate different domains. More precisely, is regularity extraction a domain-general or a domain-specific neural mechanism? Comparisons between Experiment 1A (spatial regularities) and Experiment 1B (object regularities) indicate that similar prefrontal regions support regularity learning in the spatial and object domain, thus indicating the domain-general nature of these processes. Interestingly, a domain-specific hemispheric specialization of the hippocampus was observed. Relative to a right-

lateralized activation pattern during learning spatial regularities (Experiment 1A), bilateral hippocampus activity decreased as a function of learning object regularities (Experiment 1B). This domain-specific hemispheric specialization might suggest that the hippocampus is sensitive to the perceptual properties of invariant features, indicating the bottom-up character of this mechanism.

11.4 A Tentative Model of Regularity Learning

Based on the results of the present thesis, in the following a tentative model of regularity learning will be outlined (see Figure 11.1). In all imaging experiments, regularity learning was mediated by a specific brain network, including the hippocampus, the striatum, and the prefrontal cortex. A consistent relationship between learning and activation in these three brain regions was observed: hippocampal activation decreased across trials as a function of learning, whereas the striatum and the prefrontal cortex showed a learning-related activation increase. But what is the exact division of labor between the hippocampus, the striatum, and the prefrontal cortex during regularity learning? Is there a time-limited role of one region during a particular phase of learning? What are the functional roles of these brain regions and how do they interact in the service of regularity learning? Is there a functional specialization, i.e. are there neuronal and computational constraints making one region appropriate for meeting specific functional demands?

The present model proposes that the hippocampus is crucial for relational binding of unique episodic features (11.4.1) that the striatum subserves coding of the predictability of invariant episodic features (11.4.2), and that the prefrontal cortex exhibits active maintenance of regularities by means of rule-like representations in favor of flexible task control (11.4.3). Finally, the model proposes possible functional interactions between the hippocampus, the striatum, and the prefrontal cortex during the time-course of learning (11.4.4).

A Tentative Model of Regularity Learning

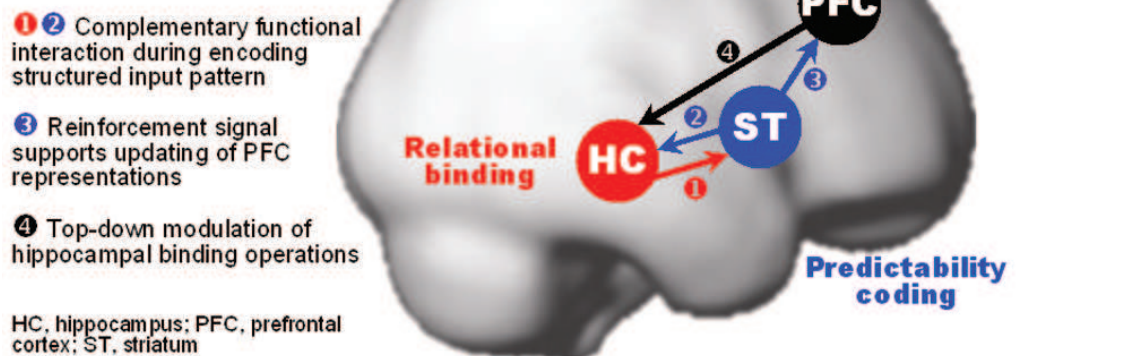


Figure 11.1. A Tentative Model of Regularity Learning. During the time-course of regularity learning, the hippocampus (HC) is involved in relational binding of unique episodic features. The striatum (ST) encodes the increased predictability of invariant episodic features. The lateral prefrontal cortex (PFC) subserves the active maintenance of regularities by means of rule-like representations. Based on these representations, the PFC might implement an efficient encoding strategy. Arrows denote possible functional interactions between the three brain regions. (1) and (2): Complementary functional interaction between the hippocampus and the striatum during encoding of structured input pattern. Hippocampal relational binding requirements decrease, once regularities are detected. In contrast, the striatum codes the increased predictability of invariant episodic features and provides a reinforcement signal to the PFC (3). Based on rule-like representations, the PFC modulates hippocampal binding operations (4). See text for further details.

11.4.1 Hippocampus and Relational Binding

The hippocampus has been implicated in relational memory binding, i.e. binding of distinct features of unique episodes (Cohen et al., 1999; Eichenbaum, 2000, in press) [C1]. The preferential role of the hippocampus for relational memory processing has been elucidated by several recent single cell recording (Bunsey & Eichenbaum, 1996; Dusek & Eichenbaum, 1997; Fortin et al., 2002) and imaging studies (Davachi & Wagner, 2002; Giovanello et al., 2004; Preston et al., 2004). It has been argued that a specific, pattern-separated representational format is required to avoid or to minimize interference between representations of different episodes, i.e. representations of unique episodes have to be kept separate from representations of other episodes (Norman & O'Reilly, 2003; O'Reilly & Rudy, 2000) [C4]. Another core abil-

ity of a neural system specialized for representing unique episodes is a high learning rate (Norman & O'Reilly, 2003) [C3]. The pattern separation mechanisms has to be fast, since episodes have to be encoded directly when they occur. Moreover, it is assumed that the hippocampal representation is sparse and conjunctive (O'Reilly & Rudy, 2001). The sparseness entails fast encoding as well as minimizing interference between representations of distinct episodes. A conjunctive representation is the result of binding different features of unique episodes (O'Reilly & Rudy, 2000).

In the present experimental framework, the sample phase of each trial (i.e. a set of sequential object-position conjunctions) is defined as an 'episode'. In each trial, specific objects have to be bound to specific positions (first-order binding). Moreover, in each trial several object-position bindings have to be bound together to entail an exhaustive, pattern-separated (O'Reilly & Norman, 2002) representation of the whole episode (second-order binding). In the CS condition of the present task, these relational binding requirements remain approximately constant across trials, since objects are variably mapped to positions. Hence, the probability for each possible object-position conjunction is approximately constant within one block. In contrast, in the IL condition relational binding requirements become reduced as a function of time within experimental blocks, since invariant objects are bound to variable positions or vice versa. By this, the probability of specific object-position conjunctions is substantially increased (e.g. a specific object presented at one of the positions held constant in a block [spatial IL condition] or one of the invariant objects presented at a specific position [object IL condition], respectively), whereas other object-position conjunctions never occur in one block, i.e. the probability for these object-position conjunctions is reduced to zero (e.g. a specific object presented at one of the remaining positions never occurring in a block [spatial IL condition] or one of the remaining objects never occurring in a block presented at a specific position [object IL condition], respectively). It is conceivable that the learning-related decrease of hippocampal activity is due to these reduced relational binding requirements during the experimental blocks, leading to a reduced hippocampal conjunctive representation. These results concur with several imaging studies investigating probabilistic cue-outcome learning (Poldrack et al., 2001), associative learning (Zeineh et al.,

2003), sequence learning (Grafton et al., 1995; Schendan et al., 2003), and artificial grammar learning (Opitz & Friederici, 2003). Similar to the present task, the reduction of binding requirements is associated with decreased hippocampal involvement in these studies.

Furthermore, the material-specific hemispheric specialization observed in the present imaging experiments might suggest that the hippocampus is characterized by a stimulus-driven processing mode, sensitive to perceptual attributes of invariant episodic features. Consistent with this view, in an artificial grammar learning study Opitz and Friederici (2004) showed that hippocampal activation was modulated by changes of superficial features of grammatical sentences, whereas prefrontal cortex activity was sensitive to the introduction of new grammatical rules. Moreover, in showing highly stimulus-specific learning-related modifications in activation pattern of single hippocampal neurons in monkeys during associative learning, Wirth et al. (2003) further support the view that hippocampal processing is sensitive to perceptual attributes of the to-be-learned stimuli. This stimulus-driven bottom-up processing mode is possibly mediated by projections from distinct visual brain regions via the rhinal cortex, providing information about various stimulus attributes to the hippocampus (cf. Murray & Bussey, 1999). Finally, the results of Experiment 3 indicate that the hippocampus is preferentially involved in instance-based learning as indexed by a learning-related decrease of activity within learning blocks. These data are consistent with the view that the hippocampus is characterized by a stimulus-driven bottom-up processing mode, since instance-based learning mechanisms have been supposed to rely on the superficial similarity between learning instances (Opitz & Friederici, 2004; Pothos, in press).

The shift from a hippocampal to a prefrontal brain system in the present thesis is in line with models of memory consolidation that postulate a gradual reorganization of the neural substrates of memory representations in the course of learning (e.g. Alvarez & Squire, 1994; McClelland, McNaughton & O'Reilly, 1995; Squire & Alvarez, 1995, but see Káli and Dayan, 2004, and Nadel et al., 2000, for alternative views). According to these models, the learning process is accomplished by means of an interaction between the hippocampal system and the neocortex in establishing

permanent cortical representations, thereby gradually losing their dependence of the hippocampus (see Haist, Gore & Mao, 2001; Hoffman & McNaughton, 2002; Maviel, Durkin, Menzaghi & Bontempi, 2004, for recent neuroimaging and neurophysiological evidence for the cortical consolidation hypothesis). Based on this evidence, it might be speculated that regularity learning in the present experimental framework might be a first step towards the consolidation of memory traces of invariant episodic features, thereby becoming independent of the hippocampus as reflected in reduced hippocampal involvement as a function of time. Furthermore, LTP (cf. Malenka & Nicoll, 1999), especially in the hippocampus, has been widely considered as a molecular mechanism of memory consolidation (McGaugh, 2000). Thus, it could be argued, though highly speculative that LTP might serve as a molecular mechanism underlying the learning-related reduction of relational binding requirements in the hippocampus and by this might support regularity learning [C6]. Taken together, the hippocampus seems to be best suited to meet the functional demands of the first learning and memory system (cf. McClelland et al., 1995; O'Reilly & Norman, 2002; Packard & Knowlton, 2002; Poldrack & Packard, 2003; Squire, 1992) [C7].

11.4.2 Striatum and Predictability Coding

In the present fMRI experiments, the striatum exhibited a robust learning-related increase of activation during the experimental blocks. One influential view posits that the midbrain dopamine system (in particular the ventral tegmental area and the substantia nigra, zona compacta) and its main target areas, especially the striatum (nucleus accumbens, putamen, and caudate nucleus), play a pivotal role during reward-related learning (Schultz et al., 1997; Schultz, 2000, 2002; Schultz et al., 2003). Based on assumptions of formal reinforcement learning theories (Rescorla & Wagner, 1972; Sutton & Barto, 1990), it has been proposed that these brain regions represent predictions of future reward delivery and by this drive learning the predictive relationship between stimuli in the environment (Schultz et al., 1997; Schultz, 2002). Schultz argues that the activation pattern of dopamine neurons reflects a prediction error (PE), i.e. a measure of the difference between the actual reward and the expected reward. During initial learning, rewards are highly unexpected. As a con-

sequence, the PE and the neuronal dopamine response is high. With experience of the specific stimulus-reward associations, the PE-related neuronal signal shifts backward in time to stimuli that predict reward (Mirenowicz & Schultz, 1994; O'Doherty et al., 2003; Schultz et al., 1997), i.e. learning chains backward in time. Thus, the striatum is involved in building expectations of upcoming reward delivery and by this supports the acquisition of stimulus-reward (McClure et al., 2003; O'Doherty et al., 2003), stimulus-response-reward (Lauwereyns et al., 2002b; O'Doherty et al., 2004) and higher-order associations (Seymour et al., 2004; Tanaka et al., 2004).

Based on this evidence, it could be argued that the striatum codes the increased predictability of invariant objects and positions, respectively and by this provides a measure of the increased probability of specific object-position conjunctions in the present experiments [C2], possibly mediated by projections from midbrain dopamine neurons [C6]. These striatal operations coding stimulus predictability evolve during the time-course of learning as reflected in a learning-related increase of activity within learning blocks. Interestingly, this learning-related striatal activation pattern is further modulated across learning blocks in Experiment 3, reflecting the adaption of predictability coding operations and hence learning across blocks. In contrast, no striatal involvement was observed in the CS condition of the present imaging experiments. This is due to the fact that a sequence of unique episodes (as in the CS condition) includes no predictive relationships among stimuli [C2]. It could be argued that the striatum is involved in learning predictive relationships of task-relevant features (i.e. invariant objects and positions) and by this is crucial for building a statistical model of the task environment (Dayan et al., 2000), coding the contingencies between episodic features. These operations are gradual in nature, since an integration across many experiences is required to establish a valid statistical model [C3]. In sum, the striatum seems to be one core structure subserving the main functions of the second learning and memory system (cf. Packard & Knowlton, 2002; Poldrack & Packard, 2003; Squire & Zola, 1996; White & McDermott, 2002) [C7].

11.4.3 Prefrontal Cortex and the Representation of Rule-like Information

Several influential theories of prefrontal function have been proposed in the last years (Duncan, 2001; Miller & Cohen, 2001; Shimamura, 2000; Wood & Grafman, 2003, see Wood and Grafman, 2003, for a detailed overview of PFC theories). In general, these theories converge to suggest that the PFC is essential for implementing cognitive control operations in the service of goal-directed behavior (cf. Koechlin, Ody & Kouneiher, 2003). It is assumed that the PFC is characterized by several general key functions subserving cognitive control, namely the selection of task-relevant information (e.g. Rowe et al., 2000), the maintenance of the selected information (cf. Smith & Jonides, 1999), the protection of information against distraction (Mecklinger et al., 2003; Miller et al., 1996; Sakai et al., 2002), and the flexible updating of representations (e.g. Badre & Wagner, 2004, see Braver and Cohen, 2000, Frank, Loughry, and O'Reilly, 2001, for computational models implementing these control mechanisms).

Another major function of the PFC is the acquisition and the implementation of task-relevant rules (Miller, 2000; Miller & Cohen, 2001; Miller & Asaad, 2002, see Freedman et al., 2001, Wallis et al., 2001, for intriguing findings in monkeys supporting this notion). Miller argues “(...) that the PFC represents not specific episodes but the regularities across them that describe task rules (...)” (Miller, 2000, p. 63). The flexible acquisition of task rules entails an appropriate goal-directed behavior. In line with this, it has been further proposed that the PFC is characterized by a specific processing mode, namely adaptive coding (Duncan, 2001). That is, prefrontal neurons have the core ability to rapidly code task-relevant information, an essential prerequisite for the flexible adaption of behavior. In computational neuroscience, the PFC has been characterized by activation-based representations, whereas the hippocampus is usually ascribed weight-based representations (Frank et al., 2001) [C5]. In contrast to weight-based representations, activation-based representations can be rapidly updated and maintained information can directly affect processing in other parts of the brain (Frank et al., 2001). Furthermore, the lateral PFC has

been posited to subserve specific control operations during memory encoding and retrieval, like the reorganization, evaluation, and manipulation of the items to be memorized and the monitoring of reinstated memory traces (Fletcher & Henson, 2001; Simons & Spiers, 2003).

Based on these findings, I argue that the PFC contributes to regularity learning by means of the (1) *extraction* and the (2) *maintenance* of the set of invariant episodic features. In a recent computational model of prefrontal memory control, Becker and Lim (2003) could show that the PFC represents internal mnemonic codes, rapidly adjusted to current task requirements. Thus, it could be further argued that the extracted regularities in the present experiments are represented as an internal mnemonic code (Becker & Lim, 2003) and that the PFC (3) *implements an efficient encoding strategy* on the basis of these mnemonic codes. These task-relevant regularities are supposed to be temporarily stored in an aggregated, rule-like abstract representational format in the prefrontal cortex (Cer & O'Reilly, in press; Miller, 2000, see Bor et al., 2003, for recent fMRI evidence for information aggregation by the PFC) [C4]. In the present experiments, the representation of regularities, aggregated across many experiences, is abstract in the sense that it is independent of a specific learning episode. Accordingly, lateral prefrontal regions showed a learning-related modulation across blocks in Experiment 3, presumably reflecting the preferential role of the lateral PFC during rule-based learning (Fletcher et al., 1999; Opitz & Friederici, 2004; Strange et al., 2001). In summary, the lateral PFC seems to be appropriate in meeting specific functional demands of the second learning and memory system [C7], especially extraction, maintenance, and strategic encoding of regularities.

11.4.4 Possible Functional Interactions between the Hippocampus, the Striatum, and the Prefrontal Cortex

In the following section, functional interactions between the hippocampus, the striatum, and the prefrontal cortex which might subserve regularity learning will be outlined.

Hippocampal-Striatal Interactions In all present fMRI experiments, a complementary activation pattern was observed in the hippocampus and the striatum. The hippocampus showed a learning-related activation decrease, whereas striatal activity increased as a function of learning. This differential involvement of the hippocampus and the striatum is in accordance with the view that the hippocampus and the basal ganglia comprise interactive memory systems, specialized for distinct memory processes, namely declarative memory (hippocampus) and the acquisition of habits (striatum), respectively (Poldrack et al., 2001; Packard & Knowlton, 2002; Poldrack & Packard, 2003, cf. White & McDermott, 2002) [C7]. As mentioned in Chapter 3, a wide variety of neuropsychological, neuroimaging, and animal lesion studies support this view. For instance, using a feedback-based probabilistic categorization task, Knowlton et al. (1996) showed that patients with striatal dysfunction (Parkinson's disease) showed impaired probabilistic learning, however they exhibited intact memory for specific details of the training phase. In contrast, patients with MTL dysfunction (amnesic patients) showed preserved probabilistic learning but impaired memory performance. In agreement with and in extension of these patient data, in a recent fMRI study Poldrack and colleagues (2001) showed that hippocampal activation decreased as a function of probabilistic learning, whereas caudate nucleus activity increased as a function of learning. Contrary, in an additional paired associate task the hippocampus showed stronger BOLD responses as compared to the probabilistic categorization task. The authors argue that the dissociable impairments of patients as well as the complementary activation pattern of the MTL and the striatum are due to a differential recruitment of declarative and non-declarative memory processes: the paired associate task (Poldrack et al., 2001) as well as the memory task in the Knowlton et al. (1996) study requires declarative memory operations, whereas learning in the probabilistic classification task is based on non-declarative memory processes.

Further support for this notion is provided by animal studies. For example, in a series of studies Packard and colleagues (e.g. Packard & McGaugh, 1996; Packard, 1999) trained rats to perform a plus-maze task, i.e. the animals learned routes from a start arm (e.g. south) to a goal arm (e.g. west; here the animals received food).

Start and goal positions were held constant during the first days of training. In a second phase of the experiments, rats were placed in a new start position (e.g. north). The authors distinguished between ‘response learners’ and ‘place learners’. Response learners select the west-goal arm, whereas place learners select the east goal-arm after a new start box was introduced. According to Packard, response learners learn a response tendency, whereas place learners build a cognitive map of the environment in memory, e.g. representations of the start and goal position and their relationships (O’Keefe & Nadel, 1978).

In the study of Packard and McGaugh (1996), normal rats exhibited place learning in initial probe trials, whereas in later probe trials response learning was predominant, thus suggesting a shift from a place-based to a response-based strategy with extended training. For rats receiving either intra-caudate and intra-hippocampal infusions of local anesthetic lidocaine before the second phase of the experiment, a double dissociation was observed (control rats received intra-hippocampal and intra-caudate infusion of saline). Rats with intra-caudate infusions of lidocaine exhibited place learning in the early as well as in the late trials, whereas place learning was blocked in initial trials for rats with intra-hippocampal lidocaine injections. In contrast, these rats showed response learning in late trials. These data suggest that place learning is associated with the hippocampus, whereas response learning seems to be mediated by the caudate nucleus (see Hartley et al., 2003, for similar findings in humans). Most interestingly, learning is accompanied by a shift from a (fast) hippocampal-based learning system to a (slow) caudate-based learning system [C3,C7].

In a follow-up study, Packard (1999) could show that the posttraining infusion of glutamate in the hippocampus and the caudate nucleus selectively strengthens a specific learning system. Glutamate infusions in the hippocampus blocked the shift from a place to a response learning strategy, whereas rats with glutamate infusions in the caudate nucleus exhibited an accelerated shift to the response learning tendency.

Taken together, the results from different areas of behavioral and cognitive neuroscience suggest that separate brain systems mediate learning during differential

phases of learning, i.e. a fast hippocampal system at the beginning of learning and a slow striatal system at the end of learning. These results are in accordance with the findings from the present studies. Whereas hippocampal relational binding requirements diminish, striatal operations coding stimulus predictability evolve during the time-course of regularity learning in the IL condition.

Prefrontal-Hippocampal Interactions The results of the present thesis point to a dynamic interplay between medial temporal and lateral prefrontal brain regions during regularity learning. As described in Chapter 4, there is neuroanatomical evidence that the PFC and the MTL are interconnected by bidirectional projections (Laroche et al., 2000). Simons and Spiers (2003) propose that prefrontal control operations influence hippocampal processing during memory encoding and retrieval. For instance, the reorganization and selection of episodic features could bias hippocampal binding operations during encoding. During retrieval, an appropriate retrieval cue specification mediated by the PFC might support the reinstantiation of prior episodes by the hippocampus (Simons & Spiers, 2003). In the computational model of prefrontal memory control referred to above (Becker & Lim, 2003), the authors showed that the prefrontal cortex represents internal mnemonic codes. Becker and Lim's (2003) simulations reveal that "(t)hese self-organized mnemonic codes act as selective retrieval cues to the medial temporal lobe memory system" (p. 822). Evidence for this notion was provided by a study combining single-cell recordings and lesions in monkeys (Tomita et al., 1999). These data suggest that the PFC provides top-down signals to MTL structures during memory retrieval. Moreover, rats with PFC lesions are less able to suppress irrelevant spatial cues and thereby place fields of hippocampal place cells become less stable (Kyd & Bilkey, 2003). These results support the relevance of PFC-MTL interactions during memory formation. Based on these findings it could be speculated that the PFC modulates hippocampal processing by means of rule-like representations of task-relevant regularities (possibly in the form of mnemonic codes), leading to facilitated hippocampal binding.

Prefrontal-Striatal Interactions In the fMRI experiments of the present thesis, the lateral PFC and the striatum showed a strikingly similar learning-related activation pattern. Activity in both regions increased as a function of learning. This parallel pattern might suggest a cooperative functional interaction during the time-course of regularity learning. It has been argued that the striatum is involved in coding the increased predictive relationship among invariant episodic features, whereas the PFC represents these task-relevant regularities by means of an internal mnemonic code and by this facilitates the encoding of the current episode (i.e. a specific set of object-position conjunctions). It could be speculated that these predictive codes represented in the striatum serve as internal reinforcement signals (Koechlin et al., 2002; Schultz, 2002) provided to the PFC. It has been suggested that such (dopaminergic) reinforcement signals play a crucial role in maintaining and updating of prefrontal memory representations (Cohen, Braver & Brown, 2002; Miller, 2000; Miller & Cohen, 2001). Thus, it is conceivable that the PFC might implement a rule-like representation of regularities and by this an efficient encoding strategy on the basis of these reinforcement signals. Two recent computational models (Braver & Cohen, 2000; Frank et al., 2001) provided evidence for this notion. Both models showed that the PFC maintains task-relevant information in the face of interference. In both models, the basal ganglia (Frank et al., 2001) and the dopamine system (Braver & Cohen, 2000) contributed a gating mechanism. This gating mechanism entailed the flexible updating of PFC representations in the service of ongoing performance. In one model (Braver & Cohen, 2000), the gating signal is triggered by a reward prediction error, thus suggesting that basic dopaminergic signals are involved in updating prefrontal representations and by this might modify prefrontal control operations.² Braver and Cohen argue that dopamine “(...) plays a unified role in motivation and cognition by regulating cognitive processes to increase the frequency with which rewards are obtained.” (Braver & Cohen, 2000, p. 730) [C6]. Taken together, these simulations might indicate that the parallel learning-related activation pattern in the striatum and the PFC is due to a functional interaction. Striatal

²It is noteworthy that the authors used the same temporal difference learning algorithm as in the studies referred to above, e.g. O’Doherty et al., 2003, 2004.

reinforcement signals might be involved in updating information about regularities maintained in the PFC, by this stabilizing prefrontal rule-like representations.

11.5 Conclusions and Open Issues

In conclusion, the present thesis identified a hippocampal-striatal-prefrontal brain network supporting learning regularities across multiple episodes. A learning-related decrease of activation was observed in the hippocampus, whereas striatal and prefrontal activity increased as a function of learning. The hippocampus mediates regularity learning by means of relational binding of episodic features. During the time-course of learning, these binding requirements get reduced as reflected in decreased hippocampal activation. In contrast, the striatum seems to code the increased predictability of invariant episodic features and by this contributes to the development of a statistical model of the environment. Moreover, the present data indicate that the prefrontal cortex is involved in active maintenance of task-relevant regularities, possibly by means of rule-like aggregated representations. Based on these representations, the PFC might implement an efficient encoding strategy which entails increased performance.

The current model has important limitations. For instance, the issue of awareness was not directly addressed in the present experiments. It is still controversially discussed whether artificial grammar, category, or sequence learning is predominantly implicit or explicit in nature (Cleeremans et al., 1998; Frensch, 1998; Kinder & Shanks, 2003; Reber, 1989, 2002; Shanks & St. John, 1994) and whether distinct or overlapping brain networks might subserve either mechanism (Aizenstein et al., 2004; Chun & Phelps, 1999; Reber & Squire, 1998; Reber et al., 2003; Schendan et al., 2003). Are there differences with respect to the learning process and the underlying brain network, when subjects are (1) explicitly informed about the task-relevant regularities at the beginning of the experiment or when (2) an additional attentional demanding interference task is introduced, making an explicit learning strategy rather unlikely? It could be speculated that prefrontal operations involved in learning might be either facilitated (1) or hampered (2) by these experimental

manipulations, respectively.

Furthermore, future research should address the role of feedback and reward during regularity learning in more detail, e.g. by selectively manipulating the magnitude and salience of reward as well as the status of feedback, e.g. no feedback or misleading feedback (cf. Shohamy et al., 2004; Zink et al., 2004).

Moreover, the proposed functional interactions between the hippocampus, the striatum, and the prefrontal cortex are based on the comparison between differential learning-related activation pattern in these regions. To validate these assumptions experimentally, further analysis of functional connectivity are necessary (e.g. Dynamic Causal Modelling, DCM; Friston et al., 2003; Friston, in press). In the same vein, studies with neurological patients would provide new insights into the underlying brain network. Patients with selective damage the hippocampus (amnesics), the basal ganglia (Parkinson's, Chorea Huntington), and the prefrontal cortex should exhibit differential impairments during regularity learning, possibly restricted to specific phases of learning. With respect to such a lesion approach, it is conceivable that one region might compensate for the damage in another system (cf. Packard & Knowlton, 2002; Voermans et al., 2004).

Finally, beyond the learning functions used as regressors in the parametric fMRI designs of the present thesis, more fine-grained parameters derived from computational models simulating the underlying learning process could provide new insights into the neural basis of regularity learning (cf. O'Doherty et al., 2004; Seymour et al., 2004).

Even in the light of the limitations of the current model, it provides an account for the dynamic interaction between hippocampal, striatal, and prefrontal brain structures, specialized for distinct and even complementary functions during the time-course of learning. Thus, the current model may lead to significant challenges for further theoretical and experimental work.

References

- Aggleton, J. P. & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences*, 22, 425–489.
- Agster, K. L., Fortin, N. J., & Eichenbaum, H. (2002). The hippocampus and disambiguation of overlapping sequences. *The Journal of Neuroscience*, 22, 5760–5768.
- Aguirre, G. & D’Esposito, M. (1999). Experimental design for brain fMRI. In C. T. W. Moonen & P. A. Bandettini (Eds.), *Functional MRI* (pp. 369–380). Berlin: Springer.
- Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., & Carter, C. S. (2000). Complementary category learning systems identified using event-related functional MRI. *Journal of Cognitive Neuroscience*, 12, 977–987.
- Aizenstein, H. J., Stenger, V. A., Cochran, J., Clark, K., Johnson, M., Nebes, R. D., & Carter, C. S. (2004). Regional brain activation during concurrent implicit and explicit sequence learning. *Cerebral Cortex*, 14, 199–208.
- Alkire, M. T., Haier, R. J., Fallon, J. H., & Cahill, L. (1998). Hippocampal, but not amygdala, activity at encoding correlates with long-term, free recall of nonemotional information. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 14506–14510.
- Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 899–912.
- Alvarez, P. & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 7041–7045.
- Alvarez, P., Zola-Morgan, S., & Squire, L. R. (1995). Damage limited to the hippocampal region produces long-lasting memory impairment in monkeys. *The Journal of Neuroscience*, 15, 3796–3807.
- Amaral, D. G. (1999). What is where in the medial temporal lobe? *Hippocampus*, 9, 1–6.

- Amaral, D. G. & Insausti, R. (1990). The hippocampal formation. In G. Paxinos (Ed.), *The Human Nervous System* (pp. 711–755). San Diego: Academic Press.
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B. M., & Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *The Journal of Comparative Neurology*, *412*, 319–341.
- Andersson, J. L. R., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling geometric deformations in EPI time series. *Neuroimage*, *13*, 903–919.
- Ashburner, J. & Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, *7*, 254–266.
- Ashburner, J. & Friston, K. J. (2000). Spatial transformation. In C. T. W. Moonen & P. A. Bandettini (Eds.), *Functional MRI* (pp. 205–220). Berlin: Springer.
- Ashby, F. G. & Casale, M. B. (2003). The cognitive neuroscience of implicit category learning. In L. Jiménez (Ed.), *Attention and implicit learning* (pp. 109–141). Amsterdam: John Benjamins Publishing Company.
- Ashby, F. G. & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*, 204–210.
- Ashby, F. G. & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*, 124–150.
- Ashby, F. G. & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.
- Badre, D. & Wagner, A. D. (2004). Selection, integration, and conflict monitoring: assessing the nature and generality of prefrontal cognitive control mechanisms. *Neuron*, *41*, 473–487.
- Bandettini, P. A. (2000). The temporal resolution of functional MRI. In C. T. W. Moonen & P. A. Bandettini (Eds.), *Functional MRI* (pp. 205–220). Berlin: Springer.
- Becker, S. & Lim, J. (2003). A computational model of prefrontal control in free recall: strategic memory use in the California Verbal Memory Task. *Journal of Cognitive Neuroscience*, *15*, 821–832.
- Bellgowan, P. S. F., Saad, Z. S., & Bandettini, P. A. (2003). Understanding neural system dynamics through task modulation and measurement of functional MRI amplitude, latency, and width. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 1415–1419.
- Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *The Journal of Neuroscience*, *21*, 2793–2798.

- Birn, R. M., Cox, R. W., & Bandettini, P. A. (2002). Detection versus estimation in event-related fMRI: choosing the optimal stimulus timing. *Neuroimage*, *15*, 252–264.
- Birn, R. M., Saad, Z. S., & Bandettini, P. A. (2001). Spatial heterogeneity of the nonlinear dynamics in the fMRI BOLD response. *Neuroimage*, *14*, 817–826.
- Bischoff-Grethe, A., Goedert, K. M., Willingham, D. T., & Grafton, S. T. (2004). Neural substrates of response-based sequence learning using fMRI. *Journal of Cognitive Neuroscience*, *16*, 127–138.
- Bor, D., Duncan, J., Wiseman, R. J., & Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron*, *37*, 361–367.
- Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience*, *16*, 4207–4221.
- Brasted, P. J., Bussey, T. J., Murray, E. A., & Wise, S. P. (2003). Role of the hippocampal system in associative learning beyond the spatial domain. *Brain*, *126*, 1202–1223.
- Braver, T. S. & Brown, J. W. (2003). Principles of pleasure prediction: specifying the neural dynamics of human reward learning. *Neuron*, *38*, 150–152.
- Braver, T. S. & Cohen, J. D. (2000). On the control of control: the role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. Driver (Eds.), *In control of cognitive processes: Attention and performance XVIII* (pp. 713–737). Cambridge: MIT Press.
- Brett, M., Penny, W., & Kiebel, S. (in press). An introduction to random field theory. In R. S. J. Frackowiak, K. J. Friston, C. Frith, R. J. Dolan, C. J. Price, J. Ashburner, W. D. Penny & S. Zeki (Eds.), *Human brain function (2nd ed.)*. San Diego: Elsevier.
- Brewer, J. B. & Moghekar, A. (2002). Imaging the medial temporal lobe: exploring new dimensions. *Trends in Cognitive Sciences*, *6*, 217–223.
- 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.
- Brown, M. W. & Aggleton, J. P. (2001). Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, *2*, 51–61.
- Büchel, C., Coull, J. T., & Friston, K. J. (1999). The predictive value of changes in effective connectivity for human learning. *Science*, *283*, 1538–1541.
- Büchel, 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.

- Büchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage*, 8, 140–148.
- Buckner, R. L. (1998). Event-related fMRI and the hemodynamic response. *Human Brain Mapping*, 6, 373–377.
- Buckner, R. L., Bandettini, P. A., O’Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., & Rosen, B. R. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 14878–14883.
- Bunge, S. A., Ochsner, K. N., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (2001). Prefrontal regions involved in keeping information in and out of mind. *Brain*, 124, 2074–2086.
- Bunsey, M. & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, 379, 255–257.
- Burgess, N., Maguire, E. A., & O’Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, 35, 625–641.
- Burgess, N., Maguire, E. A., Spiers, H. J., & O’Keefe, J. (2001). A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *Neuroimage*, 14, 439–453.
- Buxton, R. B. (2001). The elusive initial dip. *Neuroimage*, 13, 953–958.
- Buxton, R. B. & Frank, L. R. (1997). A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *Journal of Cerebral Blood Flow and Metabolism*, 17, 64–72.
- Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the Balloon model. *Magnetic Resonance in Medicine*, 39, 855–864.
- Cabeza, R., Rao, S. M., Wagner, A. D., Mayer, A. R., & Schacter, D. L. (2001). Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 4805–4810.
- Cameron, K. A., Yashar, S., Wilson, C. L., & Fried, I. (2001). Human hippocampal neurons predict how well word pairs will be remembered. *Neuron*, 30, 289–298.
- Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying encoding and retrieval of source memory. *Cerebral Cortex*, 12, 1048–1056.

- Cer, D. M. & O'Reilly, R. C. (in press). Neural mechanisms of binding in the hippocampus and neocortex: insights from computational models. In H. Zimmer, A. Mecklinger & U. Lindenberger (Eds.), *Binding in human memory: a neurocognitive approach*. Oxford: Oxford University Press.
- Chein, J. M. & Schneider, W. (in press). Designing effective fMRI experiments. In J. Grafman & I. Robertson (Eds.), *Handbook of neuropsychology*. Amsterdam: Elsevier.
- Chomsky, N. & Miller, G. A. (1958). Finite state languages. *Information and Control*, 1, 91–112.
- Chun, M. M. & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, 2, 844–847.
- Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences*, 2, 406–416.
- Cleeremans, A. & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120, 235–253.
- Clegg, B. A., DiGirolamo, G. J., & Keele, S. W. (1998). Sequence learning. *Trends in Cognitive Sciences*, 2, 275–281.
- Cocosco, C. A., Kollokian, V., Kwan, R. K. S., & Evans, A. C. (1997). Brainweb: online interface to a 3D MRI simulated brain database. *Neuroimage*, 5, 425.
- Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 17–30.
- Cohen, J. D., Braver, T. S., & Brown, J. W. (2002). Computational perspectives on dopamine function in prefrontal cortex. *Current Opinion in Neurobiology*, 12, 223–229.
- Cohen, M. S. (1996). Rapid MRI and functional applications. In A. W. Toga & J. C. G. Mazziotta (Eds.), *Brain mapping: the methods* (pp. 223–255). New York: Academic Press.
- Cohen, M. S. (2000). Echo-planar imaging and functional MRI. In C. T. W. Moonen & P. A. Bandettini (Eds.), *Functional MRI* (pp. 137–148). Berlin: Springer.
- Cohen, N. J. & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. Cambridge, MA: MIT Press.
- Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., & Nash, C. (1999). Hippocampal system and declarative (relational) memory: summarizing the data from functional neuroimaging studies. *Hippocampus*, 9, 83–98.

- Corbetta, M. & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3, 201–215.
- Curran, T. (1997). Higher-order associative learning in amnesia: evidence from the serial reaction time task. *Cerebral Cortex*, 9, 522–533.
- Curran, T. & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 189–202.
- Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8, 109–114.
- Dale, A. M. & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5, 329–340.
- Dale, A. M. & Halgren, A. M. (1999). Spatiotemporal mapping of brain activity by integration of multiple imaging modalities. *Current Opinion in Neurobiology*, 11, 202–208.
- Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, 47, 109–121.
- Davachi, L. (2004). The ensemble that plays together, stays together. *Hippocampus*, 14, 1–3.
- Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 2157–2162.
- Davachi, L. & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *Journal of Neurophysiology*, 88, 982–990.
- Dayan, P. & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, 36, 285–298.
- Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, 3, 1218–1223.
- D’Esposito, M., Postle, B. R., Jonides, J., & Smith, E. E. (1999). The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 7514–7519.
- Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 875–887.

- Dobbins, I. G., Rice, H. J., Wagner, A. D., & Schacter, D. L. (2003). Memory orientation and success: separable neurocognitive components underlying episodic recognition. *Neuropsychologia*, *41*, 318–333.
- Dobbins, I. G., Simons, J. S., & Schacter, D. L. (2004). fMRI evidence for separable and lateralized prefrontal memory monitoring processes. *Journal of Cognitive Neuroscience*, *16*, 908–920.
- Doeller, C. F. & Opitz, B. (2004). Cognitive and neural mechanisms subserving regularity learning. In A. Mecklinger, H. Zimmer & U. Lindenberger (Eds.), *Bound in memory: insights from behavioral and neuropsychological studies* (pp. 129–157). Aachen: Shaker Verlag.
- Doeller, C. F., Opitz, B., Krick, C. M., Mecklinger, A., & Reith, W. (in preparationa). Differential hippocampal and prefrontal-striatal contributions to instance-based and rule-based learning.
- Doeller, C. F., Opitz, B., Krick, C. M., Mecklinger, A., & Reith, W. (in preparationb). Domain-specific lateralization of the hippocampus during learning regularities across episodes.
- Doeller, C. F., Opitz, B., Krick, C. M., Mecklinger, A., & Reith, W. (in press). Prefrontal-hippocampal dynamics involved in learning regularities across episodes. *Cerebral Cortex*.
- Doeller, C. F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., & Schröger, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *Neuroimage*, *20*, 1270–1282.
- Dominey, P. F. (1998). Influences of temporal organization on sequence learning and transfer: comments on Stadler (1995) and Curran and Keele (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 234–248.
- Dominey, P. F., Lelekov, T., Ventre-Dominey, J., & Jaennerod, M. (1998). Dissociable processes for learning the surface structure and abstract structure of sensorimotor sequences. *Journal of Cognitive Neuroscience*, *10*, 734–751.
- Donaldson, D. I. & Buckner, R. L. (2001). Effective paradigm design. In P. Jefferard, P. M. Matthews & S. M. Smith (Eds.), *Functional magnetic resonance imaging: an introduction to methods* (pp. 177–196). Oxford: Oxford University Press.
- Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgement: how conscious and how abstract? *Journal of Experimental Psychology: General*, *113*, 541–555.

- Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1985). On consciousness in syntactic learning and judgment: a reply to Reber, Allen, and Regan. *Journal of Experimental Psychology: General*, 114, 33–49.
- Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience*, 2, 820–829.
- Dusek, J. A. & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 7109–7114.
- Düzel, E., Habib, R., Rotte, M., Guderian, S., Tulving, E., & Heinze, H.-J. (2003). Human hippocampal and parahippocampal activity during visual associative recognition memory for spatial and nonspatial stimulus configurations. *The Journal of Neuroscience*, 23, 9439–9444.
- Düzel, E., Vargha-Khadem, F., Heinze, H. J., & Mishkin, M. (2001). Brain activity evidence for recognition without recollection after early hippocampal damage. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 8101–8106.
- Eichenbaum, H. (1997). Declarative memory: insights from cognitive neurobiology. *Annual Review of Psychology*, 48, 547–572.
- Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nature Reviews Neuroscience*, 1, 41–50.
- Eichenbaum, H. (in press). Mechanisms of relational representation. In H. Zimmer, A. Mecklinger & U. Lindenberger (Eds.), *Binding in human memory: a neurocognitive approach*. Oxford: Oxford University Press.
- Eichenbaum, H., Alvarez, P., & Ramus, S. J. (2000). Animal models of amnesia. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology (Vol. 2) (2nd ed.)* (pp. 1–24). Amsterdam: Elsevier.
- Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The hippocampus, memory, and place cells: is it spatial memory or a memory space? *Neuron*, 23, 209–226.
- Eichenbaum, H., Otto, T., & Cohen, J. (1994). Two component functions of the hippocampal memory system. *Behavioral and Brain Sciences*, 17, 449–517.
- Eichenbaum, H., Schoenbaum, G., Young, B., & Bunsey, M. (1996). Functional organization of the hippocampal memory system. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 13500–13507.

- 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. *Nature Neuroscience*, *3*, 1149–1152.
- Erickson, M. A. & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*, 107–140.
- Erickson, M. A. & Kruschke, J. K. (2002). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review*, *9*, 160–168.
- Feenan, K. & Snodgrass, J. G. (1990). The effect of context on discrimination and bias in recognition memory for pictures and words. *Memory & Cognition*, *18*, 517–527.
- Fell, J., Klaver, P., Lehnertz, K., Grunwald, T., Schaller, C., Elger, C. E., & Fernández, G. (2001). Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nature Neuroscience*, *4*, 1259–1264.
- Ferbinteanu, J. & Shapiro, M. L. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron*, *18*, 1227–1239.
- 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., Klaver, P., Fell, J., Grunwald, T., & Elger, C. E. (2002). Human declarative memory formation: segregating rhinal and hippocampal contributions. *Hippocampus*, *12*, 514–519.
- Fernández, G., Weyerts, H., Schrader-Bölsche, M., Tendolkar, I., Smid, H. G., Tempelmann, C., Hinrichs, H., Scheich, H., Elger, C. E., Mangun, G. H., & Heinze, H.-J. (1998). Successful verbal encoding into episodic memory engages the posterior hippocampus: a parametrically analyzed functional magnetic resonance imaging study. *The Journal of Neuroscience*, *18*, 1841–1847.
- Fletcher, P., Büchel, C., Josephs, O., Friston, K., & Dolan, R. J. (1999). Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebral Cortex*, *9*, 168–178.
- Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., Papadakis, N., & Bullmore, E. T. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nature Neuroscience*, *4*, 1043–1048.
- Fletcher, P. C. & Henson, R. N. A. (2001). Frontal lobes and human memory: insights from functional neuroimaging. *Brain*, *124*, 849–881.

- Fortin, N. J., Agster, K. L., & Eichenbaum, H. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*, 5, 458–462.
- Fortin, N. J., Wright, S. P., & Eichenbaum, H. (2004). Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature*, 431, 188–191.
- Fox, P. T. & Raichle, M. E. (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proceedings of the National Academy of Sciences of the United States of America*, 83, 1140–1144.
- Fox, P. T., Raichle, M. E., Mintun, M. A., & Dence, C. (1988). Nonoxidative glucose consumption during focal physiologic neural activity. *Science*, 241, 462–464.
- Frackowiak, R. S. J., Friston, K. J., Frith, C., Dolan, R. J., Price, C. J., Ashburner, J., Penny, W. D., & Zeki, S. (in press). *Human brain function (2nd ed.)*. New York: Academic Press.
- Frank, M. J., Loughry, B., & O'Reilly, R. (2001). Interaction between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, & Behavioral Neuroscience*, 1, 137–160.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291, 312–316.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2002). Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. *Journal of Neurophysiology*, 88, 929–941.
- Frensch, P. A. (1998). One concept, multiple meanings: how to define the concept of implicit learning. In M. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 47–104). Thousand Oaks, CA: Sage Publications.
- Frensch, P. A., Buchner, A., & Lin, J. (1994). Implicit learning of unique and ambiguous serial transitions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 567–584.
- Frensch, P. A. & Miner, C. S. (1994). Effect of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition*, 22, 95–110.
- Fried, I., MacDonald, K. A., & Wilson, C. L. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron*, 18, 753–765.
- Frisk, V. & Milner, B. (1990). The role of the left hippocampal region in the acquisition and retention of story content. *Neuropsychologia*, 28, 349–359.

- Friston, K. J. (2002a). Bayesian estimation of dynamical systems: an application to fMRI. *Neuroimage*, *16*, 513–530.
- Friston, K. J. (2002b). Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience*, *25*, 221–250.
- Friston, K. J. (in press). Statistical parametric mapping. In R. S. J. Frackowiak, K. J. Friston, C. Frith, R. J. Dolan, C. J. Price, J. Ashburner, W. D. Penny & S. Zeki (Eds.), *Human brain function (2nd ed.)*. San Diego: Elsevier.
- Friston, K. J. & Büchel, C. (2000). Attentional modulation of effective connectivity from V2 to V5/MT in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 7591–7596.
- Friston, K. J., Büchel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, *6*, 218–229.
- Friston, K. J., Fletcher, P., Josephs, O., Holmes, A. P., Rugg, M. D., & Turner, R. (1998a). Event-related fMRI: characterizing differential responses. *Neuroimage*, *7*, 30–40.
- Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. J. (1991). Comparing functional (PET) images: the assessment of significant change. *Journal of Cerebral Blood Flow and Metabolism*, *11*, 690–699.
- Friston, K. J., Frith, C. D., Turner, R., & Frackowiak, R. S. J. (1995a). Characterizing evoked hemodynamics with fMRI. *Neuroimage*, *2*, 157–165.
- Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., Ashburner, J., Hinton, G., & Ashburner, J. (2002a). 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., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., & Turner, R. (1995b). Analysis of fMRI time series revisited. *Neuroimage*, *2*, 45–53.
- Friston, K. J., Holmes, A. P., Price, C. J., Büchel, C., & Worsley, K. J. (1999a). Multi-subject fMRI studies and conjunction analyses. *Neuroimage*, *10*, 385–396.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. B., Frith, C. D., & Frackowiak, R. S. J. (1995c). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, *2*, 189–210.

- Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time series. *Human Brain Mapping, 1*, 153–171.
- Friston, K. J., Josephs, O., Rees, G., & Turner, R. (1998b). Non-linear event-related responses in fMRI. *Magnetic Resonance in Medicine, 39*, 41–52.
- Friston, K. J., Josephs, O., Zarahn, E., Holmes, A. P., Rouquette, S., & Poline, J.-B. (2000a). To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *Neuroimage, 12*, 196–208.
- Friston, K. J., Mechelli, A., Turner, R., & Price, C. J. (2000b). Nonlinear responses in fMRI: the Balloon model, Volterra kernels and other hemodynamics. *Neuroimage, 12*, 466–477.
- Friston, K. J., Penny, W., Kiebel, S., Hinton, G., & Ashburner, J. (2002b). Classical and Bayesian inference in neuroimaging: theory. *Neuroimage, 16*, 465–483.
- Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A., & Dale, A. M. (1999b). Stochastic designs in event-related fMRI. *Neuroimage, 10*, 607–619.
- Fujii, N. & Graybiel, A. M. (2003). Representation of action sequence boundaries by macaque prefrontal cortical neurons. *Science, 301*, 1246–1249.
- Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology, 49*, 87–115.
- 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.
- Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage, 15*, 870–878.
- Giovanello, K. S., Schnyder, D. M., & Verfaellie, M. (2004). A critical role for the anterior hippocampus in relational memory: evidence from an fMRI study comparing associative and item recognition. *Hippocampus, 14*, 5–8.
- Golby, A. J., Poldrack, R. A., Brewer, J. B., Spencer, D., Desmond, J. E., Aron, A. P., & Gabrieli, J. D. E. (2001). Material-specific lateralization in the medial temporal lobe and prefrontal cortex during memory encoding. *Brain, 124*, 1841–1854.
- Goldman-Rakic, P. S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational memory. In F. Plum & V. Mountcastle (Eds.), *Handbook of physiology: Section 1. The nervous system (Vol. 5)* (pp. 373–417). Bethesda, MD: American Physiological Society.

- Gomez, R. & Schvaneveldt, R. W. (1994). What is learned from artificial grammars? Transfer tests of simple association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 396–410.
- Gottfried, J. A., O'Doherty, J. P., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*, 301, 1104–1107.
- Gottlieb, J. (2002). Parietal mechanisms of target representation. *Current Opinion in Neurobiology*, 12, 134–140.
- Grafton, S. T., Hazeltine, E., & Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience*, 7, 497–510.
- Grön, G., Wunderlich, A. P., Spitzer, M., Tomczak, R., & Riepe, M. W. (2000). Brain activation during human navigation: gender-different neural networks as substrate of performance. *Nature Neuroscience*, 3, 404–408.
- Grunwald, T., Pezer, N., Münte, T. F., Kurthen, M., Lehnertz, K., Van Roost, D., Fernández, G., Kutas, M., & Elger, C. E. (2003). Dissecting out conscious and unconscious memory (sub)processes within the human medial temporal lobe. *Neuroimage*, 20, S139–S145.
- Hagberg, G. E., Zito, G., Patria, F., & Sanes, J. N. (2001). Improved detection of event-related functional MRI signals using probability functions. *Neuroimage*, 14, 1193–1205.
- Haist, F., Gore, J. B., & Mao, H. (2001). Consolidation of human memory over decades revealed by functional magnetic resonance imaging. *Nature Neuroscience*, 4, 1139–1145.
- Hampson, R. E., Simalal, J. D., & Deadwyler, S. A. (1999). Distribution of spatial and nonspatial information in dorsal hippocampus. *Nature*, 402, 610–614.
- Hartley, T., Maguire, E. A., Spiers, H. J., & Burgess, N. (2003). The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron*, 37, 877–888.
- Hazeltine, E., Grafton, S. T., & Ivry, R. (1997). Attention and stimulus characteristics determine the locus of motor sequence encoding: a PET study. *Brain*, 120, 123–140.
- Heckers, S., Weiss, A. P., Alpert, N. M., & Schacter, D. L. (2002). Hippocampal and brain stem activation during word retrieval after repeated and semantic encoding. *Cerebral Cortex*, 12, 900–907.
- Heeger, D. J. & Ress, D. (2002). What does fMRI tell us about neuronal activity? *Nature Reviews Neuroscience*, 3, 142–151.

- Helmstaedter, C., Brosch, T., Kurthen, M., & Elger, C. E. (2004). The impact of sex and language dominance on material-specific memory before and after left temporal lobe surgery. *Brain*, *127*, 1518–1525.
- Henke, K., Buck, A., Weber, B., & Wieser, H. G. (1997). Human hippocampus establishes associations in memory. *Hippocampus*, *7*, 249–256.
- Henson, R. N. A. (in press). Analysis of fMRI timeseries: linear time-invariant models, event-related fMRI, and optimal experimental design. In R. S. J. Frackowiak, K. J. Friston, C. Frith, R. J. Dolan, C. J. Price, J. Ashburner, W. D. Penny & S. Zeki (Eds.), *Human brain function (2nd ed.)*. San Diego: Elsevier.
- Henson, R. N. A., Büchel, C., Josephs, O., & Friston, K. J. (1999). The slice-timing problem in event-related fMRI. *Neuroimage*, *9*, 125.
- Henson, R. N. A., Price, C. J., Rugg, M. D., Turner, R., & Friston, K. J. (2002). Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentation. *Neuroimage*, *15*, 83–97.
- Hoffman, K. L. & McNaughton, B. L. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science*, *297*, 2070–2073.
- Hollerman, J. R. & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, *1*, 304–309.
- Holmes, A. P., Blair, R. C., Watson, J. D. G., & Ford, I. (1992). Nonparametric analysis of statistic images from functional mapping experiments. *Journal of Cerebral Blood Flow and Metabolism*, *16*, 7–22.
- Holmes, A. P. & Friston, K. J. (1998). Generalizability, random effects and population inference. *Neuroimage*, *7*, 754.
- Homa, D., Dunbar, S., & Nohre, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 418–439.
- Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, *3*, 284–291.
- Howseman, A. M. & Bowtell, R. W. (1999). Functional magnetic resonance imaging: imaging techniques and contrast mechanism. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences*, *354*, 1179–1194.
- Huettel, S. C., Mack, P. B., & McCarthy, G. (2002). Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. *Nature Neuroscience*, *5*, 485–490.

- Huxter, J., Burgess, N., & O'Keefe, J. (2003). Independent rate and temporal coding in hippocampal pyramidal cells. *Nature*, *425*, 828–832.
- Iaria, G., Petrides, M., Dagher, A. Pike, B., & Bohbot, V. D. (2003). Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *The Journal of Neuroscience*, *23*, 5945–5952.
- Insausti, R., Herrero, M. T., & Witter, M. P. (1997). Entorhinal cortex of the rat: cytoarchitectonic subdivisions and the origin and distribution of cortical efferents. *Hippocampus*, *7*, 146–183.
- Jackson, O. & Schacter, D. L. (2004). Encoding activity in anterior medial temporal lobe supports subsequent associative recognition. *Neuroimage*, *21*, 456–462.
- Jog, M. S., Kubota, Y., Connolly, C. I., Hillegart, V., & Graybiel, A. M. (1999). Building neural representations of habits. *Science*, *286*, 1745–1749.
- Johnstone, T. & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemanns and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 524–531.
- Josephs, O. & Henson, R. N. A. (1999). Event-related functional magnetic resonance imaging: modelling, inference, and optimization. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences*, *354*, 1215–1228.
- Josephs, O., Turner, R., & Friston, K. (1997). Event-related fMRI. *Human Brain Mapping*, *5*, 243–248.
- Káli, S. & Dayan, P. (2004). Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nature Neuroscience*, *7*, 286–294.
- Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review*, *110*, 316–339.
- 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.
- Kim, D.-S., Duong, T. Q., & Kim, S.-G. (2000a). High-resolution mapping of iso-orientation columns by fMRI. *Nature Neuroscience*, *3*, 164–169.
- Kim, S.-G., Lee, S. P., Goodyear, B., & Silva, A. C. (2000b). Spatial resolution of BOLD and other fMRI techniques. In C. T. W. Moonen & P. A. Bandettini (Eds.), *Functional MRI* (pp. 195–203). Berlin: Springer.

- Kinder, A. & Shanks, D. R. (2001). Amnesia and the declarative/nondeclarative distinction: a recurrent network model of classification, recognition and repetition priming. *Journal of Cognitive Neuroscience*, 13, 648–669.
- Kinder, A. & Shanks, D. R. (2003). Neuropsychological dissociations between priming and recognition: a single-system connectionist account. *Psychological Review*, 110, 728–744.
- Kirchhoff, B. A., Wagner, A. D., Maril, A., & Stern, C. E. (2000). Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *The Journal of Neuroscience*, 20, 6173–6180.
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, 273, 1399–1402.
- Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, 3, 172–179.
- Knowlton, B. J. & Squire, L. R. (1993). The learning of categories: parallel brain systems for item memory and category knowledge. *Science*, 262, 1747–1749.
- Knowlton, B. J. & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 79–91.
- Knowlton, B. J. & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 169–181.
- Koch, I. & Hoffmann, J. (2000). The role of stimulus-based and response-based spatial information in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 863–882.
- Koechlin, E., Danek, A., Burnod, Y., & Grafman, J. (2002). Medial prefrontal and subcortical mechanisms underlying the acquisition of motor and cognitive action sequences in humans. *Neuron*, 35, 371–381.
- Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302, 1181–1185.
- Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Kyd, R. J. & Bilkey, D. K. (2003). Prefrontal cortex lesions modify spatial properties of hippocampal place cells. *Cerebral Cortex*, 13, 444–451.

- Laroche, S., Davis, S., & Jay, T. M. (2000). Plasticity at hippocampal to prefrontal cortex synapses: dual roles in working memory and consolidation. *Hippocampus*, *10*, 438–446.
- Lauwereyns, J., Takikawa, Y., Kawagoe, R., Kobayashi, S., Koizumi, M., Coe, B., Sakagami, M., & Hikosaka, O. (2002a). Feature-based anticipation of cues that predict reward in monkey caudate nucleus. *Neuron*, *33*, 463–473.
- Lauwereyns, J., Watanabe, K., Coe, B., & Hikosaka, O. (2002b). A neural correlate of response bias in monkey caudate nucleus. *Nature*, *418*, 413–417.
- Lavenex, P. & Amaral, D. G. (2000). Hippocampal-neocortical interaction: a hierarchy of associativity. *Hippocampus*, *10*, 420–430.
- Lavenex, P., Suzuki, W. A., & Amaral, D. G. (2002). Perirhinal and parahippocampal cortices of the macaque monkey: projections to the neocortex. *The Journal of Comparative Neurology*, *447*, 394–420.
- Leon, M. I. & Shadlen, M. N. (1999). Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron*, *24*, 415–425.
- Liang, Z.-P. & Lauterbur, P. C. (2000). *Principles of magnetic resonance imaging: a signal processing perspective*. New York: IEEE Press.
- Liao, C. H., Worsley, K. J., Poline, J.-B., Aston, J. A. D., Duncan, G. H., & Evans, A. C. (2002). Estimating the delay of the fMRI response. *Neuroimage*, *16*, 593–606.
- Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience*, *16*, 427–438.
- Lisman, J. E. (1999). Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron*, *22*, 233–242.
- Liu, T. T., Frank, L. R., Wong, E. C., & Buxton, R. B. (2001). Detection power, estimation efficiency, and predictability in event-related fMRI. *Neuroimage*, *13*, 759–773.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigations of the basis of the fMRI signal. *Nature*, *412*, 150–157.
- Luck, S. J. (1999). Direct and indirect integration of event-related potentials, functional magnetic resonance images, and single-unit recordings. *Human Brain Mapping*, *8*, 115–120.
- Magistretti, P. J. & Pellerin, L. (1999). Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences*, *354*, 1155–1163.

- Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S. J., Frith, C. D., & O'Keefe, J. (1998). Knowing where and getting there: a human navigation network. *Science*, *280*, 921–924.
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 4398–4403.
- Maguire, E. A., Valentine, E. R., Wilding, J. M., & Kapur, J. N. (2003). Routes to remembering: the brains behind superior memory. *Nature Neuroscience*, *6*, 90–95.
- Malenka, R. C. & Nicoll, R. A. (1999). Long-term potentiation - a decade of progress? *Science*, *285*, 1870–1874.
- Mandeville, J. B., Marota, J. J., Ayata, C., Zarachuk, G., Moskowitz, M. A., Rosen, B. R., & Weiskoff, R. M. (1999). Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *Journal of Cerebral Blood Flow and Metabolism*, *19*, 679–689.
- Mandeville, J. B. & Rosen, B. R. (2002). Functional MRI. In A. W. Toga & J. C. G. Mazziotta (Eds.), *Brain mapping: the methods* (pp. 315–349). New York: Academic Press.
- Manns, J. R., Hopkins, R. O., Reed, J. M., Kitchener, E. G., & Squire, L. R. (2003a). Recognition memory and the human hippocampus. *Neuron*, *37*, 171–180.
- Manns, J. R., Hopkins, R. O., & Squire, L. R. (2003b). Semantic memory and the human hippocampus. *Neuron*, *38*, 127–133.
- Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. *Journal of Physics C*, *10*, L55–L58.
- Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). The role of implicit and explicit processes in learning from examples: a synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1083–1100.
- Maviel, T., Durkin, T. P., Menzaghi, F., & Bontempi, B. (2004). Sites of neocortical reorganization critical for remote spatial memory. *Science*, *305*, 96–99.
- Mayes, A. R., Holdstock, J. S., Isaac, C. L., Montaldi, D., Grigor, J., Gummer, A., Cariga, P., Downes, J. J., Tsivilis, D., Gaffan, D., Gong, Q., & Norman, K. A. (2004). Associative recognition in a patient with selective hippocampal lesions and relatively normal item recognition. *Hippocampus*, *14*, 763–784.

- Mayr, U. (1996). Spatial attention and implicit sequence learning: evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 350–364.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. (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. *Psychological Review*, 102, 419–457.
- McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, 38, 339–346.
- McGaugh, J. L. (2000). Memory - a century of consolidation. *Science*, 287, 248–251.
- Mecklinger, A. (2000). Interfacing mind and brain: a neurocognitive model of recognition memory. *Psychophysiology*, 37, 565–582.
- Mecklinger, A., Bosch, V., Gruenewald, C., Bentin, S., & von Cramon, D. Y. (2000). What have Klingon letters and faces in common? An fMRI study on content-specific working memory systems. *Human Brain Mapping*, 11, 146–161.
- Mecklinger, A., Weber, K., Gunter, T. C., & Engle, R. W. (2003). Dissociable brain mechanisms for inhibitory control: effects of interference content and working memory capacity. *Cognitive Brain Research*, 18, 26–38.
- Menon, R. S. & Kim, S.-G. (1999). Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends in Cognitive Sciences*, 3, 207–216.
- Menon, R. S., Luknowsky, D. C., & Gati, J. S. (1998). Mental chronometry using latency-resolved functional fMRI. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 10902–10907.
- Metcalfe, J. & Fisher, R. P. (1986). The relation between recognition memory and classification learning. *Memory & Cognition*, 14, 164–173.
- Meulemans, T. & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1007–1028.
- Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., & Buckner, R. L. (2000). Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage*, 11, 735–759.
- Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience*, 1, 59–65.

- Miller, E. K. & Asaad, E. F. (2002). The prefrontal cortex: conjunction and cognition. In J. Grafman (Ed.), *Handbook of neuropsychology* (pp. 29–54). New York: Elsevier.
- Miller, E. K. & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.
- Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanism of visual working memory in prefrontal cortex of the macaque. *The Journal of Neuroscience*, 16, 5154–5167.
- Mirenowicz, J. & Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *Journal of Neurophysiology*, 72, 1024–1027.
- Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. *Nature*, 273, 297–298.
- Mishkin, M., Malamut, B., & Bachevalier, J. (1984). Memories and habits: two neural systems. In J. L. McGaugh, G. Lynch & H. M. Weinberger (Eds.), *Neurobiology of learning and memory* (pp. 65–77). New York: Guilford Press.
- Mishkin, M., Vargha-Khadem, F., & Gadian, D. G. (1998). Amnesia and the organization of the hippocampal system. *Hippocampus*, 8, 212–216.
- Mitchell, K. J., Johnson, M. K., Raye, C. L., & Greene, E. J. (2004). Prefrontal cortex activity associated with source monitoring in a working memory task. *Journal of Cognitive Neuroscience*, 16, 921–934.
- Morris, R. G., Garrud, P., Rawlins, J. N., & O’Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature*, 297, 681–683.
- Mugler, J. P. & Brookeman, J. R. (1990). Three-dimensional magnetization-prepared rapid gradient-echo imaging (3-D MP RAGE). *Magnetic Resonance in Medicine*, 15, 152–157.
- Muller, R. (1996). A quarter of a century of place cells. *Neuron*, 17, 979–990.
- Muller, R. U., Kubie, J. L., & Ranck, J. B. (1987). Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *The Journal of Neuroscience*, 7, 1935–1950.
- Murray, E. A. & Bussey, T. J. (1999). Perceptual-mnemonic functions of the perirhinal cortex. *Trends in Cognitive Sciences*, 3, 142–151.
- Murray, E. A. & Mishkin, M. (1998). Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *The Journal of Neuroscience*, 18, 6568–6582.

- Murray, E. A. & Richmond, B. J. (2001). Role of perirhinal cortex in object perception, memory, and associations. *Current Opinion in Neurobiology*, 11, 188–193.
- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature Neuroscience*, 8, 774–781.
- Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus*, 10, 352–368.
- Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Sun, L. D., Kato, A., Carr, C. A., Johnston, D., Wilson, M. A., & Tonegawa, S. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science*, 297, 211–218.
- Nichols, T. E. & Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, 15, 1–25.
- Nissen, M. J. & Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology*, 19, 1–32.
- Norman, K. A. & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary learning systems approach. *Psychological Review*, 110, 611–646.
- Norris, D. G. (2003). High field human imaging. *Journal of Magnetic Resonance Imaging*, 18, 519–529.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
- O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, 38, 329–337.
- O'Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304, 452–454.
- O'Doherty, J. P., Deichmann, R., Critchley, H., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron*, 33, 815–826.
- O'Doherty, J. P., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representation in the human orbitofrontal cortex. *Nature Neuroscience*, 4, 96–102.

- Ogawa, S. & Lee, T. M. (1990). Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation. *Magnetic Resonance in Medicine*, 16, 9–18.
- Ogawa, S., Lee, T. M., Nayak, A. S., & Glynn, P. (1990a). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, 14, 68–78.
- Ogawa, S., Lee, T. M., & Tank, D. W. (1990b). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 9868–9872.
- Ohye, C., Kimura, M., & McKenzie, J. S. (1996). *The basal ganglia V*. New York: Plenum.
- O'Keefe, J. (1999). Do hippocampal pyramidal cells signal non-spatial as well as spatial information? *Hippocampus*, 9, 352–364.
- O'Keefe, J. & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, 381, 425–428.
- O'Keefe, J. & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34, 171–175.
- O'Keefe, J. & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Oxford University Press.
- Ollinger, J. M., Shulman, G. L., & Corbetta, M. (2001a). Separating processes within a trial in event-related functional MRI I. The methods. *Neuroimage*, 13, 210–217.
- Ollinger, J. M., Shulman, G. L., & Corbetta, M. (2001b). Separating processes within a trial in event-related functional MRI II. Analysis. *Neuroimage*, 13, 218–229.
- Opitz, B. & Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *Neuroimage*, 19, 1730–1737.
- Opitz, B. & Friederici, A. D. (2004). Brain correlates of language learning: the neural dissociation of rule-based vs similarity-based learning. *The Journal of Neuroscience*, 24, 8436–8440.
- Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., & Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *Neuroimage*, 15, 167–174.
- O'Reilly, R. C. & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: avoiding a tradeoff. *Hippocampus*, 4, 661–682.

- O'Reilly, R. C. & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.
- O'Reilly, R. C. & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems approach. *Trends in Cognitive Sciences*, 6, 505–510.
- O'Reilly, R. C. & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus*, 10, 389–397.
- O'Reilly, R. C. & Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychological Review*, 108, 311–345.
- 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.
- Packard, M. G. (1999). Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 12881–12886.
- Packard, M. G. & Knowlton, B. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, 25, 563–593.
- Packard, M. G. & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory*, 65, 65–72.
- Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience*, 5, 97–98.
- Paivio, A. (1986). *Mental representations: a dual coding approach*. Oxford: Oxford University Press.
- Paller, K. A. & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, 6, 93–102.
- Pauling, L. & Coryell, C. D. (1936). The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxy-hemoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, 22, 210–216.
- Pavlov, I. P. (1927). *Conditioned reflexes*. Oxford: Oxford University Press.
- Perruchet, P. (1994). Defining the knowledge units of a synthetic language: commentary on Vokey and Brooks (1992). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 223–228.

- Perruchet, P. & Pacteau, C. (1990). Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge. *Journal of Experimental Psychology: General*, *119*, 264–275.
- 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. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences*, *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. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences*, *354*, 1261–1281.
- Petrides, M. & Pandya, D. N. (1994). Comparative architectonic analysis of the human and the macaque frontal cortex. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (pp. 17–57). Amsterdam: Elsevier Science.
- Petrides, M. & Pandya, D. N. (1999). Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *European Journal of Neuroscience*, *11*, 1011–1036.
- Petrides, M. & Pandya, D. N. (2002). Association pathways of the prefrontal cortex and functional observations. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 31–50). Oxford: Oxford University Press.
- Pfeuffer, J., McCullough, J. C., Van de Moortele, P.-F., Ugurbil, K., & Hu, X. (2003). Spatial dependence of the nonlinear BOLD response at short stimulus duration. *Neuroimage*, *18*, 990–1000.
- Poldrack, R. A., Clark, J., Paré-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. & Packard, M. G. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*, *41*, 245–251.
- Posner, M. I. & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.
- Posner, M. I. & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*, 304–308.
- Postle, B. R., Berger, J. S., & D’Esposito, M. (1999). Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working

- memory performance. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 12959–12964.
- Pothos, E. M. (in press). The rules versus similarity distinction. *Behavioral and Brain Sciences*.
- Pothos, E. M. & Bailey, T. M. (2000). The role of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 847–862.
- Prabhakaran, V., Narayanan, K., & Gabrieli, J. D. E. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience*, 3, 85–90.
- Preston, A. R., Shrager, Y., Dudukovic, N. M., & Gabrieli, J. D. E. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, 14, 148–152.
- Rajkowska, G. & Goldman-Rakic, P. S. (1995a). Cytoarchitectonic definition of prefrontal areas in the normal human cortex: I. Remapping of areas 9 and 46 using quantitative criteria. *Cerebral Cortex*, 5, 307–322.
- Rajkowska, G. & Goldman-Rakic, P. S. (1995b). Cytoarchitectonic definition of prefrontal areas in the normal human cortex: II. Variability in locations of areas 9 and 46 and relationship to the Talairach coordinate system. *Cerebral Cortex*, 5, 323–337.
- Ranganath, C., Cohen, M. X., Dam, C., & D'Esposito, M. (2004a). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *The Journal of Neuroscience*, 24, 3917–3925.
- Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004b). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, 42, 2–13.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855–863.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118, 219–235.
- Reber, P. J. (2002). Attempting to model dissociations of memory. *Trends in Cognitive Sciences*, 6, 192–194.
- Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M.-M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience*, 15, 574–583.

- Reber, P. J. & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning & Memory*, 2, 1–13.
- Reber, P. J. & Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience*, 10, 248–263.
- Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998a). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory*, 5, 420–428.
- Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998b). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 747–750.
- Reingold, E. M. & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*, 44, 563–575.
- Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II. Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Richardson, M. P., Strange, B. A., Duncan, J. S., & Dolan, R. J. (2003). Preserved verbal memory function in left medial temporal pathology involves reorganisation of function to right medial temporal lobe. *Neuroimage*, 20 (Suppl 1), S112–S119.
- Robson, M. D., Dorosz, J. L., & Gore, J. C. (1998). Measurements of the temporal fMRI response of the human auditory cortex to trains of tones. *Neuroimage*, 7, 185–198.
- Rolls, E. T. (2000). Memory systems in the brain. *Annual Review of Psychology*, 51, 599–630.
- Rose, M., Haider, H., Weiller, C., & Büchel, C. (2002). The role of medial temporal lobe structures in implicit learning: an event-related fMRI study. *Neuron*, 36, 1221–1231.
- Rosen, B. R., Buckner, R. L., & Dale, A. M. (1998). Event-related functional MRI: past, present, and future. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 773–780.
- Röther, J., Knab, R., Hamzei, F., Fiehler, J., Reichenbach, J. R., Büchel, C., & Weiller, C. (2002). Negative dip in BOLD fMRI is caused by blood-flow oxygen consumption uncoupling in humans. *Neuroimage*, 15, 98–102.
- Rouder, J. N. & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General*, 133, 63–82.

- Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S. J., & Passingham, R. E. (2000). The prefrontal cortex: response selection or maintenance within working memory? *Science*, *288*, 1656–1660.
- Rugg, M. D., Otten, L. J., & Henson, R. N. A. (2002). The neural basis of episodic memory: evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences*, *357*, 1097–1110.
- Rugg, M. D. & Yonelinas, A. P. (2003). Human recognition memory: a cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *7*, 313–319.
- Ryan, L., Nadel, L., Keil, K., Putnam, K., Schnyer, D., Trouard, T., & Moscovitch, M. (2001). Hippocampal complex and retrieval of recent and very remote autobiographical memories: evidence from functional magnetic resonance imaging in neurologically intact people. *Hippocampus*, *11*, 707–714.
- Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience*, *5*, 479–484.
- Schacter, D. L., Alpert, N. M., Savage, C. R., Rauch, S. L., & Albert, M. S. (1996). Conscious recollection and the human hippocampal formation: evidence from positron emission tomography. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 321–325.
- 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.
- Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An fMRI study of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, *37*, 1013–1025.
- Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature Neuroscience*, *1*, 155–159.
- Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews Neuroscience*, *1*, 199–207.
- Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*, 241–263.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.
- Schultz, W., Tremblay, L., & Hollerman, J. R. (2003). Changes in behavior-related neuronal activity in the striatum during learning. *Trends in Neurosciences*, *26*, 321–328.

- Schultz, W., Trembley, L., & Hollerman, J. R. (2000). Reward processing in the primate orbitofrontal cortex and basal ganglia. *Cerebral Cortex*, 10, 272–283.
- Scoville, W. B. & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20, 11–21.
- Servan-Schreiber, E. & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 592–608.
- Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., & Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, 429, 664–667.
- Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge: Cambridge University Press.
- Shanks, D. R. & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17, 367–447.
- Shastri, L. (2002). Episodic memory and cortico-hippocampal interactions. *Trends in Cognitive Sciences*, 6, 162–168.
- Sherry, D. F. & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review*, 94, 439–454.
- Shidara, M. & Richmond, B. J. (2002). Anterior cingulate: single neuronal signals related to the degree of reward expectancy. *Science*, 296, 1709–1711.
- Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology*, 28, 207–218.
- Shohamy, D., Myers, C. E., Grossman, S., Sagem, J., Gluck, M. A., & Poldrack, R. A. (2004). Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. *Brain*, 127, 851–859.
- Simons, J. S. & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*, 4, 637–648.
- Smith, E. E. & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283, 1657–1661.
- Smith, M. L. & Milner, B. (1981). The role of the right hippocampus in the recall of spatial location. *Neuropsychologia*, 19, 781–793.
- Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34–50.

- Spiers, H. J., Maguire, E. A., & Burgess, N. (2001). Hippocampal amnesia. *Cortex*, 7, 357–382.
- Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195–231.
- Squire, L. R. & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology*, 5, 169–177.
- Squire, L. R., Clark, R. E., & Knowlton, B. J. (2001). Retrograde amnesia. *Hippocampus*, 11, 50–55.
- Squire, L. R. & Knowlton, B. J. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 12470–12474.
- Squire, L. R. & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 13515–13522.
- Squire, L. R. & Zola, S. M. (1997). Amnesia, memory and brain systems. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences*, 352, 1663–1673.
- Squire, L. R. & Zola, S. M. (1998). Episodic memory, semantic memory, and amnesia. *Hippocampus*, 8, 205–211.
- Squire, L. R. & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, 253, 1380–1386.
- Stadler, M. A. (1995). Role of attention in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 674–685.
- Stefanacci, L., Suzuki, W. A., & Amaral, D. G. (1996). Organization of connections between the amygdaloid complex and the perirhinal and parahippocampal cortices in macaque monkeys. *The Journal of Comparative Neurology*, 375, 552–582.
- 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. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 8660–8665.
- Strange, B. A., Fletcher, P. C., Henson, R. N. A., Friston, K. J., & Dolan, R. J. (1999). Segregating the functions of human hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 4034–4039.

- Strange, B. A., Henson, R. N. A., Friston, K. J., & Dolan, R. J. (2001). Anterior prefrontal cortex mediates rule learning in humans. *Cerebral Cortex*, *11*, 1040–1046.
- Sutton, R. S. & Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychological Review*, *88*, 135–170.
- Sutton, R. S. & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. Moore (Eds.), *Learning and computational neuroscience: foundations of adaptive networks* (pp. 497–537). Cambridge, MA: MIT Press.
- Suzuki, W. A. (1996). The anatomy, physiology and functions of the perirhinal cortex. *Current Opinion in Neurobiology*, *6*, 179–186.
- Suzuki, W. A. (1999). The long and the short of it: memory signals in the medial temporal lobe. *Neuron*, *24*, 295–298.
- Suzuki, W. A. & Amaral, D. G. (1994a). Perirhinal and parahippocampal cortices of the macaque monkey. *The Journal of Comparative Neurology*, *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. *The Journal of Neuroscience*, *14*, 1856–1877.
- Suzuki, W. A. & Amaral, D. G. (2003). Perirhinal and parahippocampal cortices of the macaque monkey: cytoarchitectonic and chemoarchitectonic organization. *The Journal of Comparative Neurology*, *463*, 67–91.
- Swanson, L. W. & Petrovich, G. D. (1998). What is the amygdala? *Trends in Neurosciences*, *23*, 323–331.
- Talairach, J. & Tournoux, P. (1988). *Co-planar stereotaxis atlas of the human brain*. New York: Thieme.
- Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., & Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience*, *7*, 887–893.
- Tettamanti, M., Alkadhi, H., Moro, A., Perani, D., Kollias, S., & Weniger, D. (2002). Neural correlates for the acquisition of natural language syntax. *Neuroimage*, *17*, 700–709.
- Thierry, A.-M., Gioanni, Y., Dégénétais, E., & Glowinski, J. (2000). Hippocampo-prefrontal cortex pathway: anatomical and electrophysiological characteristics. *Hippocampus*, *10*, 411–419.
- Thompson, J. K., Peterson, M. R., & Freeman, R. D. (2003). Single-neuron activity and tissue oxygenation in the cerebral cortex. *Science*, *299*, 1070–1072.

- de Toledo-Morrell, L., Dickerson, B., Sullivan, M. P., Spanovic, C., Wilson, R., & Bennett, D. A. (2000). Hemispheric differences in hippocampal volume predict verbal and spatial memory performance in patients with Alzheimer's disease. *Hippocampus*, *10*, 136–142.
- Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, *401*, 699–703.
- Toni, I., Ramnani, N., Josephs, O., Ashburner, J., & Passingham, R. E. (2001). Learning arbitrary visuomotor associations: temporal dynamic of brain activity. *Neuroimage*, *14*, 1048–1057.
- Tremblay, L. & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, *398*, 704–708.
- Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron*, *41*, 281–292.
- Tsukiura, T., Fujii, T., Takahashi, T., Xiao, R., Sugiura, M., Okuda, J., Iijima, T., & Yamadori, A. (2002). Medial temporal lobe activation during context-dependent relational processes in episodic retrieval: an fMRI study. *Human Brain Mapping*, *17*, 203–213.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.
- Tulving, E. (1985). Memory and consciousness. *Canadian Journal of Psychology*, *26*, 1–12.
- Tulving, E. & Markowitsch, H. J. (1998). Episodic and declarative memory: role of the hippocampus. *Hippocampus*, *8*, 198–204.
- Turner, R. (2002). How much cortex can a vein drain? Downstream dilution of activation-related cerebral blood oxygenation changes. *Neuroimage*, *16*, 1062–1067.
- Ugurbil, K. (2002). High-field magnetic resonance. In A. W. Toga & J. C. G. Mazziotta (Eds.), *Brain mapping: the methods* (pp. 291–313). New York: Academic Press.
- Ugurbil, K., Adriany, G., Andersen, P., Chen, W., Gruetter, R., Hu, X., Merkle, H., Kim, D.-S., Kim, S.-G., Strupp, J., Zhu, X. H., & Ogawa, S. (2000). Magnetic resonance studies of brain function and neurochemistry. *Annual Review of Biomedical Engineering*, *2*, 633–660.
- Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

- Vanzetta, I. & Grinvald, A. (1999). Increased cortical oxidative metabolism due to sensory stimulation: implications for functional brain imaging. *Science*, *286*, 1555–1558.
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Conelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, *277*, 376–380.
- Vazquez, A. L. & Noll, D. C. (1998). Nonlinear aspects of the BOLD response in functional MRI. *Neuroimage*, *7*, 108–118.
- 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 the caudate nucleus during route recognition. *Neuron*, *43*, 427–435.
- Vokey, J. R. & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammar. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 328–344.
- Vokey, J. R. & Brooks, L. R. (1994). Fragmentary knowledge and the processing-specific control of structural sensitivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1504–1510.
- Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*, 43–48.
- Wagner, A. D. (2002). Cognitive control and episodic memory: contributions from prefrontal cortex. In L. R. Squire & D. L. Schacter (Eds.), *Neuropsychology of memory* (pp. 174–192). New York: Guilford Press.
- Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R., & Buckner, R. L. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, *281*, 1188–1191.
- Wallenstein, G. V., Eichenbaum, H., & Hasselmo, M. E. (1998). The hippocampus as an associator of discontiguous events. *Trends in Neurosciences*, *21*, 317–323.
- Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*, 953–956.
- Wan, H., Aggleton, J. P., & Brown, M. W. (1999). Different contributions of the hippocampus and perirhinal cortex to recognition memory. *The Journal of Neuroscience*, *19*, 1142–1148.
- Warner, J. J. (2001). *Atlas of neuroanatomy*. Boston: Butterworth-Heinemann.
- Watanabe, M. (1996). Reward expectancy in primate prefrontal neurons. *Nature*, *382*, 629–632.

- White, N. M. & McDermott, R. J. (2002). Multiple parallel memory systems in the brain of the rat. *Neurobiology of Learning and Memory*, 77, 125–184.
- Willingham, D. B., Salidis, J., & Gabrieli, J. D. E. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. *Journal of Neurophysiology*, 88, 1451–1460.
- Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003). Single neurons in the hippocampus and learning of new associations. *Science*, 300, 1578–1581.
- Wood, E. R., Dudchenko, P. A., & Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, 397, 613–616.
- Wood, E. R., Dudchenko, P. A., Robitsek, R. J., & Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, 27, 623–633.
- Wood, J. N. & Grafman, J. (2003). Human prefrontal cortex: processing and representational perspectives. *Nature Reviews Neuroscience*, 4, 139–147.
- Worsley, K. J., Evans, A. C., Marrett, S., & Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12, 900–918.
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4, 58–73.
- Yeshurun, Y. & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396, 72–75.
- Yonelinas, A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences*, 356, 1363–1374.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.
- 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. *Nature Neuroscience*, 5, 1236–1241.
- Zarahn, E., Aguirre, G., & D'Esposito (1997). A trial-based experimental design for fMRI. *Neuroimage*, 6, 122–138.

- Zeineh, M. M., Engel, S. A., Thompson, P. M., & Bookheimer, S. Y. (2001). Unfolding the human hippocampus with high resolution structural and functional MRI. *The Anatomical Record (New Anat.)*, 265, 111–120.
- Zeineh, M. M., Engel, S. A., Thompson, P. M., & Bookheimer, S. Y. (2003). Dynamics of the hippocampus during encoding and retrieval of face-name pairs. *Science*, 299, 577–580.
- Zimmer, H., Mecklinger, A., & Lindenberger, U. (in press). *Binding in human memory: a neurocognitive approach*. Oxford: Oxford University Press.
- Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., & Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron*, 42, 509–517.
- Zola, S. M., Squire, L. R., Teng, E., Stefanacci, L., Buffalo, E. A., & Clark, R. E. (2000). Impaired recognition memory in monkeys after damage limited to the hippocampal region. *The Journal of Neuroscience*, 20, 451–463.
- Zola-Morgan, S. & Squire, L. R. (1985). Medial temporal lesions in monkeys impair memory on a variety of tasks sensitive to human amnesia. *Behavioral Neuroscience*, 99, 22–34.
- Zola-Morgan, S., Squire, L. R., & Mishkin, M. (1982). The neuroanatomy of amnesia: amygdala-hippocampus versus temporal stem. *Science*, 218, 1337–1339.
- Zola-Morgan, S., Squire, L. R., & Ramus, S. J. (1994). Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system. *Hippocampus*, 94, 483–494.

List of Figures

2.1. Network Architecture of the MTL Cortex and the Hippocampal Model (Complementary Learning Systems Approach; O'Reilly & Norman, 2002). . .	17
2.2. Design and Results of the Fortin et al. (2002) Study.	21
2.3. Results of the Fell et al. (2001) and Davachi et al. (2003) Study.	27
3.1. Learning-related Activity of a Dopamine Neuron and Brain Regions In- volved in Reward Processing.	42
3.2. Stimuli and Results of the Freedman et al. (2001) Study and Example Stim- uli of the Prototype Distortion Task.	49
3.3. Finite State Grammar and Sequence Learning.	54
3.4. The Finite State Grammar and Results of the Opitz and Friederici (2003) Study.	58
4.1. Neuroanatomy of the Medial Temporal Lobe (MTL).	68
4.2. Cytoarchitectonic Maps of the Prefrontal Cortex (PFC) Following the No- tation of Petrides and Pandya (1999).	71
4.3. Medial Temporal-prefrontal Interconnections.	73
5.1. Nuclear Spin Systems in the Magnet.	78
5.2. T_1 and T_2 Relaxation Curves.	79
5.3. Examples of MR Sequences.	82
5.4. The Physiological Basis and the Main Characteristics of the BOLD Response.	86
5.5. Modeling the BOLD Response by Different Temporal Basis Functions.	93
5.6. Results of the Dale and Buckner (1997) Study.	96
6.1. Session, Block, and Trial Structure of the Experiments.	104
6.2. Experimental Conditions.	105
7.1. Experiment 1A: Trial and Block Structure of the Experiment.	115
7.2. Experiment 1A: Behavioral Results.	121
7.3. Experiment 1A: Direct Comparisons between Conditions.	122
7.4. Experiment 1A: Brain Regions Showing a Learning-related Activation Pat- tern.	127

8.1. Experiment 1B: Experimental Procedure and Behavioral/Imaging Results. .	140
8.2. Experiment 1B: Learning-related Activation Pattern of the Hippocampus. .	142
8.3. Experiment 1B: Across Experiment Comparison.	143
9.1. Experiment 2A: Stimuli and Results.	154
9.2. Experiment 2B: Stimuli and Results.	156
9.3. Experiment 2C: Results.	160
9.4. Experiment 2D: Results.	162
9.5. Experiment 2E: Stimuli and Results.	164
9.6. Experiment 2F: Results.	166
10.1. Experiment 3A/3B: Session, Block, and Trial Structure.	175
10.2. Experiment 3A/3B: Behavioral Results - Within-block Learning.	181
10.3. Experiment 3A/3B: Behavioral Results - Across-block Learning.	183
10.4. Experiment 3A/3B: Imaging Results - Within-block Learning.	185
10.5. Experiment 3A/3B: Imaging Results - Across-block Learning.	187
10.6. Experiment 3A/3B: Imaging Results: Slope-related Activation Pattern. . . .	188
11.1. A Tentative Model of Regularity Learning.	206

List of Tables

7.1. Experiment 1A: fMRI Activation Foci - Direct Contrasts (All Trials).	123
7.2. Experiment 1A: fMRI Activation Foci - Direct Contrasts (Correct Trials).	124
7.3. Experiment 1A: fMRI Activation Foci - Learning-related Activity.	127
9.1. Experiment 2A-2F: Task and Design Parameters.	151
10.1. Experiment 3A/3B: Imaging Results - Within-block Learning.	184
10.2. Experiment 3A/3B: Imaging Results - Across-block Learning.	186
11.1. Two Separate Learning and Memory Systems: Main Characteristics and Differences.	199

Symbols and Abbreviations

γ	gyromagnetic ratio
$\vec{\mu}$	magnetic moment
χ	magnetic susceptibility
ω_0	Larmor frequency
$\vec{B}_0$	magnetic field strength
$\vec{J}$	angular momentum
$\vec{M}$	bulk magnetization
$\vec{M}_{xy}$	transverse magnetization
$\vec{M}_z$	longitudinal magnetization
T_1	longitudinal relaxation/spin-lattice relaxation
T_2	transverse relaxation/spin-spin relaxation
T_A	acquisition time
T_E	echo time
T_R	repetition time

ACC	anterior cingulate cortex
AGL	artificial grammar learning
ANOVA	analysis of variance
BA	Brodmann area
BOLD	blood oxygen level dependent
CA	cornu ammonis
CN	caudate nucleus
CoS	collateral sulcus
CR	conditioned response
CS	conditioned stimulus
CS cond.	context-specific condition
DCT	discrete cosine transform
DG	dentate gyrus
dHb	deoxyhemoglobin
DNMS	delayed nonmatch to sample task
EPI	echo-planar imaging
ERC	entorhinal cortex

ERP	event-related potential
FDR	false discovery rate
FG	fusiform gyrus
FIR	finite impulse response
fMRI	functional magnetic resonance imaging
HC	hippocampus
HRF	hemodynamic response function
IFG	inferior frontal gyrus
IFS	inferior frontal sulcus
IL cond.	invariant learning condition
IPL	inferior parietal lobule
ISI	interstimulus interval
ITI	intertrial interval
LG	lingual gyrus
LTI	linear time invariant
LTP	long term potentiation
MFG	middle frontal gyrus
MNI	Montreal Neurological Institute
MPRAGE	magnetization-prepared rapid gradient-echo imaging
MRI	magnetic resonance imaging
MTL	medial temporal lobe
PCC	posterior cingulate cortex
PE	prediction error
PEB	parametrical empirical Bayes
PET	positron emission tomography
PFC	prefrontal cortex
PHC	parahippocampal cortex
PRC	perirhinal cortex
PST	peristimulus time
PUT	putamen
rCBF	regional cerebral blood flow
RF	radio frequency
ROI	region of interest
RT	reaction time
SFG	superior frontal gyrus
SLF	superior longitudinal fasciculus
SMA	supplementary motor area
SnPM	Statistical Non-Parametric Mapping
SOA	stimulus onset asynchrony
SOS	superior occipital sulcus
SPECT	single photon emission computerized tomography

SPM	Statistical Parametric Mapping
SRT	serial reaction time task
ST	striatum
SVC	small volume correction
TD	temporal difference
UR	unconditioned response
US	unconditioned stimulus

Index

- T_1 , 78
- T_1 -weighted, 81
- T_2 , 79
- T_2 -weighted, 81
- T_2' , 79
- T_A , 89
- T_E , 81
- T_R , 81
- T_2^* , 79
- k -space, 81
- 3-D MP-RAGE, 82
- acquisition time, 89
- across-block learning, 173
- AGL, 51
- ALCOVE, 47
- AMTL-N400, 24
- amygdala, 67
- amygdalar complex, 67
- angular momentum, 76
- anterior cingulate gyrus, 70
- anterior thalamic nuclei, 69
- anterior/frontopolar PFC, 69
- arcuate fasciculus, 71
- artificial grammar learning, 51
- ATRIUM, 47
- BA, 69
- Balloon model, 85
- basolateral complex, 67
- bed nuclei of the stria terminalis, 67
- Bloch equation, 80
- blocked designs, 94
- blocking effect, 40
- blocking paradigm, 40
- blood oxygen level dependent, 83
- BOLD, 83
- Bonferroni correction, 93
- boxcar, 92
- Brodmann area, 69
- bulk magnetization, 76
- CA1, 68
- CA2, 68
- CA3, 68
- canonical HRF, 92
- category learning, 46
- caudate nucleus, 186
- central amygdalar nucleus, 67
- chunk strength, 52
- chunk-based learning, 52
- cingulate cortex, 68, 72
- cingulum bundle, 72
- classical conditioning, 38
- CN, 186
- collateral sulcus, 67
- conditioned reaction, 38
- conditioned stimulus, 38
- context-specific condition, 103
- contrast, 93
- contrast vector, 93
- contrast-to-noise ratio, 87
- coregistration, 90
- cornu ammonis, 68
- CoS, 68
- CR, 38
- CS, 38, 103
- data vector, 91
- DCT, 92

decision-bound models, 47
 delayed nonmatch to sample task, 17
 dentate gyrus, 67
 deoxyhemoglobin, 84
 design matrix, 91
 detection power, 98
 deterministic designs, 97
 DG, 68
 dHb, 84
 dipole analysis, 96
 discrete cosine transform, 92
 DNMS, 17
 dopamine, 39
 dorsal limbic pathway, 72
 dorsolateral PFC, 69
 Dynamic Causal Modelling, 88

 echo time, 81
 echo-planar imaging, 82
 entorhinal cortex, 67
 EPI, 82
 Ernst angle, 81
 ERP, 24
 error vector, 91
 estimation efficiency, 98
 Euler characteristic, 94
 event-related designs, 95
 event-related potentials, 24
 exemplar-based learning, 47, 55
 extinction, 39
 extreme capsule, 71

 false discovery rate, 94
 false-positive rate, 93
 familiarity, 30
 FDR, 94
 FG, 68
 finite-state grammar, 51
 FIR basis set, 93
 first system, 6
 flip angle, 81
 fornix, 69

 Fourier basis set, 93
 Fourier transform, 80
 fragment-based learning, 52, 53, 55
 free precession, 78
 frontomarginal gyrus, 70
 frontopolar gyrus, 70
 fusiform gyrus, 67

 Gaussian priors, 90
 Gaussian random field theory, 94
 General Recognition Theory, 47
 Generalized Context Model, 47
 gradient-echo, 83
 gyromagnetic ratio, 76
 gyrus rectus, 70

 Hahn echo, 83
 half-sine, 92
 HC, 68
 hemispheric asymmetry, 138
 hemispheric specialization, 138
 hemodynamic response function, 87, 92
 hemoglobin, 83
 hippocampal formation, 12, 67
 hippocampal region, 67
 hippocampal system, 67
 hippocampus, 67
 hippocampus proper, 67
 HRF, 87, 92
 hyperparameters, 90

 IFG, 122
 IFS, 122
 IL, 103
 inferior frontal gyrus, 120
 inferior frontal sulcus, 120
 inferior longitudinal fasciculus, 71
 inferior parietal lobule, 69, 71, 121
 information-integration tasks, 46
 initial dip, 85
 instance-based learning, 55, 173
 insula, 68

prototype, 47
 prototype-distortion tasks, 46
 PST, 92
 PUT, 186
 putamen, 141

 radial maze task, 18
 radio frequency, 77
 rCBF, 83
 realignment, 89
 recollection, 30
 regional cerebral blood flow, 83
 reinforcement learning, 38
 remember know procedure, 23, 29
 ReML, 117
 repetition time, 81
 Rescorla-Wagner formula, 38
 resolution elements, 94
 retrosplenial cortex, 69, 72
 rule-based learning, 47, 52, 173
 rule-based tasks, 46

 Schaffer collateral pathway, 69
 second system, 6
 sequence learning, 57
 serial reaction time, 57
 SFG, 69
 single photon emission computerized tomography, 86
 SLF, 70
 slice timing, 89
 slice-selection gradient, 80
 slope, 186
 SMA, 71
 smoothing, 90
 SnPM, 94
 SOA, 97
 SOS, 122
 SPECT, 86
 spin, 76
 spin density-weighted, 81
 spin-echo, 83

 spin-lattice relaxation, 78
 spin-spin relaxation, 78
 SPL, 70
 SRT, 57
 stationary stochastic designs, 97
 Statistical Non-Parametric Mapping, 94
 statistical parametric map, 93
 Statistical Parametric Mapping, 88
 stimulus model, 91
 stimulus onset asynchrony, 97
 striatum, 121
 Sub, 68
 subcallosal gyrus, 70
 subicular complex, 67
 subiculum, 67
 subsequent memory paradigm, 24
 substantia nigra, 40
 superior frontal gyrus, 69
 superior longitudinal fasciculus, 70
 superior occipital sulcus, 121
 superior parietal lobe, 70
 superior temporal gyrus, 68
 supplementary motor area, 71
 supramarginal gyrus, 71
 Sylvian fissure, 70

 TD, 39
 temporal basis function, 92
 temporal difference, 39
 theory of competitive chunking, 52
 transition matrix, 97
 transverse magnetization, 78
 transverse relaxation, 78
 triangular part of the IFG, 70

 uncinate fasciculus, 71
 unconditioned reaction, 38
 unconditioned stimulus, 38
 unwarping, 89
 UR, 38
 US, 38

ventral limbic pathway, 72
ventral tegmental area, 40
ventrolateral PFC, 69
Volterra hemodynamic model, 85
VTA, 40

Windkessel model, 85
within-block learning, 173

Zusammenfassung

Einleitung

Zahlreiche experimentalpsychologische und neurowissenschaftliche Modelle nehmen an, dass qualitativ unterschiedliche Lern- und Gedächtnissysteme existieren (McClelland et al., 1995; Mishkin et al., 1984; O’Keefe & Nadel, 1978; O’Reilly & Norman, 2002; Sherry & Schacter, 1987; Squire, 1992; Squire & Zola, 1996, vgl. Packard & Knowlton, 2002; Poldrack & Packard, 2003; Rolls, 2000; White & McDonald, 2002). Diese Modelle gehen davon aus, dass ein System für das schnelle Lernen unique Episoden spezialisiert ist (erstes System). Ein weiteres System ist für das graduelle Lernen episoden-übergreifender Regularitäten verantwortlich (zweites System). Die vorliegende Arbeit untersuchte in mehreren fMRT- und behavioralen Experimenten die neuronalen und kognitiven Grundlagen des hier skizzierten zweiten Systems, d.h. des Erwerbs episoden-übergreifender Regularitäten. Eine Kernfunktion des ersten Systems ist die Bildung episodischer Gedächtniseinträge (Tulving, 1983). Zahlreiche Tierläsionsstudien (Fortin et al., 2002; Zola et al., 2000) und tierphysiologische Studien (Wirth et al., 2003; Wood et al., 1999), Patientenstudien (Mishkin et al., 1998; Yonelinas et al., 2002) und elektrophysiologische (Fell et al., 2001; Fernández et al., 1999) und Bildgebungsstudien am Menschen (Brewer et al., 1998; Davachi et al., 2003; Eldrige et al., 2000; Gabrieli et al., 1997; Otten et al., 2001) haben gezeigt, dass der Hippocampus eine zentrale Rolle bei Enkodierung und Abruf episodischer Erinnerungen spielt. Zwei einflussreiche Theorien über die Rolle des Hippocampus wurden in den letzten Jahren aufgestellt, das Eichenbaum-Modell und das O’Reilly-Modell. Basierend auf zahlreichen tierexperimentellen Befunden geht Eichenbaum (Eichenbaum et al., 1994, 1996; Eichenbaum, 1997, in press, vgl. Cohen et al., 1999) davon aus, dass der Hippocampus wesentlich an einer Reihe von gedächtnisrelevanten relationalen Bindungsprozessen beteiligt ist, z.B. der Bindung von Objekt- und Rauminformationen zu einem kohärenten Ereignis und dem Binden von Ereignissen in ihrer zeitlichen Sequenz (Bunsey & Eichenbaum, 1996; Dusek & Eichenbaum, 1997; Fortin et al., 2002; Wood et al., 1999, vgl. Davachi & Wagner, 2002; Giovanello et al., 2004; Preston et al., 2004). In ihrem computationalen Modell nehmen O’Reilly und Mitarbeiter (McClelland et al., 1995; Norman & O’Reilly, 2003; O’Reilly & Munakata, 2000; O’Reilly & Rudy, 2000, 2001; O’Reilly & Norman, 2002) an, dass der Hippocampus durch ein spezielles Repräsentationsformat charakterisiert ist: ein sparsames, merkmals-separierendes Format (pattern-separated). Dieses

Format ermöglicht das schnelle Enkodieren und den schnellen Abruf von Einzelepisoden. Weiterhin gewährleistet dieses spezielle hippocampale Repräsentationsformat, dass Einzelepisoden separat von anderen Episoden repräsentiert werden können, um dadurch die Gefahr von Interferenzeffekten zu verringern. Beide Modelle haben unterschiedliche Annahmen bzgl. der neuronalen Grundlagen des Lernens von episoden-übergreifenden Regularitäten. Eichenbaum nimmt an, dass spezielle hippocampale Neurone Repräsentationen distinkter Episoden binden, um episoden-übergreifende Merkmale zu extrahieren (Agster et al., 2002; Eichenbaum, in press; Wallenstein et al., 1998). Auf der anderen Seite geht O'Reilly davon aus, dass der rhinale und parahippocampale Kortex überlappende Merkmale multipler Episoden repräsentiert.

Die Lernprozesse, die dem oben skizzierten zweiten Lern- und Gedächtnissystem zugrunde liegen, wurden in zahlreichen experimentalpsychologischen Studien und Experimenten, die sich moderner Bildgebungsmethoden bedienen untersucht. Drei zentrale Paradigmen, die verschiedene Aspekte des Lernens untersuchen, spielen in diesem Zusammenhang eine zentrale Rolle: Das Lernen künstlicher Grammatiken (Gomez & Schvaneveldt, 1994; Meulemans & Van der Linden, 1997; Reber, 1967; Perruchet & Pacteau, 1990), Kategorienlernen (Ashby & Perrin, 1988; Erickson & Kruschke, 1998; Nosofsky, 1986; Posner & Keele, 1968, 1970) und Sequenzlernen (Cleeremans & McClelland, 1991; Cohen et al., 1990; Nissen & Bullemer, 1987; Stadler, 1995). In diesen Paradigmen lernen Versuchspersonen (VP) überlappende Merkmale vieler Einzelepisoden, z.B. Buchstabenfolgen, die grammatikalischen Regeln folgen, überlappende Exemplare, die einer gleichen Kategorie angehören oder sich wiederholende Sequenzen von Stimuli (und Reaktionen). Es herrscht eine rege Diskussion, ob das Lernen in diesen Aufgaben auf dem Erwerb abstrakter Regeln oder relativ einfacher Oberflächeneigenschaften basiert (Ashby & Perrin, 1988; Dulany et al., 1984, 1985; Kruschke, 1992; Nosofsky, 1986; Perruchet & Pacteau, 1990; Perruchet, 1994; Pothos, in press; Reber, 1989; Servan-Schreiber & Anderson, 1990; Shanks & St. John, 1994). In Studien mit Amnestikern konnte gezeigt werden, dass das Lernen künstlicher Grammatiken, Kategorienlernen und Sequenzlernen weitestgehend unabhängig von mesiotemporalen Strukturen (einschließlich des Hippocampus) stattfindet (Curran, 1997; Knowlton & Squire, 1993, 1994, 1996; Reber & Squire, 1994, 1998; Squire & Knowlton, 1995). Zahlreiche Bildgebungsstudien zeigen zudem, dass der präfrontale Kortex (prefrontal cortex, PFC) eine entscheidende Rolle bei diesen Aufgaben spielt (Aizenstein et al., 2004; Bischoff-Grethe et al., 2004; Fletcher et al., 1999; Freedman et al., 2001; Hazeltine et al., 1997; Opitz & Friederici, 2003, 2004; Reber et al., 2003; Schendan et al., 2003; Strange et al., 2001).

Experimente

Zentrale Fragestellung

Die Experimente der vorliegenden Arbeit hatten zum Ziel, die neuronalen und kognitiven Grundlagen des Lernens episodien-übergreifender Regularitäten zu untersuchen. Zu diesem Zweck wurde ein Paradigma entwickelt, welches es gestattet, diese Lernprozesse abzubilden. Es wurde davon ausgegangen, dass hippocampale Bindungsprozesse initial zum Lernen benötigt werden. Im späteren Verlauf des Lernens sollten hippocampale Prozesse an Bedeutung verlieren, wohingegen die präfrontale Aktivierung als Funktion des Lernens ansteigen sollte. Es wird angenommen, dass der präfrontale Kortex sowohl an der Extraktion von Regularitäten als auch an der effizienten Enkodierung von Stimuli im Lichte der extrahierten Regularitäten beteiligt ist.

Methode

Procedere und Versuchsplan der Experimente In jedem Durchgang der Experimente mussten die VP sequentiell verschiedene Objekte an verschiedenen Positionen enkodieren (Sample-Phase; Mit Ausnahme von Experiment 2F; hier wurden die Objekt-Positions-Verknüpfungen in der Sample-Phase simultan präsentiert.). Objekte wurden in einer 4×4 Matrix präsentiert. Nach einer kurzen Pause wurde ein Objekt an einer Position präsentiert (Probe-Stimulus) und die VP mussten angeben, ob genau dieses Objekt an genau dieser Position in der Sample-Phase gezeigt wurde. Visuelles Feedback wurde direkt nach der Darbietung des Probe-Stimulus präsentiert (500 ms; grün = korrekt, rot = falsch, gelb = zu langsam). Probes waren zu 50% alte und zu 50% neue Objekt-Positions-Verknüpfungen (3 gleichhäufige Kategorien: altes Objekt/neue Position, neues Objekt/alte Position, neues Objekt/neue Position). Verschiedene Parameter, wie Stimulus-Typ, Anzahl an Objekt-Positions-Verknüpfungen in der Sample-Phase, Dauer der Stimulus-Präsentation, Inter-Stimulus Intervall (ISI), Dauer des Delays zwischen Sample- und Probe-Phase, Dauer der Probe-Präsentation, Inter-Trial Intervall (ITI), Anzahl der Durchgänge pro Block und Anzahl der Blöcke pro Bedingung wurden in den Experimenten selektiv manipuliert (siehe Tabelle S1).

Im vorliegenden experimentellen Ansatz wurde die Menge an sequentiellen Objekt-Positions-Verknüpfungen in der Sample-Phase als eine "Episode" definiert (vgl. Eichenbaum, 2000; O'Reilly & Norman, 2002). Das experimentelle Design umfasste 2 Bedingungen, eine kontext-spezifische Bedingung (context-specific, CS) und eine invariante Lern-Bedingung (invariant learning, IL). Mit Ausnahme der Experimente 2C-2F wurden in allen übrigen Experimenten Blöcke der CS-Bedingung und Blöcke der IL-Bedingung zusammen in einer Sitzung präsentiert. In der IL-Bedingung wurden Regularitäten innerhalb von experimentellen Blöcken eingeführt. Zwei Varianten der IL-Bedingung wurden verwendet. In einer räumlichen IL-Bedingung wurden innerhalb eines experimentellen Blocks variable Objekte an invarianten Positionen präsentiert, d.h. Objekte wurden immer an denselben Positionen

präsentiert. In einer Objekt IL-Bedingung hingegen wurden invariante Objekte an variablen Positionen präsentiert. Im Gegensatz zu diesen konstanten Positionen bzw. Objekten in der IL-Bedingung wurden in der CS-Bedingung sowohl Positionen als auch Objekte variabel gehalten.

Tabelle S1.
Design-Parameter der Experimente

Exp.	Reg.	Stimuli	Stimuli/ Sample- Phase	Stimulus- Dauer [ms]	ISI [ms]	Delay [ms]	Probe- Dauer [ms]	ITI [ms]	Durch- gänge/ Block	Anzahl Blöcke/ Bed.	VP
1A (fMRT)	Raum	geom. Formen	4	800	500	1300	2000	5000 (var.)	16	7	9
1B (fMRT)	Obj.	geom. Formen	4	800	500	1300	2000	5000 (var.)	16	7	8
2A (behav.)	Obj.	geom. Formen	4	800	500	1300	2000	2000 (const.)	16	7	16
2B (behav.)	Obj.	konkrete Objekte	4	800	500	1300	2000	2000 (const.)	16	7	16
2C (behav.)	Obj.	konkrete Objekte	6	800	500	1300	2000	5000 (var.)	36	6	16
2D (behav.)	Obj.	konkrete Objekte	6	500	0	2000	1000	5000 (var.)	36	6	16
2E (behav.)	Obj.	abstrakte Objekte	6	500	0	2000	1000	5000 (var.)	36	6	16
2F (behav.)	Obj.	abstrakte Objekte	6	2000	0	1000	1000	5000 (var.)	36	8	16
3A (fMRT)	Obj.	konkrete Objekte	6	600	100	2000	1000	4000 (var.)	36	4	12
3B (fMRT)	Raum	konkrete Objekte	6	600	100	2000	1000	4000 (var.)	36	4	12

Abkürzungen: Bed. = Bedingung; behav. = behavioral; const. = constant; Exp. = Experiment; fMRT = funktionelle Magnetresonanztomographie; geom. = geometrisch; ms = Millisekunden; Obj. = Objekt; Reg. = Regularität; var. = variabel; VP = Versuchspersonen

Behaviorale Methoden Zur Analyse der Performanz- und Lernleistung der Probanden wurden Pr-Werte berechnet. Der Diskriminationsindex Pr ist ein nicht-parametrisches Äquivalent des Parameters d' (vgl. Feenan & Snodgrass, 1990; Snodgrass & Corwin, 1988) und bezeichnet die Differenz von Treffern (richtige Antworten auf alte Probes) und falschen Alarmen (falsche Antworten auf neue Probes). Pr -Werte wurden jeweils über einige aufeinanderfolgende Durchgänge gemittelt, um den Lernverlauf innerhalb von Blöcken zu analysieren. Weiterhin wurden neben Pr -Werten auch die verschiedenen Arten von falschen Alarmen analysiert (d.h. falsche Alarme auf Probes, die entweder ein neues Objekt an einer alten Position, ein altes Objekt an einer neuen Position oder ein neues Objekt an einer neuen Position umfassten).

Funktionelle Bildgebung In den zentralen Experimenten der vorliegenden Arbeit wurde die Methode der funktionellen Magnetresonanztomographie (fMRT) verwendet. Die fMRT macht sich die unterschiedlichen magnetischen Eigenschaften von sauerstoffreichem Hämoglobin (Oxyhämoglobin) und sauerstoffarmem Hämoglobin (Deoxyhämoglo-

bin) zunutze (der sog. BOLD-Effekt [blood oxygen level dependent]; vgl. Ogawa et al., 1990a,b; Ogawa & Lee, 1990). Deoxyhämoglobin ist im Gegensatz zu Oxyhämoglobin paramagnetisch. Neuronale Aktivierung zieht eine Erhöhung des regionalen Blutflusses nach sich. Dieser Blutflussanstieg ist von einem vergleichsweise geringen Anstieg der Sauerstoffextraktion begleitet (Fox & Raichle, 1986; Fox et al., 1988). Dieser Effekt resultiert in einer Netto-Abnahme des Deoxyhämoglobin-Anteils am venösen Blut. Als Folge zeigt sich ein Signal-Anstieg in T_2^* -gewichteten MR-Bildern (BOLD-Effekt). Es scheint mittlerweile gesichert, dass das BOLD-Signal neuronale Aktivierung widerspiegelt, insbesondere den synchronisierten synaptischen Input einer Neuronen-Population (Logothetis et al., 2001, siehe Heeger und Ress, 2002, für einen detaillierten Überblick über die physiologischen Grundlagen des BOLD-Signals).

Die vorliegenden fMRT-Experimente wurden an einem Siemens Sonata Gerät bei einer Feldstärke von 1.5 Tesla durchgeführt. Zur strukturellen Messung wurde eine T_1 -gewichtete MPRAGE-Sequenz (Mugler & Brookeman, 1990) mit folgenden Parametern verwendet: $T_R = 1900$ ms, $T_E = 3.93$ ms, IT = 1100 ms, Flip-Winkel = 8° , Schicht-Dicke = 1 mm, Auflösung innerhalb einer Schicht = 1×1 mm, FoV = 256 mm^2 , 176 Schichten). In Experiment 3 wurde eine zusätzliche T_1 -gewichtete strukturelle Sequenz verwendet ($T_R = 600$ ms, $T_E = 13$ ms, Flip-Winkel = 80° , Schicht-Dicke = 4 mm, interslice gap = 1 mm, Auflösung innerhalb einer Schicht = 0.9×0.9 mm, FoV = 224 mm^2 , 20 axiale Schichten). BOLD-sensitive, T_2^* -gewichtete, funktionelle MR-Bilder wurden mit folgender Sequenz gewonnen: $T_R = 1.8$ s, $T_E = 50$ ms, Flip-Winkel = 85° , Schicht-Dicke = 4 mm, interslice gap = 1 mm, Auflösung innerhalb einer Schicht = 3.5×3.5 mm, FoV = 224 mm^2 , 20 axiale Schichten.

Die Daten-Vorverarbeitung und statistische Auswertung der fMRT-Daten erfolgte mit dem Software-Paket SPM2 (<http://www.fil.ion.ucl.ac.uk/spm/spm2.html>). In den vorliegenden fMRT-Experimenten wurde eine Kombination zwischen einer epochen- und ereigniskorrelierten Modellierung des BOLD-Signals durchgeführt. In beiden Fällen wurde zur Modellierung die sog. kanonische HRF (hemodynamic response function; Friston et al., 1998a) verwendet. Aktivierung während der Sample-Phase wurde als Epoche (entsprechend der Dauer der Sample-Phase), Aktivierung während der Probe-Phase als Ereignis modelliert. Die Auswertung beschränkte sich ausschließlich auf die Sample-Phase. Die Koordinaten der Aktivierungsfoci wurden in den kanonischen Talairach-Raum transformiert (Talairach & Tournoux, 1988, vgl. <http://www.mrc-cbu.cam.ac.uk/imaging/mnispac.html>). Zentraler Bestandteil der fMRT-Auswertung war eine sog. parametrische Analyse (siehe Henson, in press; Opitz & Friederici, 2003; Poldrack et al., 2001). Zu diesem Zweck wurden Performanzdaten der VP (Pr-Werte; Feenan & Snodgrass, 1990) als parametrische Regressoren in der fMRT-Auswertung berücksichtigt. Der Vorteil einer solchen Auswertungsstrategie besteht in der Möglichkeit, lernrelatierte Aktivierungsverläufe von Gehirnstrukturen abzubilden.

Experiment 1A

Fragestellung Ziel dieser Studie war es zu untersuchen, ob das Lernen von räumlichen Regularitäten und das Lernen uniquer Episoden auf separaten Gehirnstrukturen basiert. Zu diesem Zweck wurden in Experiment 1A Blöcke der CS-Bedingung mit Blöcken in der räumlichen IL-Bedingung kontrastiert (siehe Tabelle S1 für weitere Details des experimentellen Designs). Auf behavioraler Ebene wurde ein Performanz-Anstieg innerhalb von Blöcken der IL-Bedingung erwartet, wohingegen konstante Performanz innerhalb von Blöcken der CS-Bedingung vorhergesagt wurde. Auf neuronaler Ebene wurde konstante hippocampale Aktivierung in der CS-Bedingung erwartet, da VP in dieser Bedingung in jedem Durchgang unique Episoden enkodieren mussten. Im Gegensatz dazu wurde eine Abnahme hippocampaler Aktivierung in der IL-Bedingung als Funktion des Lernens erwartet, da die Einführung von räumlichen Regularitäten eine Reduktion relationaler Verarbeitungsanforderungen nach sich ziehen sollte. Eine zunehmende Beteiligung des präfrontalen Kortex wurde als Funktion des Lernens in der IL Bedingung vorhergesagt.

Ergebnisse Im Gegensatz zur CS-Bedingung zeigten VP in der IL-Bedingung einen signifikanten Anstieg der Performanz (Pr-Werte) innerhalb experimenteller Blöcke. Auf behavioraler Ebene konnte weiterhin eine Dissoziation der False-Alarm-Rate (FA) in der IL-Bedingung beobachtet werden. Die FA-Rate für bestimmte Probe-Kategorien (neues Objekt an alter Position, neues Objekt an neuer Position) unterschied sich weder für beide Bedingungen, noch zeigte sich ein Abfall der FA-Rate für diese Probes in der IL-Bedingung. Im Gegensatz dazu zeigte sich ein signifikanter Abfall der FA-Rate für Probes, die ein altes Objekt an einer neuen Position umfassten. Weiterhin war die FA-Rate für diese Probe-Kategorie in der IL-Bedingung im Vergleich zur CS-Bedingung signifikant reduziert.

Die fMRT-Auswertung umfasste 2 Schritte: In einem ersten Schritt wurden direkte Kontraste zwischen beiden Bedingungen berechnet ($p < 0.0005$, unkorrigiert). In einem zweiten Schritt wurden parametrische, lernrelatierte Auswertungen durchgeführt. Diese parametrischen Analysen wurden auf die im ersten Schritt identifizierten Areale beschränkt ($p < 0.05$, SVC [small volume corrected]; Worsley et al., 1996). Entsprechend der Vorhersagen zeigte der rechte Hippocampus einen lernrelatierten Aktivierungsabfall in der IL-Bedingung und konstante Aktivierung in der CS-Bedingung. Weiterhin ergab sich ein lernrelatiertes Aktivierungsanstieg des linken gyrus frontalis inferior (pars triangularis) und des rechten Striatums (ventrales Putamen) und einen lernrelatierten Aktivierungsabfall des linken lobus parietalis inferior und des rechten sulcus occipitalis superior.

Diskussion Die Ergebnisse dieser Studie zeigen, dass das Lernen von räumlichen Regularitäten durch einen Übergang von einem hippocampalen zu einem präfrontal-striatalen System begleitet wird. Eine Reduktion hippocampaler Aktivierung konnte als Funktion des Lernens beobachtet werden, wohingegen ein lernrelatiertes Aktivierungsanstieg im lateralen PFC (gyrus frontalis inferior) und im Striatum auftrat. Konstante hippocampale

Aktivierung konnte in der CS-Bedingung beobachtet werden. Es wird argumentiert, dass dieses konstante Aktivierungsmuster auf die gleichbleibenden relationalen Bindungsanforderungen zurückzuführen ist (Eichenbaum, 2000), da in der CS-Bedingung in jedem Durchgang variable Objekte und variable Positionen gebunden werden müssen. In der IL-Bedingung hingegen nehmen diese relationalen Bindungsanforderungen über die Zeit hinweg ab, da variable Objekte und invariante Positionen gebunden werden. Dies zeigt sich an der lernrelatierten Abnahme hippocampaler Aktivierung. In Gegensatz dazu scheint das Striatum die erhöhte Vorhersagbarkeit invarianter räumlicher Merkmale innerhalb der Blöcke zu kodieren (Koechlin et al., 2002, vgl. Schultz, 2002). Wie vorhergesagt zeigte der laterale PFC eine Aktivierungszunahme als Funktion des Lernens. Zahlreiche fMRT-Studien konnten eine Beteiligung des lateralen PFC in Lernexperimenten zeigen, z.B. beim Lernen künstlicher Grammatiken (Opitz & Friederici, 2003), beim Kategorienlernen (Reber et al., 1998b) und beim Sequenzlernen (Schendan et al., 2003). Es wird angenommen, dass der laterale PFC extrahierte Regularitäten in Form eines Gedächtniscodes (Becker & Lim, 2003) repräsentiert und diese Repräsentationen aktiv gegen Distraction aufrechterhält (Sakai et al., 2002). Neu zu enkodierende Episoden werden dann anhand dieser Gedächtniscodes reorganisiert, was zu einer effektiveren Enkodierstrategie und folglich zu einer erhöhten Aktivität des präfrontalen Kortex führt.

Experiment 1B

Fragestellung In dieser Studie wurde die Fragestellung untersucht, ob das beobachtete differentielle hippocampal-striatal-präfrontale Aktivierungsmuster während des Lernens von räumlichen Regularitäten auch dem Lernen von Objekt-Regularitäten zugrunde liegt. Zu diesem Zweck wurden Blöcke in der CS-Bedingung mit Blöcken einer Objekt-IL-Bedingung kontrastiert (siehe Tabelle S1 für weitere Details des experimentellen Designs). Im Gegensatz zu Experiment 1A wurden hier Objekte, nicht Positionen in Objekt-Positions-Verknüpfungen innerhalb experimenteller Blöcke konstant gehalten. Alle anderen experimentellen Parameter blieben unverändert gegenüber Experiment 1A.

Ergebnisse Analog zu Experiment 1A zeigten VP einen Anstieg der Performanz (Pr-Werte und insbesondere Treffer-Rate) innerhalb der experimentellen Blöcke in der IL-Bedingung, aber nicht in der CS-Bedingung. Die Auswertung der FA-Rate für spezifische Probe-Kategorien ergab ein komplementäres Muster im Vergleich zu Experiment 1A. In der IL-Bedingung (relativ zur CS-Bedingung) zeigte sich eine Reduktion der FA-Rate für Probes, die ein neues Objekt an einer alten Position umfassten. Keine Bedingungsunterschiede der FA-Rate konnten für Probes beobachtet werden, die ein altes Objekt beinhalteten.

Die lernrelatierte fMRT-Auswertung wurde hypothesengetrieben auf den Hippocampus, das Striatum, und den lateralen PFC beschränkt ($p < 0.05$, SVC). Im Gegensatz zu einem rechts-lateralisierten Aktivierungsabfall im Hippocampus in Experiment 1A konnte

ein bilateraler Abfall hippocampaler Aktivierung in der IL-Bedingung von Experiment 1B beobachtet werden. Weiterhin zeigten der linke gyrus frontalis medius und das rechte Striatum eine Aktivierungszunahme als Funktion des Lernens.

Diskussion In diesem Experiment wurde die Frage der Domänenspezifität von Lernprozessen untersucht. Im Gegensatz zu einem rechts-lateralisierten Aktivierungsabfall des Hippocampus beim Lernen von räumlichen Regularitäten konnte beim Lernen von Objekt-Regularitäten eine bilaterale Aktivierungsabnahme des Hippocampus beobachtet werden. Diese Ergebnisse deuten auf eine domänenspezifische hemisphärische Spezialisierung hippocampaler Bindungsprozesse hin, wie sie für episodische Gedächtnisleistungen bereits beschrieben wurden (Golby et al., 2001; Kelley et al., 1998). Die Sensitivität gegenüber den Reizmerkmalen der episoden-übergreifenden Regularitäten unterstützt die Annahme relationaler Verarbeitung im Hippocampus auf Basis der perzeptuellen Charakteristika einer Episode. Im Gegensatz dazu zeigte sich bzgl. der Aktivierungsfoci im lateralen PFC und im Striatum keine Hemisphären-Unterschiede zwischen Experiment 1A und 1B. Aufgrund dieser Daten kann angenommen werden, dass der laterale PFC Lernen von Regularitäten im Sinne eines domänen-übergreifenden Top-down Prozesses auf der Basis abstrakter Regularitäten unterstützt (Strange et al., 2001; Wallis et al., 2001).

Experiment 2

Fragestellung Experiment 2 hatte zum Ziel, die kognitiven Grundlagen des Lernens von Objekt-Regularitäten zu untersuchen. In 6 behavioralen Studien wurde der Effekt der Objekt-Distinktheit (Experiment 2A, 2B), der Lerndauer und der Gedächtnisbelastung (Experiment 2C, 2D, 2E) und der simultanen Präsentation von Objekt-Positions-Verknüpfungen in der Sample-Phase (Experiment 2F) auf das Lernen von Objekt-Regularitäten untersucht (siehe Tabelle S1 für weitere Details des experimentellen Designs).

Ergebnisse In allen 6 Untersuchungen wurden Objekte innerhalb von experimentellen Blöcken konstant gehalten. Neben den Blöcken in der IL-Bedingung wurden in Experiment 2A und 2B zusätzlich Blöcke in der CS-Bedingung präsentiert. In den übrigen Experimenten 2C-2F wurden Blöcke der IL-Bedingung sequentiell präsentiert und auf die Präsentation von Blöcken der CS-Bedingung verzichtet. In Experiment 2A wurden dieselben geometrischen Formen wie in Experiment 1A und 1B als Stimulusmaterial verwendet. Zur Erhöhung der Distinktheit der Objekte wurden jeweils 4 Objekte in derselben Farbe präsentiert (insgesamt wurden 4 unterschiedliche Farben verwendet). In Experiment 2B wurden Farbzeichnungen von konkreten Objekten des Alltagslebens als Stimulusmaterial verwendet. Eine Erhöhung der Objektdistinktheit relativ zu Experiment 1A und 1B hatte einen geringen Effekt auf die Lernleistung in Experiment 2A und 2B (gemessen als Verbesserung der Performanz innerhalb von Lernblöcken). Relativ zu Experiment 1A und 1B (geometrische Formen) zeigte sich jedoch eine Erhöhung der Gesamtgedächtnisleistung

(gemessen als mittlere Performanz über alle Blöcke).

In Experiment 2C-2E wurde einerseits die Gedächtnisbelastung durch die Präsentation von 6 (statt 4) Objekt-Positions-Verknüpfungen in der Sample-Phase und andererseits die Lerndauer durch die Einführung von 36 (statt 16) Durchgängen pro Block erhöht. Die Manipulation von Gedächtnisbelastung und Lerndauer hatte einen Einfluss auf die Lerndynamik. In allen drei Experimenten konnten Lerneffekte innerhalb der experimentellen Blöcke beobachtet werden. Diese Effekte waren am stärksten ausgeprägt in der ersten Hälfte des Experiments; in späteren Blöcken wurde die Lernasymptote hingegen schneller erreicht. In Experiment 2D wurde gegenüber Experiment 2C die Gedächtnisbelastung weiter erhöht, indem die Stimuluspräsentationszeit, das ISI und die Darbietungsdauer verkürzt wurden (vgl. Tabelle S1). Diese Manipulation führte zu einer weiteren Verstärkung des Lerneffekts bei gleichbleibender Gesamtgedächtnisleistung. In Experiment 2E wurden anstelle von konkreten Objekten abstrakte Objekte verwendet. Alle anderen Design-Parameter wurden im Vergleich zu Experiment 2D konstant gehalten. Die Lerneffekte blieben weitestgehend unbeeinflusst von dieser experimentellen Manipulation, wohingegen ein Abfall der Gesamtgedächtnisleistung beobachtet werden konnte. Gegenüber Experiment 2E wurde in Experiment 2F das Präsentationsformat modifiziert. Anstelle sequentieller Darbietung wurden die Objekt-Positions-Verknüpfungen in der Sample-Phase simultan präsentiert. Sowohl eine Verringerung des Lerneffekts als auch eine Reduktion der Gesamtgedächtnisleistung konnte beobachtet werden.

Diskussion Die behavioralen Daten zeigen, dass die Manipulation der Objekt-Distinktheit insbesondere einen Einfluss auf die Gedächtnisleistung hatte. Relativ zu Experiment 1B zeigte sich ein Anstieg der Gesamtperformanz in Experiment 2A und 2B. Weiterhin führte die Erhöhung der Gedächtnisbelastung und der Lerndauer zu einer Vergrößerung des Lerneffekts (Experiment 2C-2E). Der Vergleich zwischen Experiment 2C-2D (erhöhte Lerndauer und distinkte Objekte) mit Experiment 2E (erhöhte Lerndauer und nicht-distinkte Objekte) deutet darauf hin, dass Objekt-Distinktheit einen modulierenden Effekt auf die Lerndynamik hat, wenn die Lerndauer ausreichend groß ist. Die Daten von Experiment 2F zeigen, dass das Lernen episodens-übergreifender Regularitäten vermutlich das Vorhandensein distinkter Repräsentationen von Objekt-Positions-Verknüpfungen voraussetzt. Schließlich deuten insbesondere die Ergebnisse von Experiment 2C und 2D darauf hin, dass zwei separate Lernprozesse unterschieden werden können: Ein Prozess innerhalb von Lernblöcken und ein Prozess über Blöcke hinweg. Der erste Prozess scheint mit der Extraktion von Regularitäten assoziiert zu sein, wobei der zweite Prozess eher den Transfer des ersten Prozesses auf neue Exemplare widerzuspiegeln scheint.

Experiment 3

Fragestellung In Experiment 3 wurde der Frage nachgegangen, ob unterschiedliche Gehirnstrukturen dem Lernen von Regularitäten innerhalb von Lernblöcken und dem Ler-

nen über Lernblöcke hinweg zugrunde liegen. Die in Experiment 2 beobachtete Dissoziation, d.h. spätes Erreichen der Lernasymptote in den initialen Blöcken bzw. das frühe Erreichen der Lernasymptote in den finalen Blöcken, scheint dafür zu sprechen, dass dem Lernen innerhalb von Blöcken und dem Lernen über Blöcke hinweg unterschiedliche Prozesse zugrunde liegen. Blöcke in der CS-Bedingung wurden mit Blöcken in einer Objekt-IL-Bedingung (Experiment 3A) bzw. mit Blöcken einer räumlichen IL-Bedingung (Experiment 3B) kontrastiert. Im Gegensatz zu Experiment 1A und 1B wurden die Blöcke beider Bedingungen getrennt voneinander präsentiert. Als weitere Modifikationen gegenüber Experiment 1A und 1B wurden konkrete Objekte als Stimulusmaterial verwendet und sowohl die Gedächtnisbelastung als auch die Lerndauer erhöht (siehe Tabelle S1 für weitere Details des experimentellen Designs). Ein Performanzanstieg innerhalb von Blöcken sollte ähnlichkeitsbasiertes Lernen widerspiegeln, wohingegen Lernen über Blöcke hinweg auf regelbasierten Prozessen beruhen sollte (vgl. Opitz & Friederici, 2004).

Ergebnisse Auf behavioraler Ebene konnte sowohl für die Objekt-IL-Bedingung (Experiment 3A) als auch für die räumliche IL-Bedingung (Experiment 3B) ein Performanzanstieg (Pr-Werte) innerhalb von Lernblöcken beobachtet werden. In der entsprechenden CS-Bedingung in beiden Experimenten blieb die Performanz konstant innerhalb der Blöcke. Analog zu Experiment 1A und 1B zeigte sich eine Dissoziation der FA-Rate in Abhängigkeit von der spezifischen Probe-Kategorie für die jeweilige IL-Bedingung relativ zur entsprechenden CS-Bedingung. In Experiment 3A (Objekt-Regularitäten) verringerte sich die FA-Rate für (1) Probes, die ein neues Objekt beinhalteten, nicht hingegen die FA-Rate für (2) Probes, die eine neue Position beinhalteten. In Experiment 3B zeigte sich ein inverses Muster: Eine Verringerung der FA-Rate für Probe-Kategorie (2), nicht aber für Probe-Kategorie (1). In beiden Experimenten zeigte sich auch ein Lerneffekt über Blöcke hinweg (gemessen als Veränderung der gemittelten Performanz [Pr-Werte] der ersten beiden Blöcke gegenüber den letzten beiden Blöcken). Ein Anstieg der Pr-Werte konnte in der Objekt-IL-Bedingung (Experiment 3A) und der räumlichen IL-Bedingung (Experiment 3B) von Block 1-2 zu Block 3-4 beobachtet werden. Im Gegensatz dazu blieb die Performanz über Blöcke konstant in den entsprechenden CS-Bedingungen. Die fMRT-Auswertung zeigte einen lernrelatierten Abfall hippocampaler und einen Anstieg lateral-präfrontaler und striataler Aktivierung innerhalb von Lernblöcken in der IL-Bedingung beider Experimente. Weiterhin konnte eine Aktivierungsmodulation distinkter präfrontal-striataler Areale in der IL-Bedingung beider Experimente über Lernblöcke hinweg beobachtet werden.

Diskussion Die Ergebnisse dieser Studie legen eine Dissoziation von zwei Lerneffekten nahe: Das Lernen innerhalb von Blöcken und das Lernen über Blöcke hinweg. Sowohl in Experiment 3A als auch in Experiment 3B zeigten sich beide Lerneffekte, nicht jedoch in den entsprechenden CS-Bedingungen. Weiterhin konnte eine Dissoziation der FA-Rate in beiden Experimenten beobachtet werden. Diese Daten sprechen für eine Zunahme der

Verarbeitungssensitivität für invariante episodische Merkmale als Funktion des Lernens. Der erste Lerneffekt war mit einem lernrelatierten Abfall hippocampaler und einem Anstieg lateral-präfrontaler und striataler Aktivierung verknüpft. Dieses Muster steht im Einklang mit den Ergebnissen aus Experiment 1A und 1B. Weiterhin scheint der Lerneffekt über Blöcke hinweg mit distinkten präfrontal-striatalen Arealen assoziiert zu sein. Dieses Ergebnismuster deutet darauf hin, dass der Hippocampus eine spezifische Rolle beim ähnlichkeitsbasierten Lernen spielt, wohingegen präfrontal-striatale Strukturen sowohl ähnlichkeitsbasiertes als auch regelbasiertes Lernen unterstützen (Opitz & Friederici, 2004; Lieberman et al., 2004).

Allgemeine Diskussion

Theorien aus der Experimentalpsychologie und der tierexperimentellen, kognitiven und computationalen Neurowissenschaft legen nahe, dass qualitativ unterschiedliche Lern- und Gedächtnissysteme existieren (McClelland et al., 1995; Mishkin et al., 1984; O'Keefe & Nadel, 1978; O'Reilly & Norman, 2002; Sherry & Schacter, 1987; Squire, 1992; Squire & Zola, 1996, vgl. Packard & Knowlton, 2002; Poldrack & Packard, 2003; Rolls, 2000; White & McDonald, 2002). Ein erstes System scheint für das schnelle Lernen uniquer Episoden verantwortlich sein, wohingegen ein zweites System für den graduellen Erwerb von episoden-übergreifenden Regularitäten spezialisiert ist. Die hier beschriebenen Experimente hatten zum Ziel, die neuronalen und kognitiven Mechanismen zu untersuchen, die dem Lernen von episoden-übergreifenden Regularitäten zugrunde liegen (d.h. dem hier skizzierten zweiten Lern- und Gedächtnissystem).

In mehreren fMRT- und behavioralen Experimenten mussten VP mehrere Objekt-Positions-Verknüpfungen in einer Sample-Phase enkodieren und bei anschließender Präsentation einer Objekt-Positions-Verknüpfung (Probe-Stimulus) entscheiden, ob diese Verknüpfung in der Sample-Phase präsentiert wurde. Im hier vorgestellten experimentellen Ansatz wurde die Menge an Objekt-Positions-Verknüpfungen der Sample-Phase als 'Episode' definiert. In einer Bedingung wurden entweder Objekte (Experiment 1B, 2 und 3A) oder Positionen (Experiment 1A, 3B) in Objekt-Positions-Verknüpfungen innerhalb experimenteller Blöcke hinweg konstant gehalten (IL-Bedingung). Durch die Einführung invarianter episodischer Merkmale sollte das Lernen von Regularitäten ermöglicht werden. In einer weiteren Bedingung wurden hingegen in jedem Durchgang unique Objekt-Positions-Verknüpfungen präsentiert, d.h. es existierten keine episoden-übergreifenden Regularitäten (CS-Bedingung). Die behavioralen Ergebnisse der Experimente legen nahe, dass VP von der Einführung invarianter Merkmale profitieren, d.h. einen Performanzanstieg innerhalb von experimentellen Blöcken in der IL-Bedingung, jedoch nicht in der CS-Bedingung zeigen. Die fMRT-Daten sprechen für einen graduellen Übergang von einem hippocampalen zu einem präfrontal-striatalen System während des Lernens episoden-übergreifender

Regularitäten, wohingegen das Lernen unique Episoden mit konstanter hippocampaler Aktivierung assoziiert war.

Zwei Separate Lern- und Gedächtnis-Systeme: Schnelles Lernen Unique Episoden vs Graduelles Lernen episoden-übergreifender Regularitäten

Das erste System ist spezialisiert für das Lernen unique Episoden (O'Reilly & Norman, 2002; Squire, 1992). Diese Episoden sind charakterisiert durch nicht-prädiktive Beziehungen zwischen episodischen Merkmalen. Das zweite System ist verantwortlich für das Lernen episoden-übergreifender Regularitäten. Hier existieren prädiktive Beziehungen zwischen episodischen Merkmalen. Das erste System zeichnet sich durch eine hohe Verarbeitungsgeschwindigkeit aus, um merkmals-separierte (pattern-separated) Repräsentationen (O'Reilly & Norman, 2002) unique Episoden zu speichern. Das zweite System hingegen ist verantwortlich für den graduellen Erwerb von Regularitäten, die möglicherweise in einer aggregierten Repräsentationsform gespeichert werden (O'Reilly & Norman, 2002; Sherry & Schacter, 1987). Auf computationaler Ebene scheint Informationsspeicherung im ersten System in einer gewichts-basierten, im zweiten System hingegen in einer aktivations-basierten Form stattzufinden (vgl. Frank et al., 2001). Auf molekularer Ebene basiert Informationsverarbeitung und -speicherung im ersten System vermutlich auf Langzeit-Potenzierung und im zweiten System auf dopaminergen Mechanismen (Malenka & Nicoll, 1999; Schultz, 2002). Schließlich legen u.a. die Daten der vorliegenden Arbeit nahe, dass das erste System v.a. mit der Hippocampusformation, das zweite System hingegen mit dem Striatum und dem PFC assoziiert ist.

Eine Definition des Lernens episoden-übergreifender Regularitäten

Es wurden mehrere Kriterien aufgestellt, um das Lernen episoden-übergreifender Regularitäten zu beschreiben. (1) Grundlage von Lernen in den hier beschriebenen Experimenten ist das Vorhandensein überlappender Merkmale multipler Episoden. (2) Lernen episoden-übergreifender Regularitäten ist ein gradueller Prozess. Die Aggregation von Wissen erfolgt langsam über viele Einzelepisoden hinweg. (3) Im Verlauf eines Lernblocks müssen extrahierte Regularitäten aktiv aufrechterhalten werden. (4) Schließlich wird davon ausgegangen, dass die episoden-übergreifenden Regularitäten in einem aggregierten Format repräsentiert werden (im Gegensatz zu einer episodenspezifischen Repräsentation).

Ein vorläufiges Modell des Lernens episodens-übergreifender Regularitäten

Der Hippocampus und relationales Binden episodischer Merkmale Die hier beschriebenen Experimente legen nahe, dass der Hippocampus am Lernen episodens-übergreifender Regularitäten beteiligt ist. In der CS-Bedingung zeigte sich konstante hippocampale Aktivierung. Dies kann auf weitestgehend konstante relationale Bindungsprozesse zurückgeführt werden. In der CS-Bedingung müssen in jedem Durchgang variable Objekte und variable Positionen gebunden werden. In der IL-Bedingung nehmen diese relationalen Bindungsanforderungen als Funktion des Lernens ab, da variable Objekte und invariante Positionen bzw. vice versa gebunden werden. Diese Abnahme relationaler Bindungsanforderungen zeigt sich in einer Reduktion hippocampaler Aktivierung im Verlauf des Lernens. Dieses Datenmuster ist konsistent mit Befunden von fMRT-Studien, die probabilistisches Lernen (Poldrack et al., 2001), assoziatives Lernen (Zeineh et al., 2003), Sequenzlernen (Schendan et al., 2003) und Lernen künstlicher Grammatiken (Opitz & Friederici, 2003) untersuchten. In diesen Studien scheint eine Abnahme hippocampaler Aktivierung ebenfalls an eine Reduktion relationaler Bindungsanforderungen gekoppelt zu sein. Schließlich legt die hemisphärische Spezialisierung beim Lernen von räumlichen im Vergleich zum Lernen objektrelatierter Regularitäten nahe, dass der Hippocampus sensitiv ist für die spezifische Domäne invarianter episodischer Merkmale (Bottom-up Mechanismus), eine Voraussetzung für ähnlichkeitsbasierte Lernprozesse.

Das Striatum und die Kodierung der Vorhersagbarkeit episodischer Merkmale Das Striatum zeigte innerhalb von Blöcken einen lernrelatierten Anstieg an Aktivierung. Die Basalganglien als Zielregion dopaminerger Projektionen aus dem ventralen Tegmentum sind beteiligt am Verstärkungslernen (Schultz et al., 1997; Schultz, 2000, 2002; Schultz et al., 2003) und am Stimulus-Response Lernen (Poldrack et al., 2001; Packard & Knowlton, 2002; Poldrack & Packard, 2003; White & McDermott, 2002). Zahlreiche Studien mit Einzelzelleableitungen (Lauwereyns et al., 2002b) und fMRT-Experimente (McClure et al., 2003; O'Doherty et al., 2003, 2004) konnten zeigen, dass v.a. das Putamen (klassisches Konditionieren) bzw. der nucleus caudatus (operantes Konditionieren) einen sog. Vorhersagefehler kodieren und damit die neuronale Grundlage für Lernen darstellen. Der Vorhersagefehler reflektiert die Differenz zwischen dem vorhergesagten Belohnungsreiz (Erwartung) und dem tatsächlich erhaltenen Reiz. Basierend auf diesen Ergebnissen kann davon ausgegangen werden, dass das Striatum die erhöhte Vorhersagbarkeit invarianter episodischer Merkmale kodiert und damit am Aufbau eines statistischen Modells der Umgebung beteiligt ist (Dayan et al., 2000). Weiterhin kann angenommen werden, dass diese striatalen Prozesse ein Verstärkungssignal an den PFC senden, das maßgeblich an der Bildung anhaltender präfrontaler Repräsentationen der episodens-übergreifenden Regularitäten beteiligt ist (vgl. Cohen et al., 2002; Braver & Cohen, 2000; Frank et al., 2001; Koechlin et al., 2002; Miller, 2000; Miller & Cohen, 2001).

Der präfrontale Kortex und die Repräsentation regel-basierter Information Distinkte präfrontale Areale zeigten einen Aktivierungsanstieg als Funktion des Lernens innerhalb von Blöcken und über Blöcke hinweg. Der PFC spielt eine zentrale Rolle bei verschiedenen kognitiven Kontrollfunktionen (Duncan, 2001; Miller & Cohen, 2001; Shimamura, 2000; Wood & Grafman, 2003), z.B. der Selektion aufgabenrelevanter Information (Rowe et al., 2000), der Aufrechterhaltung selektierter Information (vgl. Smith & Jonides, 1999), dem Schutz dieser Information vor interferierenden Stimuli (Miller et al., 1996; Sakai et al., 2002) und dem flexiblen Aktualisieren von Repräsentationen (Badre & Wagner, 2004). Das vorliegende Datenmuster legt nahe, dass der PFC eine wesentliche Rolle bei der Extraktion episoden-übergreifender Regularitäten und der Aufrechterhaltung dieser Regularitäten spielt. In Übereinstimmung mit anderen Studien (Becker & Lim, 2003) wird angenommen, dass diese Regularitäten in Form von Gedächtniscodes repräsentiert sind. Neu zu enkodierende Episoden werden dann anhand dieses Gedächtniscodes reorganisiert, was zur Realisierung einer effektiveren Enkodierstrategie und damit einhergehend zu erhöhter präfrontaler Aktivierung führt (Top-down Mechanismus). Man kann davon ausgehen, dass der PFC zum Lernen episoden-übergreifender Regularitäten in Form von regel-basierten Repräsentationen der kritischen invarianten episodischen Merkmale beiträgt (vgl. Bor et al., 2003; Cer & O'Reilly, in press; Miller, 2000; Opitz & Friederici, 2003; Strange et al., 2001; Wallis et al., 2001).

Zusammengefasst konnten die vorliegenden Experimente zeigen, dass das Lernen episoden-übergreifender Regularitäten von einem graduellen Übergang von einem hippocampalen zu einem präfrontal-striatalen System begleitet ist.

Danksagung

An erster Stelle möchte ich Herrn Dr. Bertram Opitz danken. Er hat in mir das große Interesse an den neuronalen Grundlagen kognitiver Prozesse geweckt und mir die Möglichkeit gegeben, die hier beschriebenen Experimente im Rahmen seines Projektes innerhalb der DFG-Forschergruppe 448 durchzuführen.

Weiterhin möchte ich Herrn Prof. Dr. Axel Mecklinger danken, der mein konzeptionelles Auge geschult hat und mich desöfteren auf den experimentalpsychologischen Boden der Tatsachen zurückgeführt hat. Ihm möchte ich auch für die sehr exakte Ausbildung im Publizieren danken.

Zu großem Dank bin ich Herrn Prof. Dr. Wolfgang Reith und Herrn Dr. Christoph Krick verpflichtet. Durch die sehr erfolgreiche, unbürokratische Saarbrücken-Homburg Kooperation hatte ich die Möglichkeit, zahlreiche fMRT-Experimente in der Abteilung für Neuro-radiologie in Homburg durchzuführen.

Den Herren Dipl.-Psych. Patric Meyer und Dipl.-Psych. Ben Eppinger möchte ich für die zahlreichen wertvollen Diskussionen und Anregungen danken. Allen Mitgliedern der Abteilung für Experimentelle Neuropsychologie möchte ich an dieser Stelle für die sehr angenehme Arbeitsatmosphäre danken, insbesondere Lea Meyer und Henning Löbbbecke für ihre unermüdliche Unterstützung bei den Messungen und Frau Dipl.-Psych. Daniela Czernochowski für das Korrekturlesen dieser Arbeit.

Meiner Frau Tanja und meinen Eltern danke ich für die Unterstützung in allen Lebenslagen.