NIM : a situated computational memory model

NIM:
A Situated Computational Memory Model

NIM:
A Situated Computational Memory Model

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op donderdag 20 september 2007 om 14:00 uur

door

Joyca Lacroix

UNIVERSITAIRE
PERS MAASTRICHT

Promotores: Prof. dr. J.M.J. Murre
Prof. dr. E.O. Postma
Prof. dr. H.J. van den Herik

Beoordelingscommissie:
Prof. dr. A.J. van Zanten (voorzitter)
Prof. dr. G.W. Cottrell (University of California San Diego)
Prof. dr. R. Goebel
Prof. dr. J. van Heerden
Prof. dr. J-J.Ch. Meyer (Universiteit Utrecht)

Netherlands Organisation for Scientific Research

The research reported in this thesis has been funded by NWO, the Netherlands Organization for Scientific Research, in the framework of the Cognition Program. It is part of the larger project: Events in memory and environment: modelling and experimentation in humans and robots, project number: 051.02.2002.

SIKS Dissertation Series No. 2007-15
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Cover design: Steven de Jong
ISBN 978 90 5278 648 3

Preface

Memory is fundamental to natural cognition. Everything we perceive and every action we perform is influenced by our previous experiences that are laid down in various forms in memory. As a psychology student I became fascinated with the wide-ranging capacities of natural memory. Throughout our lives, we continuously acquire new information which is integrated effortlessly with previously learned information stored in memory. In sharp contrast, neurally inspired connectionist models of memory are severely limited in their storage capacity; new information frequently overwrites memory containing previously learned information. Neuropsychological, physiological, and behavioural studies suggest that the brain is able to deal with the accumulation of an almost endless number of memories by gradually consolidating newly acquired memories into long-term memory during sleep. In my M.Sc. research, I studied a computational model that simulated the memory processes in operation during the various stages of sleep. There, I found that the model was able to mimic some of the basic characteristics of memory consolidation on the basis of a restricted set of highly simplified input patterns. Obviously, these patterns are unrepresentative for the complex patterns of natural environmental input.

Inspired by the ability of human memory to deal with an enormous amount of highly complex natural sensory input, I decided to pursue a Ph.D. study at Maastricht University examining models of memory in a realistic environment. Soon I discovered that despite the vast volume of research in the domain of cognitive modelling, few models have been developed that meet natural environmental demands. The research presented in this thesis extends the memory models from the computational psychology tradition to enable the coping with natural input. The result is a situated computational memory model that provides an initial step towards a plausible model of natural cognition.

For the accomplishment of this scientific work I received support from many people who deserve my sincerest acknowledgements. First of all, I would like to thank my supervisors Jaap Murre, Eric Postma, and Jaap van den Herik. Their creative, optimistic, and thorough guidance was essential to this thesis. Next, many colleagues and friends encouraged me through inspiring discussions and by creating a socially enjoyable working environment. I would like to thank Michel van Dartel, Ben Torben-Nielsen, Rens Kortmann, Guido de Croon, Sander Spek, Sjoerd Bakker, Steven de Jong, Marc Ponsen, Jahn Saito, Igor Berezhnoy, Evgueni Smirnov, Guillaume Chaslot, and Andra Waagmeester. Furthermore, I owe many thanks to Ida Sprinkhuizen-Kuyper and Niek Bergboer for sharing their mathematical insights,

to Jeroen Donkers for his LaTeX support, to Peter Geurtz for his technical assistance, and to Joke Hellemons and the people from the secretariat for their help with administrative matters.

Over the years, I was fortunate to be surrounded by several people that helped me relax and focus on the important things in life. As a significant supplement to those already mentioned, I would like to give my thanks and appreciation to Judith te Brake, Linda Daas, Marieke Lansbergen, Jaimie Luermans, and Esther Vermeer. In particular, I would like to express my gratitude to Simon Muller for his love and support. Finally, my respect and thankfulness go out to my parents and brothers who provided me with their unconditional support.

Joyca Lacroix

Contents

Chapter 1

Introduction

Since the early days of the cognitive sciences, theories of cognitive functions have been framed in terms of computational models. The two main advantages of computational cognitive models are: (1) that they enable the fully transparent specification of the computations that underlie the different cognitive processes and their interactions, and (2) that they allow for the simulation and prediction of behavioural responses in various experimental situations. Computational cognitive models benefit from the abstraction of cognitive processes to the level of computations. However, the abstraction comes at a cost. The cognitive processes operate on symbols rather than real-world input. For instance, a model of memory for images may represent an image of a chair by a proposition (see, e.g., Anderson, 1993) or a binary or discrete valued vector (see, e.g., Rumelhart, McClelland, and The PDP Research Group, 1986; Shiffrin and Steyvers, 1997). Virtually all present-day cognitive models are disconnected from the real world because they rely on highly simplified representations of the visual input (but see Kohonen, Oja, and Lehtiö, 1981; Cottrell, Bartell, and Haupt, 1990, for notable exceptions). In this thesis we deal with the main limitation of computational cognitive models which is their lack of a connection with the real world.

In section 1.1, we provide an overview of the development of computational memory modelling throughout the years. Subsequently, in section 1.2, the problem statement and the research questions are formulated. This is followed, in section 1.3, by a description of the research methodology that was applied to address the problem statement. Finally, section 1.4 presents an outline of the thesis.

1.1 Computational memory modelling

The development of computational memory modelling can be characterized by three types of models: the data-description models, the mechanism-based models, and the situated models. The data-description models can be considered as the first computational models. They comprised straightforward mathematical descriptions of behaviourally obtained data from learning and memory experiments (e.g., Ebbinghaus,

1885/1964; Thorndike, 1913)[1]. With the advent of the computer and the cognitive sciences, mechanism-based models became widespread. These models were more elaborate, describing the mechanisms and making specific predictions for new situations (e.g., Raaijmakers and Shiffrin, 1981; Eich, 1982; Murdock, 1982; Shiffrin and Steyvers, 1997). Situated models represent a recent modelling trend that attempts to take realistic environmental demands into account.

Below, we discuss the data-description models (1.1.1), the mechanism-based models (1.1.2), and the situated models (1.1.3) in more detail.

1.1.1 Data-description models

Ebbinghaus (1885/1964) and Thorndike (1913) performed the first endeavours to gain insight into learning and memory by describing behaviourally obtained data from memory tasks in mathematical terms. In a series of learning and memory studies, Ebbinghaus (1885/1964) obtained data about the learning and forgetting of items on a study list as a function of (1) the retention interval, (2) the amount and temporal distribution of study time, and (3) the serial position of the item on the study list. Subsequently, he attempted to capture the results in mathematical descriptions such as the forgetting curve, the learning curve, and the serial position curve (Ebbinghaus, 1885/1964).While Ebbinghaus (1885/1964) focussed mainly on the retention of information, the learning and memory experiments performed by Thorndike (1913) addressed the effect of reward on producing a particular response in a distinctive situation and the degree of transfer between learning trials. In a similar way as Ebbinghaus, Thorndike (e.g., 1913) tried to capture his experimental results on learning in mathematical descriptions. He proposed different types of learning curves for trial-and-error learning and learning based on insights.

The data-description models developed by Ebbinghaus and Thorndike provided the first formal models of memory in cognitive psychology. The advantage of the data-description models is that they give precise descriptions of the relation between the examined memory characteristic and the manipulated variable. The main drawback of the data-description models is that they do not provide insight into the mechanisms underlying the described data. Consequently, these models are of limited value for making predictions for new situations (Raaijmakers and Shiffrin, 2002). A new type of computational memory models addressed this limitation by focussing more on the underlying memory mechanisms.

1.1.2 Mechanism-based models

The mechanism-based models deal with the main limitation of the data-description models by attempting to simulate the memory mechanisms underlying performances on memory tasks. Mechanism-based models formalize and simulate the dynamics

[1]It should be noted that sometimes a distinction is made between learning models that concentrate on what is retained and memory models that focus somewhat more on what is forgotten. However, it is extremely difficult to distinguish the first from the latter as most models describe the combination of learning and memory. Therefore, we discuss the models that were originally called learning models along with the memory models.

of (one or more of) the different mechanisms underlying memory. Memory processes are conceived of as computations on a multi-dimensional representation space. Item representations are formalized as points (or vectors) in the multi-dimensional space, the dimensions of which correspond to various features of the items. Examples of mechanism-based models include the SAM model (Raaijmakers and Shiffrin, 1981; Gillund and Shiffrin, 1984), the TODAM model (Murdock, 1982), the CHARM model (Eich, 1982; Eich, 1985), the MATRIX model (Pike, 1984), the MINERVA2 model (Hintzman, 1986), the REM model (Shiffrin and Steyvers, 1997), the model of differentiation (McClelland and Chappell, 1998), and the BCDMEM model (Dennis and Humphreys, 2001). The mechanism-based models elaborate on the early data-description models by describing the storage and retrieval processes quite precisely. Therefore, these models are suitable to make predictions for various situations. For example, the REM model (Shiffrin and Steyvers, 1997) has been successfully applied to a range of different memory tasks such as old-new recognition (Shiffrin and Steyvers, 1997), perceptual identification (Schooler, Shiffrin, and Raaijmakers, 2001), associative recognition (Criss and Shiffrin, 2004), and judgements of frequency (Malmberg, Holden, and Shiffrin, 2004). Despite these successes, the mechanism-based models suffer from an important limitation. They lack a connection with the real world, because they fail to explain how sensory input is translated into memory representations. Rather than incorporating a perceptual mechanism to transform real-world input into representations, the mechanism-based memory models assume a predefined artificial representation space. In such a space, representations are predefined in terms of abstract features or in terms of other representations, rather than derived from the physical features of real-world stimuli themselves.

The mechanism-based models can be said to descend from two closely related traditions that enjoy a long-standing support in the field of cognitive psychology, the information-processing approach (e.g., Lindsay and Norman, 1977) and the cognitivist approach (e.g., Fodor, 1980). Inspired by the invention of the computer, these approaches drew a parallel between a natural cognitive system and a computer, where, crudely speaking, the brain was considered as the hardware and the mental processes as the software. In order to understand mental processes, the information-processing and cognitivist approaches disregard the relations with the real world (Fodor, 1980). They claim that the processes can be understood best by abstracting away from the hardware (i.e., the brain and also the body) and the environment in which it is embedded (Fodor, 1980). The idea to treat cognition as internal computation that is isolated from a body and its environment is now losing support (e.g., Clark, 1999; Roy, 2005a; Pecher and Zwaan, 2005), because it is hampered by two interrelated problems: the symbol grounding problem (Harnad, 1990) and the transduction problem (Barsalou, 1999). The symbol grounding problem refers to the failure to map representations onto real world stimuli (Harnad, 1990). It concerns how representations acquire meaning by relating directly to something in the real world without the need for an external interpreter. The transduction problem refers to the failure to provide an explanation for the origin of the abstract representations in a cognitive system (Barsalou, 1999). It concerns the lack of a mechanism for translating the input from the real world into representations in the cognitive system. The grounding and transduction problems are essentially

two ways of looking at the same problem, i.e., the establishment of a direct relation between representations and the objects that they refer to in the external world. The mechanism-based models suffer from the grounding and transduction problems, because they lack an encoding process that creates a representation space directly from real-world stimuli.

1.1.3 Situated models

The grounding and transduction problems can be solved by improving the ecological validity of memory models. Contrary to the mechanism-based memory models, natural systems ground representations in their physical interaction with the real world. The way natural systems understand and represent the world is a direct consequence of their perceptions and actions within the world (e.g., Barsalou, 1999; Pecher and Zwaan, 2005). Inspired by natural cognition, a shift is taking place from studying cognition as an isolated computational system towards studying cognition as an acting system within a real-world environment (i.e., a 'situated' model; Clancey, 1997). The idea of emphasizing the environment in perception was first proposed by Gibson (e.g., 1979) in his ecological approach to perception. At the time Gibson introduced his ecological approach, it was a radical departure from the conventional information-processing approach in cognitive psychology. While the information-processing approach focussed on internal computations, Gibson underscored the information originating from the interaction with the environment. He argued that information does not reside in the static retinal image, but instead is available from the spatial and temporal optic patterns that emerge from acting in the world. Gibson was fiercely criticized by cognitive scientists and artificial intelligence researchers. The main criticism was that his views did not lend themselves to an algorithmic approach. In this respect, Gibson's work is often contrasted to that of Marr (1982), who did present a detailed computational account of vision. The recent trend towards more realism in cognitive psychology, artificial intelligence, and robotics, leads to renewed interest in the ecological approach. Cognitive systems are increasingly developed in the context of realistic environmental demands (e.g., Atkeson *et al.*, 2000; de Croon, Postma, and van den Herik, 2006a; Sprague, Ballard, and Robinson, in press). However, the artificial systems developed so far are limited to fairly simple behaviours and lack a plausible memory system. In contrast, as elucidated in subsection 1.1.2, the existing computational memory models from the field of cognitive psychology provide quite detailed descriptions of the functioning of the human memory processes, but ignore the role of the environment.

Acknowledging the essential role of the environment for natural cognition and the fundamental problems in the traditional cognitive theories that assume abstract representation spaces, cognitive modelling developments are increasingly moving from an isolated approach towards a situated approach. An essential element of situatedness is the implementation of sensor-grounded representations (e.g., Pecher and Zwaan, 2005), i.e., representations derived directly from the physical features of real-world stimuli. Extending the mechanism-based memory models with a perceptual front-end that translates real-world stimuli into representations would constitute an important initial step towards the development of a situated computational memory model. It is the aim of this thesis to do so.

1.2 Problem statement and research questions

As exemplified in the previous section, it is clear that computational modelling so far has mainly produced models that lack a connection with the real world. Currently, the field of cognitive science is entering a new stage that underscores the importance of studying cognition within a realistic environment.

The existing mechanism-based memory models as outlined in section 1.1, provide a starting point for the extension into more plausible models of natural memory within a realistic environment. Their lack of a connection with the real world leads us to the following problem statement.

PS: How can computational memory models be extended to solve the grounding and transduction problems?

While the real-world input of natural cognitive systems consists of different types of perceptions, we restrict our scope to real-world *visual* input. Furthermore, we focus our modelling studies on two types of memory-based cognitive processes: recognition and classification. We attempt to approach the problem statement by proposing a perceptual preprocessing front-end that constitutes a connection between a computational memory back-end and the real world. The perceptual front-end transforms natural visual input into representations that can be operated on by the computational memory back-end. By combining the perceptual front-end with a computational memory back-end we obtain a situated memory model. On the basis of the situated model we aim to answer the following three research questions.

RQ 1: To what extent can a situated model produce human responses to individual natural visual stimuli?

RQ 2: To what extent can a situated model produce recognition-memory effects on the basis of natural visual stimuli?

RQ 3: To what extent can a situated model classify natural visual stimuli?

1.3 Research methodology

We employ an empirical methodology in which we perform the following two steps: (1) model construction using insights from biology and computer vision, and (2) model validation using behavioural data. Below we discuss the model construction (1.3.1) and validation (1.3.2).

1.3.1 Model construction

We address our problem statement by aiming to construct a computational model of human memory that operates on real-world visual input. To meet our objective, we propose a perceptual front-end that transforms real-world visual input (i.e., natural images) into memory representations that can be operated on by computational memory processes. In chapter 2, we formulate three guiding principles for creating a representation space from real-world input which are derived from psychophysical and biological knowledge about the main characteristics of the human visual system. For the realization of a perceptual front-end that fulfils these guiding principles, we borrow elements from influential approaches in the field of visual object recognition (i.e., computer vision). In computer vision, several methods have been proposed for the construction of representation spaces for visual input. Our approach relies on these methods to develop a perceptual front-end that aims at building visual representations from the real world. The front-end will realize a connection between a computational memory model and the real world.

1.3.2 Model validation

We will validate our proposed situated model by using data from behavioural experiments. We perform three validation studies with three different model variants. The model variants are obtained by connecting our proposed perceptual front-end with three variants of a computational memory back-end. The first model is a model for recognition of natural stimuli called the Natural Input Memory model (NIM), which forms the basis for the second and third models. The second model extends NIM into NIM-REM, a model for recognition of natural stimuli, the memory back-end of which is based on the powerful computational model Retrieving Effectively from Memory (REM) proposed by Shiffrin and Steyvers (1997). The third model extends NIM into a model for classification of natural stimuli, called NIM-CLASS. The models will be validated on the basis of human data from behavioural studies.

NIM

NIM realizes a situated model for recognition memory by combining our proposed perceptual front-end with a computational memory back-end that is based on the Generalized Context Model (GCM; see, e.g., Nosofsky, 1986, 1987); the latter is an exemplar-similarity based model for recognition. We validate NIM by comparing model similarity judgements and recognition rates for a set of natural stimuli with human similarity judgements and recognition rates obtained in behavioural studies (e.g., Busey and Tunnicliff, 1999).

NIM-REM

NIM-REM combines our perceptual front-end with a computational memory back-end based on the REM model (Shiffrin and Steyvers, 1997). We validate NIM-REM by testing the model for producing a number of general recognition memory effects often obtained in human recognition-memory studies. A comparison is made between the

pattern of model recognition results and human recognition results for the different effects.

Nim-Class

Nim-Class extends Nim into a classification model for natural stimuli. The model combines our perceptual front-end with a computational memory back-end for classification based on a nearest-neighbour classifier. We validate Nim-Class by assessing its classification performance and comparing it with human classification performance. Moreover, we extend Nim-Class into two Nim-Class variants, Nim-Class A and B, that employ an active vision mechanism for the selection of visual input on the basis of its relevance. We test to what extent the active vision extensions enhance performance on the classification task as compared to the original Nim-Class performance and to what extent it corresponds with human classification performance.

1.4 Outline of the thesis

In chapter 2, we start by reviewing the nature of the representations in existing computational memory models. Then, we formulate three guiding principles for building a veridical representation space from real-world visual input and show how the principles can be fulfilled computationally in a perceptual front-end. Subsequently, in chapter 3, we present a situated computational recognition-memory model called Nim that combines the perceptual front-end with a computational memory back-end.

Thereafter, in chapters 4, 5, and 6, different variants of Nim are proposed and validated using data from behavioural studies. Chapter 4 validates Nim by comparing model responses to individual natural stimuli with human responses to the same set of stimuli obtained in behavioural studies. Chapter 5 proposes Nim-Rem and validates the model by testing to what extent it can produce recognition-memory effects robustly obtained in behavioural studies. Finally, chapter 6 proposes Nim-Class and validates the model by comparing model and human classification results. Moreover, chapter 6 introduces Nim-Class A and B that employ different types of top-down gaze-control mechanisms to select relevant visual input. We evaluate to what extent classification performance improves by employing top-down gaze control and to what extent it improves the agreement with human results.

The model validation chapters are followed in chapter 7 by a discussion of our approach by placing it in the context of existing computational models of object recognition, by identifying points for the improvement of the psychological and biological plausibility of our models, and by relating it to the global developments in cognitive modelling. Finally, chapter 8 provides the thesis conclusions.

Chapter 2

Connecting memory models with the real world

As stated in chapter 1, our study aims to solve the symbol grounding and transduction problems by connecting memory models directly with the real world. This chapter shows how veridical memory representations can be derived directly from real-world visual input. It introduces a perceptual mechanism that is based on the main characteristics of the human visual system for transforming natural visual input into memory representations. The perceptual mechanism can function as a perceptual front-end to a computational memory model.

The chapter starts in section 2.1 with a review of the nature of representations in existing memory models and the identification of the main limitation of these memory models: the absence of a connection between the representations employed and their real-world referents. Two factors can help connecting memory models to real-world visual input: (1) existing approaches to map visual input to a representation space and (2) knowledge about the processing in the human visual system. Section 2.2 presents four approaches to map visual input to representations that have been influential in the domain of computer vision. This is followed in section 2.3 by an inventory of the main characteristics of the human visual system, which are taken as guiding principles for creating a veridical representation space from real-world input. Section 2.4 shows how the guiding principles can be fulfilled computationally by introducing a perceptual front-end that constructs a veridical representation space from natural visual input. The perceptual front-end connects the visual real world with a computational memory model. Our approach combines the perceptual front-end that operates on the visual real world with a computational memory model to obtain a situated memory model, which is called the Natural Input Memory model (NIM). The model forms the basis for our studies in the chapters 3 to 6.

2.1 Representations in existing memory models

Memory models are often conceptualized as representation-process pairs (Anderson, 1976). In such models, perceived or remembered items are represented in some way. Then, a variety of memory processes may act upon these representations. As stated in chapter 1, existing memory models focus on the memory processes rather than on the origin of the representations (e.g., Raaijmakers and Shiffrin, 1981; Eich, 1982; Murdock, 1982; Gillund and Shiffrin, 1984; Pike, 1984; Eich, 1985; Hintzman, 1986; Shiffrin and Steyvers, 1997; McClelland and Chappell, 1998; Dennis and Humphreys, 2001). While the memory processes are described quite precisely, the translation of sensory input into memory representations is mostly left unspecified. The memory models thus lack a direct connection with the real world.

Rather than specifying how an adequate representation can be derived from real-world input, the memory models rely on different types of abstracted representations from real-world objects. Two main representation types can be distinguished: (1) the rule-based type, and (2) the similarity-based type. The rule-based type consists of a set of variables and propositions that can be combined to generate an infinite number of representations (see, e.g., Newell and Simon, 1972; Sun and Zhang, 2006). The similarity-based type employs items that are organised in a representation space according to the similarities between the items (see, e.g., Shepard, 1957; Goldstone and Son, 2005). Theories of categorization, generalization, and learning and memory often employ similarity-based representations. Representations of the rule-based type are explicit descriptions of real-world objects but lack a straightforward connection with their real-world instantiations. As outlined in subsection 1.1.2 this lack of a connection is generally referred to by two interrelated problems: the symbol grounding problem (Harnad, 1990) and the transduction problem (Barsalou, 1999). Similarity-based representations can be considered to indirectly solve the grounding and transduction problems, since they are connected indirectly with the real world by reflecting the functional and perceptual similarities between their referents. Therefore, similarity-based models are to be preferred for grounding memory representations. Fundamental to many theories of memory and cognition that assume a similarity-based representation space is the feature-space approach, which uses a multi-dimensional metric space to represent items along several discrete or continuous dimensions (e.g., Brooks, 1978; Medin and Schaffer, 1978; Nosofsky, 1986; Shepard, 1987; Nosofsky, 1987, 1988, 1989, 2003). Hence, we focus on this type of representation space.

The following three types of representational methods have been employed in existing memory models for the construction of a feature space of real-world stimuli: (1) representation by statistical assumptions about feature values of real-world stimuli, (2) representation by context-dependent co-occurrences of items, and (3) representation by human judgements. The three methods will be discussed in subsections 2.1.1, 2.1.2, and 2.1.3, respectively. Subsequently, we discuss the main limitation of the existing memory models in subsection 2.1.4.

2.1.1 Representation by statistical assumptions

A number of memory models have attempted to build their feature-vector representations by making straightforward assumptions about the distribution of feature values in the real world. For example, in the TODAM model (Murdock, 1982) and the CHARM model (Eich, 1982, 1985), feature values are drawn from the normal distribution, because real-world features are assumed to be normally distributed. In contrast, in other models, such as the original REM model, feature values are drawn from the geometric distribution (Shiffrin and Steyvers, 1997), the parameters of which depend on the frequency of occurrence in the real world of the features that together represent the item. The statistical assumptions provide a straightforward means to build a feature space. However, they do not reflect the true similarity relations of items in the real world, but at best form a crude approximation.

2.1.2 Representation by context-dependent co-occurrences

Building representation spaces by means of context-dependent co-occurrences is based on the idea that similar items occur in similar contexts. While this method can be used for perceptually based stimuli, such as objects and faces, as well as for conceptually based stimuli, such as words, it has mostly been applied to the latter (e.g., Landauer and Dumais, 1997). The method of building a similarity space on the basis of context-dependent co-occurrences assumes that the meaning of a word and the similarity of one word to another is largely determined by all the contexts in which the word does and does not appear. Two operationalizations of this method are the Latent Semantic Analysis (LSA; e.g., Landauer and Dumais, 1997) and the Hyperspace Analogue to Language (HAL; e.g., Burgess, Livesay, and Lund, 1998). Using LSA and HAL, similarity spaces are constructed in which the representations of items are a product of the contexts in which the items are found. Representation by context-dependent co-occurrences establishes a more direct connection with the real world than representation by statistical assumptions. Nevertheless, it achieves only an approximate relation between representations and their referents, because its representations come about through averaging over many instances. Moreover, texts are a biased and imperfect reflection of the real world.

2.1.3 Representation by human judgements

Using a multi-dimensional scaling (MDS) analysis or a singular value decomposition (SVD), a representation space is obtained that is based on human similarity judgements (e.g., Shepard, 1957; Caramazza, Hersch, and Torgerson, 1976; Busey, 1998; Busey, 2001), human free-association data (e.g., Steyvers, Shiffrin, and Nelson, 2004), or other types of judgements that reveal the perceived similarity of stimuli. These methods have often been used to build a representation space that accurately reflects similarities as perceived by humans, i.e., a psychologically plausible representation space. For example, in a similarity-rating experiment, Busey (1998) asked subjects to rate the similarity of two faces on a scale from 1 to 9. Across subjects, he collected at least six ratings for each of the 5356 possible pairs of faces from a set of 104 faces. The similarity ratings were the input to an MDS analysis. Fig. 2.1

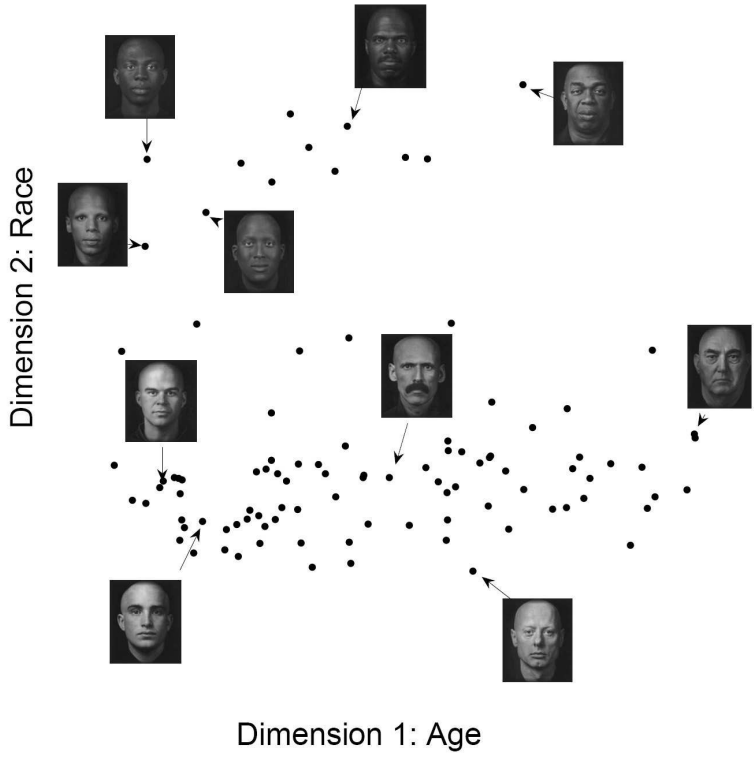


Figure 2.1: Faces plotted in a two-dimensional space obtained by applying a multi-dimensional scaling analysis to the similarity judgements of human subjects. Reproduced from Busey (2001).

shows several of the faces plotted in a two-dimensional representation space acquired by an MDS analysis of the similarity ratings for pairs of faces from the face-image set (Busey, 1998). The figure is reproduced from Busey (2001), but is based on the similarity ratings obtained by Busey (1998), and discussed and analysed by Busey and Tunnicliff (1999). After plotting the faces in MDS space, Busey and Tunnicliff (1999) interpreted the two dimensions shown here as age and race. In a similar way, free-association data can form the input to an MDS analysis or an SVD method in order to create a psychologically plausible representation space. Steyvers *et al.* (2004) applied MDS and SVD methods to free-association data obtained by Nelson, McEvoy, and Schreiber (1999). The free-association data were collected by asking subjects to write down the first word that came to mind that was meaningfully related or strongly associated to a presented word. By applying the scaling methods on the free-association data, a representation space is obtained such that words with similar associative patterns reside in similar regions of the space.

The main advantage of using human judgements is that the obtained representa-

tion space has a considerable degree of psychological plausibility. While this method achieves the best connection between the real world and memory representations, it is still based on indirect evidence, i.e., the human judgements.

2.1.4 A limitation in existing memory models

Existing memory models often employ a representation space created by using one of the three methods discussed in subsections 2.1.1, 2.1.2, and 2.1.3. The three methods that aim at building memory representations for real-world stimuli all suffer from the limitation that they rely on an analysis of the entire stimulus set rather than on an analysis of the individual stimulus to define its representation. While these methods can produce psychologically plausible representation spaces, the representations are not derived directly from the physical features of each of the stimuli, i.e., they are not grounded in the real world.

In the following section we examine four existing and influential methods from the domain of computer vision to realize the grounding of memory representations.

2.2 From the real world to memory representations

In order to acquire representations from real-world (i.e., natural) input, a transformation is needed that maps items to a suitable representation space. Since visual input constitutes a major part of human sensory input, we focus on natural visual input (i.e., images). In the domain of computer vision, several researchers have been working on methods to construct veridical visual representations for object recognition (e.g., Marr, 1982; Biederman, 1985, 1987; Swain and Ballard, 1991; Edelman and Intrator, 1997; Mel, 1997). In this section we provide a brief (and incomplete) historical overview of methods that have been developed to map visual input to a multi-dimensional representation space for object recognition. Four important approaches are: (1) the computational approach (e.g., Marr, 1982), (2) the recognition-by-components approach (e.g., Biederman, 1985; 1987), (3) the similarity-space approach (Edelman, 1998), and (4) the feature-space approach (e.g., Mel, 1997). Although we will discuss the approaches separately in subsections 2.2.1, 2.2.2, 2.2.3, and 2.2.4, respectively, we would like to emphasize that they share many characteristics with each other. In subsection 2.2.5, we briefly discuss how we will move towards a new approach for translating the visual real world into memory representations.

2.2.1 The computational approach

Marr's (1982) computational approach to visual object recognition demarcated the start of computational models of vision. The computational approach focussed on defining the stages involved in extracting 3D model representations from images. Marr's model assumes that the visual system builds up a full 3D object representation from the retinal image via several progressively complex representations. In order to construct the sequence of representations, Marr (1982) proposed a computational algorithm that distinguishes three stages of processing: (1) the construction

of the (raw and full) primal sketch (i.e., an image-based representation), (2) the construction of the 2.5D sketch (i.e, a viewer-dependent surface-based representation), and (3) the construction of the 3D model representation (i.e., a viewer-independent object-based representation). The first stage realizes the translation from an image into the raw primal sketch: a representation in terms of a set of intensity-based image features (i.e., primitives), including edges, lines, blobs, and terminations. Subsequently, the raw primal sketch is extended into the full primal sketch by connecting and grouping the primitives into larger structures and segmenting regions that differ from each other in texture. The second stage builds on the primal sketch to construct the 2.5D sketch that represents the orientation and depth of the visible surfaces as well as the discontinuities. The 2.5D sketch contains information on the local surface orientation primitives, the distance from the viewer, and the discontinuities in depth and surface orientation. The last stage of Marr's framework translates the viewer-centred 2.5D sketch into the viewer-independent object-centred 3D model representation.

The representations constructed in Marr's framework start as image-based representations; subsequently they are transformed into surface-based and finally into object-based representations. The Marr paradigm is a compelling theoretical framework that describes how representations can be derived directly from an image using a computational algorithm. Since Marr (1982), many psychophysical studies (e.g., Farah, 1985; Biederman, 1987; Logothetis and Sheinberg, 1996) and neuroscientific studies (e.g., Tanaka, 1996) have revealed knowledge about human object recognition (for an overview, see, e.g., Peissig and Tarr, 2007). Based on advances in the study of object recognition, more effective vision applications (for various tasks, e.g., recognition and navigation) have been developed than the one proposed by Marr. Yet, Marr's framework has been very influential in the world of computational visual object recognition, as a result of which several theories have been built on Marr's framework, e.g., Biederman's (1985, 1987) recognition-by-components theory.

2.2.2 The recognition-by-components approach

While Marr's computational approach focussed on defining the stages involved in extracting view-invariant 3D model representations from images, Biederman's (1985, 1987) recognition-by-components theory aims at identifying objects by recognizing their basic view-invariant constituent components, called 'geons'. Examples of geons include arcs, spheres, blocks, and cylinders. According to Biederman (1985, 1987) object identification proceeds by recognizing spatial configurations of geons through their properties (so-called non-accidental properties) and matching these with stored object representations. Analogously to the processing stages in Marr's framework, Biederman assumes a number of processing stages that construct a sequence of representations that start as image-based representations and are then transformed into surface-based, and finally view-invariant object-based representations (e.g., Palmer, 1999). Biederman (1985, 1987) proposes that representations are constructed and matched to stored representations in four consecutive stages: (1) edge extraction, (2) detection of non-accidental properties and parsing at regions of concavity, (3) determination of geons and their spatial relations, and (4) matching of components

to stored representations. During the first stage, a representation of the edge information in the visual scene is constructed on the basis of luminance gradients. Then, in the second stage, the edge information is used to detect essential view-invariant properties (non-accidental properties) for geon identification. Simultaneously, the second stage attempts to parse objects into the constituent geons based on regions of concavity. Stage three combines the results on non-accidental property detection and object parsing of stage two to specify the geons at the different locations and the spatial relations that hold among them. Finally, stage four matches the definitive geon-description representation with stored representations.

Both Biederman's object representation in terms of geons and Marr's 3D model representation rely on a reconstruction of the physical object that they represent. Such first-order isomorphism was characteristic of many models of visual perception and computer vision in the early days of artificial intelligence. Although both Marr's and Biederman's representational approaches were quite influential, there is not much evidence for a view-invariant first-order isomorphism in human visual object recognition. More recent models often employ another type of isomorphism that represents objects in terms of their perceptual and functional similarities to other objects.

2.2.3 The similarity-space approach

The similarity-space approach proposed by Edelman (see, e.g., Edelman, 1995a; Edelman and Duvdevani-Bar, 1997a; Edelman and Intrator, 1997; Edelman, 1998) is based on the assumption that the human brain attempts to represent the perceptual and functional similarities between objects in the real world. This assumption is partly justified by the fact that similar objects and events tend to have similar characteristics and tend to behave in a similar way (e.g., Goldstone and Son, 2005). Therefore, our judgement of similarity allows us to generalize from one situation to a similar situation and from the behaviour and characteristics of one object to a similar object. The similarity structure of the real world has long been recognized as a fundamental element of cognitive processes such as learning, categorization, memory, and representation (e.g., Shepard, 1957, 1964; Shepard and Chipman, 1970; Nosofsky, 1986; Edelman, 1995a; Busey, 2001; Nosofsky and Zaki, 2003; Markman and Gentner, 2004). Acknowledging the importance of similarity for generalization, Arkadev and Braverman (1966) formulated already in the 1960s the compactness hypothesis. This hypothesis states that similar objects in the real world are close in their representations and that there is no ground for any generalization on representations that do not obey this demand. In agreement with Arkadev and Braverman's (1966) hypothesis, the similarity-space approach states that the similarities between the representations in a representational system reflect the similarities between objects in the real world (Shepard, 1957, 1964; Shepard and Chipman, 1970; Edelman, 1995a; Edelman and Duvdevani-Bar, 1997a; Edelman and Intrator, 1997; Edelman, 1998). An example of such a space is shown in Fig. 2.2, which is reproduced from Edelman (1995b). The isomorphism between similarities of objects in the real world and the similarities of their representations in a representation space is referred to as a second-order isomorphism (Shepard and Chipman, 1970; Shepard, 1987). Similar-

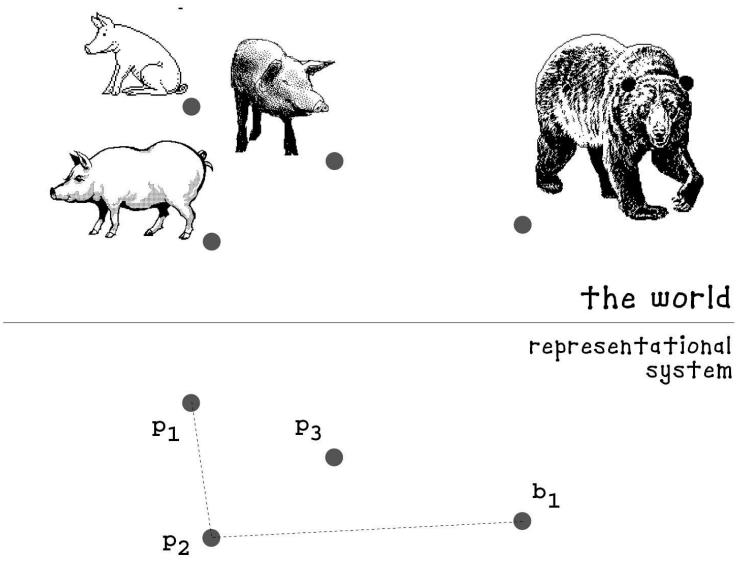


Figure 2.2: Example of a similarity space representing items that are highly similar in close proximity of each other. Reproduced from Edelman (1995b).

ities of objects are represented rather than the objects themselves (as in first-order isomorphism).

Edelman (1998) provides a computational basis for the representation of similarity by postulating a number of adaptive templates (or prototypes) each matching a particular instance of a class. The template representations are activated to a degree that is proportional to the match with the input. An incoming stimulus is represented in terms of the responses of a restricted number of such adaptive templates. The ensemble of k adaptive templates, called the 'chorus of prototypes' (Edelman, 1995a, 1998), maps the stimuli from the real world onto a k-dimensional representation space in which the response of each adaptive template represents the contribution of the associated prototype to the recognized object. The stimulus representations obtained in this way have been shown to be in agreement with human classification and recognition responses (Edelman and Duvdevani-Bar, 1997a).

2.2.4 The feature-space approach

The feature-space approach has been mentioned several time in this chapter already. Subsections 2.1.1, 2.1.2, and 2.1.3 described the indirect construction of feature-space representations on the basis of general statistical assumptions, context-dependent co-occurrences, and human judgements. Moreover, subsection 2.2.3 outlined Edelman's computational basis for building a representation space with his 'chorus of prototypes' that can be conceived as a feature space in which each template constitutes a

high-level feature. However, none of the previously discussed methods constructed a feature space directly from natural images (but see Edelman, 1995a, who employed a wavelet-based transform for encoding the templates). Such a feature space can be based on any type of feature set that is assumed to be relevant for the task at hand and that can be extracted readily from the image.

The feature-space approach may be based on the method of image histogramming to obtain a suitable representation space for object recognition (e.g., Swain and Ballard, 1991; Mel, 1997). Image histogramming entails counting the occurrence of feature values in an image given a certain feature, such as colour, shape, or texture. Colour histogramming has been shown to be quite successful for object identification and localization. Using a colour-histogramming approach, Swain and Ballard (1991) tested a visual system on object identification and on object localization. To identify an object, the system matched the colour histogram of an incoming image containing the object to the colour histograms in the database based on the number of corresponding pixels of the same colour. Localization of an object is based on a backpropagation network that finds the location in the image that contains the colours of the object being looked for. Using the colour-histogramming approach Swain and Ballard (1991) showed that robust and efficient object identification as well as object localization can be achieved. Since colour histogramming is a pixel-based method, it does not contain any spatial information. Hence, it is insensitive to any permutation of the pixels in an image.

The success of the feature-extraction approach for object recognition was also demonstrated by Mel (1997), who used the histogramming method to construct a visual model for object recognition inspired by the feedforward feature-extraction hierarchy assumed to underlie vision in natural systems. Mel's (1997) visual model SEEMORE extends the colour-histogramming-based system developed by Swain and Ballard (1991) using spatially extended features that capture spatial information. SEEMORE employed a set of feature channels that were relatively sensitive to changes in the image that correspond to a change in object identity, such as colour, shape, and texture, and relatively insensitive to changes in the image that do not correspond to changes in identity, such as pose and orientation. The shape and texture-related channels employed by Mel (1997) included, feature detectors for edges, edge pairs, texture, colour, and simple shapes. The feature-space representations resulting from SEEMORE's feature channels were mapped onto a low-dimensional feature space and were subsequently used in a classifier to determine the recognition response. Consistent with the results of Swain and Ballard (1991), SEEMORE's results show that when using solely the colour features, the system achieves a rather successful recognition performance (Mel, 1997). However, the performance significantly improved when features relating to shape and texture were employed.

2.2.5 Towards a new approach

The four approaches to constructing veridical visual representations for real-world objects discussed in this section have been quite influential in the field of visual object recognition. Each of the approaches relies on neurocognitive and psychophysical ideas about the human visual-processing and representational system. Inspired by

these approaches and by psychophysical and biological knowledge about the human visual system, the next section identifies appropriate guiding principles for a veridical mapping from images to feature-space representations and proposes how the guiding principles can be fulfilled.

2.3 Guiding principles for a veridical mapping

Psychophysical and biological findings demonstrate that processing in the human visual system relies on three main characteristics. First, the individual cells of the primary visual cortex V1 respond most strongly to edges in a visual scene (see, e.g., Hubel and Wiesel, 1959, 1962, 1968). Second, the cells of the V1 cortical area perform a local multi-scale analysis of the visual input (see, e.g., Palmer, 1999). Third, the human eye takes selected samples in a visual scene by means of eye fixations (see, e.g., Yarbus, 1967; Henderson, 2003).

We consider the basic processing characteristics of the human visual system as guiding principles for achieving a veridical mapping from images to a feature space. Below we discuss the guiding principles in more detail: (1) an edge-based analysis (2.3.1), (2) a local multi-scale analysis (2.3.2), and (3) a fixation-based analysis (2.3.3).

2.3.1 Edge-based analysis

The detection of discontinuities in intensity, colour, or texture, often referred to as 'edges' and their orientation, is performed by the cells in the primary visual processing area V1 of the human cortex. The functional roles of these cells were discovered in a series of famous experiments performed by Hubel and Wiesel (1959, 1962, 1968). Hubel and Wiesel tried to determine the tuning characteristics of cells in V1 by visually presenting stimuli varying from spots of light to more complex patterns of light. They found that the majority of the cells fired most strongly when an oriented line or an edge was presented on a spatially confined part of the retina. Originally, these cells were referred to as 'edge detectors' by Hubel and Wiesel. Inspired by the findings of Hubel and Wiesel (e.g., 1959, 1962, 1968), many computational edge-detector models have been developed (e.g., Marr and Hildreth, 1980; Fukushima, 1980; Marr, 1982; Canny, 1986).

Acknowledging the importance of edge detection for human vision, several algorithms have been developed to detect edges in an image (i.e., edge operators). We mention three classes of algorithms: differential operators, spatial filters, and statistical algorithms.

The first class of algorithms relies on first-order and second-order spatial derivatives of the luminance values (e.g., Marr and Hildreth, 1980; Canny, 1986). First-order differential operators perform differencing operations on adjacent (pixel-based) luminance values, while second-order differential operators compute the difference between the adjacent outcomes of the first-order operators.

The second class of computational algorithms that attempted to model the function of the edge-detector cells that Hubel and Wiesel discovered, include the detection of edges by means of local spatial frequency filters such as Gabor filters. A

Gabor function is a sinusoid weighted by a Gaussian. An example of the resulting response profile is presented in Fig. 2.3(a). The response profiles of many V1 cells resemble those of Gabor filters.

The third class of algorithms performs a statistical analysis of the intensity values in an image. Two examples of statistical methods are principal component analysis (PCA) and independent component analysis (ICA). PCA and ICA are methods that exploit the statistical structure of natural images to create compact visual representations (Field, 1994; Lee, 1998b; Pearson, 1901). PCA finds a linear transformation of the array of constituent pixel values onto a space spanned by n orthogonal principal components. The components are ordered in terms of the variance they explain. Since most of the variance in natural images is explained by oriented edges at multiple scales, the principal components can be said to detect edges. ICA is related to PCA, and tries to find a linear transformation that yields components that are independent and not normally distributed (e.g., Comon, 1994; Field, 1994; Lee, 1998b). When applied locally, both PCA and ICA yield filters that are similar to Gabor filters.

Inspired by the neurophysiological findings on edge detection in natural vision and by the fruitful computational approaches to edge detection from the field of object recognition, the first guiding principle reads: an adequate mechanism to build a veridical mapping from images to feature space performs an edge-based analysis.

2.3.2 Local multi-scale analysis

Psychophysical and biological findings indicate that the early processing stages in the human visual system detect features, such as edges, in the retinal image at a variety of scales. According to the spatial frequency theory of vision formulated by Blakemore and Campbell (1969), the visual system incorporates a large number of overlapping 'channels', each selectively tuned to a particular range of spatial frequencies and orientations. The spatial frequency theory has a mathematical basis and is supported by psychophysical and biological findings.

From a mathematical point of view, every image can be decomposed into a set of sinusoidal gratings from different spatial frequencies using Fourier analysis. The linear combination of the set of component gratings contains all the information present in the original image. This led vision researchers to believe that the cells of the primary visual processing area might be performing something similar to a global Fourier analysis (see, e.g., Palmer, 1999).

Blakemore and Campbell (1969) collected psychophysical evidence for a spatial frequency analysis in the human visual system. They showed that when subjects viewed a sinusoidal grating for a prolonged time, their visual system selectively adapted to the frequency and orientation of the presented grating, while retaining responsive to gratings of another frequency or orientation. The results by Blakemore and Campbell (1969) were considered as strong evidence for the existence of the spatial frequency channels in vision.

More support for the idea that human vision performs a multi-scale analysis by means of cells that selectively respond to a particular spatial frequency, comes from biological studies that examined the receptive field structure of V1 cells. While

Hubel and Wiesel mainly found receptive fields that were strongly responsive to edges, later studies showed that the receptive fields often have a more complex structure (de Valois and de Valois, 1988). In their study, de Valois and de Valois (1988) found that many cells which are tuned to a bar in a specific orientation and position, show a pattern of activation and inhibition when the bar is shifted laterally away from the centre of the cell's receptive field. Moreover, they found a variety of receptive field sizes, the larger ones responding most strongly to coarse spatial scales and the smaller ones responding most strongly to fine spatial scales.

Many vision researchers now assume that the receptive field structures that were studied by Hubel and Wiesel (1959, 1962, 1968) and later by de Valois and de Valois (1988) and by many others form the biological implementation of the frequency channels in the visual nervous system. Biological studies indicate that processing of visual input in V1 is local in the sense that the receptive fields of V1 cells are limited to a few degrees of visual angle (Palmer, 1999). In contrast to a global spatial frequency analysis (i.e., a Fourier analysis), the cells are assumed to perform a so-called 'local spatial frequency analysis'.

The spatial-frequency theory of natural vision is supported by psychophysical and biological findings and leads us to formulate our second guiding principle: an adequate mechanism to build a veridical mapping from images to feature space performs a local multi-scale analysis.

2.3.3 Fixation-based analysis

The extraction of visual input is performed through a fixation-based sampling of information contained in a visual scene, by means of sequences of eye fixations. Humans make approximately three to five eye fixations per second (see, e.g., Henderson, 2003; McSorley and Findlay, 2003). Eye fixations allow parts of the visual scene to be sampled with high resolution. Visual acuity is highest in the centre (i.e., fovea) of the retina and decreases towards the periphery. By moving the eyes around, parts of the visual scene can be 'foveated' to obtain detailed information at the centre of gaze. Psychological studies have shown that both bottom-up (stimulus-based) and top-down (concept-based) processes contribute to the selection of eye-fixation locations.

Numerous studies showed that bottom-up processes draw the eyes towards salient visual features such as contours (e.g., Yarbus, 1967; Norman, Phillips, and Ross, 2001), regions with high edge density (e.g., Mannan, Ruddock, and Wooding, 1996), and local contrast (e.g., Parkhurst and Niebur, 2003). These preferences can be explained in terms of the principle of maximizing information (e.g., Wainwright, 1999; Petrov and Zhaoping, 2003). Natural images are characterized by a high degree of redundancy, mainly caused by intensity correlations among adjacent pixels. At contours, at locations of high edge density, and at locations with high local contrast, intensity correlations between adjacent pixels are low, which makes these locations highly informative. PCA and ICA can be interpreted as methods to reduce the redundancy of visual images.

In addition to image saliency, the selection of fixation locations is based on top-down processes. Top-down processes rely on stored knowledge to select the most

informative location to orient the eyes (see, e.g., Henderson, 2003). The types of stored knowledge that play a role include: (1) knowledge about previous encounters with the currently viewed stimuli, (2) knowledge about the spatial arrangement of the currently viewed visual scene, and (3) task-related knowledge. These types of knowledge are discussed in more detail in chapter 6.

Motivated by psychological finding on the gathering of visual input in human vision by means of eye fixations, our third guiding principle is phrased: an adequate mechanism to build a veridical mapping from images to feature space performs a fixation-based analysis.

2.3.4 Fulfilling the three guiding principles

Having identified three guiding principles for grounding memory representations in the real visual world, we now turn to discuss ways to fulfil the guiding principles. The first two guiding principles, i.e., that of performing an edge-based analysis and that of performing a local, multi-scale analysis can be fulfilled by using the wavelet-transform approach. The third guiding principle, i.e., that of performing a fixation-based analysis, can be fulfilled by applying the wavelet-transform approach to spatially confined image regions.

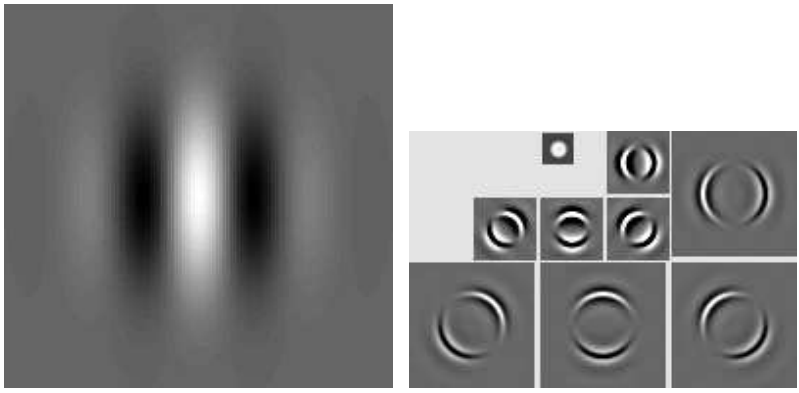
The next section discusses a specification of our front-end for memory models that fulfils the three guiding principles and borrows elements from the four approaches discussed in section 2.2.

2.4 A front-end for memory models

Wavelet transforms provide a mathematical and computational realization of both the edge-based and the local multi-scale analysis characteristics of the human visual system (we recall that they were identified as the first and the second guiding principles for a veridical mapping from images to a feature space). Wavelets can be used to model the receptive field structure discovered by Hubel and Wiesel (1959, 1962, 1968) and de Valois and de Valois (1988), which underlies the local spatial frequency analysis in human vision. Suitable wavelets are Gaussian derivatives and Gabor functions, which can be regarded as approximations of high-order Gaussian derivatives.

The steerable pyramid transform is one of several possible approaches that adheres to the guidelines of an edge-based, local, multi-scale image transform. The steerable filter pyramid implements a number of wavelets at multiple orientations and at multiple scales (Freeman and Adelson, 1991). The wavelets are oriented Gaussian derivatives. Fig. 2.3(b) shows an example of an image of a white disk that is processed (i.e., filtered) with the steerable pyramid at four orientations and two scales. To fulfil the guiding principle of the fixation-based spatial selection, our front-end applies the steerable pyramid to image regions centred around a fixation point. Each sample results in a fixation-based feature vector representing the image region.

Our front-end is a perceptual mechanism that employs a fixation-based steerable pyramid transform. The front-end connects memory models with the visual world.



(a) (b)

Figure 2.3: (a) The resulting pattern of a Gabor function. (b) An example steerable pyramid decomposition of an image of a white disk on a black background. Retrieved from http://www.cns.nyu.edu/ eero/steerpyr/ that describes the Steerable Pyramid (Freeman and Adelson, 1991)

In chapter 3, we introduce the situated memory model called the Natural Input Memory model (Nim) that combines a perceptual front-end with an exemplar-similarity-based memory model (see, e.g., Nosofsky, 1986, 1987, 1988, 1989, 2003). Nim is the focus of our studies in chapter 4. Then, chapters 5 and 6 introduce two Nim variants. The variant described in chapter 5 extends Nim's memory stage to obtain a natural-input variant of the powerful Rem model (Shiffrin and Steyvers, 1997) and tests the variant on several well-known recognition-memory effects that are obtained in behavioural memory studies. The variant described in chapter 6 extends Nim's memory stage to obtain a natural-input model for classification and introduces a fixation-selection mechanism to select relevant fixation locations.

2.5 Chapter summary

We identified the lack of a connection with the real world as the main limitation of existing memory models. We aim to develop a memory model that operates on natural images. Therefore, we investigated the guiding principles for realizing a veridical mapping from images to a feature space and identified three guiding principles that should be fulfilled. The fixation-based multi-scale wavelet transform fulfils all three guiding principles and is therefore selected as a front-end to ground memory representations.

Chapter 3

The Natural Input Memory Model

This chapter is based on[1]:

1. Lacroix, J. P. W., Murre, J. M. J., Postma, E. O., and van den Herik, H. J. (2006a). Modeling recognition memory using the similarity structure of natural input. *Cognitive Science*, Vol. 30, pp. 121–145.

In chapter 2 we discussed that existing computational memory models suffer from the limitation that they are not grounded in the real world. It was observed that: (1) they lack a mechanism that derives representations directly from the perceptual features of the stimuli, and (2) they rely on indirect methods by which an abstract feature space is predefined. Since the computational models operate on predefined abstract representations, they are unable to make predictions for natural individual stimuli in behavioural memory experiments.

In this chapter, we propose a situated computational recognition-memory model that operates on natural images: the Natural Input Memory model (NIM). NIM combines a perceptual front-end with an exemplar-similarity-based memory model. NIM's perceptual front-end is a biologically informed perceptual preprocessing method that translates a natural image into a similarity-space representation (i.e., a feature-space representation). The exemplar-similarity-based memory model operates on this similarity space in order to make a memory decision. The exemplar-similarity-based memory model that we propose in this chapter is a recognition version of the Generalized Context Model (e.g., Nosofsky, 1986, 1987). However, we would like to emphasize that NIM's perceptual front-end can be combined with other types of exemplar-similarity-based memory models and classification models (see, e.g., chapters 5 and 6).

The chapter is organized as follows. In section 3.1, we provide an overview of the model. This is followed, in section 3.2 and 3.3, by a detailed exposition of NIM's

[1]The author would like to thank her co-authors and the publisher of *Cognitive Science* for their kind permission to reuse relevant parts of the article in this thesis.

two stages: the perceptual preprocessing stage and the memory stage, respectively. Subsequently, section 3.4 compares NIM with other models. Finally, section 3.5 presents the chapter summary.

3.1 The NIM essentials

NIM encompasses the following two stages.

1. A perceptual preprocessing stage that translates a natural image into feature vectors.
2. A memory stage comprising two processes:
   (a) a storage process that stores feature vectors in a straightforward manner;
   (b) a recognition process that compares feature vectors of the image to be recognized with previously stored feature vectors.

Fig. 3.1 presents a schematic overview of NIM. The face image is an example of a natural image. The left and right side of the figure correspond to NIM's two stages: the perceptual preprocessing stage (left) and the memory stage (right). The preprocessing stage selects eye-fixation locations and extracts perceptual input (i.e., a feature vector) at each eye-fixation location. Then, it reduces the dimensionality of the feature vectors (not shown in the figure), which leads to a low-dimensional feature-vector representation. The low-dimensional feature-vector representation forms the input of the memory stage. During storage, the memory stage stores the feature-vector representation. During recognition, the memory stage matches the feature-vector representation with previously stored representations. Below, the perceptual preprocessing stage and the memory stage will be explained in more detail.

3.2 The perceptual preprocessing stage

The perceptual preprocessing stage is inspired by: (1) biological knowledge about the processing of information in the human visual system (see, e.g., Hubel, 1988; Palmer, 1999) and (2) certain computational considerations (see, e.g., Bellman, 1961; Bishop, 1995; Edelman and Intrator, 1997). Both are discussed in subsection 3.2.1. In subsection 3.2.2, we provide some relevant implementation details.

3.2.1 Biological inspiration and computational considerations

It is well known that early visual processing in the brain leads to the activation of millions of optic nerve cells (Palmer, 1999). The nerve-cell activations may be conceived as a high-dimensional vector. The high dimensionality enables the representation of a large amount of information without suffering from interference (Rao and Ballard, 1995), but it also hampers the memory performance, as the number of examples that is necessary for a reliable generalization performance grows exponentially with the number of dimensions. This phenomenon is known as the 'curse

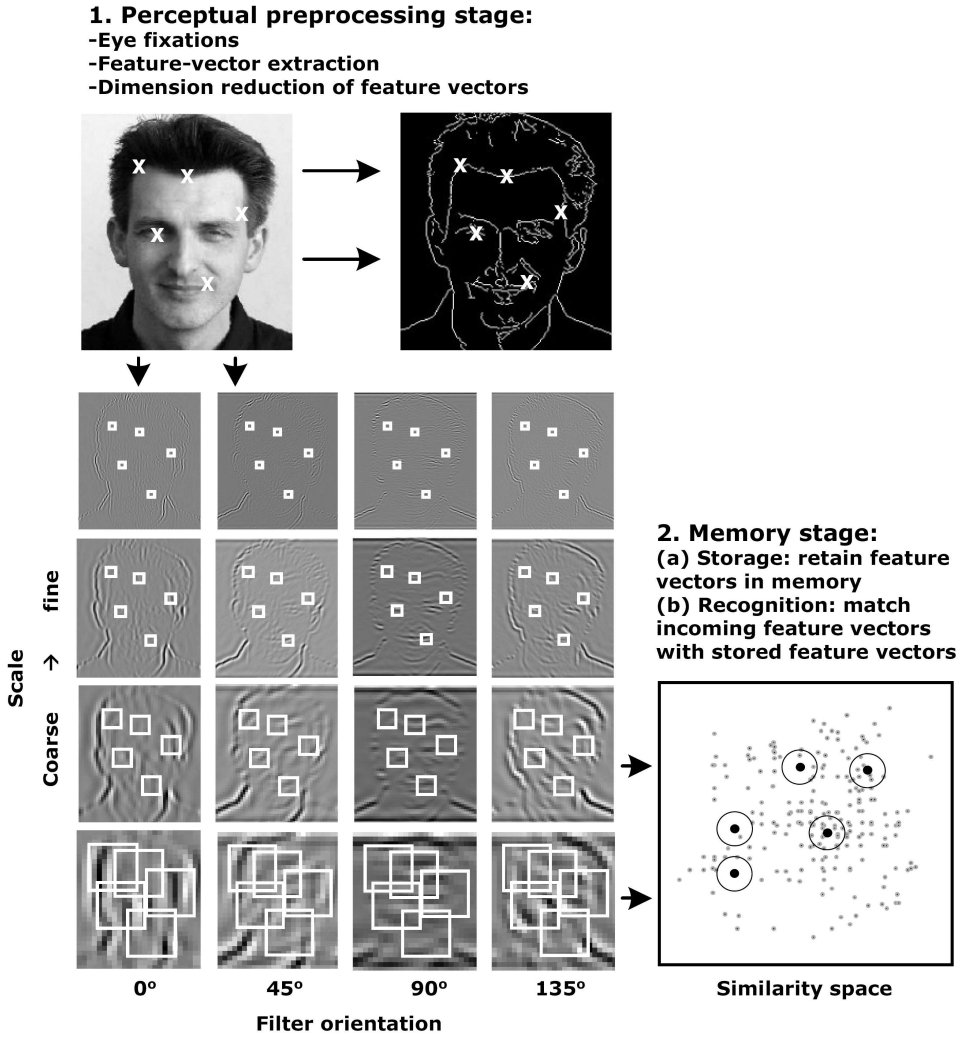


Figure 3.1: The Natural Input Memory model (NIM).

of dimensionality' (Bellman, 1961; Bishop, 1995; Edelman and Intrator, 1997). In coping with this phenomenon, subsequent stages in the visual system are assumed to reduce the dimensionality of the high-dimensional input (see, e.g., Hubel, 1988; Barlow, 1989; Tenenbaum, de Silva, and Langford, 2000). This assumption is supported by the findings of Edelman and Intrator (1997), who showed that the human visual system performs a dimensionality reduction that extracts the low-dimensional structure of high-dimensional visual input.

In NIM, dimension reduction of a high-dimensional natural image is achieved by a biologically informed feature-vector extraction method (Freeman and Adelson, 1991). A natural digitized image has a high dimensionality because it is treated as a vector, the elements of which are the constituent pixel values. Motivated by eye fixations in human vision, the feature-vector extraction method takes local samples from selected locations in the image. To emphasize the parallel with human vision we refer to these samples as 'fixations'. Human fixations tend to cluster at or near contours (see, e.g., Yarbus, 1967; Norman *et al.*, 2001). The preference for contours can be explained in terms of the principle of maximizing information (see, e.g., Wainwright, 1999; Petrov and Zhaoping, 2003), since (1) natural images are characterized by a high degree of redundancy, mainly caused by intensity correlations among adjacent pixels, and (2) at contours intensity correlations between adjacent pixels are low, which makes contours highly informative. NIM's feature-vector extraction method places fixations along the contours of the image.

For each fixation, features (i.e., a feature vector) are extracted from the image area centred at the fixation location. NIM's feature-vector extraction mimics the visual processing in area V1. The responses of neurons in V1 are modelled by a multi-scale wavelet decomposition (to be described in detail in subsection 3.2.2). Multi-scale wavelet decomposition models are biologically plausible (see, e.g., Lee, 1998a; Palmeri and Gauthier, 2004). Moreover, several studies showed that distances between representations in the similarity space that results from preprocessing input with the biologically informed multi-scale wavelet decomposition agree well with dissimilarities as perceived by humans (see, e.g., Kalocsai, Zhao, and Biederman, 1998; Lyons and Akamatsu, 1998; Lyons, 2000; Dailey *et al.*, 2002; Lacroix *et al.*, 2006a). The extracted feature vectors contain information on oriented edges at multiple scales and form an efficient basis for object recognition (see, e.g., Rao and Ballard, 1995). In order to reduce the dimensionality of the feature vectors we use the method of principal component analysis (PCA). PCA discovers a linear transformation of n-dimensional vectors onto a space spanned by n orthogonal principal components (see also subsection 2.3.1). The components are ordered in terms of the variance they explain. The dimensionality of the extracted feature vectors is reduced by taking the projection onto the first p ($p < n$) principal components. The p-dimensional feature vectors so obtained reside in a similarity space where visual similarity translates to proximity of feature vectors.

Translating a two-dimensional image by using a multi-scale wavelet decomposition followed by a dimension reduction by means of a principal component analysis, is an often applied method in the domain of visual object recognition to model the first stages of processing of information in the human visual system (i.e., retina/LGN, V1/V2, V4; cf., Dailey *et al.*, 2002; Palmeri and Gauthier, 2004).

3.2.2 Implementation of the perceptual preprocessing stage

The features extracted from the face image consist of the responses of different derivatives-of-Gaussian filters. By applying a set of these filters at multiple scales and orientations, a representation of the face image area centred at the fixation location is obtained (Freeman and Adelson, 1991). To extract the feature vectors, the entire input image is transformed into a multi-scale representation at four spatial scales. At every scale, the image is processed by four oriented filters in the orientations 0°, 45°, 90°, and 135° using the steerable-pyramid transform (Freeman and Adelson, 1991). The choice for four orientations is assumed to be sufficient for the task at hand. The steerable-pyramid transform with four orientations is overcomplete, which means that it contains more information than necessary to reconstruct the original image. The human visual system appears to 'filter' retinal images with much more orientations (Palmer, 1999).

The steerable-pyramid processing results in the 16 (four scales times four orientations) filtered images shown on the lower left-hand corner of Fig. 3.1. The pixel values represent filter coefficients. Brighter pixels correspond to larger filter responses. From each of the 16 images a 7×7 window is selected centred at a fixation location and the 7×7 coefficients of the 16 images are placed in a vector. The sampling of 7×7 coefficients at each scale and orientation of the steerable pyramid is based on the following two considerations. First, it is assumed to reflect the way the human visual system samples the local neighbourhood of fixated locations at multiple scales (Postma, van den Herik, and Hudson, 1997). Second, it corresponds roughly to the estimates of the resolution of spatial attention (Nakayama, 1990). In addition to the filter coefficients, the pixel values of a 7×7 window of a low-resolution version of the image, centred at the fixation location, are appended to the vector. The 7×7 low-resolution sub-image is included in the feature vector, because it contains absolute brightness information at the fixation location. This information is absent in the other 7×7 windows that contain the coefficients reflecting filter responses. Taken together, each fixation yields a 833-dimensional ($16 \times 7 \times 7 + 1 \times 7 \times 7$) feature vector that contains information ranging from visual details (high-scale features) to coarse visual characteristics (low-scale features). It is worth noting that though information about the spatial locations and arrangement of fixations is not represented explicitly, it is represented implicitly by the low-scale features.

Fixation locations are randomly drawn from the contours of the face image. To detect the contours in an image we used an edge detector, which is based on the standard Canny edge detector with an adaptive threshold (i.e., the minimum and maximum intensity values) and $\sigma = 1$ (Canny, 1986)[2]. In Fig. 3.1 the contours of the input image are shown to the right of the input image.

After feature-vector extraction, we reduce the dimensionality of the feature vectors extracted by the multi-scale wavelet decomposition. To find the principal components that cover most of the variance of the 833-dimensional vectors, we apply PCA to a large set of feature vectors (i.e., about 60,000 feature vectors extracted from the images used in the simulations). Then, in our simulations, we project the feature vectors onto the first $p = 50$ of these principal components. It has been

[2]In principle, the steerable pyramid can be used to detect the contours.

shown that approximately 50 components are sufficient to represent faces accurately (Hancock, Burton, and Bruce, 1996; Calder *et al.*, 2001; Dailey *et al.*, 2002).

3.3 The memory stage

NIM's memory stage is based on the Generalized Context Model (GCM; e.g., Nosofsky, 1986, 1987). While the GCM was initially introduced as an exemplar-similarity model for explaining categorization and identification of multi-dimensional perceptual items (e.g., Nosofsky, 1986, 1987), it has successfully been applied to old-new recognition tasks as well (see, e.g., Busey and Tunnicliff, 1999; Zaki and Nosofsky, 2001; Nosofsky and Zaki, 2003). GCM-based recognition models rely on the assumption that individual exemplars are stored in memory and that recognition decisions are based on the similarity of new inputs to previously stored inputs (see, e.g., Busey and Tunnicliff, 1999; Busey, 2001). In section 3.3.1, we briefly discuss NIM's GCM-based memory stage that is a recognition version of the GCM. Subsequently, we describe the implementation details in subsection 3.3.2.

3.3.1 The GCM-based memory stage

We distinguish two processes in the memory stage: the storage process and the recognition process.

The storage process

NIM retains (i.e., 'stores') exemplars of natural images in a similarity space. The storage process stores the preprocessed natural images. A preprocessed natural image is represented by a number of fixations, i.e., low-dimensional feature vectors in a similarity space, each corresponding to the preprocessed image contents at the fixation location.

The recognition process

Recognition decisions in the GCM are based on the familiarity of an item. Familiarity is assessed by summing the similarities of a test image to the individual exemplars stored in memory (e.g., Nosofsky and Zaki, 2003). Similarity between two items is a decreasing function of distance in the representation space. The recognition process of NIM's GCM-based memory stage determines the familiarity of an item to be recognized (i.e., a test image) by summing the similarities of the feature vectors of the image to previously stored feature vectors. The similarity of two feature vectors is defined as a decreasing function of their Euclidean distance in the similarity space. The similarity summed over all previously stored feature vectors serves as a basis for old-new recognition decisions. Although, in the GCM it is suggested that similarities are changed systematically by selective attention (Nosofsky, 1986), we, for now, ignore the role of attention. This means that each dimension in the similarity space receives an equal weight.

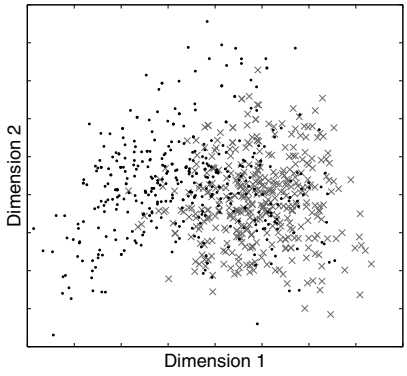


Figure 3.2: Two-dimensional projection of the feature-vector representations of two natural images in a similarity space.

3.3.2 Implementation of the GCM-based memory stage

Below, we discuss the implementation of the two processes of the memory stage: the storage process and the recognition process.

The storage process

In each unit of presentation time, the storage process stores s feature vectors for an image (corresponding to s fixations). Fig. 3.2 is an illustration of the feature-vector representations of two natural images in a two-dimensional projection of a similarity space. The x-axis and the y-axis show the first and second dimensions (i.e., principal components), respectively. In the illustration, each image is represented by a cluster of 400 feature vectors, i.e., fixations (indicated by black dots ($\cdot$) and grey crosses ($\times$)). Note that this is an example of a representation based on 400 eye fixations. In our recognition simulations, however, we stored 10 fixations for each image. This corresponds to about two to three seconds of viewing time in a recognition-memory experiment, as humans make approximately three to five eye fixations per second (see e.g., Henderson, 2003; McSorley and Findlay, 2003).

The recognition process

The recognition process defines the familiarity of a test image as the average familiarity across its feature vectors. The familiarity of a feature vector is based on the summed similarity to the previously stored feature vectors. The GCM typically uses a Gaussian or an exponential decay function for relating stimulus similarity to distance in the similarity space. In a preliminary pilot study, we obtained the best results with the step decay function. The step decay function is 1 when the Euclidean distance is smaller than or equal to a certain radius r and 0 otherwise. In

other words, the step function simply calculates the number of previously stored feature vectors that reside within a hypersphere with radius r around a feature vector. Formally, the familiarity of a test feature vector, j, is denoted as:

$$fam_j = \sum_{i=1}^{T} s(d_{i,j}), \tag{3.1}$$

where

$$s(d_{i,j}) = \begin{cases} 1 & \text{if } d_{i,j} \leq r \\ 0 & \text{otherwise} \end{cases}$$

with T the total number of previously stored feature vectors and $d_{i,j}$ the Euclidean distance between feature vector i and j. The familiarity of a test image J is then defined as:

$$fam_J = \frac{1}{N} \sum_{j=1}^{N} fam_j \tag{3.2}$$

with N the total number of feature vectors of test image J.
We use the logistic function to transform the familiarity values into a recognition probability between 0 and 1 (see, e.g., Busey and Tunnicliff, 1999; Busey, 2001). The logistic function is defined as:

$$P(recognized|J) = \frac{1}{1 + \beta e^{-\theta fam_J}}, \tag{3.3}$$

with β and θ two free parameters, and fam_J the familiarity of face J (Busey and Tunnicliff, 1999).

3.4 Comparison with other models

The main difference between NIM and other existing memory models is that NIM encompasses a biologically informed perceptual front-end that operates on natural images. In subsection 3.4.1, we discuss the benefit of encompassing such a perceptual front-end and in subsection 3.4.2 we compare the perceptual front-end with other image-processing techniques.

3.4.1 The benefit of a perceptual front-end

NIM's perceptual preprocessing applies a transformation that yields a perceptual similarity structure of natural images. So far, existing memory models have been tested with artificial data (e.g., the REM model, Shiffrin and Steyvers, 1997), with predefined similarity spaces (e.g., the GCM, Nosofsky, 1987; the SIMSAMPLE model, Busey, 2001; the REM model, Steyvers, 2000), and with synthesized texture images (the NEMO model, Kahana and Sekuler, 2002). For conceptually based stimuli, many methods for defining the similarity space are based on word co-occurrence counts in texts (LSA, e.g., Derweester *et al.*, 1990; Landauer and Dumais, 1997; HAL, e.g., Burgess *et al.*, 1998), on experimentally obtained free association data

(WAS, e.g., Steyvers *et al.*, 2004) or on experimentally obtained similarity ratings (MDS, Caramazza *et al.*, 1976). For perceptually based stimuli, such as faces, the latter approach has also often been used (e.g., Busey, 2001). Since the resulting perceptual similarity spaces accurately reflect similarities as perceived by humans, this approach leads to useful representational models of memory. However, because representations are not derived directly from the individual input patterns, these models fall short in constructing a representation that is grounded in the real world. Because NIM's preprocessing stage can be conceived as an image-processing front-end, it can be applied to other models of memory to realize grounded representations. Therefore, NIM's front-end complements rather than replaces existing computational memory models such as REM. The strength of NIM thus lies in the fact that it is able to operate directly on natural images. Hence, NIM can be used to predict recognition judgements for novel stimuli. We consider this a clear benefit of a perceptual front-end.

3.4.2 Image processing

The approaches to modelling the cognitive elements of recognition have largely evolved independently of the image-processing approaches to recognition. The latter are mainly concerned with how visual input can be mapped onto a certain representation despite variations in viewpoint and lighting conditions. NIM introduces advanced image preprocessing in the cognitive modelling of memory. The advantages of combining image-processing techniques with cognitive approaches have been emphasized by several researchers (Edelman, 1995a; Burton, Bruce, and Hancock, 1999; Steyvers and Busey, 2000; Calder *et al.*, 2001; Dailey *et al.*, 2002). The method of PCA is widely applied in the domain of image processing. Below, we consider how our preprocessing method relates to PCA. Then, we discuss how NIM's preprocessing stage relates to that of Dailey *et al.*'s (2002) model called EMPATH.

PCA applied to the pixel values of natural images yields a compact representation of the images (Hancock, Baddeley, and Smith, 1992). A number of face-recognition models (e.g., Turk and Pentland, 1991; O'Toole *et al.*, 1993; Burton *et al.*, 1999) apply PCA to the entire (shape-standardized) image to obtain so-called 'eigenfaces' (e.g., Turk and Pentland, 1991), i.e., the principal components that account for most of the variance in a number of face images. Analogously to NIM's similarity space, the similarity space spanned by the eigenfaces forms the psychological abstract notion of a similarity space for faces that is generally assumed to underlie face memory (Valentine, 1991; O'Toole, Wenger, and Townsend, 2001). Models based on principal components have been shown to perform successfully different recognition tasks, such as old-new recognition memory, identification, and recognition of facial expressions (e.g., O'Toole *et al.*, 1993; Burton *et al.*, 1999; Calder *et al.*, 2001). Although PCA applied to the entire face image extracts important visual features from a set of face images, the nature of the features depends on the specific expressions or shapes of the faces. Therefore, the use of a different set of images requires recomputation of the principal components so that they fit the new set. In order to obtain the general features of natural images, a PCA should be applied to an extensive collection of natural images containing a wide variety of objects and scenes instead of a limited set

of training images. Hancock *et al.* (1992) found that the principal components that resulted from performing PCA on a large number of natural images approximated derivatives of two-dimensional Gaussian functions (or Gabor functions). Such functions, which are used for feature extraction in NIM, form an appropriate basis for building a similarity-space representation from natural images that is independent of the specific set of images used. Therefore, we prefer their use over standard PCA applied to the raw or shape-standardized image. After feature vectors have been extracted using Gaussian derivatives, we still apply a PCA to the extracted feature vectors in order to reduce computational demands.

Our preprocessing stage resembles that in EMPATH developed by Dailey *et al.* (2002): multi-scale wavelet filtering followed by PCA. EMPATH relies on the preprocessing to explain psychological findings on the perception of facial expressions. Moreover, Dailey, Cottrell, and Busey (1999) used EMPATH as a preprocessing front-end to different memory models in an attempt to account for the experimental recognition data obtained by Busey and Tunnicliff (1999) (in a similar way, we evaluate to what extent NIM can account for the results by Busey and Tunnicliff (1999) in one of our studies in chapter 4 and provide a comparison with the EMPATH results in subsection 4.2.5). By showing that EMPATH simulates a variety of psychological results on categorization, similarity, discrimination, and recognition difficulty related to facial expression perception, Dailey *et al.* (2002) validate the similarity structure extracted from the input by the preprocessing method. Although the image-processing methods employed by NIM and EMPATH are quite similar, they differ in two ways. The first difference is a minor difference. For feature extraction, Dailey *et al.* (2002) use Gabor wavelets at different scales and orientations. In contrast, we use the steerable pyramid (a set of Gaussian derivatives) at different scales and orientations (see section 2.4 and subsection 3.2.2). While Dailey *et al.* (2002) use a slightly different set of filters, both feature-extraction methods are multi-scale wavelet decompositions that contain information about oriented edges at different scales and orientations. The second and main difference between NIM and EMPATH concerns the way in which image locations are selected for feature extraction. In EMPATH, images are filtered at 29×35 grid points, evenly distributed over the image (in Dailey *et al.*, 1999, only 64 grid points are used). In contrast, NIM selects a few eye-fixation locations randomly along the contours in the image (moreover, in chapter 6 we will introduce fixation-selection mechanisms for the selection of relevant fixations) and extracts the filter responses from a small image area surrounding each fixation. This corresponds to the way human subjects attend different parts of a visual scene by means of eye fixations. NIM's mechanism of selectively fixating different parts of the image provides a natural way of dealing with overt spatial attention. Moreover, since the number of stored fixations corresponds to viewing time, the timing of stimuli can be modelled. Simulations that manipulate NIM's viewing time show that the model's recognition performance increases with the number of stored fixations (Lacroix *et al.*, 2004).

3.5 Chapter summary

In this chapter, we described the details of the Natural Input Memory model (NIM). NIM differs from existing memory models in that it incorporates a perceptual pre-processing method that builds a similarity space directly from natural visual input. Recognition is based on a matching of incoming similarity-space representations with previously stored representations. Complementing a memory model with a perceptual front-end allows for automatic predictions of recognition memorability for individual natural stimuli.

Chapter 4

Validation of Nim on individual natural stimuli

This chapter is based on[1]:

1. Lacroix, J. P. W., Postma, E. O., and Murre, J. M. J. (2005). Predicting experimental similarity ratings and recognition rates for individual natural stimuli with the Nim model. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society (CogSci 2005)* (eds. B. G. Bara, L. Barsalou, and M. Bucciarelli), pp. 1225–1230, Lawrence Erlbaum Associates, Mahwah, NJ.

2. Lacroix, J. P. W., Murre, J. M. J., Postma, E. O., and van den Herik, H. J. (2006a). Modeling recognition memory using the similarity structure of natural input. *Cognitive Science*, Vol. 30, pp. 121–145.

In this chapter, we show that Nim can predict similarity ratings and recognition rates for individual natural stimuli (Lacroix, Postma, and Murre, 2005; Lacroix *et al.*, 2006a)[2]. The model is tested on a similarity-rating task and a face-recognition task using the same natural face images that were used in behavioural experiments (Busey, 1998; Busey and Tunnicliff, 1999; Busey and Arici, in preparation).

The chapter is organized as follows. In section 4.1, Nim is tested on a similarity-rating task in order to assess the psychological plausibility of the similarity space that Nim builds from the natural input. Subsequently, in section 4.2 Nim is validated on a face-recognition task. Finally, section 4.3 provides the chapter conclusion.

[1]The author would like to thank her co-authors and the publisher of the CogSci 2005 proceedings and of *Cognitive Science* for their kind permission to reuse relevant parts of the article in this thesis.

[2]Dr. Busey of Indiana University Bloomington is gratefully acknowledged for providing us with his data set of facial images and helpful comments. Moreover, we wish to thank the reviewers for their helpful comments on earlier versions of the article on which this chapter is partly based.

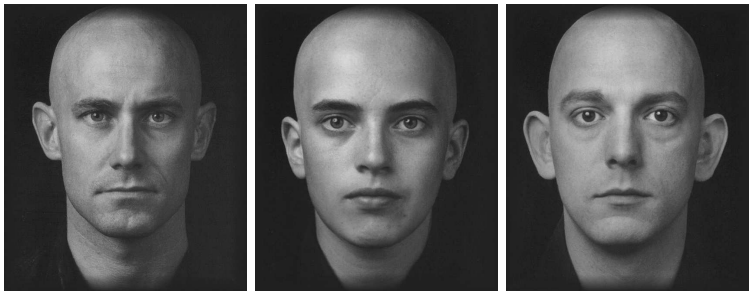


Figure 4.1: Three examples of faces contained in the sets of faces used in the similarity-rating experiments.

4.1 The similarity-rating task

Since NIM's key characteristic is that it transforms natural images into similarity-space representations, the psychological plausibility of the constructed similarity space is crucial for NIM's behaviour. Therefore, we assess the psychological plausibility of the similarity space that NIM builds from natural input. In order to do so, we compare the NIM similarity ratings with experimentally obtained human similarity ratings. Human similarity ratings were acquired from two studies: Busey and Arici (in preparation) [3] and Busey (1998).

In this section, we review the two behavioural similarity-rating experiments (4.1.1), describe the similarity-rating stimulations with NIM (4.1.2), provide the simulation results (4.1.3), discuss the results (4.1.4), and compare our results with those obtained in other similarity-rating modelling studies (e.g., Dailey *et al.*, 1999; Steyvers and Busey, 2000) (4.1.5).

4.1.1 Behavioural similarity-rating experiments

In both the experiments by Busey and Arici (in preparation) and by Busey (1998), subjects were repeatedly confronted with two face images and were instructed to rate the similarity by assigning a number ranging from 1 (most similar) to 9 (least similar). This resulted in human similarity ratings for all possible pairs of faces. The sets of stimuli used in the experiments consisted of grey-scale images of bald male faces without glasses. Three examples of these faces are shown in Fig. 4.1.

In Busey and Arici (in preparation), 238 subjects were tested on pairs of faces from a set of 60 faces (8 of the faces contained in the set were morphs, created by averaging 2 of the 52 other faces). Across subjects, this resulted in a total of about 25 ratings for each of the 1770 possible pairs of faces.

In Busey (1998), 343 subjects were tested on pairs of faces from a set of 104 faces (16 of which were morphs, created by averaging two of the 88 other faces), resulting in at least six ratings for each of the 5356 possible pairs of faces. Within

[3] also referred to in Steyvers and Busey (2000)

each experiment, the similarity ratings of individual subjects were translated into z-scores and averaged across subjects.

4.1.2 Similarity-rating simulation with NIM

NIM bases a similarity rating for two faces on a matching of the similarity-space representations that result from preprocessing those faces. Since this task involves no memory processes, the memory stage is not employed in these simulations.

We presented NIM with all possible face pairs from the sets of face images used in the experiments by Busey and Arici (in preparation) and by Busey (1998). For the assessment of the similarity of two faces, we used a method similar to the assessment of familiarity as defined in equations 3.1 and 3.2 in chapter 3. For each face of a pair (A, B), 100 feature vectors were extracted. Then, for each feature vector of face A, we determined the summed similarity to the feature vectors of face B using a step function as defined for familiarity with radius parameter r (see subsection 3.3.2). The average summed similarity defined the NIM similarity value s for face A and B. To compare the similarity values to the human similarity ratings, we linearly transformed the human similarity ratings (i.e., the average z-scores of the human similarity ratings) to cover the range from 0 to 1 and logistically transformed the NIM similarity values into similarity ratings SR using the logistic function as defined previously in subsection 3.3.2.

In our simulations, we varied the radius r from 2.0 to 6.0 to obtain the value for which correlations between the human similarity ratings and the NIM similarity ratings SR were optimal. Results were averaged across 1000 simulation runs.

4.1.3 Similarity-rating simulation results

The highest correlations between human and NIM similarity ratings were found when the radius had a value ranging from 4.0 to 5.4. Fig. 4.2 presents the correlations between human and NIM similarity ratings (SR) as a function of the radius r. As can be seen in Fig. 4.2, for the experiment by Busey and Arici (in preparation), the simulations with $r = 5.1$ produced the highest correlation of $R = 0.65$. For the experiment by Busey (1998), the simulations with $r = 4.9$ resulted in the highest correlation of $R = 0.70$. Fig. 4.3(a) shows the NIM similarity ratings SR for a radius of $r = 5.1$ as a function of the human similarity ratings that were obtained experimentally by Busey and Arici (in preparation). Fig. 4.3(b) shows the NIM similarity ratings for for $r = 4.9$ as a function of the human similarity ratings obtained in the experiment by Busey (1998).

4.1.4 Discussion of the similarity-ratings results

Below, we discuss four points that provide some perspective on the correlations obtained in our experiments: (1) the limited consistency of similarity judgements across subjects, (2) the logistic transform of similarity values into similarity ratings, (3) the selection of eye-fixation locations, and (4) the contribution from the different spatial scales.

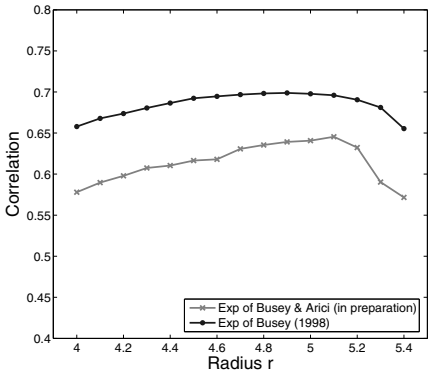
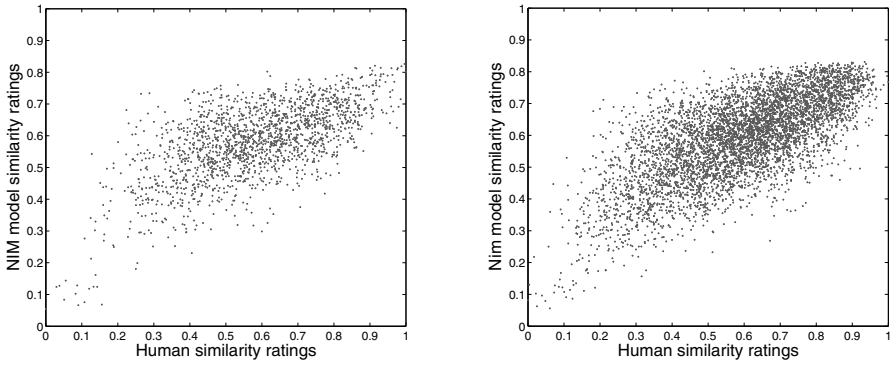


Figure 4.2: Correlations between experimentally obtained human similarity ratings and NIM similarity ratings (SR) as a function of the radius r.



(a) (b)

Figure 4.3: The NIM similarity ratings (SR) as a function of the experimentally obtained human similarity ratings (linearly transformed to the range 0 to 1). (a) The experiment by Busey and Arici (in preparation), $R = 0.65$. (b) The experiment by Busey (1998), $R = 0.70$.

The first point concerns the limited consistency of the human similarity judgements. An analysis of the human similarity ratings obtained in the experiment by Busey and Arici (in preparation) showed that there was considerable variation across the different subjects' similarity ratings. When the subjects' similarity ratings were divided into 5 groups (each consisting of some 48 subjects), correlations between the different group averages ranged from 0.72 to 0.75. This low correlation value demonstrates the limited consistency of the similarity judgements across subjects. It is highly unlikely that correlations between the similarity ratings produced by NIM and experimental similarity ratings will exceed the correlation value of 0.75. Given this consideration, NIM's similarity-space representations then can be said to correlate reasonably well with the experimentally obtained similarity ratings.

The second point is the use of a non-linear function to map similarity values onto similarity ratings. Our motivation for using the non-linear logistic function is that it constrains similarity ratings to the unit interval (Busey and Tunnicliff, 1999). Since the logistic function produces a constrained range of outputs, we applied it to translate the NIM similarity values into similarity ratings. However, a linear transform of the similarity values produced comparable correlations that were 0.64 and 0.70 for the experiments by Busey and Arici (in preparation) and Busey (1998), respectively.

The third point is about the role of the eye-fixation locations. The correlations obtained can be explained partly by the selection of eye-fixation locations. In a separate study, we selected eye-fixation locations randomly along the image (as opposed to selection along the contours in the image). Using this procedure, correlations between model and human similarity ratings are significantly lower than when fixations are selected along the contours in the image. The selection of eye-fixation locations plays an important role in human vision (e.g., Rajashekar, Cormack, and Bovik, 2002, see also chapter 6).

Finally, the fourth point concerns the contribution from the various spatial scales. To test this contribution, we ran similarity-rating simulations that selectively ignored the features at one of the four spatial scales. The results demonstrate that correlations between model and human similarity judgements decrease significantly more when low-scale visual features (i.e., coarse visual information) are removed than when high-scale visual features (i.e., visual details) are removed. It is very well possible that coarse visual information plays a more important role than detailed visual information, when judging the similarity of two faces.

4.1.5 Comparison with other modelling studies

Two other studies have compared the representation spaces with the human similarity ratings obtained in the behavioural studies described above. Dailey *et al.* (1999) performed a comparison with the 104 face images used in the experiment by Busey (1998) and Steyvers and Busey (2000) with the 60 face images used in the experiment by Busey and Arici (in preparation). Below, we provide a brief overview of both studies.

The study by Dailey *et al.* (1999) examined the ability of three different memory models operating on three types of representation spaces to account for the experi-

mental face-recognition results by Busey and Tunnicliff (1999). Two of these three types of representation spaces were generated using PCA and Gabor filtering followed by PCA (similar to the preprocessing in Dailey *et al.*'s (2002) EMPATH; see the discussion in subsection 3.4.2 for a detailed description of the differences between NIM's and EMPATH's preprocessing). Dailey *et al.* (2002) compared the resulting representations with the distances between pairs of faces in the MDS space for the 104 face images. They found a correlation of 0.39 for the principal-component representation space and a correlation of 0.52 for the Gabor-filter representation space, which are considerably smaller than the correlation of 0.70 obtained with NIM.

Higher correlations were obtained by Steyvers and Busey (2000). In their feature-mapping model, Steyvers and Busey related the features extracted using various preprocessing mechanisms to the dimensions of a multi-dimensional scaling solution based on human similarity judgements for the 60 face images (Busey and Arici, in preparation). They investigated three different preprocessing mechanisms: principal component analysis, Gabor-filter based preprocessing, and geometric information extraction (i.e., describing the face by a set of distances between landmark points in the face). The feature-mapping model differs from NIM in an important respect: in the feature-mapping model the pre-processed images were fed into an optimization network to enhance the fit with behavioural data. The optimization network learned the mapping between the input features and the psychological dimensions of the MDS solution of the human similarity ratings (Steyvers and Busey, 2000). Considering that Steyvers and Busey (2000) specifically optimized their representations with respect to the human similarity ratings, it is not surprising that they found correlations ranging from 0.45 to 0.86. In contrast to their approach, NIM changes only three parameters, whereas Steyvers and Busey (2000) adapt at least 40 weights to optimize the fit with the human data. The study by Steyvers and Busey (2000) provides important clues about how the dimensions of feature vectors extracted with a certain preprocessing mechanism should be weighed in a psychologically plausible way. Future NIM versions should analogously address the weighing of feature-vector dimensions in a more psychologically plausible way.

Several other studies have compared the representation spaces resulting from various image preprocessing schemes with the outcomes of psychological experiments. For instance, Calder *et al.* (2001) and Dailey *et al.* (2002) did this for coding facial expressions, and Hancock, Bruce, and Burton (1998), Kalocsai *et al.* (1998), and Lyons (2000) did this for face identity. Hancock *et al.* (1998) compared human similarity judgements with similarity-space representations based on a principal component analysis or on a graph-matching system. They found small correlations of about 0.20 and argued that it was caused by the noisiness of the human data. Kalocsai *et al.* (1998) used a samedifferent judgement task to compare the performances of a preprocessing scheme based on a Gabor-filtering and a global template-matching classifier with human data. In the samedifferent judgement task, participants had to judge whether two sequentially presented images were of the same individual. Kalocsai *et al.* (1998) found correlations up to 0.91. However, these were based on a small sample of 16 judgements. Moreover, the similarity judgements were obtained using a much simpler binary classification task. Lyons (2000) found a correlation of 0.71 between human and model-based face similarity ratings, which is slightly

higher than the correlations we found. However, he employed a very small set of 10 facial images and did not apply PCA to reduce the dimensionality of the representation space. In our studies, we employed sets of 60 and 104 faces and used feature representations of a much smaller dimensionality.

4.2 The recognition task

In order to validate NIM's recognition predictions for individual natural stimuli, we compare the NIM recognition rates with the human recognition rates that were obtained in two behavioural experiments. The first behavioural recognition experiment is the face-recognition experiment by Busey and Arici (in preparation), employing 60 faces and 238 subjects (as described in subsection 4.1.1). The second experiment is the face-recognition experiment 1 by Busey and Tunnicliff (1999), employing 104 faces and 180 subjects. The face images used in these experiments were identical to those used in the similarity-rating experiments.

Below, we review the behavioural face-recognition experiments (4.2.1), describe the face-recognition simulations with NIM (4.2.2), present the simulation results (4.2.3), discuss the results (4.2.3), and compare our modelling results with those obtained in other face-recognition modelling studies (Busey and Tunnicliff, 1999; Dailey *et al.*, 1999) (4.2.5).

4.2.1 Behavioural recognition experiments

In their recognition experiments, Busey and Arici (in preparation) and Busey and Tunnicliff (1999) assessed the recognition rates (i.e., hit rates and false-alarm rates) for different types of faces. The sets of 60 and 104 faces employed in the experiments contained two types of faces: (1) normal faces, and (2) morph faces. Each morph face was the average of two normal faces, so-called parent faces.

In the face-recognition experiment by Busey and Arici (in preparation), the set of 60 faces was subdivided into two sets of 30 faces in such a way that each set contained 26 normal faces, 8 of which were defined as parent faces, and 4 morph faces (each of which was the average of 2 of the 8 parent faces). When a morph was in one of the two sets, its parents were in the other. Half of the subjects were presented with a study list of the faces from set 1 and tested for old-new recognition for the faces from both set 1 (i.e., targets) and set 2 (i.e., lures). For the other half of the subjects this was reversed. In this way, hit and false-alarm rates were obtained for each of the 60 faces.

In the face-recognition experiment by Busey and Tunnicliff (1999) that employed a set of 104 faces, subjects were presented with a study list of 68 normal faces, 32 of which were defined as parent faces. Then, old-new recognition was tested for the 68 normal faces from the study list (i.e., 36 normal targets and 32 parent targets) along with 20 new normal faces (i.e., normal lures) and 16 new morph faces (i.e., morph lures). The morph lures were the average of two parent targets that were either dissimilar or similar to each other. Dissimilar and similar morph lures resulted from dissimilar and similar morph parents, respectively.

4.2.2 Recognition simulation with NIM

In the simulations of Busey and Arici's (in preparation) and Busey and Tunnicliff's (1999) face-recognition experiments, NIM was presented with a set of study-list faces and then tested for old-new recognition of the set of faces from the study list (i.e., targets) along with a set of new faces (i.e., lures). The simulations with NIM used study lists and test lists identical to those in the behavioural experiments. In other words, NIM was provided with input similar to the input that the human subjects had received in the behavioural experiments.

For half of the subjects in the experiment by Busey and Arici, the faces from set 1 were the targets and the faces from set 2 were the lures; for the other half, this was reversed. In a similar way, in half of the simulations of Busey and Arici's (in preparation) face-recognition experiment (henceforth referred to as simulation set A), the faces from set 1 were the targets and the faces from set 2 were the lures. For the other half of the simulations (henceforth referred to as simulation set B) this was reversed. See subsection 4.2.1 for the composition of the face sets 1 and 2.

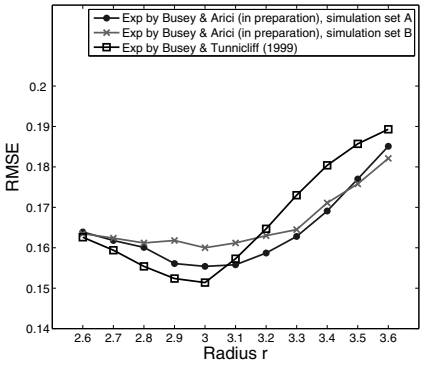
In the simulations of the face-recognition experiment by Busey and Tunnicliff (1999), NIM was presented with the faces from the study list of the behavioural experiment (see subsection 4.2.1). Then recognition was tested for the faces of the test list of the behavioural experiment. Again, see subsection 4.2.1 for the composition of the study and test lists.

In the behavioural experiments, the images were presented for 1500 ms, followed by a two-second delay. In our simulations, the number of fixations selected and stored for each face was set to 10 which corresponds to approximately two seconds of viewing time (see, e.g., Henderson, 2003; McSorley and Findlay, 2003). For recognition, NIM calculated the familiarity of each target and lure on the basis of 100 fixations using equation 3.1. Familiarity values were transformed into recognition probabilities using the logistic transform as defined in equation 3.3. In the simulations the radius r was varied from 2.0 to 6.0 to determine the r value that produced (1) the smallest root-mean-square errors ($RMSE$) and (2) the largest correlation values between the experimentally obtained human recognition rates (i.e., hit rates and false-alarm rates) and the NIM recognition rates.

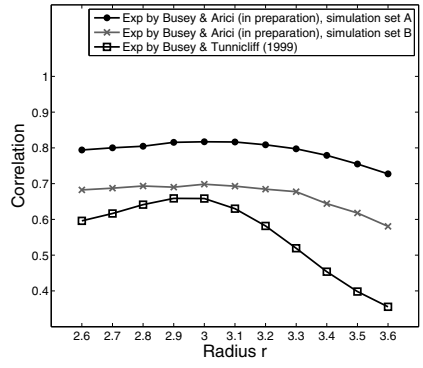
4.2.3 Recognition simulation results

The smallest $RMSE$s and the largest correlation values between the human and the NIM recognition rates were obtained when the radius r had a value ranging from 2.6 to 3.6. Fig. 4.4(a) presents the $RMSE$s between the human and the NIM recognition rates and Fig. 4.4(b) the correlations between the human and the NIM recognition rates as a function of the radius r.

For the simulations of the face-recognition experiment by Busey and Arici (in preparation), the smallest $RMSE$s between the human and the NIM recognition rates were obtained for $r = 3.0$ ($RMSE = 0.155$ for simulation set A and $RMSE = 0.160$ for simulation set B). The correlations for $r = 3.0$ were $R = 0.82$ and $R = 0.70$ for simulation sets A and B, respectively. Using this radius, Figs. 4.5(a) and 4.5(b) present the NIM recognition rates as a function of the human recognition rates for simulation sets A and B, respectively. Fig. 4.5(c), to be discussed in subsection



(a)



(b)

Figure 4.4: A comparison of the human and the NIM recognition rates for the experiments by Busey and Arici (in preparation) and by Busey and Tunnicliff (1999). (a) The $RMSE$s between the human and the NIM recognition rates as a function of the radius r. (b) The correlation between the human and the NIM recognition rates as a function of the radius r.

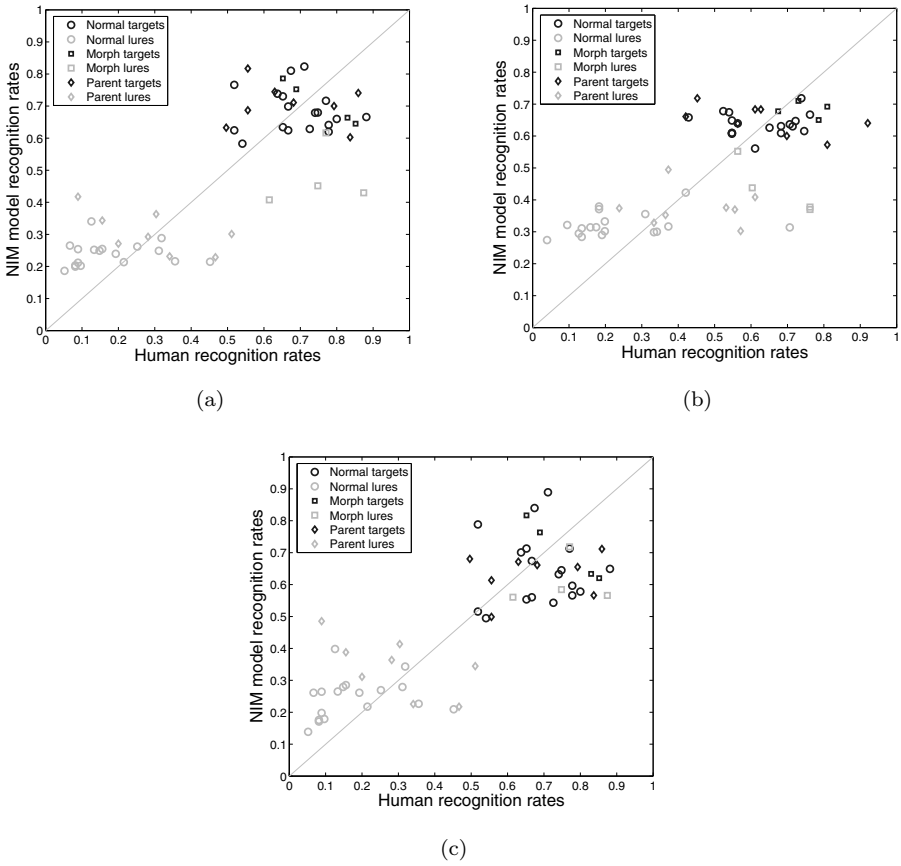


Figure 4.5: The NIM recognition rates as a function of the human recognition rates (Busey and Arici, in preparation). (a) Simulation set A, for $r = 3.0$, $RMSE = 0.155$, $R = 0.82$. (b) Simulation set B, for $r = 3.0$, $RMSE = 0.160$, $R = 0.70$. (c) Simulation set A, for $r = 3.3$, $RMSE = 0.162$, $R = 0.80$.

4.2.4 presents the NIM recognition rates for a radius of $r = 3.3$ as a function of the human recognition rates for simulation set A.

For the simulations of the face-recognition experiment by Busey and Tunnicliff (1999), the smallest $RMSE$ and the highest correlation between the human recognition rates and the NIM recognition rates obtained for $r = 3.0$ are $RMSE = 0.151$ and $R = 0.66$.

Fig. 4.6(a) presents the NIM recognition rates for $r = 3.0$ as a function of the human recognition rates (Busey and Tunnicliff, 1999). Fig. 4.6(b), which will be discussed in subsection 4.2.4, does this for $r = 3.2$.

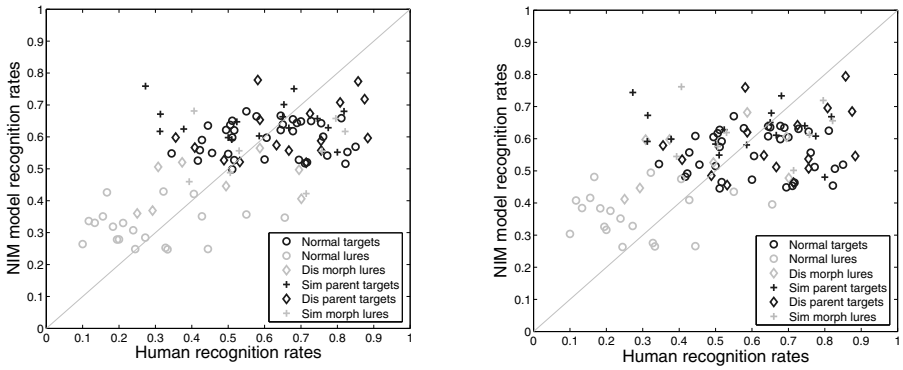


Figure 4.6: The NIM recognition rates as a function of the human recognition rates (Busey and Tunnicliff, 1999). (a) For $r = 3.0$, $RMSE = 0.151$, $R = 0.66$, (b) For $r = 3.2$, $RMSE = 0.164$, $R = 0.58$.

4.2.4 Discussion of the recognition results

The recognition rates produced by NIM agree rather well with experimentally obtained human recognition rates. Below, the main differences and agreements between NIM's recognition results and the experimentally obtained human recognition results will be described. We start by identifying three main behavioural effects that are successfully replicated by NIM. Subsequently, we briefly touch upon NIM's ability to explain the within-class variability of human recognition rates. Finally, we discuss to what extent our results depend on the number and locations of fixations, and on the different spatial scales.

Three replicated effects

The first effect found in the experimental data is that morph lures are falsely recognized more often than normal lures. This effect was successfully produced by the NIM simulations of both the experiments by Busey and Arici (in preparation) and Busey and Tunnicliff (1999).

The second effect is that in the experiment by Busey and Tunnicliff (1999), which employed similar and dissimilar morph lures, false-alarm rates for the similar morph lures were higher than those for the dissimilar morph lures. This effect was also produced successfully by the NIM simulations of the experiment by Busey and Tunnicliff (1999).

The third effect concerns the false-alarm rates for lures compared to the hit rates for their parents. In the behavioural experiments, the false-alarm rate for the morph lures approached the hit rates for their parents. In the experiment by Busey and Tunnicliff (1999) the false-alarm rates even marginally exceeded the hit rates

for the similar lures (and not for the dissimilar lures)[4]. This effect is called the morph-inversion effect (Dailey *et al.*, 1999). As can be seen in Figs. 4.5(a), 4.5(b), and 4.6(a), NIM produced substantially smaller false-alarm rates for the morph lures compared to the hit rates for their parents for $r = 3.0$. By increasing NIM's radius r, a better agreement with experimental findings for the morphs can be obtained. This is at the expense, however, of a decrease in overall correlation and an increase in the $RMSE$s between experimental and model recognition rates. Fig. 4.5(c) presents the results for simulations of the experiment by Busey and Arici (in preparation) for simulation set A, using a radius of $r = 3.3$. Fig. 4.6(b) presents the results for simulations of the experiment by Busey and Tunnicliff (1999) with a radius of $r = 3.2$. Compared to the results in Figs. 4.5(a) and 4.6(a), the false-alarm rates for the morph lures have increased, while the hit rates for their parents have decreased. Obviously, the increase in r selectively benefits the recognition rates for the different types of faces. With the increased radius, the NIM findings agree with the experimental findings by Busey and Tunnicliff (1999); a morph-inversion effect is produced for the similar-morph lures, but not for the dissimilar morph lures. This is a direct consequence of the face similarity structure reflected in NIM's similarity space. With a larger radius r, more stored fixations of the parent faces contribute to the familiarity of the morph lures, in particular when the target parents are highly similar to the morph lures (i.e., their feature-vector representations are close together in the similarity space). Since the feature vectors of lures that are dissimilar to the stored targets are farther apart in the similarity space, the familiarity of these faces is only marginally affected by an increase in the radius r.

The within-class variability

While quite high overall correlations between the NIM and human recognition rates were obtained, there is an important discrepancy between them. The discrepancy concerns the within-class variability of recognition rates (for example, within the class of normal targets). The human recognition rates show noticeably larger within-class variability than the NIM recognition rates. Busey and Tunnicliff (1999) report the same discrepancy between the human data and their SIMSAMPLE model predictions based on an MDS space. They argue that the discrepancy is likely to be due to memory characteristics that operate independently from the overall similarity structure of the faces (e.g., a small striking facial feature, such as a birth mark, may make the face highly memorable, but may be insignificant for the overall similarity of faces). Similarly, a different dependency on the similarity structure of faces might explain why the radius r, for which the correspondence between the model and the human results is optimal, differs for the similarity-rating task and for the recognition task.

[4]One more point should be noted about the third effect. While a morph-inversion effect was obtained for the similar morph lures (and not for the dissimilar morph lures) in the behavioural experiment by Busey and Tunnicliff (1999), the effect was not statistically significant ($\alpha > 0.05$) and the average recognition rates for the similar morph lures only marginally exceeded the average recognition rates for their parents.

Fixations and spatial scales

The correlations between the NIM and human recognition rates appear to be relatively independent of the number of fixations that are selected and stored (i.e., the storage strength). In our original simulations we used a storage strength of 10 fixations, since this approximately corresponds to the number of fixations per face in a face-recognition experiment. In order to test to what extent the results depend on the storage strength, we ran simulations with various storage strengths. For both the experiments by Busey and Arici (in preparation) and by Busey and Tunnicliff (1999), the different storage strengths (varying from 3 to 30 fixations) gave similar results as the simulations with 10 fixations. While, the correspondence between the NIM and the human recognition performance is unaffected by the number of stored fixations, the location of the fixations seems to play an important role. Randomly selecting fixations across the image rather than along the contours significantly reduces correlations between the NIM and the human recognition performance. The reduction mainly results from a substantial decrease in the ability to account for the variation of recognition rates across the different types of lures and across the different types of targets. As for the similarity rating results, we analysed the contribution from visual information at the different spatial scales. The decrease in correlations between model and human recognition rates that results from removing visual information from one scale is approximately equal for the different spatial scales. In contrast, for the similarity-rating task, information from the lower spatial scales seems to play a more important role than fine visual details. Apparently, the amount of visual detail used to make recognition decisions differs from the amount of detail used to make similarity judgements. This complies with Busey and Tunnicliff's (1999) suggestion that striking visual details can play an important role for memory, while being unimportant for judging similarity. Moreover, a number of studies showed that observers attend information at different spatial scales depending on the task (see, e.g., Goffaux *et al.*, 2005; Sowden and Schyns, 2006).

4.2.5 Comparison with other modelling studies

The face-recognition results obtained by Busey and Tunnicliff (1999) have been modelled previously by Busey and Tunnicliff (1999) themselves and by Dailey *et al.* (1999). In order to explain their experimental recognition data, Busey and Tunnicliff (1999) tested several models among which were a recognition version of the GCM model, their SIMSAMPLE model, and the SIMSAMPLE model extended with two different versions of a prototype mechanism. They obtained $RMSE$s of 0.1800, 0.1462, 0.1441, and 0.1411 for the four models, respectively. The $RMSE$s are somewhat smaller than the $RMSE$ of 0.151 obtained with NIM. This is likely to be due to the use of a weighed MDS space based on human similarity ratings instead of using the face images as input. In contrast, Dailey *et al.* (1999) used the face images as input for a principal component analysis and a Gabor-filter preprocessing mechanism in their EMPATH model (see subsection 3.4.2, for a detailed description of EMPATH). The resulting representations were used to test the ability of different memory models to account for the recognition data by Busey and Tunnicliff (1999). Using the Gabor-filter preprocessing method, they obtained $RMSE$s of 0.1624 with a version of the GCM.

4.3 Chapter conclusions

Since NIM operates directly on natural images, it is suitable for making automatic predictions of recognition memorability for individual novel stimuli. In this chapter, NIM was validated on individual natural stimuli. The model was tested on a similarity-rating task and a face-recognition task using the same face images as were used in behavioural studies.

From the results obtained in the similarity-rating task, we may conclude that the NIM similarity ratings correlate reasonably well with the human similarity ratings obtained in the behavioural studies. An additional analysis of the contribution of the visual information from the various spatial scales suggested that coarse visual information plays a more important role than detailed information for judging the similarity of two faces. Moreover, extracting visual information from the contours rather than randomly across the images, produced significantly larger correlations between model and humans.

From the results obtained in the face-recognition task, we may conclude that the NIM recognition rates agree rather well with the behaviourally obtained human recognition rates. NIM successfully replicated the main effects in the human data and absolute deviations from the human recognition rates were relatively small. An additional analysis of the contribution of the visual information from the various spatial scales revealed that visual details play a more important role for making recognition decisions than for making similarity judgements. Moreover, as for the similarity-rating task, the extraction of visual input along the contours rather than randomly across the image produced larger correlations between model and humans.

Overall, we may conclude that NIM quite reliably produces the human similarity ratings and recognition performances for individual natural stimuli.

Chapter 5

Nim-Rem and the recognition-memory effects

This chapter is based on[1]:

1. Lacroix, J. P. W., Murre, J. M. J., Postma, E. O., and van den Herik, H. J. (2004). The natural input memory model. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society (CogSci 2004)* (eds. K. Forbus, D. Gentner, and T. Regier), pp. 773-778, Lawrence Erlbaum Associates, Mahwah, NJ.

2. Lacroix, J. P. W., Postma, E. O., Murre, J. M. J., and van den Herik, H. J. (in preparation). Modelling recognition-memory effects with Nim-Rem.

The modelling studies presented in chapter 4 revealed that Nim quite reliably predicts behaviourally obtained similarity ratings and recognition performances for *individual* natural stimuli (i.e., face images). In this chapter, we investigate Nim's natural input recognition properties for *general* recognition-memory effects (see, e.g., Lacroix *et al.*, 2004; Lacroix *et al.*, in preparation). In order to do so, we introduce a variant of the Nim model, called Nim-Rem, which realizes a natural input version of the powerful Retrieving Effectively from Memory model (Rem) proposed by Shiffrin and Steyvers (1997). Nim-Rem generalizes Rem in such a way that it can operate directly on natural input. The validity of a memory model is generally assessed by testing the model's ability to replicate a number of well-known recognition-memory effects that are found in behavioural recognition-memory studies. In a similar way, we examine Nim-Rem's ability to replicate the recognition-memory effects found in behavioural experiments. The original Rem successfully produced the behaviourally found recognition-memory effects using a representation space based on oversimplified statistical assumptions about the features of natural input (e.g., Shiffrin and Steyvers, 1997; also, see subsection 2.1.1). Below, we examine to what extent Nim-

[1] The author would like to thank her co-authors and the publisher of the CogSci 2004 proceedings for their kind permission to reuse relevant parts of the article in this thesis.

REM produces these effects using a representation space that is derived directly from natural visual input.

The chapter is organized as follows. First, we discuss NIM-REM in section 5.1. This is followed by a brief description of the behavioural recognition-memory experiments and the paradigms that they used to study recognition memory. Then, sections 5.3, 5.4, 5.5, and 5.6 address four recognition-memory effects. For each of the recognition-memory effects, we present the behavioural findings, the simulations with NIM-REM, the simulation results, and a discussion of the results. Subsequently, in section 5.7, NIM-REM is compared with dual-process models. Moreover, the relation between faces and words as experimental items is discussed. Finally, section 5.8 gives our chapter summary.

5.1 NIM-REM

Below, the NIM-REM essentials are described. Since the preprocessing stage is identical to the one described in chapter 3, we here consider only the memory stage. As mentioned in chapter 3, we distinguish two processes in the memory stage: the storage process (5.1.1) and the recognition process (5.1.2).

5.1.1 The storage process

As in the original NIM model, the storage process in NIM-REM stores preprocessed samples of natural images (i.e, fixations). Both NIM and NIM-REM store s fixations for an image per unit of presentation time (where s represents the storage strength, i.e., the number of stored fixations). The original NIM stores fixations without labels not explicitly distinguishing between fixations coming from different items. However, REM handles representations coming from different items separately during recognition by means of one separate feature vector for each stored item (Shiffrin and Steyvers, 1997). Via a slight adaptation of this idea, NIM-REM is enabled to store labelled feature vectors. As a result, an image is represented by the set of stored feature vectors that are similarly labelled.

REM makes two main assumptions about the stored representations. Below we present the assumptions and show how we adhere to these assumptions in NIM-REM.

REM's two main representational assumptions are: (1) representations are incomplete copies of the input, and (2) representations are error-prone copies of the input (Shiffrin and Steyvers, 1997). The representations that result from NIM's preprocessing stage (adopted by NIM-REM) automatically adhere to the first assumption. NIM's preprocessing stage is based on samples (fixations) of the input. Each sample represents a fragment of the original input. Therefore, NIM (and NIM-REM) has the characteristic of incomplete storage. The representations that result from NIM's preprocessing stage do not automatically seem to adhere to REM's second assumption of error-proneness. Below, we will argue that NIM (and NIM-REM) does in fact incorporate a mechanism that generates errors. In NIM, REM's rather artificial assumption of error-proneness is replaced by a more natural assumption of noisy sampling of natural input. We assume that this noisy sampling has effects that are comparable to the artificial error-proneness characteristic. For natural images it is

known that displacements yield errors that are inversely proportional to the size of the displacement. More importantly, the errors are larger for locations typically fixated by human observers viewing natural images (Reinagel and Zador, 1999). Since NIM and NIM-REM are likely to select 'human-like' fixation locations, we assume that fixational displacements yield samples that are significantly different. As a consequence of the variations in fixation locations, different encounters with the same image give rise to different representations.

5.1.2 The recognition process

NIM-REM fully adopts REM's recognition process. Below, we start by describing REM's recognition process. Subsequently, we describe how NIM-REM adopts REM's recognition process.

In REM, the familiarity of a test item is determined by performing a Bayesian comparison of all stored representations to the test representation in terms of likelihood ratios. For each of the stored representations, the likelihood ratio expresses the probability that it represents the test item (i.e., the test item is old) as opposed to that it does not represent the test item (i.e., the test item is new). The overall odds in favour of an old over a new test item are obtained by averaging likelihood ratios across all stored representations. These overall odds are used to produce a recognition decision (see, e.g., Shiffrin and Steyvers, 1997; Steyvers, 2000). In REM the overall odds are defined as:

$$\phi = \frac{1}{n} \sum_{J=1}^{n} \lambda_J, \tag{5.1}$$

where n represents the number of stored representations and λ_J is the likelihood for the J-th stored representation that is defined as the ratio of the probability that it represents the test item (i.e., a 'same' (S) judgement) over the probability that it represents an item different from the test item (i.e., a 'different' (D) judgement) (Steyvers, 2000). Formally, the likelihood is defined as:

$$\lambda_J = \frac{P(S)}{P(D)}, \tag{5.2}$$

where S and D represent 'same' and 'different' judgements, respectively. The calculation of each likelihood ratio λ follows directly from a Bayesian principle that uses the available information to determine the match between the test item and the stored representation (for details, see Shiffrin and Steyvers, 1997).

NIM-REM employs a similar method. As REM, NIM-REM calculates the likelihood ratio λ for each of the stored representations and determines the overall odds ϕ in favour of a target image over a lure image by averaging over the likelihood ratios. In order to apply REM's recognition process to the representations produced by NIM-REM, we need to redefine the calculation of the probabilities $P(S)$ and $P(D)$ for NIM-REM. Our Bayesian calculations deviate somewhat from the approach followed by REM, because NIM-REM employs continuous valued features, rather than discrete valued features (as employed by REM).

In NIM-REM, the probabilities $P(S)$ and $P(D)$ are related directly to the 'familiarity' of a test image as defined originally for NIM (see equation 3.1 and 3.2). As noted by Barrington, Marks, and Cottrell (2007), the familiarity calculation can be interpreted as a kernel density estimate that centres a Parzen window at each stored feature vector (see, e.g., Duda, Hart, and Stork, 2001). The kernels are hyperspheres with radius r and uniform density functions. For each test fixation, the Parzen window computes the probability density of the extracted feature vector. Since NIM originally stored unlabelled feature vectors that were handled similarly as exemplars of the class 'old', the sum of the probability densities formed the estimation of the probability that the test image was 'old'. In contrast, NIM-REM stores labelled feature vectors and handles each stored representation (consisting of the set of similarly labelled feature vectors) separately. Therefore, NIM-REM sums probability densities over the set of (similarly labelled) feature vectors of a stored representation J, in order to estimate the probability $P(S)$ that J represents the test image. In the same way, $P(D)$ can be estimated by computing the probability densities of a set of feature vectors that represent images different from J (in our simulations we extracted 1,000 feature vectors from each of the images different from the stored representation J).

As in the original REM (cf., Shiffrin and Steyvers, 1997), NIM-REM produces a positive recognition decision when the overall odds ϕ in favour of a target over a lure exceed the value 1.0; it produces a negative recognition decision otherwise. Thus, by comparing the obtained ϕ values (that were first normalized for each list type by dividing them by the mean ϕ values of targets and lures of the specific list type) to a fixed recognition decision criterion of $c = 1.0$, recognition rates (i.e., hit rates and false-alarm rates) are obtained.

Some of the recognition-memory effects addressed in this chapter are expressed in terms of overall recognition-score patterns rather than in terms of hit rate and false-alarm rate patterns. Therefore, in addition to the calculation of the recognition rates (i.e., the hits and false alarms), overall recognition scores are calculated for each list. The recognition scores are calculated directly from the overall odds values ϕ in a signal detection analysis. The signal-detection calculation of a recognition score is based on the difference between the average ϕ value for the targets and the average ϕ value for the lures. When the standard deviations of the two ϕ-value distributions are equal, recognition performance, d', is generally expressed as the difference between the average ϕ value for the targets, $\overline{\phi_T}$, and the average ϕ value for the lures, $\overline{\phi_L}$, normalized by the standard deviation of the distribution of the ϕ values for the lures:

$$d' = \frac{\overline{\phi_T} - \overline{\phi_L}}{\sigma_{(\phi_L)}}. \tag{5.3}$$

However, in the more general case (when it is unknown whether the standard deviations of the two distributions are equal), the appropriate measure for the recognition score is expressed as (cf. Simpson and Fitter, 1973; Swets, 1986a, 1986b):

$$d_a = \frac{\overline{\phi_T} - \overline{\phi_L}}{\sqrt{\frac{\sigma^2_{(\phi_T)} + \sigma^2_{(\phi_L)}}{2}}}. \tag{5.4}$$

We used the more general measure, d_a, to calculate the recognition scores from the obtained ϕ values.

We based calculations of each d_a value on the ϕ_T values and the ϕ_L values of 100 single recognition simulations. For each list type we ran 100,000 single recognition simulations.

5.2 Behavioural experiments

In order to reveal the workings of the different episodic recognition-memory processes, numerous behavioural experiments have been conducted. The various experiments employ a straightforward list paradigm in which subjects are presented with a study list of items and are tested for their old-new recognition memory of the studied items (targets) and a number of non-studied items (lures) (see, e.g., the studies described in subsection 4.2.1). The behavioural recognition-memory experiments perform various list manipulations to examine the effect on recognition memory, e.g., strengthening (some of) the items on the study list (see the list-strength effect in subsection 5.2.1 and the item-strength effect in subsection 5.2.3), lengthening the study list (see the list-length effect in subsection 5.2.2), and increasing the similarity of targets and lures (see the false-memory effect in subsection 5.2.4). Examples of strengthening items are: increasing the amount of study time for the item and increasing the number of times that the item occurs on the study list (see, e.g., Ratcliff, Clark, and Shiffrin, 1990; McClelland and Chappell, 1998; Norman, 2002). Lengthening of the study list is performed by increasing the number of items on the list (e.g., Ratcliff *et al.*, 1990; Ohrt and Gronlund, 1999; Cary and Reder, 2003). Increasing the similarity between targets and lures is achieved straightforwardly by using lures on the test list that are highly similar to the studied targets.

In our investigations, we address four recognition-memory effects that have been found consistently across the behavioural recognition-memory experiments. Therefore, these are now considered to be robust effects (see, e.g., Ratcliff *et al.*, 1990; Murnane and Shiffrin, 1991; Yonelinas, Hockley, and Murdock, 1992; Roediger and McDermott, 1995; McClelland and Chappell, 1998; Stretch and Wixted, 1998; Busey and Tunnicliff, 1999; Cabeza *et al.*, 1999; Ohrt and Gronlund, 1999; Stewart and McAllister, 2001; MacAndrew *et al.*, 2002; Norman, 2002; Roark, O'Toole, and Abdi, 2003; Dewhurst and Farrand, 2004; Lamont, Stewart-Williams, and Podd, 2005; Nega, 2005). The four effects are: (1) the (null/negative) list-strength effect, (2) the list-length effect, (3) the item-strength effect, and (4) the false-memory effect. Subsections 5.2.1, 5.2.2, 5.2.3, and 5.2.4 provide the descriptions of the four effects and the behavioural paradigms employed to examine them.

5.2.1 The list-strength effect

The list-strength effect depicts the effect of strengthening a subset of the items of the study list on recognition memory for the non-strengthened items.

In order to test for a list-strength effect, many list-strength studies have used the mixed/pure paradigm, first proposed by Ratcliff *et al.* (1990). In this paradigm three types of lists serve as study lists: pure weak lists (with weak items only), pure strong

lists (with strong items only), and mixed lists (with both strong and weak items). Weak and strong (i.e., strengthened) items differ in the amount of presentation time on the study list or the number of times the item appears on the list. A list-strength effect is then said to occur when the strengthened items on the study list (i.e., the strong items) would suppress recognition of the non-strengthened items (the weak items), which entails that recognition memory for weak items is better in the pure weak list than in the mixed list (Ratcliff *et al.*, 1990). Likewise, a list-strength effect entails that memory for strong items is better in the mixed list than in the pure strong list.

A somewhat simpler paradigm has also been used for testing the list-strength effect (Norman, 2002). The simpler paradigm employs two types of study lists: pure weak lists and mixed lists. A list-strength effect is then said to occur when recognition memory for weak items is better in the pure weak list than in the mixed list (Norman, 2002).

5.2.2 The list-length effect

The list-length effect represents the effect of the lengthening of the study list on the recognition memory for the study list items.

In order to test for a list-length effect, behavioural studies use a paradigm that employs two (or more) types of study lists that differ in their lengths (i.e., the number of items on the list): (1) short lists, and (2) long lists. A list-length effect is then said to occur when recognition memory for the items on a short list is better than recognition memory for the items on a long list. The recognition-memory studies examine the list-length effect either in terms of an overall performance measure, such as the performance measure d_a, or examine it in more detail by analysing the hit and false-alarm patterns. Often the specific pattern of a decrease in the hit rate and an increase in the false-alarm rate appears to underlie the list-length effect (see, e.g., Cary and Reder, 2003). The expression of the list-length effect in terms of the opposing effect on the hit and false-alarm patterns is referred to as the length-mirror effect (see, e.g., Cary and Reder, 2003).

5.2.3 The item-strength effect

The item-strength effect reflects the effect of strengthening the items of the study list on the recognition memory for these items.

In order to test for an item-strength effect, behavioural studies use a paradigm that employs two (or more) types of study lists that differ in the strength of items on the list: (1) weak lists (with weak items only), and (2) strong lists (with strong items only). Weak and strong items differ in the presentation time of an item or the number of times the item occurs on the list. An item-strength effect is then said to occur when recognition memory for the items on the strong list is better than recognition memory for items on the weak list. As the list-length effect, the item-strength effect can be examined either in terms of an overall performance measure or, in terms of a detailed analysis of the hit and false-alarm patterns. When the specific pattern of an increase in the hit rate and a decrease in the false-alarm rate underlies the

item-strength effect, this is referred to as a strength-mirror effect (e.g., McClelland and Chappell, 1998; Stretch and Wixted, 1998; Cary and Reder, 2003; Nega, 2005).

5.2.4 The false-memory effect

The false-memory effect portrays the effect of the similarity of a lure to (one of) the studied items on the probability of falsely recognizing the lure.

In order to test for a false-memory effect, behavioural studies use a paradigm that employs two types of lures on the test lists that differ in their similarity to the targets on the study list: (1) dissimilar lures, and (2) similar lures. A false-memory effect is then said to occur when similar lures are recognized falsely more often than dissimilar lures.

5.2.5 The behavioural experiments and the simulations with Nim-Rem

In our model simulations we employed the same paradigms as those used in the behavioural experiments. We used the set of 104 face images that was introduced in chapter 4. The set consisted of grey-scale images of bald male faces without glasses (see subsection 4.1.1). Nim-Rem was repeatedly first provided with a study list of images and subsequently tested for old-new recognition of a number of the studied images (i.e., targets) plus a number of non-studied images (i.e., lures). In the following four sections, the behavioural findings on the four recognition-memory effects are discussed along with theNim-Rem simulation studies that test Nim-Rem on the four effects.

5.3 List-strength studies

A list-strength effect is defined as a decrease in recognition-memory performance for a given subset of study-list items when other study-list items are strengthened (Ratcliff *et al.*, 1990).

In this section, we discuss well-known findings from behavioural recognition-memory studies on the list-strength effect (5.3.1). Subsequently, we present the Nim-Rem simulations by which we tested Nim-Rem's ability to replicate the behavioural list-strength findings (5.3.2). Then, we present the list-strength simulation results (5.3.3). Finally, we discuss the predictions of other models on the list-strength effect and outline how the list-strength effect varies with the similarity of the items (5.3.4).

5.3.1 Behavioural list-strength findings

Several studies showed that strengthening some items on the study list did not deteriorate recognition performance for the non-strengthened items (i.e., a null list-strength effect) or even significantly increased the performance for the non-strengthened items (i.e., a negative list-strength effect) (e.g., Shiffrin, Ratcliff, and Clark, 1990a; Ratcliff *et al.*, 1990; Murnane and Shiffrin, 1991; Yonelinas *et al.*, 1992). These results contradicted the expectations of a positive list-strength effect

for recognition memory for three reasons. First, positive list-strength effects were obtained for free recall, which led to the expectation that positive list-strength effects would be obtained for recognition memory also; in contrast, null and negative list-strength effect were obtained. Second, several studies showed that lengthening the study list by adding new items to the study list deteriorated recognition memory for the items on the list (i.e., a list-length effect, see subsection 5.4.1 that describes results on the list-length effect), which led to the expectation that lengthening the study list by repeating items on the study list (i.e., strengthening these items) would deteriorate recognition memory for the non-strengthened items (in a similar way as recognition memory for items deteriorated when items are added to the list); this is not what happens. Third, the null list-strength findings and the negative list-strength findings contradicted the predictions of the major existing memory models of that time (for a detailed outline, see, Shiffrin, R. Ratcliff, and Clark, 1990b). To adhere to the list-strength findings, new models were developed among which was REM that was discussed in the previous section (Shiffrin and Steyvers, 1997).

The list-strength results, which contradicted the expectations, seem to reflect robust effects; the null and the negative list-strength effect have been obtained in many studies (e.g., Ratcliff *et al.*, 1990; Murnane and Shiffrin, 1991; Yonelinas *et al.*, 1992). In an elaborate study Ratcliff *et al.* (1990) tested for a list-strength effect for recognition in seven experiments employing the mixed/pure paradigm described in subsection 5.2.1. They obtained null-list-strength effects in four of the experiments, and significantly negative list-strength effects in three of the experiments. Where non-significant (i.e., null) list-strength effects were obtained, the effects were mostly slightly negative. Since the experiments by Ratcliff *et al.* (1990), several list-strength studies have failed to obtain a list-strength effect for recognition (e.g., Murnane and Shiffrin, 1991; Yonelinas *et al.*, 1992).

Taken together, the list-strength literature shows that a list-strength effect does not occur for recognition and often a negative list-strength effect is observed.

5.3.2 List-strength simulations with NIM-REM

In our simulations, we investigated the list-strength effect. In particular, we tested to what extent NIM-REM is able to predict a null list-strength effect or even a negative list-strength effect as obtained in the behavioural experiments (e.g., Ratcliff *et al.*, 1990). To assess the robustness of the effect, we tested for the list-strength effect across several list-length levels (i.e., study lists that differ in the number of images on the list).

Below we describe (1) the paradigm that was employed for assessing the list-strength effect, and (2) the experimental procedure.

The paradigm

For assessing the occurrence of a list-strength effect, we employed the paradigm proposed by Norman (2002) as mentioned in subsection 5.2.1. This paradigm uses two types of study lists: weak lists and mixed lists. Both types of study lists contain N targets. On the weak lists all targets are presented once. On the mixed lists half of the targets are presented once and half of the targets are presented twice (i.e., they

are strengthened). The strengthened targets serve to manipulate the list strength and are not tested for recognition. A list-strength effect is said to occur when the recognition score for weak lists is higher than the recognition score for mixed lists. As mentioned previously, list-strength effects were assessed across different list-length levels (i.e., we employed short lists, long lists, and extra long lists). The *Cohen's d* effect size estimate (Cohen, 1988) was used to quantify the difference in recognition scores for weak lists and mixed lists. The effect size is calculated as follows:

$$Cohen's \;\; d = \frac{\overline{d_a(weak)} - \overline{d_a(mixed)}}{\sqrt{\frac{\sigma^2_{d_a(weak)} + \sigma^2_{d_a(mixed)}}{2}}} \tag{5.5}$$

with $d_a(weak)$ and $d_a(mixed)$ the recognition scores for the weak and mixed lists, respectively. The effect size is a measure for the degree to which a list-strength effect has occurred.

The experimental procedure

For each simulation run, NIM-REM was presented with a study list of targets. Then, old-new recognition was tested for a (non-strengthened) target and a lure. Targets and lures were randomly selected from the set of face images. We presented NIM-REM with: (1) weak study lists, and (2) mixed study lists. On the weak lists all targets were stored with storage strength $S = 10$ (i.e., 10 feature vectors were stored, corresponding to 10 fixations). On the mixed lists, half of the targets were stored with storage strength $S = 10$ (i.e., the non-strengthened targets) and half of the targets were stored with storage strength $S = 20$ (i.e., the strengthened targets that were shown twice). NIM-REM was tested on study lists of different lengths N: (1) short lists ($N = 8$), (2) long lists ($N = 12$), and (3) extra long lists ($N = 18$)[2]. Recognition tests were performed using the radius parameter $r = 5.0$.

5.3.3 List-strength simulation results

Table 5.1 presents the results. The rows show the results for the weak and mixed lists of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$). Since the list-strength effect is defined in terms of an overall recognition score rather than in terms of hit and false-alarm patterns, we present the d_a recognition scores. Also, the size of the obtained list-strength effects are presented. Overall, recognition scores for mixed lists are higher than those for weak lists. This indicates that negative list-strength effects occurred[3].

[2]List lengths were chosen in such a way that the results would reveal the difference between the effect on recognition memory of lengthening the study list by repeating items on the study list (i.e., strengthening these items) and the effect on recognition memory of lengthening the study list by adding new items to the study list. Therefore, the mixed short study lists contained an equal number of images (i.e., 4 non-strengthened targets and 4 strengthened targets (shown twice) which added up to 12 targets) as the weak long list and the mixed long list contained an equal number of images (i.e., 6 non-strengthened targets and 6 strengthened targets (shown twice) which added up to 18 targets) as the weak extra long list.

[3]We would like to note that, while lengthening the study list by repeating some of the targets on the study list improved recognition memory (i.e., a worse performance for weak lists compared

	weak	*mixed*	effect size
	d_a	d_a	*Cohen's d*
short ($N = 8$)	0.93	1.02	0.60
long ($N = 12$)	0.77	0.86	0.63
extra long ($N = 18$)	0.64	0.70	0.41

Table 5.1: Recognition scores (d_a) for weak lists and mixed lists of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$). *Cohen's d* effect sizes quantify the obtained list-strength effects.

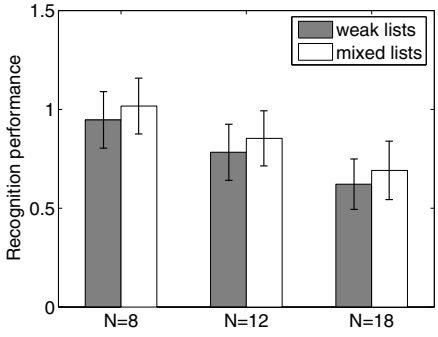


Figure 5.1: The list-strength effect. Recognition performance, d_a, for weak lists (grey bars) and mixed lists (white bars) for lists of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$). The error bars indicate the standard deviations of the recognition scores.

Fig. 5.1 shows bar graphs of the recognition scores for lists of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$), for weak lists (grey bars) and mixed lists (white bars). The error bars indicate the standard deviations of the recognition scores. ANOVAs support the statistical significance of the negative list-strength effects for short lists ($N = 8$), $F(1, 1999) = 182.54$, $p < 0.05$, long lists ($N = 12$), $F(1, 1999) = 195.67$, $p < 0.05$, and extra long lists ($N = 18$), $F(1, 1999) = 83.81$, $p < 0.05$.

It should be noted that for d_a to be a suitable performance measure, the target and lure familiarity values should be distributed normally. Fig. 5.2 presents

to that for mixed lists), lengthening the study list by adding targets to the study list decreased recognition performance (i.e., a worse performance for long lists compared to that for short lists and a worse performance for extra long lists compared to that for long lists). See also section 5.4 about the list-length effect.

familiarity-value distributions for targets and lures on weak and mixed lists for short lists ($N = 8$) and extra long lists ($N = 18$). From Fig. 5.2, it can be seen that familiarity-value distributions become more normally-distributed when the study list is lengthened. For short lists ($N = 8$) (corresponding to a relatively small number of stored fixations), familiarity values were slightly skewed, due to their truncation at a familiarity value of zero (see Fig. 5.2(b) and (d)). The truncation somewhat distorts the standard deviations and therefore the d_a values. The magnitude of the distortion depends on the skewedness of the distribution. However, we believe that NIM-REM reliably produces a negative list-strength effect, because the simulations for longer study lists produce a similar pattern of results based on more normally distributed familiarity values.

5.3.4 Discussion of the list-strength results

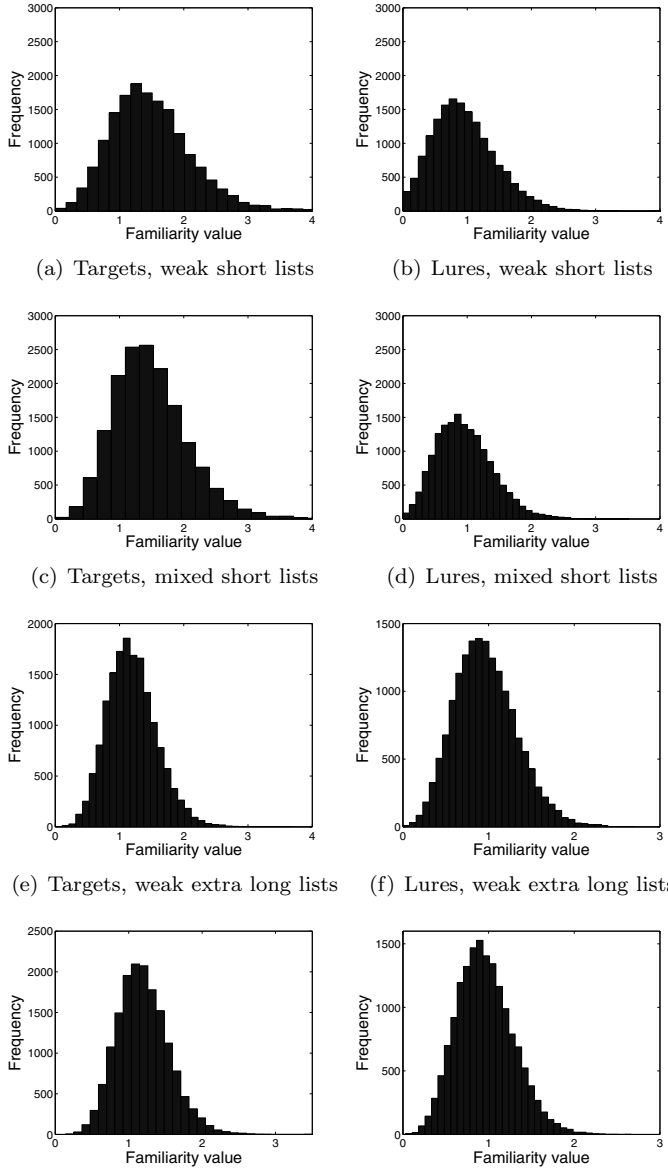
Below we clarify why some models predict a null or even negative list-strength effect, whereas others fail to do so. Then, we discuss how similarity affects the size of the list-strength effect.

List-strength predictions of memory models

At the time Ratcliff *et al.* (1990) published their results on the null and negative list-strength effect, the major existing memory models (e.g., the SAM model (Raaijmakers and Shiffrin, 1981; Gillund and Shiffrin, 1984), the TODAM model (Murdock, 1982), the CHARM model (Eich, 1982; Eich, 1985), the MATRIX model (Pike, 1984), the MINERVA2 model (Hintzman, 1986)) were unable to predict a negative list-strength effect. The models had the characteristic that strengthening some items on the study list had a similar qualitative effect as adding new items to the study list. Therefore, the models predicted positive list-strength effects (in a similar way as they predicted positive list-length effects). Shiffrin *et al.* (1990a) demonstrated that for a model to accommodate a null or negative list-strength effect, the model should incorporate the characteristic that when an item is strengthened it becomes less confusable with a test item that is different from the stored item. This phenomenon is called *differentiation* (McClelland and Chappell, 1998). The original REM produced an effect of differentiation, because more features are stored for strengthened items, which makes their representations less confusable with test words other than the strengthened item. In a similar way, NIM-REM stores more feature vectors for an image when it is strengthened, which makes the feature-vector representation more refined and less confusable with test images other than the strengthened image. As a consequence of the refinement of the representations, the variability in the familiarity values for targets and lures decreases, which results in higher recognition scores.

Similarity and the list-strength effect

While in most studies a null list-strength effect or a negative list-strength effect was obtained for recognition, in a more recent study it was shown that a positive list-strength effect can be obtained when there is a high degree of similarity between



(a) Targets, weak short lists (b) Lures, weak short lists

(c) Targets, mixed short lists (d) Lures, mixed short lists

(e) Targets, weak extra long lists (f) Lures, weak extra long lists

(g) Targets, mixed extra long lists (h) Lures, mixed extra long lists

Figure 5.2: Target and lure familiarity-value distributions for weak and mixed lists for short lists ($N = 8$) and extra long lists ($N = 18$) based on 50,000 familiarity values.

the items on the study and the test list (Norman, 2002). Norman adheres to a dual-processing approach to explain his results, which assumes that two processes underlie recognition: (1) a familiarity process, i.e., a context-insensitive automatic process that determines the familiarity of an item, and (2) a recollection process, i.e., a context-sensitive strategic process that recollects specific studied details (Norman, 2002; Norman and O'Reilly, 2003). Consistent with the fact that a list-strength effect is found for recall and not for recognition (Ratcliff *et al.*, 1990), Norman (2002) argues that the magnitude of the list-strength effect depends on the extent to which recollection drives recognition. In a few experiments, Norman (2002) manipulated the degree to which subjects had to rely on recollection to make recognition decisions by varying the similarity between the items on the study list and the test list. Norman (2002) argued that when the items on the test list are highly similar to the items on the study list, all test items produce strong ratings of familiarity. Therefore, the subject will rely more on recollection of specific details when using similar study items (i.e., targets) and test items (i.e., lures) than when using dissimilar targets and lures. As a consequence, a larger list-strength effect will be obtained for similar targets and lures than for dissimilar targets and lures. His results supported the prediction that the size of the list-strength effect varies with similarity; while he found a null list-strength effect using similar targets and lures, a positive list-strength effect was obtained for similar study and test items.

In order to test whether Nim-Rem is able to predict that the similarity between studied images (i.e., targets) and non-studied images (i.e., lures) affects the size of the list-strength effect, we ran an additional simulation in which list-strength effects were assessed for targets vs. dissimilar lures (TD) and for targets vs. similar lures (TS). We presented the model with weak and mixed study lists of length $N = 8$. After the last image of the study list had been stored, three images were presented for recognition: (1) a (non-strengthened) target, (2) a lure dissimilar to the targets, and (3) a lure similar to one of the (non-strengthened) targets. Similarity was based on the similarity ratings that were obtained in the studies of chapter 4 (also see subsection 5.6.2 which explains in more detail when two images are considered to be dissimilar or similar). Analogously to the behavioural results obtained by Norman (2002), the Nim-Rem results showed that similarity affects the size of the list-strength effect; significantly smaller list-strength effects were obtained for TD than for TS. However, Nim-Rem produced negative and slightly negative list-strength effects for TD and TS, respectively, while Norman (2002) obtained a null list-strength effect and a positive list-strength effect for TD and TS, respectively. In its current form Nim-Rem is unable to produce positive list-strength effects since it produces an effect of differentiation. Strengthening some of the targets will make the stored representations less confusable, which will benefit recognition. While the beneficial effect is less pronounced and can even be a null effect when target-lure similarity increases, it will not have a detrimental effect on recognition. When the positive list-strength effects for similar targets and lures that were obtained by Norman (2002) turn out to be robust recognition-memory effects, model adjustments should be made to accommodate the findings.

5.4 List-length studies

A list-length effect is defined as a decrease in memory performance for the study-list items when the study list is lengthened (Ratcliff *et al.*, 1990).

In this section, we discuss well-known findings from behavioural recognition-memory studies on the list-length effect (5.4.1). Then, we present the NIM-REM simulations by which we tested NIM-REM's ability to replicate the behavioural findings (5.4.2). Subsequently, we present the list-length results (5.4.3). Finally, we discuss why NIM-REM produces a list-length effect (5.4.4).

5.4.1 Behavioural list-length findings

The list-length effect has often been studied for recognition memory. Although many studies that tested for the effects did not control the variables that normally vary with the list length (e.g., the retention interval, the loss of attention, the displaced rehearsal, the failure of contextual reinstatement, and the recognition load), a number of studies have been conducted that controlled one or more of these confounding variables. While some of the controlled studies failed to find a list-length effect (e.g., Dennis and Humphreys, 2001), most did obtain it (e.g., Ohrt and Gronlund, 1999; Cary and Reder, 2003). List-length studies that analysed the list-length effect in terms of the hit and false-alarm patterns, revealed that subjects produce more hits and fewer false alarms for shorter lists than for longer lists (see, e.g., Ratcliff *et al.*, 1990; Cary and Reder, 2003). In other words, a length-mirror effect (i.e., an opposing effect on the hit and false-alarm patterns) underlies the list-length effect. Generally, the list-length effect is considered a well-established effect in the recognition-memory literature (e.g., Podd, 1990; Ratcliff *et al.*, 1990; Ohrt and Gronlund, 1999; MacAndrew *et al.*, 2002; Metzger, 2002; Cary and Reder, 2003; Lamont *et al.*, 2005).

5.4.2 List-length simulations with NIM-REM

Our list-length simulation tested NIM-REM's ability to produce the list-length effect. The list-length effect was analysed both in terms of the overall performance measure, and in terms of the hit and false-alarm patterns (to reveal the occurrence of a length-mirror effect, see subsection 5.2.2). In order to assess the robustness of the effect, we tested for the list-length effect across several item-strength levels (i.e., across study lists that differ in the strength of the images on the list). In the same way, subsection 5.5.2 will address NIM-REM's ability to produce the item-strength effect across several list-length levels (i.e., across study lists that differ in the number of images on the list). Considering our aim to test for a list-length effect at different item-strength levels and to test for an item-strength effect at different list-length levels, we decided to merge the tests into one design employing nine study list types of three different list-length levels and three different item-strength levels.

Below we describe (1) the paradigm that was employed for assessing the list-length effect, and (2) the experimental procedure.

The paradigm

Although the second simulation employed a design that was suitable for assessing the list-length effect as well as the the item-strength effect, here we focus on the paradigm that underlies the assessment of a list-length effect. The paradigm uses three types of lists that differ in their lengths N: (1) short lists, (2) long lists, and (3) extra long lists. As mentioned previously, these list types were presented at three different item-strength levels (i.e., weak, strong, and extra strong, see subsection 5.5.2). A list-length effect is said to occur if the recognition score decreases when the study list is lengthened. NIM-REM recognition scores were compared for the short, long, and extra long lists (at each item-strength level) in order to assess the occurrence of a list-length effect. Moreover, we analysed the hit and false-alarm patterns across the short, long, and extra long lists, to reveal length-mirror effects. To quantify the effect sizes, we again used *Cohen's d* effect size estimate.

The experimental procedure

Again (see subsection 5.3.2), each simulation run, a study list of targets was presented to NIM-REM and subsequently NIM-REM was tested for old-new recognition of a target and a lure. Targets and lures were randomly selected from the set of face images. Since this simulation was designed in such a way that it addressed the list-length effect and the item-strength effect, the length of the study lists as well as the strength of the images on the study lists were varied.

We presented to NIM-REM: (1) short lists ($N = 8$), (2) long lists ($N = 12$), and (3) extra long lists ($N = 18$), at three different item-strength levels: (1) weak, (2) strong, and (3) extra strong. Targets on the weak, strong, and extra strong lists were stored with storage strengths $S = 10$ (i.e., 10 feature vectors were stored, corresponding to 10 fixations), $S = 20$, and $S = 30$, respectively. After the last image of a study list had been stored, one target along with one lure were presented for recognition. As in the first simulation, recognition tests were performed using the radius parameter $r = 5.0$. As mentioned in subsection 5.1.2, the recognition decision criterion was set to $c = 1.0$. Small variations in the recognition decision criterion were not critical to the simulation results.

5.4.3 List-length simulation results

Table 5.2 presents the results. The table shows the recognition results for weak ($S = 10$), strong ($S = 20$), and extra strong ($S = 30$) lists (columns) for lists of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$) (rows). The results in Table 5.2 demonstrate that NIM-REM produces a list-length effect; recognition scores (d_a) decrease with the length, N, of the study list. ANOVAs show that the list-length effects are significant for weak lists, $F(2, 1499) = 584.09$, $p < 0.05$, strong lists, $F(2, 1499) = 1103.07$, $p < 0.05$, and extra strong lists $F(2, 1499) = 1772.95$, $p < 0.05$. The values presented in bold between the rows in Table 5.2 indicate the absolute sizes of the list-length effects. Moreover, the results in Table 5.2 show length-mirror effects: the hit rate for longer lists is lower than the

	weak ($S = 10$)			strong ($S = 20$)			extra strong ($S = 30$)		
	d_a	H	FA	d_a	H	FA	d_a	H	FA
short ($N = 8$)	0.94	0.63	0.27	1.22	0.70	0.24	1.41	0.74	0.22
	1.17			**1.51**			**1.95**		
long ($N = 12$)	0.78	0.61	0.31	0.99	0.67	0.29	1.11	0.70	0.28
	1.03			**1.43**			**1.82**		
extra long ($N = 18$)	0.63	0.59	0.35	0.77	0.63	0.33	0.84	0.66	0.32

Table 5.2: Recognition scores (d_a), hit rates (H), and false-alarm rates (FA), for weak ($S = 10$), strong ($S = 20$), and extra strong ($S = 30$) lists of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$). The absolute list-length effect sizes are presented in bold between the rows.

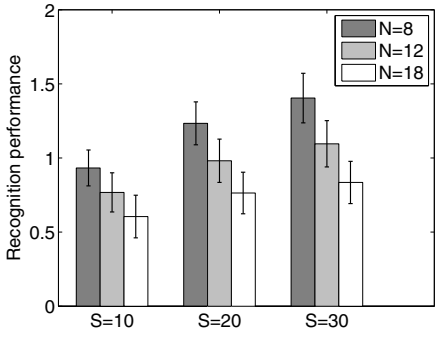


Figure 5.3: The list-length effect. Recognition scores, d_a, for (a) weak lists, (b) strong lists, and (c) extra strong lists of different lengths: short lists ($N = 8$; dark-grey bars), long lists ($N = 12$; light-grey bars), and extra long lists ($N = 18$; white bars). Error bars indicate the standard deviation of the recognition scores.

hit rate for shorter lists and the false alarm rate for longer lists is higher than the false alarm rate for shorter lists.

Fig. 5.3 presents bar graphs of the recognition scores for (a) weak lists, (b) strong lists, and (c) extra strong lists of different lengths: short lists ($N = 8$; dark-grey bars), long lists ($N = 12$; light-grey bars), and extra long lists ($N = 18$; white bars). Again, the error bars indicate the standard deviation of the recognition scores.

5.4.4 Discussion of the list-length results

Nim-Rem produces a list-length effect because storing more images decreases the average match between a tested target and the stored representations. Overall, a relatively high familiarity value is found for a test target and the previously stored representation of that test target; also, a relatively low match is found between a test target and a previously stored representation of a different image. Therefore, when the familiarity value is based on more previously stored representations (of images different from the tested target), the average match across the previously stored representations decreases for the test target. As a consequence, the difference between the average familiarity value for targets and that for lures decreases, which causes the recognition score to decrease. At the same time, lengthening the study list causes the variability in the familiarity values for targets and lures to decrease, which leads to an increase in the d_a values. Nevertheless, the decrease in the difference between the average familiarity values for targets and lures amply outweighs the decreases in variabilities of familiarity values for targets and lures. Therefore, overall, recognition scores d_a decreased when the study list was lengthened.

5.5 Item-strength studies

An item-strength effect is defined as an increase in memory performance for the study-list items when the items are strengthened (Ratcliff *et al.*, 1990).

In this section, we discuss well-known findings from behavioural recognition-memory studies on the item-strength effect (5.5.1). Then, we present the Nim-Rem simulations by which we tested Nim-Rem's ability to replicate the behavioural findings (5.5.2). After that the item-strength results are presented (5.5.3). Finally, we discuss briefly why Nim-Rem produces an item-strength effect (5.5.4).

5.5.1 Behavioural item-strength findings

Many researchers have found an item-strength effect using textual items (e.g., Ratcliff *et al.*, 1990; Cary and Reder, 2003) and faces (Laughery, Alexander, and Lane, 1971; Roark *et al.*, 2003; Nega, 2005). When examined in terms of the hit and false-alarm patterns (rather than an overall performance measure), a strength-mirror effect is often observed: when strengthening the study list, discrimination between targets and lures improves, there are more hits and fewer false alarms (e.g., McClelland and Chappell, 1998; Stretch and Wixted, 1998; Cary and Reder, 2003; Nega, 2005). As the list-length effect, the item-strength effect is generally considered to occur for recognition memory.

5.5.2 Item-strength simulations with Nim-Rem

In our simulations, we examined Nim-Rem's ability to predict the item-strength effect. The item-strength effect was analysed both in terms of the overall performance measure, and in terms of the hit and false-alarm patterns (to reveal the occurrence of a strength-mirror effect, see subsection 5.2.3).

Below we describe (1) the paradigm that was employed for assessing the item-strength effect, and (2) the experimental procedure.

The paradigm

As described in subsection 5.4.2, the second simulation employed a design that was suitable for assessing the list-length effect as well as the item-strength effect. Here, we focus on the paradigm that underlies the assessment of an item-strength effect. The paradigm uses three types of lists that differ in the item-strength (i.e., the strength of the images on the study list): (1) weak lists, (2) strong lists, and (3) extra strong lists. As explained in subsection 5.4.2, these list types were presented at three different list-length levels (i.e., short, long, and extra long). An item-strength effect is said to occur if the recognition score increases when the images on the study list are strengthened. NIM-REM recognition scores were compared for the weak, strong, and extra strong lists (at each list-length level) in order to assess the occurrence of an item-strength effect. Moreover, we analysed the hit and false-alarm patterns across the weak, strong, and extra strong lists, to reveal strength-mirror effects. To quantify the effect sizes, we again used *Cohen's d* effect size estimate.

The experimental procedure

As mentioned previously, the second simulation addressed the list-length effect and the item-strength effect. Details about the experimental procedure can be found in section 5.4.2.

5.5.3 Item-strength simulation results

Table 5.3 presents the results. The table shows the recognition results for weak ($S = 10$), strong ($S = 20$), and extra strong ($S = 30$) lists (rows) for lists of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$) (columns). Please note, that the recognition scores and the hit rates and false-alarm rates are the same ones as presented in Table 5.2. However, Table 5.3 presents the values of Table 5.2 with the rows and columns interchanged, to stress the emphasis on the item-strength effect rather than the list-length effect. Moreover, in Table 5.2 the effect sizes presented in bold between the rows in Table 5.2 and Table 5.3 represent the list-length effect sizes and the item-strength effect sizes, respectively.

Overall, recognition scores (d_a) increased with the strength, S, of the study-list images. This indicates that item-strength effects occurred. Fig. 5.4 presents bar graphs of the recognition scores for weak (dark-grey bars), strong (light-grey bars), and extra strong lists (white bars) of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$).

The values presented in bold between the rows in Table 5.3 indicate the sizes of the item-strength effects. Moreover, strength-mirror effects were obtained: hit rates for weaker lists were lower than hit rates for stronger lists and false alarm rates for weaker lists were higher than false alarm rates for stronger lists. This was the case across all list-length levels.

	short ($N = 4$)			long ($N = 12$)			extra long ($N = 18$)		
	d_a	H	FA	d_a	H	FA	d_a	H	FA
weak ($S = 10$)	0.94	0.63	0.27	0.78	0.61	0.31	0.63	0.59	0.35
	1.92			**1.42**			**0.96**		
strong ($S = 20$)	1.22	0.70	0.24	0.99	0.67	0.29	0.77	0.63	0.33
	1.21			**0.81**			**0.50**		
extra strong ($S = 30$)	1.41	0.74	0.22	1.11	0.70	0.28	0.84	0.66	0.32

Table 5.3: Recognition scores (d_a), hit rates (H), and false-alarm rates (FA), for weak ($S = 10$), strong ($S = 20$), and extra strong ($S = 30$) lists (rows) of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$) (columns). The absolute item-strength effect sizes are presented in bold between the rows.

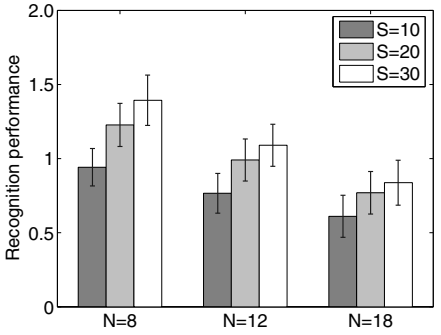


Figure 5.4: The item-strength effect. Recognition scores, d_a, for weak lists (dark-grey bars), strong lists (light-grey bars), and extra strong lists (white bars) of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$).

5.5.4 Discussion of the item-strength results

NIM-REM produces an item-strength effect because when an image is strengthened it becomes less confusable with a test image that is different from the stored image. As a consequence of the refinement of the representations, the variability in the familiarity values for targets and lures decreases, which results in higher recognition scores.

5.6 False-memory studies

The false-memory effect says that the recognition of a lure (i.e., a false memory or a false alarm) is more likely to happen when the lure is similar to (one of) the studied items.

In this section, we discuss well-known findings from behavioural recognition-memory studies on the false-memory effect (5.6.1). Then, we present the NIM-REM simulations by which we assessed NIM-REM's ability to replicate the behavioural findings (5.6.2). Subsequently, the false-memory simulation results are presented (5.6.3). Finally, we discuss the effect of similarity on the list-length effect and the item-strength effect (5.6.4).

5.6.1 Behavioural false-memory findings

False-memory effects have been found in a number of behavioural studies using textual stimuli (e.g., Postman, 1951; Roediger and McDermott, 1995; Westbury, Buchanan, and Brown, 2002; Dewhurst and Farrand, 2004), colours (Zaki and Nosofsky, 2001), object images (Koutstaal, 2003), and face images (e.g., Davies, Shepherd, and Ellis, 1979; Solso and McCarthy, 1981; Vokey and Read, 1992; Busey and Tunnicliff, 1999; Cabeza *et al.*, 1999; Stewart and McAllister, 2001). In one task in the experiment by Koutstaal (2003) younger and older adults were presented with three types of items: (1) previously-studied items (targets), (2) items that were perceptually and conceptually similar to studied items but were not presented during study (similar lures), and (3) new unrelated items (dissimilar lures). Items were coloured photographs or detailed line drawings of common objects or animals. Their results showed that for both younger and older adults, the number of false alarms was significantly higher for similar lures than for dissimilar lures. Similarly, the false-memory effect has been found in studies using faces. For example, Busey and Tunnicliff (1999) found a false-memory effect in their recognition experiment using face images (in our simulations we used the same face images). In the experiment there were three types of faces: (1) normal faces, (2) parent faces, and (3) morph faces. Each morph face was the average of two parent faces. A study list with 36 normal faces and 32 parent faces was presented to the subjects. Then, old-new recognition was tested for these 36 normal faces (i.e., normal targets) and 32 parent faces (i.e., parent targets) along with 20 new normal faces (i.e., normal lures) and 16 new morph faces (i.e., morph lures). The morph lures were the average of two parent targets that were either dissimilar or similar to each other. Dissimilar and similar morph lures resulted from dissimilar and similar morph parents, respectively. They found that

morph lures were falsely recognized much more often than normal lures. Moreover, similar morph lures were falsely recognized more often than dissimilar morph lures.

5.6.2 False-memory simulations with Nim-Rem

In our simulations, we examined Nim-Rem's ability to predict the false-memory effect. In particular, we investigated whether Nim-Rem is able to predict that similarity affects the false recognition of a lure as was found experimentally (see, e.g., Busey and Tunnicliff, 1999; Koutstaal, 2003). The robustness of the effect was assessed by testing for the false-memory effect across several list-length levels and across several item-strength levels. Below we describe (1) the paradigm that was employed for assessing the false-memory effect, and (2) the experimental procedure.

The paradigm

For assessing the occurrence of a false-memory effect, we employed a paradigm that uses two types of lures on the test list: (1) a lure that is dissimilar to the images on the study list, and (2) a lure that is similar to (at least one of) the images on the study list. A higher false-alarm rate for similar lures than for dissimilar lures is said to indicate the occurrence of a false-memory effect. We compared the false-alarm rate for dissimilar and for similar lures to assess the occurrence of a false-memory effect.

The experimental procedure

We varied the similarity of targets and lures, while keeping the similarity between the targets constant, distinguishing recognition performance for: (1) target vs. dissimilar lures (TD) and (2) targets vs. similar lures (TS).

In the similarity-rating studies described in section 4.1.2, similarity ratings for all possible pairs of images from the face image set were obtained. These were used to select dissimilar and similar images in the false-memory effect simulation. In the simulation, two images are considered dissimilar when they have a similarity rating larger than 0.50 and two images are considered similar when they have a similarity rating smaller than 0.25.

For each simulation run, Nim-Rem was provided with a study list of targets, which were selected randomly from the face image set. Then, old-new recognition was tested for: (1) a target, (2) a lure dissimilar to all of the targets, and (3) a lure similar to (at least) one of the targets. As for the list-length and item-strength simulations, the false-memory effect simulation provided Nim-Rem with 9 lists across three list-length levels: (1) short lists ($N = 8$), (2) long lists ($N = 12$), and (3) extra long lists ($N = 18$), and across three item-strength levels: (1) weak ($S = 10$), (2) strong ($S = 20$), and (3) extra strong ($S = 30$). Targets on the weak, strong, and extra strong lists, were stored with storage strengths $S = 10$, $S = 20$, and $S = 30$, respectively. As for the other simulations, recognition tests were performed using the radius parameter $r = 5.0$ and a recognition decision criterion $c = 1.0$.

5.6.3 False-memory simulation results

The false-memory simulation results are presented in Tables 5.4 and 5.5. The tables shows the recognition results for weak ($S = 10$), strong ($S = 20$), and extra strong ($S = 30$) lists of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$). Although this subsection addresses the false-memory effect in general, for reasons of readability Tables 5.4(a) and (b) and Tables 5.5(a) and (b) adhere to similar table structures as used in Table 5.2 and Table 5.3, respectively. We reiterate that the recognition scores, the hit rates, and the false-alarm rates in Tables 5.4(a) and (b) are the same as those in the Tables 5.5(a) and (b) with the rows and columns appropriately interchanged. The effect sizes are presented in bold between the rows in Tables 5.4(a) and (b) and in Tables 5.5(a) and (b). They represent the list-length effect sizes and the item-strength effect sizes, respectively. A detailed discussion of the effect of the similarity on the size of the list-length effect and the size of the item-strength effect is given in subsection 5.6.4.

A false-memory effect occurred for each list type; false-alarm rates were higher for similar lures than for dissimilar lures for each list type. The false-memory effects were significant for each list-length level and each item-strength level, F values ranged from $F(1, 999) = 6605$, $p < 0.05$, for short weak lists ($N = 8$, $S = 10$) to $F(1, 999) = 27807$, $p < 0.05$, for extra long extra strong lists ($N = 18$, $S = 30$). So, in accordance with the behavioural findings (e.g., Busey and Tunnicliff, 1999), NIM-REM predicts that the target-lure similarity modifies the false recognition of a lure.

As Table 5.2, Tables 5.4(a) and (b) present the sizes of the list-length effects in bold between the rows. Also, as Table 5.3, Tables 5.5(a) and (b) present the sizes of the item-strength effects in bold between the rows. Tables 5.4(a) and (b) illustrate that the list-length effect sizes vary with the similarity between the images on the study list and the test list. Also, Tables 5.5(a) and (b) illustrate that the item-strength effect sizes vary with the similarity between the images on the study list and the test list. Subsection 5.6.4 reflects on the effect of similarity on the sizes of the list-length and item-strength effects.

5.6.4 Discussion of the false-memory results

Below, we discuss the effect of similarity on the sizes of the list-length effects and the sizes of the item-strength effects as obtained in the false-memory simulation.

Similarity and the list-length effects

List-length effects were smaller for TD than for TS and were even non-significant for TD. The interaction between list length and similarity (TD or TS) was confirmed in the two-way ANOVAs, in which F values ranged from $F(2, 2999) = 204.44$, $p < 0.05$ for weak lists to $F(2, 2999) = 214.11$, $p < 0.05$ for extra strong lists. NIM-REM produces smaller list-length effects for TD than for TS, because there is more variation in the familiarity values for tested similar lures than for tested dissimilar lures and the variation declines less rapidly with list length for TS than for TD. For TD, additional study list images vary less in their similarity to the lures than for TS.

(a) Targets vs. dissimilar lures (TD)

	weak ($S = 10$)			strong ($S = 20$)			extra strong ($S = 30$)		
	d_a	H	FA	d_a	H	FA	d_a	H	FA
short ($N = 8$)	1.30	0.69	0.19	1.78	0.78	0.14	2.07	0.82	0.12
	0.02			**0.35**			**0.48**		
long ($N = 12$)	1.30	0.68	0.20	1.70	0.75	0.15	1.96	0.79	0.12
	0.01			**0.17**			**0.45**		
extra long ($N = 18$)	1.29	0.67	0.20	1.66	0.72	0.15	1.86	0.75	0.12

(b) Targets vs. similar lures (TS)

	weak ($S = 10$)			strong ($S = 20$)			extra strong ($S = 30$)		
	d_a	H	FA	d_a	H	FA	d_a	H	FA
short ($N = 8$)	0.63	0.69	0.46	0.82	0.78	0.48	0.94	0.82	0.49
	1.08			**1.68**			**1.87**		
long ($N = 12$)	0.47	0.68	0.51	0.58	0.75	0.54	0.67	0.79	0.55
	0.99			**1.57**			**1.96**		
extra long ($N = 18$)	0.33	0.67	0.55	0.35	0.72	0.60	0.38	0.75	0.62

Table 5.4: Recognition scores (d_a), hit rates (H), and false-alarm rates (FA), for weak ($S = 10$), strong ($S = 20$), and extra strong ($S = 30$) lists (columns) of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$) (rows). The values in bold between the rows represent the absolute list-length effect sizes. (a) Targets vs. dissimilar lures (TD). (b) Targets vs. similar lures (TS).

(a) Targets vs. dissimilar lures (TD)

	short ($N = 8$)			long ($N = 12$)			extra long ($N = 18$)		
	d_a	H	FA	d_a	H	FA	d_a	H	FA
weak ($S = 10$)	1.30	0.69	0.19	1.30	0.68	0.20	1.29	0.67	0.20
	2.49			**2.02**			**1.67**		
strong ($S = 20$)	1.78	0.78	0.14	1.70	0.75	0.15	1.66	0.72	0.15
	1.42			**1.18**			**0.82**		
extra strong ($S = 30$)	2.07	0.82	0.12	1.96	0.79	0.12	1.86	0.75	0.12

(b) Targets vs. similar lures (TS)

	short ($N = 8$)			long ($N = 12$)			extra long ($N = 18$)		
	d_a	H	FA	d_a	H	FA	d_a	H	FA
weak ($S = 10$)	0.63	0.69	0.46	0.47	0.68	0.51	0.33	0.67	0.55
	1.34			**0.75**			**0.14**		
strong ($S = 20$)	0.82	0.78	0.48	0.58	0.75	0.54	0.35	0.72	0.60
	0.84			**0.58**			**0.22**		
extra strong ($S = 30$)	0.94	0.82	0.49	0.67	0.79	0.55	0.38	0.75	0.62

Table 5.5: Recognition scores (d_a), hit rates (H), and false-alarm rates (FA), for weak ($S = 10$), strong ($S = 20$), and extra strong ($S = 30$) lists (rows) of different lengths: short lists ($N = 8$), long lists ($N = 12$), and extra long lists ($N = 18$) (columns). The values in bold between the rows represent the absolute item-strength effect sizes. (a) Targets vs. dissimilar lures (TD). (b) Targets vs. similar lures (TS).

This is so because, for TD the similarity to all of the study list images is controlled, while for TS the similarity to one of the study list images is larger than a certain value and the similarity to the other images varies.

Although there is no behavioural study we know of that *directly* examines how target-lure similarity affects the list-length effect, existing behavioural findings seem to support our Nim-Rem predictions. In a study examining the effect of phonological similarity on recognition memory, MacAndrew *et al.* (2002) provided subjects with a study list of four letters and study lists of six letters (that were presented either sequentially or simultaneously). Some of the study lists contained phonologically similar letters while others contained phonologically dissimilar letters. For recognition, either a letter from the study list (target) or a new letter (lure) was presented. For phonologically similar study lists, phonologically similar lures were used and for phonologically dissimilar study lists, phonologically dissimilar lures were used. A signal detection analysis of the hit rates and false-alarm rates determined the recognition scores. According to Klatzky (2004), the results showed that a significantly larger list-length effect occurred for similar targets and lures than for dissimilar targets and lures. Also, overall recognition scores were higher for dissimilar targets and lures than for similar targets and lures. So, in accordance with the behavioural findings of one study (MacAndrew *et al.*, 2002), Nim-Rem predicts that target-lure similarity modifies the list-length effect. However, more behavioural research should reveal the robustness of the effect of similarity on the occurrence of a list-length effect.

Similarity and the item-strength effects

Analogously to the list-length effect, the item-strength effect varies with the similarity between the images on the study list and the test list. However, the similarity affects the list-length effect and the item-strength effect in opposite directions. While list-length effects are smaller for TD than for TS, item-strength effects are larger for TD than for TS. The interaction between list length and similarity (TD or TS) was confirmed in the two-way Anovas, in which F values ranged from $F(2, 2999) = 446.95$, $p < 0.05$ for short lists to $F(2, 2999) = 454.68$, $p < 0.05$ for extra long lists. The smaller item-strength effect for TS compared to that for TD, can also be observed indirectly from the difference in the patterns of the false-alarm rates for TS and those for TD that occur when the study list is strengthened. While for TD the false-alarm rate decreases when the images on the study list are strengthened (together with the observed increase in the hit rate this represents a strength-mirror effect), for TS, the false-alarm rate does not decrease (and even slightly increases) when the images of the study list are strengthened (i.e., no strength-mirror effect occurs). This can be explained as follows. When study list images are strengthened, representations refine. Therefore, the variability in the familiarity values for targets and lures decreases. However, the refinement is less pronounced (or even absent) for similar lures, because by definition the representations of similar images overlap more than the representations of dissimilar images. Hence, Nim-Rem produces larger item-strength effects for TD than for TS.

Behavioural results seem to support our Nim-Rem predictions. Although a rise in

the hit rate and a decrease in the false-alarm rate is generally observed when items on the study list are strengthened (i.e., a strength-mirror effect; e.g., McClelland and Chappell, 1998; Stretch and Wixted, 1998; Cary and Reder, 2003; Nega, 2005), several researchers found that when strengthening the study-list items no decrease in the false-alarm rate for similar lures occurred (although judgements of frequency for both targets and similar lures steadily increased) (see, e.g., Hintzman, Curran, and Oppy, 1992; Sheffert and Shiffrin, 2003; Malmberg *et al.*, 2004). For example, the results of Malmberg *et al.* (2004) showed that when the number of presentations of the targets increased from one to twelve, the hit rate for targets increased, but the false-alarm rate for similar lures decreased neither for low frequency words nor for high frequency words, and even increased. Light *et al.* (2006) obtained similar results; when targets were strengthened, the hit rate as well as the false-alarm rate for similar lures (the switched plurality of a target word) increased. Hintzman *et al.* (1992), too, found the same results when using (1) nouns, (2) photographs ranging from pictures of landscapes to pictures of single objects, and (3) bit-mapped drawings in the experiments.

5.7 General discussion

In this section, we compare NIM-REM with dual-process memory models that have been proposed to explain the results from behavioural recognition-memory studies (5.7.1). Moreover, we discuss the differences between conceptually based and perceptually based similarity spaces (5.7.2).

5.7.1 Single-process models versus dual-process models

Our results show that a straightforward single recognition-process approach to recognition memory accommodates a wide range of findings from recognition-memory studies. Several memory models attempt to explain behavioural recognition results on the basis of two distinct types of memory processes (see Yonelinas (2002), for a review). These dual-processing models assume that recognition involves: (1) a familiarity process, i.e., a context-insensitive automatic process, and (2) a recollection process, i.e., a context-sensitive strategic process. Results from cognitive, neuropsychological, and neuroimaging studies of human memory seem to support that two types of memory processes underlie recognition memory performance (again, see Yonelinas (2002), for a review). Therefore, findings are increasingly elucidated within a dual-process framework. For example, Norman (2002) explains his behavioural findings on the effect of similarity on the list-strength effect (described in subsection 5.3.4) by a dual-processing approach (see also Norman and O'Reilly (2003)). He argues that the magnitude of the list-strength effect depends on the extent to which recollection drives recognition. When targets and lures are similar, both targets and lures are assumed to produce strong ratings of familiarity. Therefore, the subject relies on recollection of specific details to produce a recognition decision. Consequently, Norman (2002) argues, a positive list-strength effect occurs for similar targets and lures and a null or negative list-strength effect occurs for dissimilar targets and lures. Our results show that a single straightforward process for

recognition suffices to produce the effect of similarity on the size of the list-strength effect that was obtained by Norman (2002). However, in its current form NIM-REM will not produce positive list-strength effects, because of the effect of differentiation (as explained in subsection 5.3.4). While Norman and O'Reilly's (2003) dual-process model produces a positive list-strength effect for similar targets and lures and at the same time a null or negative list-strength effect for dissimilar targets and lures, it does not to produce the effects on the basis of natural input. We showed that the single-process NIM-REM model that operates on natural input, can accommodate the findings of the abundant majority of list-strength studies. NIM-REM also produces the effect of similarity on the list-strength effect (as obtained by Norman, 2002) and produces a wide range of other recognition-memory findings, all on the basis of natural input. Although the large amount of research on recognition memory does not provide a conclusive answer to the question whether more than one process is involved in recognition, neurobiological evidence seems to support the idea of more than one process (see, e.g., Yonelinas, 2002; Yonelinas *et al.*, 2005). Therefore, future work may address a dual-process variant of NIM that aims to explain Norman's and other recognition-memory findings on the basis of natural input.

5.7.2 Faces versus words

The results obtained with NIM-REM are based on the perceptual similarity structure of the input that is reflected in the similarity space. In our simulations we used face images. In contrast, many of the behavioural recognition-memory results are found mainly with textual stimuli. We are well aware of the fact that for printed words, the similarity is assumed to be based mainly on conceptual (semantic) characteristics and much less on perceptual characteristics (Stillings *et al.*, 1995). However, we argue that the effects described in this chapter generalize to different types of input, because they follow from the similarity-space representations rather than from the modality that gave rise to the representations. In our view, words (concepts) and images (objects) are processed differently to yield similar types of psychological similarity spaces.

5.8 Chapter summary

In the modelling studies presented in chapter 4, we found that NIM quite reliably produces similarity ratings and recognition rates for *individual* natural stimuli (also in Lacroix *et al.*, 2006a). In this chapter, we investigated NIM's suitability as a natural-input model for *general* recognition-memory effects. We presented a NIM variant, called NIM-REM, which realizes a natural-input version of the original REM model (Shiffrin and Steyvers, 1997) by complementing it with NIM's perceptual front end that operates directly on natural input. We tested NIM-REM by assessing its ability to produce four well-known recognition memory effects that are obtained in behavioural studies. Our results show that NIM-REM, predicts rather adequately the findings from behavioural recognition-memory experiments on the basis of natural visual input. NIM-REM produces the behavioural findings using a single-process memory model that operates on natural visual input. Future extensions of NIM-REM

may address a dual-processing approach that appears to be supported by neuropsychological and neurobiological recognition-memory studies.

Chapter 6

Classification by Nim-Class

This chapter is partly based on[1]:

1. Lacroix, J. P. W., Postma, E. O., and Murre, J. M. J. (2006b). Knowledge-driven gaze control in the Nim model. *Proceedings of the 28th Annual Meeting of the Cognitive Science Society (CogSci 2006)* (eds. R. Sun and N. Miyake), pp. 1657–1662, Lawrence Erlbaum Associates, Mahwah, NJ.

2. Lacroix, J. P. W., Postma, E. O., and van den Herik, H. J. (to appear). Modeling visual classification using bottom-up and top-down fixation selection. *The 29th Annual Meeting of the Cognitive Science Society (CogSci 2007).*

3. Lacroix, J. P. W., Postma, E. O., and van den Herik, H. J. (submitted). Toward a visual cognitive system using active top-down saccadic control.

In this chapter[2], we aim to investigate Nim's ability to classify natural images. Classification plays an important role in memory, learning, and other cognitive processes (see, e.g., Murre, 1992). We extend Nim to obtain a model of classification called Nim-Class. Classification is the process by which encountered items are grouped together based on their similarity (e.g., Medin and Schaffer, 1978; Nosofsky, 1986). It is closely related to recognition (see chapters 4 and 5), because both classification and recognition operate by assessing the similarity between an item and previously encountered items. Classification and recognition differ in that classification emphasizes the similarities of items belonging to the same class, whereas recognition emphasizes the identification of previously stored items. Nim-Class combines Nim's perceptual front-end with a new memory stage that is suitable for classification. We examine to what extent Nim-Class is able to classify natural stimuli in a classification task. Subsequently, we investigate to what extent classification performance can be improved by introducing a top-down fixation-selection mechanism to select relevant fixation locations.

[1] The author would like to thank her co-authors and the publisher of the CogSci 2006 and 2007 proceedings for their kind permission to reuse relevant parts of the articles in this thesis.

[2] The results presented in this chapter were partly presented at the workshop *Towards Cognitive Humanoid Robots of the IEEE-RAS International Conference on Humanoid Robots 2006*

The outline of the remainder of this chapter is as follows. In section 6.1, we present NIM-CLASS, a model for the classification of natural visual input. This is followed in section 6.2 by a description of the classification experiment that was used for our classification studies. Subsequently, in section 6.3, the classification performance of NIM-CLASS is evaluated on a face-classification task. Then, section 6.4 presents two NIM-CLASS variants that demonstrate how NIM-CLASS can be adapted to incorporate top-down fixation selection to direct the gaze towards relevant spatial locations in an image. After that, the classification performances of the NIM-CLASS variants extended with top-down fixation selection are assessed in sections 6.5 and 6.6. In section 6.7, we discuss bottom-up and top-down gaze control models, examine the scalability of the NIM-CLASS variants, and provide a comparison with existing models that have been tested for classification with the same stimuli. Finally, in section 6.8, we summarize the results and draw conclusions on the feasibility of NIM-CLASS as a cognitive model for visual classification of natural input.

6.1 Adapting NIM for classification

As outlined in chapter 3, the original NIM encompasses two stages: the perceptual preprocessing stage and the memory stage. While NIM is a model for recognition of natural images, here we show that it can readily be adapted into a model for classification of natural images which we call NIM-CLASS. The feasibility of adapting NIM for classification has been shown recently by Barrington *et al.* (2007) who combined NIM's preprocessing stage to transform fixated image parts into feature vectors with a Bayesian version of the memory stage in their NIMBLE model and applied it successfully to face classification. NIM-CLASS uses a slightly different approach that also adopts NIM's preprocessing stage, but introduces a different memory stage based on a nearest-neighbour classifier that has been demonstrated to be highly suitable for object classification (see, e.g., Mattern and Denzler, 2004). Since NIM-CLASS differs from NIM in the design of the memory stage only, we discuss the two processes of the NIM-CLASS memory stage: the storage process (6.1.1) and the classification process (6.1.2). The storage and classification processes correspond to the training and the testing stages that are commonly distinguished in supervised learning (see, e.g., Duda *et al.*, 2001).

6.1.1 The storage process

The NIM-CLASS storage process, retains (i.e., stores) preprocessed samples of natural images (i.e., fixations) that belong to a certain class. As in the original NIM, each natural image is represented by a number of low-dimensional feature vectors (one for each fixation) in a similarity space. For NIM-CLASS, each image represents an instance of a class. Therefore, in contrast to the original NIM that stores unlabelled feature vectors, NIM-CLASS stores class labels with each feature vector corresponding to the class associated with the image (i.e., '1' for class 1, '2' for for class 2, and so forth). The storage of labelled feature vectors was introduced previously for NIM-REM that was presented in chapter 5.

6.1.2 The classification process

The Nim-Class classification process employs a naive Bayesian method that is based on an incremental estimate of the class-dependent probabilities (Duda *et al.*, 2001). In the classification process, each fixation of the test image (i.e., each test feature vector) contributes to an n-bin histogram, the bins of which represent the 'beliefs' in the n different classes. For each test feature vector, the bin that corresponds to the label of its nearest neighbouring stored labelled feature vector (acquired in the storage process) is incremented (e.g., if the stored labelled feature vector that is closest to the test feature vector has label '1', bin 1 is incremented). Finally, upon the last fixation, the class with the largest bin (i.e., belief) determines the classification decision. This heuristic classification process could readily be extended into a Bayesian approach in which each fixation updates class-conditional probabilities according to the Bayes update rule.

6.2 Classification experiment

In our experiments, we evaluate the ability of Nim-Class to classify natural images of faces. Below, we discuss the classification task (6.2.1), the data set (6.2.2), and the experimental procedure (6.2.3).

6.2.1 The classification task

The classification task entails the identification of a natural image of a frontal face with variations in facial expression, illumination (location of the light source), and occlusion (sun glasses and scarf). For each individual, there are 13 views in total (see Fig. 6.1). Humans are generally able to identify a face after a single encounter only, despite variations in appearance (e.g., Burton *et al.*, 2005). Inspired by this fact, Nim-Class is evaluated on a task in which the training set (i.e., the study list) consists of a single image for each class and the test set (i.e., the test list) of the twelve remaining images. In this respect, our evaluation differs from most evaluations in machine learning, where the training set consists of a much larger fraction of the data set.

6.2.2 The data set

For the face-classification task, a data set with different images of the same individual was needed. We chose to use the AR data set created by Martinez and Benavente (1998) that contains over 4,000 images corresponding to the faces of 126 individuals. For each individual, the AR data set includes a sequence of 13 images of frontal view faces with different facial expressions, illumination conditions, and occlusions. For the experiment, we selected the sequence of 13 images (i.e., views) of the first 10 male individuals of the AR data set as our data set. All face images were downscaled to 165×165 pixels. Fig. 6.1 shows an example of the sequence of 13 views of one individual. The first (standard) view of each individual was selected for the study list, the remaining 12 views were assigned to the test list.

Figure 6.1: Example of the 13 views of one individual from the AR data set.

6.2.3 The experimental procedure

The face-classification experiment entailed a study and a test phase. During the study phase, we presented NIM-CLASS with the images from the study list containing the first view of each of the $n = 10$ individuals (i.e., the study faces). For each study face, NIM-CLASS extracted and stored s labelled feature vectors. Then during the test phase, the model was given the images from the test list (i.e, the 12 test faces) of each of the $n = 10$ studied individuals. For each of the test faces, the model extracted t test feature vectors to classify the face as one of the $n = 10$ individuals that it had previously encountered. To assess how the NIM-CLASS classification performance varied as a function of the number of storage fixations s and the number of test fixations t, the experiment was repeated for values of s and t in the range from 10 to 100, i.e., $s, t \in \{10, 20, ...100\}$.

6.3 Classification by NIM-CLASS

Below, we present the NIM-CLASS results for the face-classification task (6.3.1). Subsequently, we compare viewing time and fixation selection by NIM-CLASS with that by humans (6.3.2).

6.3.1 Classification results

Table 6.1 presents the percentages of correctly classified test faces for a range of values of the number of storage fixations s and the number of test fixations t. To visualize the pattern of performances as a function of the number of storage fixations s and the number of test fixations t, Fig. 6.2 presents the same results as a surface plot (for the precise performance values we refer to Table 6.1, because it is difficult to read these from Fig. 6.2). The NIM-CLASS classification performances range from just above chance level (16%) for $s = t = 10$ to a good performance of 89.0% for $s = t = 100$. Evidently, NIM-CLASS is capable of exhibiting a good performance

s \ t	10	20	30	40	50	60	70	80	90	100
10	16.0	18.2	20.6	22.1	23.6	23.7	24.4	25.3	25.5	26.2
20	21.3	26.3	29.5	32.1	35.5	38.3	39.3	41.1	42.7	43.5
30	26.5	32.8	38.1	42.5	46.3	49.0	52.0	53.3	55.5	57.3
40	30.0	39.5	45.7	51.1	55.1	58.6	60.8	63.1	64.5	66.8
50	34.0	45.2	51.7	57.0	61.8	64.9	68.0	70.0	71.5	73.7
60	36.7	49.2	57.0	62.7	66.9	70.7	73.7	75.3	77.3	78.5
70	39.8	52.9	61.8	67.7	71.2	75.3	77.8	79.6	80.9	82.5
80	42.7	57.0	65.9	70.9	75.4	77.9	80.7	82.9	84.3	85.4
90	45.7	60.1	68.3	73.8	78.3	81.1	83.3	84.8	85.9	87.4
100	47.6	63.1	71.3	77.0	80.6	83.2	84.7	87.1	87.8	89.0

Table 6.1: Percentages of correctly classified faces for a range of values of the number of storage fixations s and the number of test fixations t.

provided that a sufficient number of fixations is made.

The results show, not surprisingly, that the performance increases both with the number of storage fixations and the number of test fixations. Increasing the number of stored fixations s, improves the performances more than increasing the number of test fixations t. For small s values, the performance hardly increases with t. Evidently, increasing the number of test fixations is only useful when a sufficient number of feature vectors was stored previously. From a statistical perspective this makes sense. A proper approximation of the true distribution of feature vectors in a similarity space associated with a single face requires a sufficient number of samples (fixations) of that face.

To provide some insight into the distribution of beliefs in the different classes for each of the 120 test faces (i.e., 12 test views for each of the 10 individuals in the data set), Fig. 6.3 presents an overview of the histograms for each of the 120 test faces for $s = t = 100$. Each histogram represents the belief in class 1 (leftmost bin in each histogram) to 10 (rightmost bin in each histogram). In other words, the histograms represent the frequency counts of the labels of the nearest neighbours of the test feature vectors. Each row of histograms corresponds to the view depicted to the left of that row and each column of histograms corresponds to the individual depicted at the top of that column. A face is classified correctly when the index of the largest bin corresponds to the class of the particular face. From Fig. 6.3 it can be seen that, in most cases, the largest bin corresponds to the class of the test face. Where this is not the case, the largest bin is not considerably larger than the other bins. Therefore, it can be said that the falsely classified faces were classified with less certainty than the correctly classified faces.

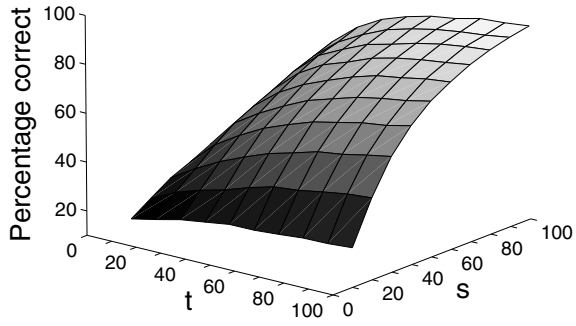


Figure 6.2: Percentages of correctly classified faces as a function of the number of storage fixations s and the number of test fixations t.

6.3.2 Comparison with humans

The results show that NIM-CLASS is able to classify faces quite accurately despite variations in facial expressions, illumination conditions, and occlusions. The model reaches a performance of 89% for $s = t = 100$ storage and test fixations. Below, we compare viewing time and fixation selection by NIM-CLASS with that by humans.

Viewing time by NIM-CLASS and by humans

The number of storage and test fixations extracted by NIM-CLASS can be interpreted as the amount of viewing time of the image during the study and test phase, respectively. Dividing the number of fixations by five provides a rough estimate of the number of seconds the image is inspected, since humans make about five fixations per second (see, e.g., Henderson, 2003; McSorley and Findlay, 2003). As the results show, the NIM-CLASS performance relies heavily on the amount of viewing time during the study phase. This accords with results from several psychological studies indicating that memory for visual information increases with the amount of viewing time during the study phase (see, e.g., Loftus, 1972; Mäntylä and Holm, 2006; Melcher, 2006). Moreover, it is interesting that a considerable percentage of faces (say $\approx 75\%$) is classified correctly after a short viewing time of about 8 seconds (40 fixations) during the test phase, provided that there was a sufficiently long viewing time of about 20 seconds (100 fixations) during the study phase.

We performed additional simulations to assess in more detail to what extent NIM-CLASS is able to classify the test faces correctly on the basis of a brief viewing time during the test phase. In these simulations, the experiment was repeated for values of s that range from 10 to 1000, i.e., $s \in \{10, 20, ...1000\}$ which corresponds to about 2 seconds to 200 seconds of viewing time during the study phase, and the number of test fixations was set to $t = 5$, which corresponds to approximately one second of

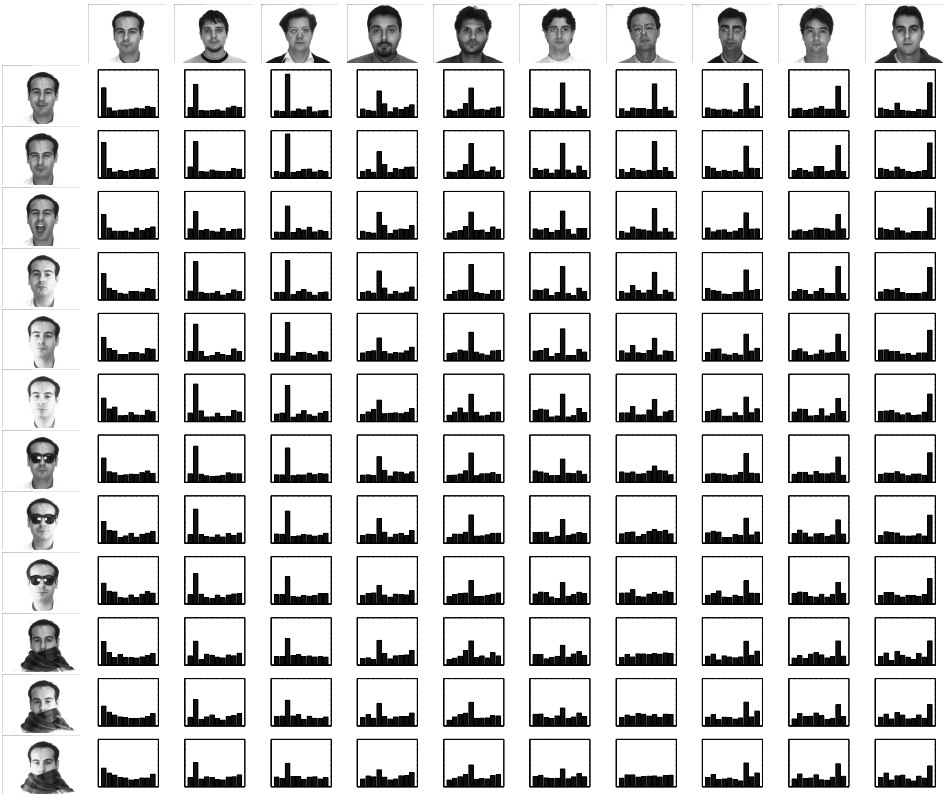


Figure 6.3: Overview of the histograms obtained in the classification experiment across the 120 test faces (i.e., 12 views of each of the 10 individuals) for $s = t = 100$.

viewing time during the test phase. Fig. 6.4 presents the NIM-CLASS performance for a fixed number of test fixations, $t = 5$, as a function of the number of storage fixations s. To illustrate the differences between the performances for the faces without and with occlusions, Fig. 6.4 shows the average performances across the six test views without occlusions (dark-grey line), the average performances across the six test views with occlusions (light-grey line), and the average performances across all the twelve test views (black line), separately. For the faces with occlusions, a

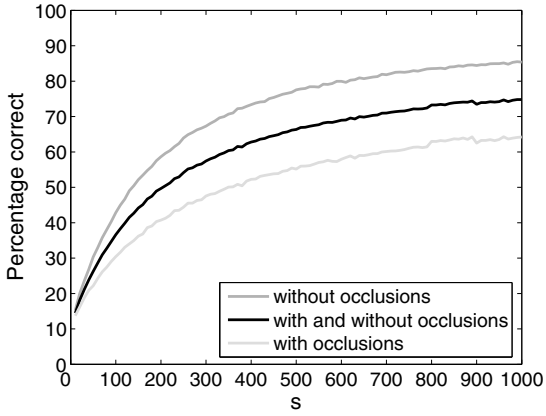


Figure 6.4: Percentages of correctly classified faces for a fixed number of test fixations, $t = 5$, as a function of the number of storage fixations s for the test views without occlusions (dark-grey line), with occlusions (light-grey line), and both with and without occlusions (black line).

degraded performance is observed compared to the performance for faces without occlusions. This is so, because the probability that a sufficient amount of relevant visual information is gathered for correct classification diminishes rapidly when one or more of the limited number of only five test fixations happen to be selected from occluded regions of the face. Still, it can be said that NIM-CLASS is able to reach a considerable average classification performance on the basis of a brief viewing time during the test phase, provided NIM-CLASS has studied the face for a sufficiently long time. The same holds for human vision, for which it is known that a brief viewing time will allow for correct identification, provided the face is sufficiently familiar to the observer (see, e.g., Findlay and Gilchrist, 2003; Burton *et al.*, 2005).

Fixation selection by NIM-CLASS and by humans

NIM-CLASS selects image samples (i.e., fixations) on the basis of their visual saliency (along the contours). Several behavioural studies showed that in human vision bottom-up processes draw the eyes toward salient visual features such as high edge density (see, e.g., Mannan *et al.*, 1996) and local contrast (see, e.g., Parkhurst and Niebur, 2003). Based on these findings, many models of gaze control employed

a bottom-up approach (e.g., Braun *et al.*, 2001; Rao *et al.*, 2002; Zhaoping and May, 2007). Often, a so-called 'saliency map' is constructed that marks those image regions that are visually distinct from their surround in one or more visual features (Itti and Koch, 2000). Then the gaze is directed to locations that are marked as highly salient on the saliency map. In fact, our contour-based selection of fixations can be regarded as a realization of the bottom-up approach in which contours are the salient features. Evidently, the bottom-up fixation-selection mechanism of NIM-CLASS, can hardly be considered to agree with the active context-dependent scanning of a visual scene that humans perform (see, e.g., Rajashekar *et al.*, 2002). In the dynamic process of actively scanning the visual scene, eye fixations are guided by both bottom-up and top-down processes (Karn and HayHoe, 2000; Henderson, 2003; Oliva *et al.*, 2003; Neider and Zelinski, 2006; Torralba *et al.*, 2006).

While bottom-up processes (based on visual saliency) have often been used to control the gaze in artificial systems, the use of top-down processes has not received as much attention. Top-down processes employ stored knowledge and the goals of the viewer to select the most relevant gaze location (see, e.g., Henderson, 2003; Henderson *et al.*, 2007; Torralba *et al.*, 2006). Several studies showed that human gaze control relies more on top-down processes than on bottom-up processes when performing an active visual task with meaningful stimuli (see, e.g., Oliva *et al.*, 2003). The top-down processes are driven by several cognitive systems, including: (1) short-term episodic memory for previously attended visual input (e.g., Chun, 2000; Henderson, 2003), (2) stored long-term knowledge about visual, spatial, and semantic characteristics of classes of items or scenes acquired through experience (e.g., Henderson, 2003), and (3) the goals and plans of the viewer (e.g., Yarbus, 1967; Land and Hayhoe, 2001; Henderson, 2003). Although bottom-up processes are important in human vision too, they are integrated with top-down processes that direct the gaze to relevant locations on the basis of cognitive systems (see, e.g., Henderson, 2003). To adhere to the use of top-down processes for fixation selection in human vision, we focus in section 6.4 on top-down fixation selection as an important element for NIM-CLASS.

6.4 Adapting NIM-CLASS for top-down fixation selection

Inspired by fixation selection in human vision, this section adapts NIM-CLASS for top-down fixation selection. We explore the use of top-down fixation selection by investigating to what extent top-down fixation selection aids performance on the classification task compared to the pure bottom-up selection. To realize top-down fixation selection, we rely on two types of knowledge known to operate in human gaze control: (1) the short-term episodic knowledge about previously attended visual input (see, e.g., Chun, 2000; Henderson, 2003; Mäntylä and Holm, 2006), and (2) the long-term knowledge about a class of items acquired through experience with instances from the class (see, e.g., Henderson, 2003). In order to assess the contribution of the different types of knowledge separately, we introduce two NIM-CLASS variants: NIM-CLASS A (6.4.1) and NIM-CLASS B (6.4.2).

6.4.1 NIM-CLASS A

NIM-CLASS A extends NIM-CLASS with a top-down fixation-selection mechanism based on episodic knowledge about previously attended visual input. NIM-CLASS A incorporates the top-down fixation selection in the classification process, while adopting the bottom-up (i.e., contour-based) fixation selection of NIM-CLASS in the storage process. Below, we discuss the two processes of the memory stage of NIM-CLASS A: (1) the storage process and (2) the classification process.

The storage process

The storage process of NIM-CLASS A is similar to that of the original NIM-CLASS, except that NIM-CLASS A stores the coordinates of each fixation along with the class label. The coordinate labels are used for top-down fixation selection by the classification process.

The classification process

The classification process of NIM-CLASS A involves a top-down fixation-selection mechanism that uses short-term episodic knowledge. In NIM-CLASS, the short-term episodic knowledge corresponds to the labelled feature vectors that were acquired during the storage process directly preceding the current classification process. The top-down fixation-selection mechanism in the classification process is inspired by the idea that episodic knowledge about attended item parts provides detailed item-specific information that may contribute to the recognition or classification of the item (see, e.g., Chun, 2000; Henderson, 2003; Mäntylä and Holm, 2006).

For the implementation of the top-down fixation-selection mechanism, we rely on the notion of Shannon's (1948) entropy. Shannon introduced entropy as a measure of uncertainty. In order to decide in the most efficient way to which class a new item belongs, a system should select new input that minimizes the entropy, i.e., the uncertainty about the class membership. In NIM-CLASS, uncertainty is represented by the histogram in which the heights of the bins represent the beliefs in the different classes. Considering the uncertainty, the top-down fixation-selection mechanism selects those locations that contain the most relevant information to decide upon the class of the face under consideration (i.e., that minimize the entropy or uncertainty about the class). In order to do so, the fixation-selection mechanism of the classification process in NIM-CLASS A uses the short-term episodic knowledge about attended parts of recently encountered faces (i.e., the stored labelled feature vectors).

To select the fixation locations that minimize the entropy (i.e., the locations that contain the most relevant information for classification) the mechanism proceeds as follows. For each fixation, it first chooses the two most likely classes, P and Q, by selecting the two highest bins in the histogram. Subsequently, it selects the fixation location that best discriminates between the two classes P and Q (i.e., contains the most relevant visual input with respect to P and Q). The selection relies on:
(1) the distances between the feature vectors of the two classes; and
(2) the distances between the spatial fixation locations from which they originated.

The idea behind the selection is that spatially adjacent fixations within one class give rise to similar feature vectors. Hence, the fixation mechanism searches for a pair of feature vectors p and q coming from classes P and Q, respectively, that originate from relatively close spatial locations and at the same time are relatively distant from each other in the representation space.

We implemented this idea heuristically. Below, we provide the steps followed by the fixation-selection mechanism:
1. Define the two classes that have the largest belief as the target classes, P and Q.
2. For each possible pair of feature vectors p and q coming from target classes P and Q, respectively, calculate the ratio $d(p,q)/d((x,y)_p,(x,y)_q)$, where $d(p,q)$ is the Euclidean distance between feature vectors p and q in the representation space and $d((x,y)_p,(x,y)_q)$ is the Euclidean distance between the spatial coordinates (x,y) of p and q.
3. Select the two feature vectors p and q for which the ratio is the highest.
4. Define the target location as the contour location that is closest to the midpoint of the line connecting the spatial coordinates of p and q.
5. Select the contour in the test image that is closest to the target location as the location to be fixated next. If this location has been fixated before, go back to step 3 and take the next highest ratio in line. Moreover, in the highly unlikely event that all locations that are selected on the basis of the ratios have been visited, select a random fixation location.

Since the bins are empty when the classification process commences, the process starts with two test fixations that are taken randomly along the contours in the image (as described in subsection 6.1.2).

6.4.2 NIM-CLASS B

NIM-CLASS B employs the top-down fixation-selection mechanism of the classification process of NIM-CLASS A, and employs in addition a top-down fixation-selection mechanism in the storage process. Below, we discuss the two processes of the memory stage of NIM-CLASS B: (1) the storage process and (2) the classification process.

The storage process

As NIM-CLASS A, NIM-CLASS B stores the coordinates of each fixation along with the class label, which are used by the classification process. In contrast to NIM-CLASS A, NIM-CLASS B employs a top-down fixation-selection mechanism in the storage process. The fixation-selection mechanism relies on long-term knowledge that is acquired through experience, i.e., encounters with many faces. Top-down fixation selection in the storage process is motivated by the idea that people have acquired long-term knowledge about which parts of the items or scenes from a particular category, e.g., the face category, are the most informative (Henderson, 2003). Through experience with numerous exemplars of faces, efficient strategies have been developed for directing the gaze to the face parts that are most relevant for discriminating among different faces (i.e., the face parts with a high entropy). The NIM-CLASS B top-down fixation-selection mechanism of the storage process relies

on the long-term knowledge about the relevance of face parts for top-down fixation selection upon the first encounter with a new face (i.e., in the storage process).

The fixation-selection mechanism employed by the storage process of NIM-CLASS B uses the acquired knowledge about the relevance R of fixation locations for the class of faces. As in NIM-CLASS and NIM-CLASS A, only locations along contours were considered as fixation locations. The relevance R of a fixation location (x, y) in an image i ($i \in 1, 2, ...n$, with n the number of images), $R_{x,y,i}$, is a direct function of the average distance across the feature vectors extracted from the (x', y') locations in the study images (where location (x', y') is the fixation location, i.e., the contour location, that is closest to (x, y)).

Formally $R_{x,y,i}$ is defined as follows.

$$R_{x,y,i} = \frac{\sum_{p=1}^{n} \sum_{p=1, q \neq p}^{n} d(p, q)}{n(n-1)}, \tag{6.1}$$

where $d(p, q)$ represents the Euclidean distance between the feature vectors extracted from coordinates (x', y') in images p and q. Fig. 6.5 shows an example of a 'relevance

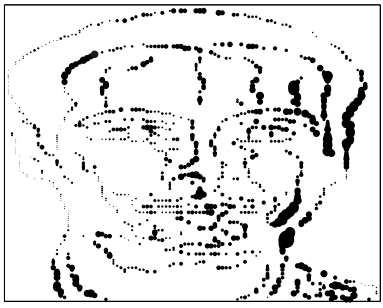


Figure 6.5: An example of the relevance of 1000 fixation locations in the face image. The sizes of the dots relate to the relevance of the location.

map' of 1000 fixation locations in which the relevance of each fixation location is represented by a disk. The size of the disk at a location is proportional to the value of R at that location. The relevance of a location is used for fixation selection upon the first encounter with a new face (i.e., in the storage process). Locations are fixated with a probability corresponding to the associated value of R. Since differences in R values were very small, R values were raised by the power 10 and divided by the sum of the R values such that they summed up to 1, to obtain the probabilities for fixation selection.

The classification process

NIM-CLASS B fully adopts the classification process of NIM-CLASS A (see subsection 6.4.1).

s \ t	10	20	30	40	50	60	70	80	90	100
10	25.2	27.8	28.8	29.1	30.9	29.9	30.1	31.1	30.0	30.6
20	32.5	40.6	43.8	45.9	48.7	50.6	50.7	51.8	53.6	53.4
30	36.9	46.3	53.3	56.6	59.3	62.0	64.0	65.4	66.7	67.1
40	41.0	51.5	58.3	63.9	66.5	68.7	71.2	72.6	73.8	75.4
50	44.1	56.2	63.7	67.8	71.3	73.5	75.7	78.1	79.3	80.9
60	46.7	59.6	66.5	72.3	75.2	77.6	79.9	81.2	82.6	84.2
70	49.1	62.7	69.8	75.0	78.5	80.8	82.2	84.2	85.8	86.7
80	51.3	64.8	72.1	77.4	80.0	83.3	85.2	85.8	86.9	88.2
90	53.4	67.0	75.3	79.3	82.7	84.7	86.6	87.6	88.7	89.8
100	54.9	69.5	77.1	81.5	84.4	85.8	87.8	88.9	90.0	91.0

Table 6.2: The NIM-CLASS A classification performance for a range of values of the number of storage fixations s and the number of test fixations t.

6.5 Classification by NIM-CLASS A

Below, we present the results for the face-classification task performed by NIM-CLASS A (6.5.1) and compare the classification performance of the original NIM-CLASS and NIM-CLASS A (6.5.2).

6.5.1 NIM-CLASS A classification results

Table 6.2 presents the percentages of correctly classified test faces as a function of the number of storage (s) and test (t) fixations for NIM-CLASS A. In addition, Fig. 6.6 visualizes the pattern of the classification performances of NIM-CLASS A as a function of s and t in a surface plot. The NIM-CLASS A classification performance ranges from 25.2% for $s = t = 10$ to a performance of 91.0% for $s = t = 100$. As for the original NIM-CLASS, the results of NIM-CLASS A show that performance increases with the number of storage fixations s and the number of test fixations t and the performance increases more with s than with t. As was demonstrated for NIM-CLASS, the results of NIM-CLASS A show that increasing the number of test fixations t becomes useful when a sufficient number of feature vectors was stored previously.

6.5.2 Discussion and comparison of classification results

Below we review and discuss the classification performances of NIM-CLASS and NIM-CLASS A.

The results show that extending NIM-CLASS with top-down fixation selection for classification to direct the gaze towards relevant locations, improves the performance on the classification task. To allow for easy comparison, Fig. 6.7 displays the performances of NIM-CLASS, and of NIM-CLASS A (and also the performances of

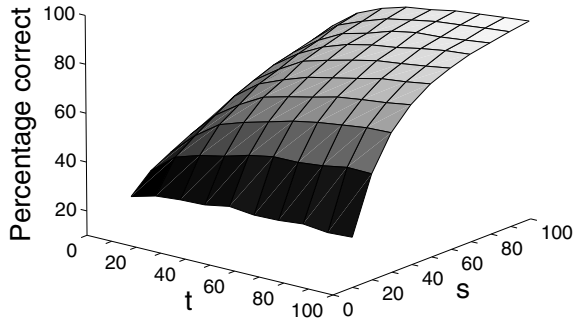


Figure 6.6: The NIM-CLASS A classification performance as a function of the number of storage fixations s and the number of test fixations t.

NIM-CLASS B, see section 6.6.1) in a comparative picture. Evidently, NIM-CLASS

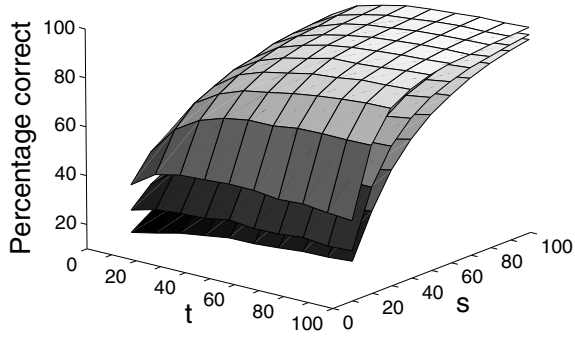


Figure 6.7: A comparison of the NIM-CLASS (lowest surface), NIM-CLASS A (middle surface), and NIM-CLASS B (top surface) classification performances as a function of the number of storage fixations s and the number of test fixations t.

A directs the gaze to locations that are more relevant to perform the classification task than those selected by the original NIM-CLASS. In NIM-CLASS A, the top-down fixation-selection mechanism in the classification process actively constructs a fixation sequence based on: (1) the task (i.e., classification), and (2) the stored episodic knowledge about previous encounters with particular faces (i.e., the stored labelled feature vectors). By combining top-down and bottom-up processes for the

selection of fixations (by considering only the fixations along contours), NIM-CLASS A acknowledges the influence of the episodic short term knowledge and the goals (i.e., classification) that are known to play a role in human gaze control (see, e.g., Henderson, 1999; Land and Hayhoe, 2001; Henderson, 2003). The active strategy employed by NIM-CLASS A in the classification process ensures that the locations are fixated that are known to discriminate well among the two most likely classes. Therefore, the model is better able to form the correct classification decision. This is particularly so, when the classification process makes a limited number of fixations. When the classification process makes a large number of fixations, a sufficient amount of relevant visual information is gathered for correct classification even when fixations are taken randomly along the contours. When the classification process makes fewer fixations, the probability that a sufficient amount of relevant visual information is gathered for correct classification decreases. Therefore, performance differences between the original NIM-CLASS and the NIM-CLASS A models are most pronounced for small t values.

6.6 Classification by NIM-CLASS B

Below, we present the results for the face-classification task performed by NIM-CLASS B (6.6.1) and compare the classification performance of NIM-CLASS A and NIM-CLASS B (6.6.2).

6.6.1 NIM-CLASS B classification results

Table 6.3 presents the percentages of correctly classified test faces as a function of the number of storage (s) and test (t) fixations for NIM-CLASS B. In addition, Fig. 6.8 visualizes the pattern of the classification performances of NIM-CLASS B as a function of s and t in a surface plot.

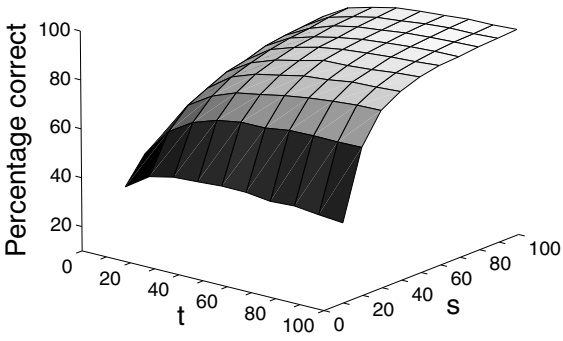


Figure 6.8: The NIM-CLASS B classification performance as a function of the number of storage fixations s and the number of test fixations t.

s \ t	10	20	30	40	50	60	70	80	90	100
10	35.5	42.2	44.4	45.0	45.6	45.5	44.3	44.8	43.7	42.8
20	45.7	57.5	63.6	67.0	68.7	68.8	70.2	70.5	70.4	70.6
30	51.4	65.1	71.3	75.1	76.9	78.9	80.3	81.5	82.3	82.6
40	54.5	68.9	76.0	79.6	81.7	83.8	84.7	86.2	86.6	87.2
50	58.2	72.5	78.7	82.3	84.5	87.0	87.4	88.2	89.3	90.1
60	60.5	74.3	81.1	84.4	86.9	88.4	89.4	90.2	90.8	91.6
70	62.1	76.9	82.7	85.6	87.8	89.4	90.4	91.3	91.9	92.1
80	63.5	78.1	83.5	86.9	89.0	90.1	91.4	92.1	92.6	93.0
90	64.7	78.9	84.8	87.6	89.8	90.8	92.0	92.8	92.9	93.4
100	66.2	79.8	85.6	88.3	90.6	91.6	92.3	93.2	93.4	94.0

Table 6.3: The NIM-CLASS B classification performance for a range of values of the number of storage fixations s and the number of test fixations t.

The NIM-CLASS B classification performance ranges from 35.5% for $s = t = 10$ and reaches a performance of 94.0% for $s = t = 100$. In correspondence with the results by the original NIM-CLASS and NIM-CLASS A, the results by NIM-CLASS B show (1) that the performance increases with the number of storage fixations s and the number of test fixations t, and (2) that the performance increases more with s than with t. Also, similarly to the NIM-CLASS and the NIM-CLASS A results, the NIM-CLASS B results show that increasing the number of test fixations t becomes useful when a sufficient number of feature vectors was stored previously.

6.6.2 Discussion and comparison of classification results

NIM-CLASS B further extends NIM-CLASS A by adding top-down fixation selection in the storage process. With this addition, NIM-CLASS B approaches the long-term spatial knowledge that people have built up about categories of items. Sufficient experience with human faces, allows for efficient strategies for directing the gaze to the face parts that are most informative for discriminating among different faces. Incorporating this type of knowledge in NIM-CLASS B resulted in the ability to fixate efficiently relevant face parts upon the first encounter with a new face. Fixating locations that contain visual input that is discriminative in terms of the different classes, leads to more discriminative memory representations. Consequently, the top-down fixation selection of the classification process that relies on finding discriminative locations based on previously stored feature vectors, can be performed more efficiently. This explains the improved performance of NIM-CLASS B compared to that of NIM-CLASS A (see Fig. 6.7 for a comparison of the performances of NIM-CLASS A and NIM-CLASS B; it also shows the performances of the original NIM-CLASS).

6.7 General discussion

Below, we discuss bottom-up and top-down gaze-control models (6.7.1). Subsequently, we discuss the scalability of the three NIM-CLASS models (6.7.2). Finally, we compare the NIM-CLASS and the NIM-CLASS A and B classification performances to the performances of existing classification models in the domain of artificial intelligence (6.7.3).

6.7.1 Bottom-up and top-down gaze-control models

Two types of gaze-control models can be distinguished: (1) bottom-up models, and (2) top-down models. Below, we first discuss bottom-up gaze-control models. Then, we compare top-down gaze control in NIM-CLASS A and B with other gaze-control models that use a top-down approach.

Bottom-up gaze-control models

Until now, the bottom-up or stimulus-driven approach has been the dominant approach to model gaze control. Bottom-up gaze-control models generally assume that fixation locations are selected in a bottom-up manner based on the image properties (see, e.g., Itti and Koch, 2000; Rao *et al.*, 2002; Zhaoping and May, 2007). These models create a saliency map that marks the saliency of each image location. Saliency is defined by the distinctiveness of a region from its surround on certain visual dimensions. Since locations with a high visual saliency are assumed to be highly informative, the gaze is directed towards highly salient locations. Often, the visual dimensions that are used to generate a saliency map are similar to the visual dimensions that are known to be processed by the human visual system such as colour, intensity, contrast, orientation, edge junctions, and motion (see, e.g., Koch and Ullman, 1985; Itti and Koch, 2000; Parkhurst, Law, and Niebur, 2002). Also, in order to discover certain important visual dimensions for generating a saliency map, a few studies analysed which visual dimensions best distinguish fixated image regions from non-fixated image regions (see, e.g., Mannan *et al.*, 1996; Parkhurst and Niebur, 2003; Henderson *et al.*, 2007).

Several studies showed that, under some conditions, fixation patterns predicted by bottom-up gaze-control models correlate well with those observed in human subjects (see, e.g., Parkhurst *et al.*, 2002). In their study, Parkhurst *et al.* (2002) recorded human scan paths when viewing a series of complex natural and artificial scenes. They found that human scan paths could be predicted quite accurately by stimulus saliency which was based on colour, intensity, and orientation. While the bottom-up approach was successful in predicting human fixation patterns for some tasks, it is inaccurate in predicting fixation patterns for an active task that uses meaningful stimuli (see, e.g., Oliva *et al.*, 2003; Turano, Geruschat, and Baker, 2003; Henderson *et al.*, 2007). For example, Turano *et al.* (2003) showed that a saliency model performed as accurately as a random model in predicting the scan paths of human subjects during a real-world activity. In contrast, they found that a model that used only top-down (i.e., knowledge-driven) gaze control outperformed the random model. Obviously, visual saliency alone cannot account for the human

fixation patterns when performing certain tasks. Similar results were found by Henderson *et al.* (2007) who analysed eye movements of subjects that viewed images of real-world scenes during an active search task. They found that a visual saliency model did not predict fixation patterns any better than a random model did. They concluded that visual saliency does not account for eye movements during active search and that top-down (i.e., knowledge-driven) processes play the dominant role.

Top-down gaze-control models

Whereas bottom-up gaze-control models use visual scene characteristics, top-down gaze-control models rely on stored knowledge and task demands to select the most relevant fixation locations (see, e.g., Henderson, 2003). In this chapter two types of top-down fixation-selection mechanisms were introduced relying on: (1) short-term episodic knowledge about previously attended image parts, and (2) long-term knowledge about relevant image parts, respectively.

The first type of top-down fixation-selection mechanism is based on short-term episodic knowledge about previously attended image parts to actively select the relevant fixation locations. This first type of top-down fixation selection mechanism that is employed by the classification process of both NIM-CLASS A and B, relates to the approach employed by probabilistic active vision models for classification (for a comparison of different active vision models see, de Croon, Sprinkhuizen-Kuyper, and Postma, 2006b). Both select visual input to reduce uncertainty about the class of a test item. The main difference between top-down gaze control in the active probabilistic models and top-down gaze control in the classification process of NIM-CLASS A and B concerns the nature and amount of knowledge that the mechanism uses to select relevant fixations. Active probabilistic models either consider all possible fixation selections at each time step (e.g., Borotschnig *et al.*, 1999; Denzler and Brown, 2002), consider all possible fixation selections on forehand (e.g., Arbel and Ferrie, 2001), or use a fixation selection policy that is acquired on the basis of an extensive training (e.g., reinforcement learning by Paletta, Prantl, and Pinz, 1998) or on the basis of an evolutionary algorithm (e.g., de Croon *et al.*, 2006a). In contrast, top-down fixation selection in the classification process of NIM-CLASS A and B relies solely on the feature vectors that were stored during one encounter with the class instance (in the storage process). A few other models have used a top-down approach to select relevant image parts for classification on the basis of limited short-term episodic knowledge (e.g., Rybak *et al.*, 1998). The main difference with these models concerns the explicit representation of movement sequences. The models often employ a separate 'where' (motor memory) system that uses fixed eye-movement programs acquired from previously viewing the image (e.g., Rybak *et al.*, 1998). In contrast, NIM-CLASS A and B actively construct fixation patterns during classification based on the short-term episodic knowledge about previous encounters with the faces (i.e., the stored labelled feature vectors), rather than relying on the eye-movement sequence that was performed during the first encounter.

The second type of top-down fixation-selection mechanism that we presented is based on stored knowledge about relevant face parts acquired through long-term experience with faces. This second type of top-down fixation selection mechanism,

that is employed by the storage process of NIM-CLASS B, relates to the probabilistic active vision models in the sense that fixation selection is based on an extensive set of built-up knowledge about the relevance of particular locations or eye movements. It is remarked that the top-down fixation selection employed by the storage process of NIM-CLASS B differs from these models because it does not consider the incremental beliefs, but rather selects fixations according to their general entropy in terms of all possible classes in the data set. According to the results presented in subsection 6.6.1, this improves performance compared to random fixation selection in the storage process.

6.7.2 Scalability of the models

In our studies we have not examined how the NIM-CLASS, the NIM-CLASS A, and the NIM-CLASS B classification performances scale up with the number of classes. Below we offer some perspective on the aspects that relate to the scalability of the models.

In our classification task, NIM-CLASS, and NIM-CLASS A and B deal with 130 objects (i.e., faces) coming from 10 different classes. Obviously, this limited number of objects can hardly be considered to be representative for the large number of objects that natural systems encounter in the real world. Ideally, a plausible classification or recognition model should be able to distinguish among large numbers of objects. However, since the different NIM-CLASS models store the complete encountered visual input, classification time is linear in the number of encountered objects (see also Bajramovic *et al.*, 2006). In order to address this problem, mechanisms can be incorporated that use the representation space in an efficient way and that ensure the maintenance of an efficient representation space. In NIM-CLASS A and B we introduced a mechanism that operates on the representation space in an efficient way by actively using the most relevant information in the representation space. In addition, in NIM-CLASS B we introduced a mechanism that maintains an efficient representation space that selects visual input on the basis of its relevance or informativeness. However, we would like to remark that the development of such relevance knowledge may be computationally challenging, since it is based on a large number of encountered items. Nevertheless, the building up of relevance knowledge may be assumed to operate over extended periods of time as it corresponds to the acquired long-term knowledge in human memory. Provided that the knowledge has been acquired and is used for fixation selection (as in NIM-CLASS B), this may lead to more discriminable class representations. Overall, we may assume that (1) NIM-CLASS A is more suitable than NIM-CLASS for upscaling to a larger number of classes because it uses the representation space in an efficient manner (in the classification process), and (2) NIM-CLASS B is even more suitable than NIM-CLASS A for upscaling to a larger number of classes, because it builds up an efficient representation space (in the storage process) and uses this space in an efficient manner (in the classification process).

Further extension may address the maintenance of an efficient representation space. For example, the storage process can be adapted in such a way that only relevant information is stored and retained. In human vision the brain continuously

makes predictions about the expected visual input at the new fixation location (see, e.g., Hawkins and Blakeslee, 2004). In a similar way a new NIM-CLASS variant may make predictions on the basis of long-term knowledge and then store only the new input that deviates significantly from the expectations (i.e., the relevant or informative input). In addition, the effectiveness of the representation space may be improved by forgetting stored information that is not addressed for a sufficiently long period of time. The neural implementation techniques described in subsection 7.2.3 can be used to realize a compact and efficient representation space even for large numbers of objects.

6.7.3 Comparison with existing classification models

Several other models have been applied for the classification of the faces of the AR data set. However, the existing models (1) generally leave out the faces with occlusions that appear to be the most difficult ones for many models (see, e.g., Lu, Wang, and Jain, 2003) or (2) are trained on more class instances than the one view that we used for training (see, e.g., Martinez and Kak, 2001; Wang *et al.*, 2005). An example of (2) is the following. Martinez and Kak (2001) compared the performances of a nearest-neighbour classifier operating on a representation space based on a Principal Component Analysis (PCA) and on a Linear Discriminant Analysis (LDA) of the pixel values of the entire AR images. Their classification task differed from ours in the number of classes and the number of training views. While their model was presented with 50 individuals (classes) rather than the 10 individuals that we selected, training was based on a larger set of two or even 13 views[3] rather than the one view that we used. Despite their larger set of training instances per class, performances were lower than the performances that we obtained with the different NIM-CLASS models. For the experiments with two training views, the maximum average classification performances of around 60% were obtained using PCA with 80 dimensions. For the experiments with 13 training views, the maximum average classification performances of around 87% were obtained using LDA. Although most models that were tested for classification on the AR data set were trained on more than one view per class or left out the unfavourable views for testing, a few studies used the same training and testing set as we did (i.e., one view for training and the remaining 12 views for testing) (e.g., Guillamet and Vitri, 2002). For example, Guillamet and Vitri (2002) used one view of each class for training and the remaining twelve for testing, when they compared the performances of different classification algorithms on the AR data set. In their study, they introduced the use of a Non-negative Matrix Factorization (NMF; see, e.g., Lee and Seung, 1999) in the context of classification and compared the performance of NMF with those of the widely applied PCA, and two influential techniques from computer vision, a feature-based technique based on Local Feature Analysis (LFA; Penev and Atick, 1996), and a Bayesian template-based technique (Moghaddam and Pentland, 1997). The techniques showed average performances for the faces without occlusions (i.e., views

[3]In addition to the series of 13 views of each individual, there was a second series of 13 views showing the same view but taken at another point in time; when Martinez and Kak (2001) used 13 training views, the 13 views of the second series were used for testing.

2 up to 7), of about 65.0% for the template-based technique, 74.0% for the PCA technique (using 150 dimensions), 85.0% for the NMF technique and 90.0% for the LFA-based technique. For a comparison, our NIM-CLASS, NIM-CLASS A, and NIM-CLASS B models showed an average performance for the faces without occlusions of 96.7%, 96.8%, and 98.5%, respectively, when we used $s = t = 100$ storage and test fixations[4] Average performances of each of the techniques tested by Guillamet and Vitri (2002) dropped substantially for the faces with occlusions, in particular for the faces with sunglasses for which the average performance was about 8.0% for the LFA-based technique, 22.0% for the PCA technique, 27.3% for the NMF technique and 32.3% for the template-based technique. For a comparison, our NIM-CLASS, NIM-CLASS A, and NIM-CLASS B models showed an average performance for the faces occluded with sunglasses of 85.5%, 87.9%, and 92.7%, respectively, when we used $s = t = 100$ storage and test fixations. The considerable drop in the classification performance of the techniques tested by Guillamet and Vitri (2002) for faces with sunglasses, demonstrates that the occluded parts contain important visual information for classification with these techniques. Occlusions are known to be problematic for techniques that construct global representations, such as PCA, rather than part-based representations (see, e.g., Lee and Seung, 1999; Fei-Fei and Perona, 2005). Although, the NMF technique and also the LFA technique are more part-based than the PCA and the template-based techniques, they still rely on global image characteristics to some degree. The performances of the NIM-CLASS variants that rely on discrete local samples across the images, appear to be less disrupted by an occlusion of the eyes even when testing under different lighting conditions. Therefore, it can be said that, despite occlusions, the three NIM-CLASS variants can make the correct classification decision on the basis of the local samples from image regions other than the occluded regions.

6.8 Chapter summary and conclusions

This chapter started with the adaptation of NIM to obtain a natural input model for classification called NIM-CLASS. NIM-CLASS was tested on a classification task that used frontal view faces with different facial expressions, illumination conditions, and occlusions. From the NIM-CLASS classification results we may conclude that NIM-CLASS is able to classify natural images of frontal faces correctly under a variety of potentially difficult conditions provided that a sufficient number of fixations is made by the storage process and by the classification process.

After assessing the NIM-CLASS classification performance, we examined how NIM-CLASS could be adapted for top-down fixation selection to select relevant fixation locations on the basis of stored knowledge. We proposed two NIM-CLASS variants, called NIM-CLASS A and NIM-CLASS B. NIM-Class A employs top-down fixation-selection mechanism in the classification process on the basis of short-term episodic knowledge about previously encountered faces. The selection mechanism selects locations for fixation that minimize the uncertainty of classification. From

[4]Since Tables 6.1, 6.2, and 6.3 show average performances across all test views, these values are not found in the tables.

the NIM-CLASS A classification results we may conclude that the top-down fixation selection employed by the classification process improves performance on the classification task compared to the NIM-CLASS performance. This is particularly so when a limited number of fixations is made by the classification process.

NIM-CLASS B adopts the top-down fixation-selection mechanism of NIM-CLASS A in the classification process and, in addition, employs a top-down fixation-selection mechanism in the storage process based on long-term stored knowledge about the relevance of different face parts. The NIM-CLASS B classification results demonstrate that the top-down fixation-selection mechanism employed by the storage process improves the performance on the classification task compared to the NIM-CLASS A performance. NIM-CLASS B stores visual input that is more relevant for classification than the visual input that NIM-CLASS A stores. This enables the top-down fixation-selection mechanism in the classification process to operate more efficiently in NIM-CLASS B than in NIM-CLASS A.

On the basis of the results presented in this chapter we may conclude that NIM-CLASS is able to classify faces correctly under a variety of unfavourable conditions on the basis of one encounter (i.e., one stored view). Moreover, NIM-CLASS A and B are able to do this on the basis of a limited number of storage and test fixations by using top-down fixation selection to extract relevant visual input. All in all, we may conclude that the NIM-CLASS A and B variants underline the importance of active top-down approaches for obtaining plausible cognitive models for the classification of natural visual input.

Chapter 7

Towards a plausible model of cognition

In this thesis we have proposed a situated computational memory model by realizing a perceptual front-end that connects a computational memory back-end with the natural visual environment. NIM provides an initial step towards the development of a plausible computational model of cognition. This chapter discusses our approach by placing it in the context of existing computational models of object recognition (7.1), by identifying points for the improvement of NIM's psychological and biological plausibility as a model of cognition (7.2), and by relating it to the global developments in cognitive modelling (7.3).

7.1 NIM and models of object recognition

By realizing a computational memory model that operates on realistic input, NIM brings together elements of two early influential approaches to object recognition: the ecological approach (e.g., Gibson, 1979) and the computational approach (e.g., Marr, 1982).

In section 1.1, we sketched a brief historical overview of computational modelling of cognition. We pointed at the recent trend towards the development of systems that can cope with realistic environmental demands. This trend has led to a reappraisal of the ecological approach formulated by Gibson (1979) in cognitive science. At the time Gibson introduced his ideas, he was criticised mainly because his views could not be translated easily into computational algorithms. In contrast, Marr's (1982) computational framework (see subsection 2.2.1) provided a clear computational formulation of the processing stages assumed to underlie object recognition.

It is clear that the object-recognition framework by Marr (1982) and the more recent recognition-by-components theory by Biederman (1985, 1987) fit nicely with the information-processing accounts of cognition. According to Marr and Biederman, the visual system performs the information-processing task that results in mental representations that are isomorphic to objects in the world, i.e., a first-order isomor-

phism (see, e.g., Shepard, 1975, 1981). In that way, processing can be directed purely to the internal representations without the need to acquire input from the external environment. Although the mental reconstruction of the environment in terms of mental representations seems highly desirable from a computational point of view, it is rather implausible from a psychological and neurophysiological point of view. Moreover, computational attempts to create first-order isomorphic representations from natural images turned out to be infeasible (see, e.g., Edelman, 2002; Edelman and Intrator, 2003).

Obviously, NIM deviates from the first-order isomorphism characteristic of Marr's and Biederman's models by transforming natural images into second-order isomorphic representations. In that respect NIM is inspired by Shepard's (e.g., Shepard, 1957, 1964; Shepard and Chipman, 1970) and Edelman's (e.g., Edelman, 1995a; Edelman and Duvdevani-Bar, 1997a; Edelman and Intrator, 1997; Edelman, 1998) similarity-based approach. The main advantage of the similarity-space approach is that it does not require a detailed reconstruction of the environment. Instead only the similarities have to be reconstructed. Edelman (see, e.g., Edelman and Duvdevani-Bar, 1997b; Edelman, 1998) and others (see, e.g., Busey, 1998; Goldstone and Steyvers, 2001) have shown the feasibility of the similarity-based approach by representing faces and objects. Below, we will elaborate on Edelman's approach and remark how he addressed the main critique on his original approach.

Edelman's original similarity-space approach is based on global shape templates. Notwithstanding its successes, this approach has been criticized to ignore the intrinsic spatial structure of objects. More specifically, the part-whole relations of an object and its constituents is not addressed by the representations. For instance, in his view-interpolation model (e.g., Edelman and Duvdevani-Bar, 1997b) objects are recognized by interpolating over multiple 2D views of the object. Each incoming object is then represented by the responses of an ensemble of stored prototypical object views (therefore this representational scheme is called the 'chorus of prototypes', see also subsection 2.2.3). Ignoring the intrinsic spatial structure of objects limits the plausibility of the interpolation model as a general model for visual representation. Edelman attempted to address this limitation with his 'chorus of fragments' (e.g., Edelman and Intrator, 2001, 2003). In this model, objects are represented in terms of templates of shapes tuned to specific locations (e.g., Edelman and Intrator, 2001) rather than in terms of templates of entire object views (as in his original 'chorus of prototypes'). The templates mimic the what and where receptive fields that support the treatment of spatial structure in the human visual system (see, e.g., Palmer, 1999). Edelman argues that a representation space that implicitly codes for parts and their spatial relations (i.e., compositionality; Bienenstock and Geman, 1995; Bienenstock, Geman, and Potter, 1997) by means of shape-specific and location-specific fragments (i.e., what and where similarity-space representations) provides an effective computational and biologically plausible alternative to the structural descriptions proposed by Marr and Biederman (see, e.g., Edelman, 2002; Edelman and Intrator, 2003). Moreover, this alternative approach is less sensitive to disruptions and variations compared with the classical structural object descriptions, because none of the individual fragments is critical for the entire representation (e.g., Nelson and Selinger, 1998; Edelman and Intrator, 2001; Edelman

and Intrator, 2003). So, while the Edelman was criticised for disregarding spatial structure in his original global shape-template approach, his 'chorus of fragments' addresses this limitation in a computational and biologically inspired manner.

Since Nim employs similarity-based representations, it may fall prey to the same criticism as Edelman's original model. However, Nim represents objects in terms of a number of preprocessed image fragments. In that respect it resembles Edelman's 'chorus of fragments'. Although Nim does not directly represent the spatial composition of objects and their parts, the fixation mechanism as discussed in chapter 6 may be considered to exploit the spatial relations for two reasons. First, in Nim spatial relations are implicitly represented in the overlap structure of fixations (see also subsection 7.2.4). Second, by retaining the fixation locations and the corresponding visual input, the spatial relations can be exploited and (oculomotor) action-perception patterns can be learned (see also subsection 7.2.2). For this reason, Nim can be extended into a model that closes the perception-action loop that is characteristic of human vision within a natural environment (chapter 6 illustrated our initial attempts to approach the interactive visual behaviour of natural systems by implementing an active vision mechanism).

In summary, Nim's perceptual front-end combines elements from the different existing approaches to realize a computational mechanism that builds a representation space from real-world input. Nim operates on image parts and builds up representations from the parts (see, e.g., Biederman, 1985; Edelman and Intrator, 2003). At the same time it adheres to the ecological approach by emphasizing the role of the environment and by incorporating a natural eye-fixation mechanism characteristic of natural vision.

Nim is inspired by neurocognitive and psychophysical findings. Still, several well-known neurocognitive and psychophysical findings about visual object recognition in the brain are not addressed by Nim and might serve as a basis for improving Nim's plausibility and realism as a model of object recognition and memory.

7.2 Nim's psychological and biological plausibility

In its present form, Nim is simple and straightforward. Enhancing Nim's psychological and biological plausibility requires the model to be extended in agreement with psychological and biological insights. In the subsections below, we discuss five of such extensions: a feature-based attentional mechanism (7.2.1), a spatial attentional mechanism (7.2.2), a neural implementation of the similarity space (7.2.3), the representation of spatial knowledge (7.2.4), and the incorporation of separate episodic and semantic representation spaces (7.2.5).

7.2.1 Feature-based attention

The first extension is the task-dependent adaptation of Nim's representations by stretching or shrinking the axes of its similarity space. It has been shown that humans attend to different features depending on task-related factors (e.g., Nosofsky, 1987; Goldstone and Steyvers, 2001; Halberstadt, Goldstone, and Levine, 2003) or depending on individual differences in perceptual or attentional processes (e.g.,

Viken *et al.*, 2002; Halberstadt *et al.*, 2003). NIM might be extended to incorporate a feature-based attentional mechanism that adjusts the similarity space by assigning the task-relevant dimensions of the similarity space a higher weight than the task-irrelevant dimensions. Feature weighting can be based on: (1) psychological knowledge about task-relevant dimensions, (2) specific human experimental data, and (3) training examples to find the weights automatically by using machine-learning techniques. We briefly discuss each of these three types of feature weighing.

The first type of feature weighing is based on psychological knowledge about task-relevant dimensions obtained in behavioural studies. Several studies examined the role of information at different spatial scales for performing a certain task (see, e.g., Goffaux *et al.*, 2005; Sowden and Schyns, 2006). For example, Sowden and Schyns (2006) found that observers actively and dynamically attend to visual information at different spatial scales depending on the task. This knowledge can be incorporated into NIM's preprocessing stage in such a way that the scale-specific parts of the vector are weighed in agreement with the psychological knowledge.

Second, the feature weights can be defined by human data obtained in a specific behavioural similarity-rating or categorization study. For example, Steyvers and Busey (2000) proposed to improve the psychological plausibility of their extracted feature-vector representations for a set of facial images by means of the human similarity-ratings for the same images. In order to do so, they proposed a feature mapping model to map their feature-vector representations to the MDS representations that were derived from the human similarity ratings. In a similar way, NIM's representations may be fed into a feature mapping model to adjust the weights according to human similarity judgements.

A third manner of feature weighing is to learn the weights from examples using feature-weighing techniques from the domain of machine learning. These learning techniques adjust the feature weights in such a way that the generalization or the classification performance of, e.g., a nearest-neighbour method is optimized (see, e.g., Lowe, 1993; Wettschereck, Aha, and Mohri, 1997; Paredes and Vidal, 1999; Stahl, 2005). Feature-weighing learning techniques may be applied to NIM's representations to optimize, for example, its classification performance of natural stimuli.

Employing feature weighing to NIM's similarity-space representations might result in more plausible representations than those obtained without considering the psychological and task-dependent relevance of the different dimensions.

7.2.2 Spatial attention

The second extension of NIM addresses the spatial selection of interesting regions for visual information extraction. Human visual selection is highly context dependent (e.g., Rajashekar *et al.*, 2002). When interacting within the environment, covert attention directs the eyes to interesting regions on the basis of bottom-up and top-down processes (Karn and HayHoe, 2000; Henderson, 2003; Oliva *et al.*, 2003) (see also section 6.4). NIM incorporates the natural characteristic of directing spatial attention by fixating different image regions (i.e., overt attention). NIM, NIM-REM, and NIM-CLASS, select eye fixations along the contours in the image, independent of the current or past states of the system. Inspired by human vision, chapter 6

introduced Nim-Class A and Nim-Class B, that employ two types of fixation-selection mechanisms for the selection of image regions on the basis of both bottom-up processes (that rely on salient features) and top-down processes (that rely on stored knowledge). Still, both models are limited in that they consider a confined subset of the types of stored knowledge known to play a role in human eye fixation selection. Future extensions should address psychologically more plausible fixation-selection mechanisms that are based on different types of short-term, long-term, and task-related knowledge. Such forms of knowledge require an extension of Nim's representation space or the addition of separate specialised representation spaces. Considering the task-dependent nature of human fixation patterns, the fixation-selection mechanisms should be tested on a variety of different tasks and be compared with fixation patterns of humans engaged in such tasks.

7.2.3 Neural implementation

The third extension of Nim deals with a neural implementation of Nim's similarity space. While Nim's feature-vector extraction is inspired by the human visual system, the memory stage retains the preprocessed input in a straightforward representation space (as is common for many computational memory models) that does not resemble the neural realization of representations in the brain. A more biologically plausible representation space may be preferred for a model of human memory. Several biologically inspired types of representation spaces may be considered as suitable alternatives for Nim's similarity space, such as self-organizing maps (e.g., Kohonen, 2001) and radial basis function networks (e.g., Edelman and Intrator, 2001; Buhmann and Ablowitz, 2003). A self-organizing map is a single-layer feedforward network where the output neurons are arranged in a low-dimensional grid. The learning algorithm of the self-organizing map ensures that adjacent neurons in the grid respond to similar inputs. The resulting 'topological maps' resemble the neural topological maps found throughout the brain (see, e.g., Knudsen, Du Lac, and Esterly, 1987). To build a neurally plausible representation space for Nim, the feature vectors extracted by Nim can be fed directly into a self-organizing map. Another way of enhancing the biological realism of Nim's feature-vector representations is to use radial basis function (Rbf) networks. These networks model the response patterns of populations of neurons throughout the brain (e.g., Buhmann and Ablowitz, 2003). An Rbf network is a similarity-based feedforward network that employs nonlinear (Gaussian or other radial) transfer functions to weigh the distance of an input vector with a stored prototype. An example of such a network is Edelman's 'chorus of fragments' (see also section 7.1) that learns the what-and-where responses on the basis of selected image fragments (see, e.g., Edelman and Intrator, 2001).

Although self-organizing maps and Rbf networks realize biologically plausible alternatives to structure Nim's feature-vector representations, more biological realism can be obtained by modelling Nim's similarity space in terms of spiking neural networks (for a review, see Gerstner, 1998). The computational advantages of such networks over the conventional neural implementations remain to be determined.

7.2.4 Representation of spatial knowledge

The fourth extension of NIM addresses the representation of spatial knowledge. In chapter 6, we introduced the NIM-CLASS A and B models that use stored visual as well as spatial knowledge from previously attended image regions in order to direct gaze towards relevant locations. NIM-CLASS A and B represent spatial knowledge explicitly by storing the coordinate labels of the fixation locations. This accords with neurobiological evidence for a separate 'where' system in the brain that keeps track of the spatial origin of visual information (see, e.g., Mishkin and Appenzeller, 1987; Courtney *et al.*, 1996; Broadbent, Squire, and Clark, 2004). However, rather than the separate storage of coordinate labels, the spatial origin of the input as well as the spatial relations between different inputs may be inferred from the input itself. Natural images, such as the face images in our studies, contain spatial regularities at two levels: (1) regularities within a natural image, and (2) regularities between the natural images of the same category (Torralba and Oliva, 2003). Within an image, visual input originating from adjacent spatial locations leads to similar feature vectors due to the correlation structure of natural images (see, e.g., Olshausen and Field, 1996; Karklin and Lewicki, 2003). Hence, the similarity-space representations provide cues for both the identity and the (relative) spatial location of the input. For images of the same category (e.g., face images), visual input coming from the same (relative) spatial location in two different images lead to similar feature vectors due to the spatial regularities of objects within one category. The knowledge about the general spatial arrangement of a certain category of stimuli or scenes is used by the top-down gaze control processes in human vision (i.e., scene-schema knowledge; Henderson, 2003). In the same way, NIM can derive scene-schema knowledge from extensive experience with objects from one category. The results obtained in a separate preliminary study with a classification variant of NIM showed that knowledge about the spatial arrangement of objects within a particular category can be extracted from the similarity-space representations (Lacroix, Postma, and Murre, 2006b). Moreover, the spatial knowledge could be used to direct the eyes towards relevant fixation locations (cf., Bergboer, Postma, and van den Herik, 2004; de Croon *et al.*, 2006a; de Croon and Postma, 2007; Bergboer, to appear).

While there might be good neurobiological evidence to support separate what and where systems, it is very well possible that in some situations the brain infers the 'where' information from the 'what' information, particularly where it concerns well-known objects or categories of objects with a high degree of spatial regularity, such as the category of faces. An extended NIM should incorporate both a separate representation of 'where' information to direct actions (fixations) and at the same time exploit the implicit spatial knowledge contained in the similarity-space representations.

7.2.5 Episodic and semantic representations in the brain

The fifth NIM extension is the incorporation of separate episodic and semantic representation spaces. NIM's perceptual front-end builds a perceptual similarity space from the visual input. The preprocessed visual inputs are stored as instances in memory on the basis of a single encounter. This type of learning is characteristic

of episodic human memory that relies on the hippocampus for fast one-shot learning (e.g., Teyler and Discenna, 1985; Zola-Morgan, 1990; Squire, 1992). It may be assumed that the similarity structure of newly encountered faces can be explained largely on the basis of instances of perceptual representations. This assumption is supported by the results from the similarity-rating experiments in chapter 4. For words or concepts, the similarity structure is assumed to be based much less on perceptual features, but instead relies mainly on conceptual features. In contrast to the one-shot learning that underlies episodic memory, the development of conceptual (i.e., semantic) memory is generally believed to result from a slow and gradual learning process based on multiple episodes (see, e.g., McClelland, McNaughton, and O'Reilly, 1995). As a consequence, semantic representations contain general knowledge that is not referenced to specific events. This contrasts with episodic memories that, by definition, refer to an event or episode. NIM may be extended to employ both a fast learning episodic memory system and slower learning semantic or conceptual memory system that builds up over time on the basis of the stored episodic representations. Such an extension can be realized by implementing a memory consolidation mechanism that is assumed to underlie the distinction and communication between fast acquired episodic memories and gradually acquired semantic memories (see, e.g., Teyler and Discenna, 1985; Alvarez and Squire, 1994; McClelland *et al.*, 1995). According to the theory of memory consolidation, recent acquired (episodic) memories are crystallized into long-term memories over time (see, e.g. Murre, 1996; Meeter and Murre, 2004; Meeter and Murre, 2005). Neuropsychological and neurophysiological studies revealed that two brain structures are involved in the process of memory-consolidation, i.e., the hippocampal structure and the neocortical structure. While memories are initially dependent on the fast-learning hippocampal structure, they are eventually transported into long-term memory that has been recognized to reside in the neocortical structure. Neurophysiological and psychological sleep studies point out that the consolidation process is based on a reactivation of stored memories in the hippocampus which occurs mainly during sleep (see, e.g., Squire, 1992; Walker *et al.*, 2003; Nelson, 2004; Stickgold, 2005). Results from connectionist memory-modelling research support the idea of two complementary learning systems in the brain and the transfer of information from one system to the other (see, e.g., Gluck and Meyers, 1993; Alvarez and Squire, 1994; McClelland *et al.*, 1995; Lacroix, 2001). An extended NIM may incorporate a consolidation mechanism that gradually generalizes the acquired episodic instances into robust conceptual memories. Several researchers argued that the conceptual representations that underlie natural language are derived from the interaction with the real world (see, e.g., Barsalou, 1999; Glenberg and Kaschak, 2002; Goldstone, Feng, and Rogosky, 2005; Roy, 2005a, 2005b).

Having discussed five biologically inspired extensions of NIM, in the next subsection we position NIM in the broader context of the global developments of models of cognition.

7.3 Global developments in modelling cognition

Traditional psychological and artificial intelligence models of natural cognition focussed mainly on symbol manipulation associated with higher-level cognitive tasks. As we argued in chapter 1, these symbol-manipulation models suffer from the symbol grounding problem (Harnad, 1990) and the transduction problem (Barsalou, 1999) by disregarding the connection between the representations and their real-world referents. For the domain of artificial intelligence, the acknowledgement of these problems led to a fundamentally different approach to the building of intelligent systems. Rather than focussing on higher-level intelligent behaviour by means of computations on knowledge represented by symbols, the new so-called 'situated' approach focusses on low-level behaviours in situated 'agents' acting within an environment (see, e.g., Beer, 1995; Clancey, 1997; Arkin, 1998; Pfeifer and Scheier, 1999). In Artificial Intelligence, the approach to cognition thus moved from the study of isolated, representation-rich, symbol manipulation systems, to the study of the dynamics of simple agents in interaction with their environments. The development in the domain of cognitive psychology shows a different pattern. After the cognitive science revolution, models addressing high-level cognitive behaviour prevailed (Gardner, 1985). This resulted in a gap between cognitive modelling in the domain of artificial intelligence and psychology. Currently, these domains are slowly moving towards each other. On the one hand situated artificial intelligence is exploiting the advantages of extending situated agents to incorporate more complex cognitive structures such as memory systems (see, e.g., the PACO-PLUS project, http://www.paco-plus.org/). On the other hand, the findings from the situated artificial intelligence approach inspire new psychological models that acknowledge the essential role of the environment for understanding cognition.

In line with these global developments in cognitive modelling, NIM constitutes a straightforward situated computational memory model. It combines insights from computational memory modelling in the domain of psychology with those from the domain of artificial intelligence by realizing a computational memory model that operates directly on the natural visual environment. In doing so, it departs from the traditional computational memory models that consider the cognitive system as detached from its environment. Rather than focussing on isolated cognitive processes, NIM brings about a coupling between cognition and the environment. The selection of visual input by means of eye fixations provides a plausible basis to model the interactive visual behaviour of natural systems.

Future extensions of NIM should address the continuous interaction between perception, cognition, and action. Similarly, other models and experimental methods in cognitive science could be extended or adapted to become more situated. The trend towards more realism is already observed in some experiments (see, e.g., Cohen and Conway, 1996; Henderson, 2005). A similar trend in cognitive modelling is needed. Acknowledging the importance of the interaction with a realistic environment may lead to a radically different conception of mental representations (see, e.g., Barsalou, 1999; van Dartel *et al.*, 2005; van Dartel, 2005) and cognitive mechanisms (see, e.g., Pfeifer and Scheier, 1999; Beer, 2003) required for the realization of a plausible model of natural cognition.

Chapter 8

Conclusion

In this thesis we addressed the main limitation of computational memory models, which is their lack of a connection with the real world. Because these models do not incorporate an encoding process that creates memory representations from real-world input, they suffer from the symbol grounding and the transduction problems. This led us to the formulation of the central problem statement of this thesis in chapter 1: *How can computational memory models be extended to solve the symbol grounding and transduction problems?* Subsequently, we approached the problem statement by attempting to construct and validate a situated computational memory model. For the construction of a situated computational memory model, we proposed to develop a perceptual front-end that transforms natural input into memory representations that can be operated on by a computational memory back-end. In order to achieve this objective, chapter 2 provided three guiding principles for creating a veridical representation space from real-world visual input and showed how the principles can be fulfilled in a perceptual front-end that transforms natural visual input into memory representations. Subsequently, in chapter 3, we presented a situated computational memory model, called Nim, that combines the perceptual front-end with a computational memory back-end.

For the validation of the model, we focussed on answering three research questions. Below, we answer our research questions on the basis of the work presented in the previous chapters. Subsequently, we formulate our conclusion on the problem statement.

8.1 Answers to our research questions

On the basis of the proposed situated model, we aimed at answering three research questions that will be addressed in subsections 8.1.1, 8.1.2, and 8.1.3.

8.1.1 RQ 1: on producing individual responses

Below, we repeat our first research question.

To what extent can a situated model produce human responses to individual natural visual stimuli?

Chapter 4 addressed the first research question. There, we compared NIM responses with human responses to individual natural stimuli in a similarity-rating task and a recognition-memory task. On the basis of the results we may conclude that we are able to produce human responses to individual natural stimuli quite reliably with our situated model. NIM produced the overall pattern of human similarity-rating data and recognition-memory data.

8.1.2 RQ 2: on producing recognition-memory effects

Below, we repeat our second research question.

To what extent can a situated model produce recognition-memory effects on the basis of natural visual stimuli?

Chapter 5 addressed the second research question by introducing a NIM variant called NIM-REM, which extended NIM into a natural input variant of the powerful REM model proposed by Shiffrin and Steyvers (1997). The chapter validated NIM-REM on four general recognition-memory effects that are typically obtained in behavioural memory studies. From the results we may conclude that NIM-REM is able to explain the recognition-memory effects to the extent that it produces the four effects successfully directly on the basis of natural visual input.

8.1.3 RQ 3: on classification

Below, we repeat our third research question.

To what extent can a situated model classify natural visual stimuli?

Chapter 6 addressed the third research question by introducing a classification variant of NIM called NIM-CLASS, which extended NIM into a classification model for natural stimuli. In addition to NIM-CLASS, the chapter introduced two variants of NIM-CLASS, called NIM-CLASS A and NIM-CLASS B, that extended NIM with a gaze control mechanism for the selection of relevant visual input based on top-down processes. NIM-CLASS A employed top-down fixation selection during classification on the basis of short-term episodic knowledge about previously encountered natural stimuli. NIM-CLASS B featured the top-down fixation selection of NIM-CLASS A during classification and, in addition, employed top-down fixation selection during storage based on long-term stored knowledge about the relevance of different image parts. NIM-CLASS and NIM-CLASS A and B, were validated in a classification task and a qualitative comparison was carried out between model and human performances. From the NIM-CLASS classification results we may conclude that NIM-CLASS is able to classify natural images correctly under a variety of potentially disturbing conditions provided that a sufficient amount of visual input was

selected for storage and classification. From the Nim-Class A classification results we may conclude that top-down fixation selection during classification improves performance on the classification task compared to the Nim-Class performance. This is particularly so when a limited number of fixations are taken during classification. The Nim-Class B classification results demonstrate that top-down fixation selection during storage yields an improved performance on the classification task when compared to the Nim-Class A performance.

8.2 Conclusion

Below we formulate our conclusion of the problem statement: *How can computational memory models be extended to solve the transduction and grounding problems?*

The situated computational memory model called Nim introduced in chapter 3 (and also the variants presented in chapters 5 and 6) provides an implicit answer to the problem statement. Nim explains how representations originate from the real world by employing a front-end that transforms real-world input into memory representations in a neurobiologically informed manner. The symbol grounding and the transduction problems address the lack of a direct relation between representations and the objects that they refer to in the external world. By creating representations from natural images, the relation between a representation and its referent is established. As a consequence, the symbol grounding and transduction problems are solved. The model validation studies demonstrate that Nim is quite successful in explaining human data from behavioural memory studies. However, the model is limited with respect to the psychological and biological realism and may be further improved as outlined in chapter 7.

Finally, we may conclude that the situated computational memory model presented in this thesis provides an efficient way to solve the symbol grounding and transduction problems. Moreover, it offers a fruitful basis for developing a psychologically plausible model of natural cognition.

References

Alvarez, P and 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*, Vol. 91, pp. 7041–7045. [105]

Anderson, J. R. (1976). *Language, memory, and thought*. Lawrence Erlbaum Associates, Hillsdale, NJ. [10]

Anderson, J. R. (1993). *Rules of the mind*. Lawrence Erlbaum Associates, Hillsdale, NJ. [1]

Arbel, T. and Ferrie, F. P. (2001). Entropy-based gaze planning. *Image and Vision Computing*, Vol. 19, pp. 779–786. [94]

Arkadev, A. G. and Braverman, E. M. (1966). *Computers and pattern recognition*. Thompson, Washington, DC. [15]

Arkin, R. C. (1998). *Behaviour-based robotics*. The MIT Press, Cambridge, MA. [106]

Atkeson, C. G., Hale, J. G., Pollick, F., Riley, M., Kotosaka, S., Schaal, S., Shibata, T., Tevatia, G., Ude, A., Vijayakumar, S., and Kawato, M. (2000). Using humanoid robots to study human behavior. *IEEE Intelligent Systems*, Vol. 15, pp. 46–56. [4]

Bajramovic, F., Mattern, F., Butko, N., and Denzler, J. (2006). A comparison of nearest neighbor search algorithms for generic object recognition. *Proceedings of the Advanced Concepts for Intelligent Vision Systems (ACIVS 2006)*, pp. 1186–1197. [95]

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, Vol. 1, pp. 295–311. [26]

Barrington, L., Marks, T. K., and Cottrell, G. W. (2007). NIMBLE: A kernel density model of saccade-based visual memory. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society (CogSci 2007)*. [52, 78]

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, Vol. 22, pp. 577–660. [3, 4, 10, 105, 106]

Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, Vol. 72, pp. 173–215. [106]

Beer, R. D. (2003). Arches and stones in cognitive architecture. *Adaptive Behavior*, Vol. 11, pp. 299–305. [106]

Bellman, R. (1961). *Adaptive control processes: A guided tour*. Princeton University Press, Princeton, NJ. [24, 26]

Bergboer, N. H. (to appear). *Context-based image analysis*. Ph.D. thesis, Maastricht University, Maastricht, The Netherlands. [104]

Bergboer, N. H., Postma, E. O., and van den Herik, H. J. (2004). A context-based model of attention. *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004)*, pp. 927–931, IOS Press, Valencia, Spain. [104]

Biederman, I. (1985). Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Understanding*, Vol. 32, pp. 29–73. [13, 14, 99, 101]

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, Vol. 94, pp. 115–147. [13, 14, 99]

Bienenstock, E. and Geman, S. (1995). Compositionality in neural systems. *Handbook of brain theory and neural networks* (ed. M. A. Arbib), pp. 223–226. The MIT Press, Cambridge, MA. [100]

Bienenstock, E., Geman, S., and Potter, D. (1997). Compositionality, MDL priors, and object recognition. *Advances in Neural Information Processing Systems (NIPS)* (eds. M. C. Mozer, M. I. Jordan, and T. Petsche), Vol. 9, pp. 838–844. The MIT Press, Cambridge, MA. [100]

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press, Oxford, UK. [24, 26]

Blakemore, C. and Campbell, F. W. (1969). On the existence of neurons in the human visual system selectively responsive to the orientation and size of retinal images. *Journal of Physiology*, Vol. 203, pp. 237–260. [19]

Borotschnig, H., Paletta, L., Prantl, M., and Pinz, A. (1999). A comparison of probabilistic, possibilistic, and evidence theoretic fusion schemes for active object recognition. *Computing*, Vol. 62, pp. 293–319. [94]

Braun, J., Koch, C., Lee, D. K., and Itti, L. (2001). Perceptual consequences of multilevel selection. *Visual attention and cortical circuits* (eds. J. Braun, C. Koch, and J. L. Davis), pp. 215–241. The MIT Press, Cambridge, MA. [85]

Broadbent, N. J., Squire, L. R., and Clark, R. E. (2004). Spatial memory, recognition memory and the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, pp. 14515–14520. [104]

Brooks, L. R. (1978). Non-analytic concept formation and memory for instances. *Cognition and categorization* (eds. E. Rosch and B. Lloyd). Lawrence Erlbaum Associates, Hillsdale, NJ. [10]

Buhmann, D. M. and Ablowitz, M. J. (2003). *Radial basis functions : Theory and implementations*. Cambridge University Press, Cambridge, UK. [103]

Burgess, C., Livesay, K., and Lund, K. (1998). Explorations in context space: Words, sentences, discourse. *Discourse Processes*, Vol. 25, pp. 211–257. [11, 30]

Burton, A. M., Bruce, V., and Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, Vol. 23, pp. 1–31. [31]

Burton, A. M., Jenkins, R., Hancock, P. J. B., and White, D. (2005). Robust representation for face recognition: The power of averages. *Cognitive Psychology*, Vol. 51, pp. 256–284. [79, 84]

Busey, T. A. (1998). Physical and psychological representations of faces: Evidence from morphing. *Psychological Science*, Vol. 9, pp. 476–482. [11, 12, 35, 36, 37, 39, 100]

Busey, T. A. (2001). Formal models of familiarity and memorability in face recognition. *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (eds. M. Wenger and J. Townsend), pp. 147–192. Lawrence Erlbaum Associates, Hillsdale, NJ. [11, 12, 15, 28, 30, 31]

Busey, T. A. and Arici, A. (in preparation). Dissociations of accuracy and confidence reveal the role of individual items and distinctiveness in recognition memory and metacognition. Manuscript in preparation. [35, 36, 37, 39, 40, 41, 42, 45, 46, 47]

Busey, T. A. and Tunnicliff, J. (1999). Accounts of blending, typicality and distinctiveness in face recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, Vol. 25, pp. 1210–1235. [6, 12, 28, 30, 32, 35, 39, 40, 41, 42, 44, 45, 46, 47, 53, 68, 69, 70]

Cabeza, R., Bruce, V., Kato, T., and Oda, M. (1999). The prototype effect in face recognition: extension and limits. *Memory and Cognition*, Vol. 27, pp. 139–151. [53, 68]

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., and Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, Vol. 41, pp. 1179–1208. [28, 31, 40]

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, pp. 679–698. [18, 27]

Caramazza, A., Hersch, H., and Torgerson, W. S. (1976). Subjective structures and operations in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, Vol. 15, pp. 103–117. [11, 31]

Cary, M. and Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language*, Vol. 49, pp. 231–248. [53, 54, 55, 62, 65, 74]

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, Vol. 4, pp. 170–178. [85, 86]

Clancey, W. J. (1997). *Situated cognition: On human knowledge and computer representations*. Cambridge University Press, Cambridge, UK. [4, 106]

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, Vol. 3, pp. 345–351. [3]

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ, 2nd edition. [57]

Cohen, G. and Conway, M. A. (1996). *Memory in the real world*. The Psychology Press, London, UK, 3rd edition. [106]

Comon, P. (1994). Independent component analysis–A new concept? *Signal Processing*, Vol. 36, pp. 287–314. [19]

Cottrell, G. W., Bartell, B. T., and Haupt, C. (1990). Grounding meaning in perception. *Proceedings of the 14th German Workshop on Artificial Intelligence*, pp. 307–321, Springer-Verlag, London, UK. [1]

Courtney, S. M., Ungerleider, L. G., Keil, K., and Haxby, J. V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cerebral Cortex*, Vol. 6, pp. 39–49. [104]

Criss, A. H. and Shiffrin, R. M. (2004). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory and Cognition*, Vol. 32, pp. 1248–1297. [3]

Dailey, M. N., Cottrell, G. W., and Busey, T. A. (1999). Facial memory is kernel density estimation (almost). *Advances in Neural Information Processing Systems (NIPS)* (eds. M. J. Kearns, S. A. Solla, and D. A. Cohn), Vol. 11, pp. 24–30. The MIT Press, Cambridge, MA. [32, 36, 39, 41, 46, 47]

Dailey, M. N., Cottrell, G. W., Padgett, C., and Adolphs, R. (2002). A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience*, Vol. 14, pp. 1158–1173. [26, 28, 31, 32, 40]

Davies, G. M., Shepherd, J. W., and Ellis, H. D. (1979). Similarity effects in face recognition. *American Journal of Psychology*, Vol. 92, pp. 507–523. [68]

de Croon, G. and Postma, E. O. (2007). Sensory-motor coordination in object detection. *Proceedings of the IEEE Symposium on Artificial Life*, pp. 147–154. [104]

de Croon, G., Postma, E. O., and van den Herik, H. J. (2006a). A situated model for sensory-motor coordination in gaze control. *Pattern Recognition Letters: Special Issue on Evolutionary Computer Vision and Image Understanding*, Vol. 27, pp. 287–314. Guest Editor G. Olague. [4, 94, 104]

de Croon, G., Sprinkhuizen-Kuyper, I. G., and Postma, E. O. (2006b). Comparing active vision models. Technical Report 06-02, MICC-IKAT, Universiteit Maastricht. [94]

de Valois, R. L. and de Valois, K. K. (1988). *Spatial vision*. Oxford University Press, New York, NY. [20, 21]

Dennis, S. and Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, Vol. 108, pp. 452–478. [3, 10, 62]

Denzler, J. and Brown, C. M. (2002). Information theoretic sensor data selection for active object recognition and state estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, pp. 145–157. [94]

Derweester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, Vol. 41, pp. 391–407. [30]

Dewhurst, S. A. and Farrand, P. (2004). Investigating the phenomenological characteristics of false recognition for categorised words. *European Journal of Cognitive Psychology*, Vol. 16, pp. 403–416. [53, 68]

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern classification*. Wiley & Sons Inc., New York, NY. [52, 78, 79]

Ebbinghaus, H. (1885/1964). *Memory: A contribution to experimental psychology*. Dover Publications, New York, NY. (Original work published in 1885). [1, 2]

Edelman, S. (1995a). Representation, similarity, and the chorus of prototypes. *Minds and Machines*, Vol. 5, pp. 45–68. [15, 16, 17, 31, 100]

Edelman, S. (1995b). Receptive fields for vision: From hyperacuity to object recognition. Technical Report CS95–29. citeseer.ist.psu.edu/article/edelman95receptive.html. [15]

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, Vol. 21, pp. 449–498. [13, 15, 16, 100]

Edelman, S. (2002). Constraining the neural representation of the visual world. *Trends in Cognitive Sciences*, Vol. 6, pp. 125–131. [100]

Edelman, S. and Duvdevani-Bar, S. (1997a). Similarity, connectionism, and the problem of representation in vision. *Neural computation*, Vol. 9, pp. 701–720. [15, 16, 100]

Edelman, S. and Duvdevani-Bar, S. (1997b). Similarity-based viewspace interpolation and the categorization of 3D objects. *Proceedings of the Edinburgh Workshop on Similarity and Categorization.* [100]

Edelman, S. and Intrator, N. (1997). Learning as extraction of low-dimensional representations. *Mechanisms of perceptual learning* (eds. R. Goldstone, D. Medin, and P. Schyns), Vol. 36, pp. 353–380. Academic press, San Diego, CA. [13, 15, 24, 26, 100]

Edelman, S. and Intrator, N. (2001). A productive, systematic framework for the representation of visual structure. *Advances in Neural Information Processing Systems (NIPS)* (ed. T. Leen), pp. 10–16. The MIT Press, Cambridge, MA. [100, 103]

Edelman, S. and Intrator, N. (2003). Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science*, Vol. 27, pp. 73–109. [100, 101]

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review*, Vol. 89, pp. 627–661. [2, 3, 10, 11, 59]

Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration and CHARM. *Psychological Review*, Vol. 92, pp. 1–38. [3, 10, 11, 59]

Farah, M. (1985). Psychophysical evidence for a shared representation medium for mental images and percepts. *Journal of Experimental Psychology: General*, Vol. 114, pp. 91–103. [14]

Fei-Fei, L. and Perona, P. (2005). A Bayesian hierarchical model for learning natural scene categories. *IEEE Computer Vision and Pattern Recognition*, pp. 524–531. [97]

Field (1994). What is the goal of sensory coding? *Neural Computation*, Vol. 6, pp. 559–601. [19]

Findlay, J. M. and Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing.* Oxford University Press, New York, NY. [84]

Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, Vol. 3, pp. 63–109. [3]

Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, pp. 891–906. [21, 26, 27]

Fukushima, K. (1980). Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, Vol. 36, pp. 193–202. [18]

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution.* Basic Books, New York, NY. [106]

Gerstner, W. (1998). Spiking neurons. *Pulsed neural networks* (eds. W. Maass and C. M. Bishop), pp. 261–295. The MIT Press, Cambridge, MA. [103]

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin, Boston, NY. [4, 99]

Gillund, G. and Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, Vol. 91, pp. 1–67. [3, 10, 59]

Glenberg, A. and Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, Vol. 9, pp. 558–565. [105]

Gluck, M. A. and Meyers, C. E. (1993). Hippocampal mediation of stimulus representation: a computational theory. *Hippocampus*, Vol. 3, pp. 491–516. [105]

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., and Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, Vol. 34, pp. 77–86. [47, 102]

Goldstone, R. L. and Son, J. Y. (2005). Similarity. *Cambridge handbook of thinking and reasoning* (eds. K. J. Holyoak and R. G. Morrison), pp. 1–29. Cambridge University Press, Cambridge, UK. [10, 15]

Goldstone, R. L. and Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, Vol. 130, pp. 116–139. [100, 101]

Goldstone, R. L., Feng, Y., and Rogosky, B. J. (2005). Connecting concepts to each other and the world. *Grounding cognition* (eds. D. Pecher and R. A. Zwaan), pp. 282–314. Cambridge University Press, Cambridge, UK. [105]

Guillamet, D. and Vitri, J. (2002). Non-negative matrix factorization for face recognition. *Lecture Notes in Computer Science*, Vol. 2504, pp. 336–344. [96, 97]

Halberstadt, J., Goldstone, R. L., and Levine, G. M. (2003). Featural processing in face preferences. *Journal of Experimental Social Psychology*, Vol. 39, pp. 270–278. [101, 102]

Hancock, P. J. B., Baddeley, R. J., and Smith, L. S. (1992). Principal components of natural images. *Network: Computation in Neural Systems*, Vol. 3, pp. 61–70. [31, 32]

Hancock, P. J. B., Burton, A. M., and Bruce, V. (1996). Face processing: human perception and principal components analysis. *Memory and Cognition*, Vol. 24, pp. 26–40. [28]

Hancock, P. J. B., Bruce, V., and Burton, A. M. (1998). A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Research*, Vol. 38, pp. 2277–2288. [40]

Harnad, S. (1990). The symbol grounding problem. *Physica D*, Vol. 42, pp. 335–346. [3, 10, 106]

Hawkins, J. and Blakeslee, S. (2004). *On intelligence*. Times Books, New York, NY. [96]

Henderson, J. M. (1999). Effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 25, pp. 210–228. [91]

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Science*, Vol. 7, pp. 498–504. [18, 20, 21, 29, 42, 82, 85, 86, 87, 91, 94, 102]

Henderson, J. M. (2005). *Visual Cognition: Special Issue on Real-World Scene Perception*. The Psychology Press, New York, NY. [106]

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., and Mack, M. (2007). Visual saliency does not account for eye movements during search in real-world scenes. *Eye movements: A window on mind and brain* (eds. R. V. G. van Gompel, M. H. Fischer, W. Murray, and R. L. Hill). Elsevier, Oxford, UK. [85, 93, 94]

Hintzman, D. L. (1986). 'Schema abstraction' in a multiple-trace memory model. *Psychological Review*, Vol. 93, pp. 411–428. [3, 10, 59]

Hintzman, D. L., Curran, T., and Oppy, B. (1992). Effects of similarity and repetition on memory: Registration without learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 18, pp. 667–680. [74]

Hubel, D. H. (1988). *Eye, brain, and vision*. W. H. Freeman, New York, NY. [24, 26]

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, Vol. 148, pp. 574–591. [18, 20, 21]

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, Vol. 160, pp. 106–154. [18, 20, 21]

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, Vol. 195, pp. 215–243. [18, 20, 21]

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, Vol. 40, pp. 1489–1506. [85, 93]

Kahana, M. J. and Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, Vol. 42, pp. 2177–2192. [30]

Kalocsai, P., Zhao, W., and Biederman, I. (1998). Face similarity space as perceived by humans and artificial systems. *Proceedings of the Third International Conference on Automatic Face and Gesture Recognition*, pp. 177–180, Nara, Japan. [26, 40]

Karklin, Y. and Lewicki, M. S. (2003). A model for learning variance components of natural images. *Advances in Neural Information Processing Systems (NIPS)* (eds. S. Becker, S. Thrun, and K. Obermayer), Vol. 9, pp. 1367–1374. The MIT Press, Cambridge, MA. [104]

Karn, K. S. and HayHoe, M. M. (2000). Memory representations guide targeting eye movements in a natural task. *Visual Cognition*, Vol. 7, pp. 673–703. [85, 102]

Klatzky, R. L. (2004). Personal communication. June 2, 2004. [73]

Knudsen, E. I., Lac, S. du, and Esterly, S. D. (1987). Computational maps in the brain. *Annual Review of Neuroscience*, Vol. 10, pp. 41–65. [103]

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, Vol. 4, pp. 219–227. [93]

Kohonen, T. (2001). *Self-organizing maps (third extended edition)*, Vol. 30. Springer-Verlag, Berlin, DE. [103]

Kohonen, T., Oja, E., and Lehtiö, P. (1981). Storage and processing of information in distributed associative memory systems. *Parallel Models of Associative Memory* (eds. G. E. Hinton and J. A. Anderson), pp. 105–143. Lawrence Erlbaum Associates, Hillsdale, NJ. [1]

Koutstaal, W. (2003). Older adults encode but do not always use perceptual details: Intentional versus unintentional effects of detail on memory judgments. *Psychological Science*, Vol. 14, pp. 189–193. [68, 69]

Lacroix, J. P. W. (2001). A connectionist model of memory processing during sleep. Master thesis. Maastricht University, Maastricht, The Netherlands. [105]

Lacroix, J. P. W., Murre, J. M. J., Postma, E. O., and van den Herik, H. J. (2004). The natural input memory model. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society (CogSci 2004)* (eds. K. Forbus, D. Gentner, and T. Regier), pp. 773–778, Lawrence Erlbaum Associates, Mahwah, NJ. [32, 49]

Lacroix, J. P. W., Postma, E. O., and Murre, J. M. J. (2005). Predicting experimental similarity ratings and recognition rates for individual natural stimuli with the NIM model. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society (CogSci 2005)* (eds. B. Bara, L. Barsalou, and M. Bucciarelli), pp. 1225–1230, Lawrence Erlbaum Associates, Mahwah, NJ. [35]

Lacroix, J. P. W., Murre, J. M. J., Postma, E. O., and van den Herik, H. J. (2006a). Modeling recognition memory using the similarity structure of natural input. *Cognitive Science*, Vol. 30, pp. 121–145. [26, 35, 75]

Lacroix, J. P. W., Postma, E. O., and Murre, J. M. J. (2006b). Knowledge-driven gaze control in the NIM model. *Proceedings of the 28th Annual Meeting of the Cognitive Science Society (CogSci 2006)* (eds. R. Sun and N. Miyake), pp. 1657–1662, Lawrence Erlbaum Associates, Mahwah, NJ. [104]

Lacroix, J. P. W., Postma, E. O., Murre, J. M. J., and van den Herik, H. J. (in preparation). Modelling recognition-memory effects with Nim-Rem. [49]

Lamont, A. C., Stewart-Williams, S., and Podd, J. (2005). Face recognition and aging: effects of target age and memory load. *Memory and Cognition*, Vol. 33, pp. 1017–1024. [53, 62]

Landauer, T. K. and Dumais, S. T. (1997). A solution to Platos problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, Vol. 104, pp. 211–240. [11, 30]

Land, M. F. and Hayhoe, M. (2001). In what ways do eye movements contribute to everday activities? *Vision Research*, Vol. 41, pp. 3559–3565. [85, 91]

Laughery, K. R., Alexander, J. F., and Lane, A. B. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology*, Vol. 55, pp. 477–483. [65]

Lee, T. S. (1998a). Image representation using 2D Gabor wavelets. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 18, pp. 959–971. [26]

Lee, T. W. (1998b). *Independent component analysis: Theory and applications*. Kluwer, Boston, MA. [19]

Lee, D. and Seung, H. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, Vol. 401, pp. 788–791. [96, 97]

Light, L. L., Chung, C., Pendergrass, R., and van Ocker, J. C. (2006). Effects of repetition and response deadline on item recognition in young and older adults. *Memory and Cognition*, Vol. 34, pp. 335–343. [74]

Lindsay, P. H. and Norman, D. A. (1977). *Human information processing*. Academic Press, New York, NY, 2nd edition. [3]

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology*, Vol. 3, pp. 525–551. [82]

Logothetis, N. K. and Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, Vol. 19, pp. 577–621. [14]

Lowe, D. (1993). Similarity metric learning for a variable-kernel classifier. *Neural Computation*, Vol. 7, pp. 295–311. [102]

Lu, X., Wang, Y., and Jain, A. K. (2003). Combining classifiers for face recognition. *Proceedings of the International Conference on Multimedia and Expo (ICME)*, Vol. 3, pp. 13–16. [96]

Lyons, M. J. (2000). A linked aggregate code for processing faces. *Pragmatics and Cognition*, Vol. 8, pp. 63–81. [26, 40]

Lyons, M. J. and Akamatsu, S. (1998). Coding facial expressions with Gabor wavelets. *Proceedings of the Third International Conference on Automatic Face and Gesture Recognition*, pp. 200–205, Nara, Japan. [26]

MacAndrew, D. K., Klatzky, R. L., Fiez, J. A., McClelland, J. L., and Becker, J. T. (2002). The phonological-similarity effect differentiates between two working memories. *Psychological Science*, Vol. 13, pp. 465–468. [53, 62, 73]

Malmberg, K. J., Holden, J. E., and Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on oldnew recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 30, pp. 319–331. [3, 74]

Mannan, S. K., Ruddock, K. H., and Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, Vol. 10, pp. 165–188. [20, 84, 93]

Mäntylä, T. and Holm, L. (2006). Gaze control and recollective experience in face recognition. *Visual Cognition*, Vol. 13, pp. 365–386. [82, 85, 86]

Markman, A. B. and Gentner, D. (2004). Nonintentional similarity processing. *The new unconscious* (eds. R. Hassin, J. A. Bargh, and J. S. Uleman). Oxford University Press, New York, NY. [15]

Marr, D. (1982). *Vision.* W. H. Freeman, New York, NY. [4, 13, 14, 18, 99]

Marr, D. and Hildreth, E. C. (1980). Theory of edge detection. *Proceedings of the Royal Society of London*, Vol. 200, pp. 269–294. [18]

Martinez, A.M. and Benavente, R. (1998). The AR Face Database. *CVC Technical Report #24.* [79]

Martinez, A. M. and Kak, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, pp. 228–233. [96]

Mattern, F. and Denzler, J. (2004). Comparison of appearance based methods for generic object recognition. *Pattern Recognition and Image Analysis*, Vol. 14, pp. 255–261. [78]

McClelland, J. L. and Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, Vol. 105, pp. 724–760. [3, 10, 53, 55, 59, 65, 74]

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocrotex: Insights from the success and failures of connectionist models of learning and memory. *Psychological Review*, Vol. 102, pp. 419–457. [105]

McSorley, E. and Findlay, J. M. (2003). Saccade target selection in visual search: Accuracy improves when more distractors are present. *Journal of Vision*, Vol. 3, pp. 877–892. [20, 29, 42, 82]

Medin, D. L. and Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, Vol. 85, pp. 207–238. [10, 77]

Meeter, M. and Murre, J. M. J. (2004). Consolidation of long-term memory: Evidence and alternatives. *Psychological Bulletin*, Vol. 130, pp. 843–857. [105]

Meeter, M. and Murre, J. M. J. (2005). Tracelink: A model of consolidation and amnesia. *Cognitive Neuropsychology*, Vol. 22, pp. 559–587. [105]

Mel, B. W. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation*, Vol. 9, pp. 777–804. [13, 17]

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, Vol. 6, pp. 8–17. [82]

Metzger, M. M. (2002). Stimulus load and age effects in face recognition: a comparison of children and adults. *North American Journal of Psychology*, Vol. 4, pp. 51–62. [62]

Mishkin, M. and Appenzeller, T. (1987). The anatomy of memory. *Scientific American*, Vol. 256, pp. 80–89. [104]

Moghaddam, B. and Pentland, A. P. (1997). Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 696–710. [96]

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, Vol. 89, pp. 609–626. [2, 3, 10, 11, 59]

Murnane, K. and Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 17, pp. 855–874. [53, 55, 56]

Murre, J. M. J. (1992). *Learning and categorization in modular neural networks*. Lawrence Erlbaum Associates, Hillsdale, NJ. [77]

Murre, J. M. J. (1996). Tracelink: A model of consolidation and amnesia. *Hippocampus*, Vol. 6, pp. 675–684. [105]

Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. *Vision: Coding and efficiency* (ed. C. Blakemore), pp. 411–422. Cambridge University Press, Cambridge, UK. [27]

Nega, C. (2005). Perceptual effects and recollective experience in face recognition. *Experimental Psychology*, Vol. 52, pp. 224–231. [53, 55, 65, 74]

Neider, M. B. and Zelinski, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, Vol. 46, pp. 614–621. [85]

Nelson, L. (2004). Neuroscience: While you were sleeping. *Nature*, Vol. 430, pp. 962–964. [105]

Nelson, R. C. and Selinger, A. (1998). Large-scale tests of a keyed, appearance-based 3D object recognition system. *Vision Research*, Vol. 38, pp. 2469–2488. [100]

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1999). The University of South Florida word association, rhyme and fragment norms. http://www.usf.edu/FreeAssociation/. [12]

Newell, A. and Simon, H. A. (1972). *Human problem solving*. Prentice-Hall, Englewood Cliffs, NJ. [10]

Norman, K. A. (2002). Differential Effects of List Strength on Recollection and Familiarity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 28, pp. 1083–1094. [53, 54, 56, 61, 74, 75]

Norman, K. A. and O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, Vol. 110, pp. 611–646. [61, 74, 75]

Norman, J. F., Phillips, F., and Ross, H. E. (2001). Information concentration along the boundary contours of naturally shaped solid objects. *Perception*, Vol. 30, pp. 1285–1294. [20, 26]

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, Vol. 115, pp. 39–57. [6, 10, 15, 22, 23, 28, 77]

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 13, pp. 87–108. [6, 10, 22, 23, 28, 30, 101]

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 14, pp. 700–708. [10, 22]

Nosofsky, R. M. and Zaki, S. R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 29, pp. 1194–1209. [10, 15, 22, 28]

Nosofsky, R. M., Clark, S. E., and Shin, H. J. (1989). Rules and Exemplars in Categorization, Identification, and Recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 15, pp. 282–304. [10, 22]

Ohrt, D. D. and Gronlund, S. D. (1999). List-length effect and continuous memory: Confounds and solutions. *On human memory: Evolution, progress, & reflections on the 30th anniversary of the Atkinson-Shiffrin model* (ed. C. Izawa), pp. 105–125. Lawrence Erlbaum Associates, Mahwah, NJ. [53, 62]

Oliva, A., Torralba, A., Castelhano, M. S., and Henderson, J. M. (2003). Top-down control of visual attention in object detection. *IEEE Proceedings of the International Conference on Image Processing*, Vol. 1, pp. 253–256. [85, 93, 102]

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, Vol. 381, pp. 607–609. [104]

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., and Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America, series A*, Vol. 10, pp. 405–411. [31]

O'Toole, A. J., Wenger, M. J., and Townsend, J. T (2001). Quantitative models of perceiving and remembering faces: precedents and possibilities. *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges* (eds. M. Wenger and J. Townsend), pp. 1–38. Lawrence Erlbaum Associates, Hillsdale, NJ. [31]

Paletta, L., Prantl, M., and Pinz, A. (1998). Reinforcement learning for autonomous three-dimensional object recognition. *Proceedings of the 6th Symposium on Intelligent Robotics Systems*, pp. 63–72, Edinburgh, UK. [94]

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. The MIT Press, Cambridge, MA. [14, 18, 19, 20, 24, 27, 100]

Palmeri, T. J. and Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, Vol. 5, pp. 291–303. [26]

Paredes, R. and Vidal, E. (1999). A nearest neighbor weighted measure in classification problems. *Proceedings of the VIII Simposium Nacional de Reconocimiento de Formas y An alisis de Im agenes*, Vol. 1, pp. 437–444. [102]

Parkhurst, D. J. and Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, Vol. 16, pp. 125–154. [20, 84, 93]

Parkhurst, D. J., Law, K., and Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, Vol. 42, pp. 107–123. [93]

Pearson, K. (1901). On lines and planes of closest fit to systems of points is space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, Vol. 2, pp. 559–572. [19]

Pecher, D. and Zwaan, R. A. (2005). *Grounding cognition*. Cambridge University Press, Cambridge, UK. [3, 4]

Peissig, J. J. and Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of psychology*, Vol. 58, pp. 75–96. [14]

Penev, P. S. and Atick, J. J. (1996). Local Feature Analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems*, Vol. 7, pp. 477–500. [96]

Petrov, Y and Zhaoping, L (2003). Local correlations, information redundancy, and sufficient pixel depth in natural images. *Journal of the Optical Society of America A*, Vol. 20, pp. 56–66. [20, 26]

Pfeifer, R. and Scheier, C. (1999). *Understanding intelligence*. The MIT Press, Cambridge, MA. [106]

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review*, Vol. 91, pp. 281–293. [3, 10, 59]

Podd, J. (1990). The effects of memory load and delay on face recognition. *Applied Cognitive Psychology*, Vol. 4, pp. 47–60. [62]

Postman, L. (1951). The generalization gradient in recognition memory. *Journal of Experimental Psychology*, Vol. 42, pp. 231–235. [68]

Postma, E. O., van den Herik, H. J., and Hudson, P. T. W. (1997). SCAN: A scalable neural model of covert attention. *Neural Networks*, Vol. 10, pp. 993–1015. [27]

Raaijmakers, J. G. W. and Shiffrin, R. M. (1981). Search of Associative Memory. *Psychological Review*, Vol. 88, pp. 93–134. [2, 3, 10, 59]

Raaijmakers, J. G. W. and Shiffrin, R. M. (2002). Model of memory. *Steven's handbook of experimental psychology, third edition* (eds. H. Pashler and D. Medin), Vol. 2, pp. 43–76. Wiley & Sons Inc., New York, NY. [2]

Rajashekar, U., Cormack, L. K., and Bovik, A. C. (2002). Visual search: Structure from noise. *Proceedings of the Eye Tracking Research & Applications Symposium 2002*, pp. 119–123, New Orleans, LA. [39, 85, 102]

Rao, R. P. N. and Ballard, D. H. (1995). An active vision architecture based on iconic representations. *Artificial Intelligence*, Vol. 78, pp. 461–505. [24, 26]

Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., and Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, Vol. 42, pp. 1447–1463. [85, 93]

Ratcliff, R., Clark, S. E., and Shiffrin, R. M. (1990). The list-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 16, pp. 163–178. [53, 54, 55, 56, 59, 61, 62, 65]

Reinagel, P. and Zador, A. M. (1999). Natural scene statistics at the center of gaze. *Computation in Neural Systems*, Vol. 1-10, pp. 763–785. [51]

Roark, D. A., O'Toole, A. J., and Abdi, H. (2003). Human recognition of familiar and unfamiliar people in naturalistic video. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003)*. [53, 65]

Roediger, H. L. III and McDermott, K. B. (1995). Creating false memories: Remembering words not presented on lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 21, pp. 803–814. [53, 68]

Roy, D. (2005a). Grounding words in perception and action: Computational insights. *Trends in Cognitive Sciences*, Vol. 9, pp. 389–396. [3, 105]

Roy, D. (2005b). Semiotic schemas: A framework for grounding language in the action and perception. *Artificial Intelligence*, Vol. 167, pp. 170–205. [105]

Rumelhart, D. E., McClelland, J. L., and PDP Research Group the (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1 and 2. The MIT Press, Cambridge, MA. [1]

Rybak, I. A., Gusakova, V. I., Golovan, A. V., Podladchikova, L. N., and Shevtsova, N. A. (1998). A model of attention-guided visual perception and recognition. *Vision Research*, Vol. 38, pp. 2387–2400. [94]

Schooler, L., Shiffrin, R. M., and Raaijmakers, J. G. W. (2001). A model for implicit effects in perceptual identification. *Psychological Review*, Vol. 108, pp. 257–272. [3]

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656. [86]

Sheffert, S. M. and Shiffrin, R. M. (2003). Auditory registration without learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 29, pp. 10–21. [74]

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, Vol. 22, pp. 325–345. [10, 11, 15, 100]

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, Vol. 1, pp. 54–87. [15, 100]

Shepard, R. N. (1975). Form, formation, and transformation of internal representations. *Information processing and cognition* (ed. R. L. Solso), pp. 87–122. Wiley & Sons Inc., New York, NY. [100]

Shepard, R. N. (1981). Psychophysical complementarity. *Perceptual organization* (eds. M. Kubovy and J.R. Pomerantz), pp. 279–341. Lawrence Erlbaum Associates, Hillsdale, NJ. [100]

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, Vol. 237, pp. 1317–1323. [10, 15]

Shepard, R. N. and Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, Vol. 1, pp. 1–17. [15, 100]

Shiffrin, R. M. and Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin & Review*, Vol. 4, pp. 145–166. [1, 2, 3, 6, 10, 11, 22, 30, 49, 50, 51, 52, 56, 75]

Shiffrin, R. M., Ratcliff, R., and Clark, S. E. (1990a). The list-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 16, pp. 179–195. [55, 59]

Shiffrin, R. M., R. Ratcliff, R., and Clark, S. E. (1990b). The list-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 16, pp. 179–195. [56]

Simpson, A. J. and Fitter, M. J. (1973). What is the best index of detectability? *Psychological Bulletin*, Vol. 80, pp. 481–488. [52]

Solso, R. L. and McCarthy, J. E. (1981). Prototype formation of faces: A case of pseudo-memory. *British Journal of Psychology*, Vol. 72, pp. 499–503. [68]

Sowden, P. T. and Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in Cognitive Sciences*, Vol. 10, pp. 538–545. [47, 102]

Sprague, N., Ballard, D. H., and Robinson, A. (in press). Modeling attention with embodied visual behaviors. *ACM Transactions on Applied Perception*. [4]

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, Vol. 99, pp. 195–231. [105]

Stahl, A. (2005). Learning similarity measures: A formal view based on a generalized CBR model. *Lecture Notes in Computer Science*, Vol. 3620, pp. 507–521. [102]

Stewart, H. A. and McAllister, H. A. (2001). One-at-a-time versus grouped presentation of mug book pictures: some surprising results. *Journal of Applied Psychology*, Vol. 86, pp. 1300–1305. [53, 68]

Steyvers, M. (2000). *Modeling semantic and orthographic similarity effects on memory for individual words*. Ph.D. thesis, Indiana University, Bloomington, IN. http://psiexp.ss.uci.edu/research/papers/dissertation_small.pdf. [30, 51]

Steyvers, M. and Busey, T. (2000). Predicting similarity ratings to faces using physical descriptions. *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (eds. M. Wenger and J. Townsend), pp. 115–146. Lawrence Erlbaum Associates, Hillsdale, NJ. [31, 36, 39, 40, 102]

Steyvers, M., Shiffrin, R. M., and Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. *Experimental cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer* (ed. A. Healy). American Psychological Association, Washington, DC. [11, 12, 31]

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, Vol. 437, pp. 1272–1278. [105]

Stillings, N. A., Weisler, S. E., Chase, C. H., Feinstein, M. H., Garfield, J. L., and Rissland, E. L. (1995). *Cognitive science*. The MIT Press, Cambridge, MA. [75]

Stretch, V. and Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 24, pp. 1379–1396. [53, 55, 65, 74]

Sun, R. and Zhang, X. (2006). Accounting for a variety of reasoning data within a cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, Vol. 18, pp. 169–191. [10]

Swain, M. J. and Ballard, D. (1991). Color indexing. *International Journal of Computer Vision*, Vol. 7, pp. 11–32. [13, 17]

Swets, J. A. (1986a). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, Vol. 99, pp. 181–198. [52]

Swets, J. A. (1986b). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, Vol. 99, pp. 100–117. [52]

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, Vol. 2, pp. 109–193. [14]

Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, Vol. 290, pp. 2319–2323. [26]

Teyler, T. J. and Discenna, P. (1985). The role of hippocampus in memory: A hypothesis. *Neuroscience and Biobehavioral Reviews*, Vol. 9, pp. 377–389. [105]

Thorndike, E. L. (1913). *Educational psychology*. Columbia University Press, New York, NY. [2]

Torralba, A. and Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, Vol. 14, pp. 391–412. [104]

Torralba, A., Oliva, A., Castelhano, M. S., and Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search. *Psychological Review*, Vol. 113, pp. 766–786. [85]

Turano, K. A., Geruschat, D. R., and Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research*, Vol. 43, pp. 333–346. [93]

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, Vol. 3, pp. 71–86. [31]

Valentine, T. (1991). Representation and process in face recognition. *Vision and visual dysfunction: Pattern recognition by man and machine* (ed. R. Watt), Vol. 14. MacMillan Press, London, UK. [31]

van Dartel, M. F. (2005). *Situated representation*. Ph.D. thesis, Maastricht University, Maastricht, The Netherlands. [106]

van Dartel, M. F, Sprinkhuizen-Kuyper, I. G., Postma, E. O., and van den Herik, H. J. (2005). Reactive agents and perceptual ambiguity. *Adaptive Bahvior*, Vol. 13, pp. 227–242. [106]

Viken, R. J., Treat, T. A., Nosofsky, R. M., McFall, R. M., and Palmeri, T. J. (2002). Modeling individual differences in perceptual and attentional processes related to bulimic symptoms. *Journal of Abnormal Psychology*, Vol. 111, pp. 598–609. [102]

Vokey, J. R. and Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory and Cognition*, Vol. 20, pp. 291–302. [68]

Wainwright, M. (1999). Visual adaptation as optimal information transmission. *Vision Research*, Vol. 39, pp. 3960–3974. [20, 26]

Walker, M. P., Brakefield, T., Hobson, J. A., and Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, Vol. 425, pp. 616–620. [105]

Wang, J., Kwok, J. T., Shen, H. C., and Quan, L. (2005). Data-dependent kernels for high-dimensional data classification. *Proceedings of the International Joint Conference on Neural Networks*, pp. 102–107, Montreal, Canada. [96]

Westbury, C., Buchanan, L., and Brown, N. R. (2002). Sounds of the neighborhood: False memories and the structure of the phonological lexicon. *Journal of Memory and Language*, Vol. 46, pp. 622–651. [68]

Wettschereck, D., Aha, D. W., and Mohri, T. (1997). A review and emprirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review*, Vol. 1, pp. 273–314. [102]

Yarbus, A. L. (1967). *Eye movements and vision.* Plenum Press, New York, NY. [18, 20, 26, 85]

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, Vol. 46, pp. 441–517. [74, 75]

Yonelinas, A. P., Hockley, W. E., and Murdock, B. B. (1992). Tests of the list-strength effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cogntion*, Vol. 18, pp. 345–355. [53, 55, 56]

Yonelinas, A. P., Otten, L. J., Shaw, K. N., and Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, Vol. 25, pp. 3002–3008. [75]

Zaki, S. R. and Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 27, pp. 1022–1041. [28, 68]

Zhaoping, L. and May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Computational Biology*, Vol. 3, pp. 616–633. [85, 93]

Zola-Morgan, S. M. (1990). The primate hippocampal formation: Evidence for a time-limited role in memory storage. *Science*, Vol. 12, pp. 288–290. [105]

Summary

Traditionally, models of natural cognition consider cognitive mechanisms as processes of symbol manipulation operating independently of the environment. The symbol-manipulation models suffer from two interrelated problems: the symbol grounding problem and the transduction problem. Both problems address the lack of a connection between representations in a cognitive system and the entities that they refer to in the real world. The relatively new situated approach to cognition deals with the two problems by viewing cognitive mechanisms as emerging from the interaction with the natural environment. The focus of this thesis is to address the symbol grounding and transduction problems for computational memory models by realizing a situated computational memory model. The situated model operates directly on the natural environment.

Chapter 1 starts with a historical overview of the development of computational memory modelling. The chapter identifies the lack of a direct connection with the real world as the main limitation of the existing computational memory models. Rather than deriving representations directly from the real world, these models rely on various types of abstract representation spaces that are only indirectly related to the real world. By assuming an abstract representation space, these models suffer from the symbol grounding problem and the transduction problem. In order to deal with these problems we phrase the following problem statement: "How can computational memory models be extended to solve the grounding and transduction problems?" The problem statement is addressed by defining a perceptual front-end that transforms natural visual input into memory representations. We propose a combination of the perceptual front-end with a computational memory back-end to obtain a situated computational memory model. To validate the situated computational memory model, three research questions are formulated: (1) to what extent can a situated model produce human responses to individual natural visual stimuli? (2) to what extent can a situated model produce recognition-memory effects on the basis of natural visual stimuli? and (3) to what extent can a situated model classify natural visual stimuli? The research methodology is presented; it consists of two steps: model construction and model validation.

Chapter 2 provides an overview of the types of abstract representation spaces used in the existing memory models. It illustrates that the representations are not derived directly from the physical features of the individual stimuli that they refer to in the real world, i.e., they are not grounded in the real world. Subsequently, two sources of inspiration for constructing veridical representations directly from visual

natural input are discussed. The first source consists of the insights taken from four influential approaches belonging to the domain of computer vision to map visual input onto representations. The second source consists of relevant knowledge about the main characteristics of the human visual system. Based on these two sources of inspiration, the chapter proceeds with the formulation of three guiding principles for grounding memory representations in the real visual world. Finally, it provides ways to fulfil the guiding principles to obtain a perceptual front-end for a situated computational memory model.

Chapter 3 introduces the situated computational memory model. The model operates on natural images and it is called: the Natural Input Memory model (NIM). NIM combines a perceptual front-end with a computational memory back-end. NIM's perceptual front-end employs a biologically informed method that selects local image samples (i.e., eye fixations) from natural images and translates these into feature-vector representations. Each feature vector contains information on oriented edges at multiple scales extracted from a small image area surrounding the fixation location. The feature-vector representations form the input to the computational memory back-end, which is an exemplar-based model that makes recognition-memory decisions on the basis of a comparison between stored and incoming similarity-space representations.

Chapter 4 validates NIM by assessing to what extent NIM can produce human similarity ratings and recognition rates for individual natural stimuli. The model is tested on a similarity-rating task and a face-recognition task using the same stimuli and tasks as those used in behavioural experiments. The NIM similarity ratings and recognition rates are compared with the behaviourally obtained human similarity ratings and recognition rates. The results demonstrate that NIM quite accurately produces the human similarity ratings and the human recognition rates for the individual natural stimuli.

Chapter 5 examines NIM's natural input recognition properties in terms of *general* recognition-memory effects. Many studies have demonstrated the success of REM (a well-established model of memory) in replicating a wide range of human recognition-memory effects. The chapter introduces a NIM variant called NIM-REM that realizes a natural input version of REM by combining NIM's perceptual front-end with a REM-based memory back-end. The remainder of the chapter focuses on validating NIM-REM by assessing its ability to explain behavioural results on four recognition-memory effects that are often studied in behavioural recognition-memory experiments: the list-strength effect, the list-length effect, the item-strength effect, and the false-memory effect. For each effect, the pattern of NIM-REM recognition results are compared with the human pattern of results obtained in behavioural experiments. The results on the four recognition-memory effects indicate that NIM-REM produces rather adequately the findings from behavioural experiments.

Chapter 6 studies NIM's ability to classify natural input. In order to test the classification ability, the chapter introduces a NIM variant called NIM-CLASS that combines NIM's perceptual front-end with a new memory back-end that is suitable for classification. The classification performance of NIM-CLASS is evaluated on a face-classification task that entails the identification of a natural image of a frontal face with variations in facial expression, illumination, and occlusion, on the basis of

a single encounter with the face. The classification results demonstrate that NIM-CLASS is able to classify natural images of frontal faces correctly under a variety of unfavourable conditions, provided that a sufficient number of fixations are made during a single encounter with the face. Subsequently, the chapter investigates to what extent the classification performance can be improved by extending NIM-CLASS with top-down fixation selection to select relevant fixation locations on the basis of stored knowledge. The chapter introduces two NIM-CLASS variants: NIM-CLASS A and NIM-CLASS B. NIM-Class A employs a top-down fixation-selection mechanism during the classification of a face that relies on short-term episodic knowledge about previously encountered faces. From the NIM-CLASS A classification results we may conclude that the short-term episodic-knowledge-based top-down fixation selection improves performance on the classification task compared to the NIM-CLASS performance. This is particularly so when a limited number of fixations is made during classification. NIM-CLASS B adopts the top-down fixation-selection mechanism of NIM-CLASS A during the classification of a face and, in addition, employs a top-down fixation-selection mechanism during the storage of a face that relies on long-term stored knowledge about the relevance of different face parts. The NIM-CLASS B classification results demonstrate that the top-down fixation-selection mechanism employed during storage improves the performance on the classification task compared to the NIM-CLASS A performance. The results obtained with the NIM-CLASS A and B variants demonstrate the beneficial effect of active top-down processes that rely on various types of stored knowledge for the classification of natural visual input.

Chapter 7 provides a discussion of our proposed situated computational memory model. First, it relates our approach to influential existing computational models of object recognition. Second, it identifies several model extensions for the improvement of the model's psychological and biological realism as a model of natural cognition. Based on psychological and biological insights, five extensions are discussed: (1) a feature-based attentional mechanism, (2) a spatial attentional mechanism, (3) a neural implementation of the similarity space, (4) the representation of spatial knowledge, and (5) the incorporation of separate episodic and semantic representation spaces. Third, the chapter places our approach in the context of the global developments in the domain of cognitive modelling. It shows that our approach departs from the traditional computational memory models and adheres to the new 'situated' approach by focussing on the interaction with a realistic environment.

Chapter 8 answers the three research questions formulated in chapter 1 and provides the thesis conclusion of the problem statement. For the first research question, the answer is that the situated model is able to produce human responses to individual natural stimuli quite reliably. For the second research question, it is stated that the situated model is able to produce the four recognition-memory effects successfully directly on the basis of natural visual input. For the third research question, we state that our situated model is able to classify natural images correctly under a variety of potentially unfavourable conditions provided that a sufficient amount of visual input is selected. Moreover, we see that the use of active top-down processes that rely on stored knowledge to select visual input enhances classification performance, in particular when classification is based on a limited amount of visual input. Finally, the chapter elaborates on the conclusion that the situated computational memory

model presented in this thesis provides a viable solution to the symbol grounding and transduction problems by relating a representation to its real-world referent in a transparent and non-trivial manner that is neurobiologically informed.

Samenvatting

In de klassieke modellen van natuurlijke cognitie worden cognitieve mechanismen beschouwd als processen van symbool-manipulatie die onafhankelijk van de omgeving werken. De symbool-manipulatiemodellen hebben last van twee aan elkaar gerelateerde problemen: het symbool-funderingsprobleem en het transductieprobleem. Beide problemen hebben betrekking op het ontbreken van een verbinding tussen representaties in het cognitieve systeem en de entiteiten waarnaar deze verwijzen in de wereld. De relatief nieuwe, gesitueerde benadering van cognitie pakt deze problemen aan door cognitieve mechanismen te beschouwen als het resultaat van de interactie met de natuurlijke omgeving. Dit proefschrift richt zich op de behandeling van het symbool-funderingsprobleem en het transductieprobleem voor computationele geheugenmodellen door de verwezenlijking van een gesitueerd computationeel geheugenmodel. Het gesitueerde model is direct werkzaam op de natuurlijke omgeving.

Hoofdstuk 1 begint met een historisch overzicht van de ontwikkeling van het computationeel modelleren van het geheugen. Het hoofdstuk identificeert het gebrek van een directe verbinding met de echte wereld als de belangrijkste beperking van de bestaande computationele geheugenmodellen. In plaats van representaties die direct van de echte wereld zijn afgeleid, zijn de bestaande modellen afhankelijk van allerlei typen abstracte representaties die slechts indirect gerelateerd zijn aan de wereld. Het gebruik van een abstracte representatieruimte in deze modellen leidt tot het symbool-funderingsprobleem en het transductieprobleem. Om deze problemen aan te pakken, formuleren we de volgende probleemstelling: "Hoe kunnen computationele geheugenmodellen worden uitgebreid zodat ze het funderings- en transductieprobleem oplossen?" De probleemstelling wordt benaderd door het definiëren van een perceptueel *front-end* dat natuurlijke visuele input vertaalt naar geheugenrepresentaties. We stellen voor om het perceptuele *front-end* te combineren met een computationeel geheugen-*back-end* om te komen tot een gesitueerd computationeel geheugenmodel. De validatie van het gesitueerde computationele geheugenmodel vindt plaats aan de hand van drie onderzoeksvragen: (1) In hoeverre kan een gesitueerd model menselijke responsen op individuele natuurlijke visuele stimuli produceren? (2) In hoeverre kan een gesitueerd model herkenningseffecten produceren op basis van natuurlijke visuele stimuli? en (3) In hoeverre kan een gesitueerd model natuurlijke visuele stimuli classificeren? De onderzoeksmethodologie wordt gepresenteerd aan de hand van twee stappen: modelconstructie en modelvalidatie.

Hoofdstuk 2 geeft een overzicht van de typen abstracte representatieruimten die

in de bestaande geheugenmodellen worden gebruikt. Het illustreert dat de representaties niet direct worden afgeleid van de fysieke eigenschappen van de individuele stimuli waarnaar ze verwijzen in de echte wereld; dit betekent dat ze niet gefundeerd zijn in de echte wereld. Vervolgens worden twee bronnen van inspiratie besproken voor de directe constructie van waarheidsgetrouwe representaties van natuurlijke visuele input. De eerste bron bestaat uit de inzichten van vier invloedrijke benaderingen uit het domein van *computer vision* voor de vertaling van visuele input in representaties. De tweede bron bestaat uit relevante kennis over de belangrijkste eigenschappen van het menselijke visuele systeem. Op basis van deze twee bronnen van inspiratie vervolgt het hoofdstuk met de formulering van drie richtlijnen voor het funderen van geheugenrepresentaties in de visuele wereld. Tenslotte geeft het methoden om deze richtlijnen te volgen en te komen tot een perceptueel *front-end* voor een gesitueerd geheugenmodel.

Hoofdstuk 3 introduceert het gesitueerde computationele geheugenmodel. Het model werkt op natuurlijke afbeeldingen en wordt het Natural Input Memory model (NIM) genoemd. NIM combineert een perceptueel *front-end* met een computationeel geheugen-*back-end*. Het perceptuele *front-end* van NIM gebruikt een biologisch geïnformeerde methode die lokale samples (oogfixaties) selecteert uit natuurlijke afbeeldingen en deze vertaalt naar vectorrepresentaties. Iedere vector bevat informatie over georiënteerde licht/donker overgangen op meerdere schalen. Deze informatie wordt verkregen van een klein gebied rondom de fixatielocatie in de afbeelding. De vectorrepresentaties vormen de input voor het computationele geheugen-*back-end*. Dit is een exemplaar-gebaseerd model dat herkenningsbeslissingen maakt op basis van een vergelijking tussen opgeslagen en inkomende vectorrepresentaties.

Hoofdstuk 4 valideert NIM door te beoordelen in hoeverre NIM in staat is om menselijke similariteitsresponsen en herkenningsresponsen op individuele natuurlijke stimuli te produceren. Het model wordt getest op een similariteitstaak en een gezichtsherkenningstaak waarbij dezelfde stimuli en taken worden gebruikt als in gedragsexperimenten. De similariteits- en herkenningsresponsen van NIM worden vergeleken met de menselijke similariteits- en herkenningsresponsen die verkregen zijn in de gedragsexperimenten. De resultaten wijzen uit dat NIM de menselijke similariteits- en herkenningsreponsen voor individuele natuurlijke stimuli vrij nauwkeurig produceert.

Hoofdstuk 5 bestudeert de herkenningseigenschappen van NIM voor natuurlijke input in termen van algemene herkenningseffecten. Vele studies hebben het succes aangetoond van REM (een gerenommeerd geheugenmodel) in het repliceren van een breed scala aan menselijke herkenningseffecten. Het hoofdstuk introduceert een NIM-variant, NIM-REM genaamd, die een REM-versie realiseert voor natuurlijke input door het perceptuele *front-end* van NIM te combineren met een REM-gebaseerd geheugen-*back-end*. Het hoofdstuk richt zich vervolgens op de validatie van NIM-REM door te beoordelen in hoeverre NIM-REM in staat is om vier herkenningseffecten uit experimenteel geheugenonderzoek te produceren: het lijststerkte-effect, het lijstlengte-effect, het itemsterkte-effect, en het pseudoherinnering-effect. Voor ieder effect wordt het patroon van NIM-REM-resultaten vergeleken met het patroon van menselijke resultaten uit experimenteel onderzoek. De resultaten tonen aan dat NIM-REM de experimentele bevindingen tamelijk adequaat benadert.

Hoofdstuk 6 bestudeert het vermogen van NIM om natuurlijke input te classificeren. Om het classificatievermogen te testen introduceert het hoofdstuk een NIM-variant, genaamd NIM-CLASS, die het perceptuele *front-end* van NIM combineert met een geheugen-*back-end* dat geschikt is voor classificatie. De classificatieprestatie van NIM-CLASS wordt beoordeeld op een gezichtsclassificatietaak waarin natuurlijke afbeeldingen van frontale gezichten, met variaties in gezichtsuitdrukking, belichting, en occlusies, moeten worden geïdentificeerd op basis van een enkele presentatie van het gezicht. De classificatieresultaten tonen dat NIM-CLASS in staat is om natuurlijke afbeeldingen van frontale gezichten, onder verscheidene ongunstige omstandigheden, correct te classificeren, gegeven dat er voldoende fixaties zijn gemaakt gedurende de enkele presentatie van het gezicht. Vervolgens onderzoekt het hoofdstuk in hoeverre de classificatie kan worden verbeterd door NIM-CLASS uit te breiden met een *top-down* fixatieselectie-mechanisme voor het selecteren van relevante fixatielocaties op basis van opgeslagen kennis. Het hoofdstuk introduceert twee NIM-CLASS-varianten: NIM-CLASS A en NIM-CLASS B. NIM-CLASS A implementeert een *top-down* fixatieselectie-mechanisme voor de classificatie van een gezicht dat gebruik maakt van episodische korte-termijn kennis van eerder gepresenteerde gezichten. Uit de classificatieresultaten mogen we concluderen dat het *top-down* fixatieselectie-mechanisme van NIM-CLASS A de prestatie op de classificatietaak verbetert vergeleken met de prestatie van NIM-CLASS. Dit is voornamelijk zo wanneer de classificatie gebaseerd is op een beperkt aantal fixaties. NIM-CLASS B gebruikt het *top-down* fixatieselectie-mechanisme van NIM-CLASS A voor de classificatie van een gezicht. Daarnaast gebruikt het een *top-down* fixatieselectie-mechanisme voor het selecteren van fixaties gedurende de opslag van een gezicht in het geheugen dat is gebaseerd op lange-termijn kennis over de relevantie van verschillende gezichtsdelen. De classificatieresultaten laten zien dat het *top-down* fixatieselectie-mechanisme voor de opslag van een gezicht, de prestatie op de classificatietaak verbetert in vergelijking met de prestatie van NIM-CLASS A. De resultaten verkregen met de NIM-CLASS A-variant en de NIM-CLASS B-variant tonen het gunstige effect van actieve *top-down* processen op de classificatie van natuurlijke visuele input, als de processen gebruik maken van verscheidene soorten van opgeslagen kennis.

Hoofdstuk 7 bediscussieert het gesitueerde computationele geheugenmodel. Ten eerste relateert het onze benadering aan invloedrijke bestaande modellen van objectherkenning. Ten tweede identificeert het verscheidene uitbreidingen van het model voor de verbetering van de psychologische en biologische plausibiliteit als model van natuurlijke cognitie. Op basis van psychologische en biologische inzichten worden vijf uitbreidingen besproken: (1) een kenmerk-gebaseerd aandachtsmechanisme, (2) een spatieel aandachtsmechanisme, (3) een neurale implementatie van de representatieruimte, (4) de representatie van spatiële kennis, en (5) de implementatie van aparte episodische en semantische representatieruimten. Ten derde plaatst het hoofdstuk onze benadering in de context van de globale ontwikkelingen op het gebied van cognitief modelleren. Het toont aan dat onze benadering verschilt van de traditionele computationele geheugenmodellen en aansluit bij de nieuwe 'gesitueerde' benadering door zich te richten op de interactie met een realistische omgeving.

Hoofdstuk 8 beantwoordt de drie onderzoeksvragen die in hoofdtsuk 1 geformuleerd zijn en geeft onze conclusie ten aanzien van de probleemstelling. Het ant-

woord op de eerste onderzoeksvraag luidt dat het gesitueerde model de menselijke responsen op individuele natuurlijke stimuli tamelijk nauwkeurig kan produceren. Het antwoord op de tweede onderzoeksvraag luidt dat het gesitueerde model de vier herkenningseffecten succesvol kan produceren op basis van natuurlijke visuele input. Het antwoord op de derde onderzoeksvraag luidt dat ons gesitueerde model in staat is om natuurlijke afbeeldingen onder een variëteit aan ongunstige omstandigheden correct te classificeren mits er voldoende visuele input geselecteerd is. Bovendien zien we dat de classificatieprestatie verbetert door het gebruik van *top-down* processen die gebruik maken van opgeslagen kennis voor het selecteren van visuele input. Dit is voornamelijk zo wanneer classificatie gebaseerd is op een beperkte hoeveelheid visuele input. Tenslotte, mogen we concluderen dat het gesitueerde computationele model dat in dit proefschrift wordt gepresenteerd een vruchtbare oplossing biedt voor het funderings- en transductieprobleem. Dit gebeurt door het bewerkstelligen van een relatie tussen een representatie en zijn referent in de wereld op een transparante en niet-triviale manier die geïnformeerd is vanuit de neurobiologie.

Curriculum vitae

Joyca Lacroix was born in Eindhoven, The Netherlands, on March 23, 1977. From 1988 to 1994, she attended secondary school (VWO) at the Eckart College in Eindhoven. In 1996, she received her propadeutics in Mathematics from Eindhoven University of Technology. In the same year she started a study Psychology at Maastricht University in Maastricht, to be completed in 2001 with a M.Sc. degree in Cognitive Psychology. In addition to the Psychology courses, she followed courses in Econometrics, Health Sciences, and Knowledge Engineering during her study time. After her graduation, she worked as a scientific teacher at Erasmus University in Rotterdam. Subsequently, in 2002, the focus of her career was redirected towards research, when she accepted a position as a Ph.D. researcher at the Department of Computer Science at Maastricht University. In the framework of the NWO Cognition Program project 'Events in Memory and Environment' (project number: 051.02.2002), she examined computational models of natural cognition under the auspices of the research school SIKS. As of November 2006, Joyca Lacroix is employed as a postdoc at the Department of Psychology at Leiden University in Leiden, where she is investigating models of perception and memory for cognitive robots in the context of the EU-funded Cognitive Systems project 'PACO-PLUS'. Besides the scientific work in the domain of psychology and artificial intelligence, she likes to be engaged in outdoor activities including, running, cycling, and snowboarding.

SIKS Dissertation Series[1]

1998

1 Johan van den Akker (CWI) *DEGAS - An Active, Temporal Database of Autonomous Objects*

2 Floris Wiesman (UM) *Information Retrieval by Graphically Browsing Meta-Information*

3 Ans Steuten (TUD) *A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective*

4 Dennis Breuker (UM) *Memory versus Search in Games*

5 Eduard Oskamp (RUL) *Computerondersteuning bij Straftoemeting*

1999

1 Mark Sloof (VU) *Physiology of Quality Change Modelling; Automated Modelling of Quality Change of Agricultural Products*

2 Rob Potharst (EUR) *Classification using Decision Trees and Neural Nets*

3 Don Beal (UM) *The Nature of Minimax Search*

4 Jacques Penders (UM) *The Practical Art of Moving Physical Objects*

5 Aldo de Moor (KUB) *Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems*

6 Niek Wijngaards (VU) *Re-Design of Compositional Systems*

7 David Spelt (UT) *Verification Support for Object Database Design*

8 Jacques Lenting (UM) *Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation*

2000

1 Frank Niessink (VU) *Perspectives on Improving Software Maintenance*

2 Koen Holtman (TU/e) *Prototyping of CMS Storage Management*

3 Carolien Metselaar (UvA) *Sociaal-Organisatorische Gevolgen van Kennistechnologie; een Procesbenadering en Actorperspectief*

4 Geert de Haan (VU) *ETAG, A Formal Model of Competence Knowledge for User Interface Design*

[1] Abbreviations: SIKS – Dutch Research School for Information and Knowledge Systems; CWI – Centrum voor Wiskunde en Informatica, Amsterdam; EUR – Erasmus Universiteit, Rotterdam; RUG – Rijsuniversiteit Groningen; KUB – Katholieke Universiteit Brabant, Tilburg; KUN – Katholieke Universiteit Nijmegen; RUL – Rijksuniversiteit Leiden; RUN – Radboud Universiteit Nijmegen; TUD – Technische Universiteit Delft; TU/e – Technische Universiteit Eindhoven; UL – Universiteit Leiden; UM – Universiteit Maastricht; UT – Universiteit Twente, Enschede; UU – Universiteit Utrecht; UvA – Universiteit van Amsterdam; UvT – Universiteit van Tilburg; VU – Vrije Universiteit, Amsterdam.

5 Ruud van der Pol (UM) *Knowledge-Based Query Formulation in Information Retrieval*
6 Rogier van Eijk (UU) *Programming Languages for Agent Communication*
7 Niels Peek (UU) *Decision-Theoretic Planning of Clinical Patient Management*
8 Veerle Coupé (EUR) *Sensitivity Analyis of Decision-Theoretic Networks*
9 Florian Waas (CWI) *Principles of Probabilistic Query Optimization*
10 Niels Nes (CWI) *Image Database Management System Design Considerations, Algorithms and Architecture*
11 Jonas Karlsson (CWI) *Scalable Distributed Data Structures for Database Management*

2001

1 Silja Renooij (UU) *Qualitative Approaches to Quantifying Probabilistic Networks*
2 Koen Hindriks (UU) *Agent Programming Languages: Programming with Mental Models*
3 Maarten van Someren (UvA) *Learning as Problem Solving*
4 Evgueni Smirnov (UM) *Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets*
5 Jacco van Ossenbruggen (VU) *Processing Structured Hypermedia: A Matter of Style*
6 Martijn van Welie (VU) *Task-Based User Interface Design*
7 Bastiaan Schonhage (VU) *Diva: Architectural Perspectives on Information Visualization*
8 Pascal van Eck (VU) *A Compositional Semantic Structure for Multi-Agent Systems Dynamics*
9 Pieter Jan 't Hoen (RUL) *Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes*
10 Maarten Sierhuis (UvA) *Modeling and Simulating Work Practice BRAHMS: a Multiagent Modeling and Simulation Language for Work Practice Analysis and Design*
11 Tom van Engers (VU) *Knowledge Management: The Role of Mental Models in Business Systems Design*

2002

1 Nico Lassing (VU) *Architecture-Level Modifiability Analysis*
2 Roelof van Zwol (UT) *Modelling and Searching Web-Based Document Collections*
3 Henk Ernst Blok (UT) *Database Optimization Aspects for Information Retrieval*
4 Juan Roberto Castelo Valdueza (UU) *The Discrete Acyclic Digraph Markov Model in Data Mining*
5 Radu Serban (VU) *The Private Cyberspace Modeling Electronic Environments Inhabited by Privacy-Concerned Agents*
6 Laurens Mommers (UL) *Applied Legal Epistemology; Building a Knowledge-Based Ontology of the Legal Domain*
7 Peter Boncz (CWI) *Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications*
8 Jaap Gordijn (VU) *Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas*
9 Willem-Jan van den Heuvel (KUB) *Integrating Modern Business Applications with Objectified Legacy Systems*

10 Brian Sheppard (UM) *Towards Perfect Play of Scrabble*
11 Wouter Wijngaards (VU) *Agent Based Modelling of Dynamics: Biological and Organisational Applications*
12 Albrecht Schmidt (UvA) *Processing XML in Database Systems*
13 Hongjing Wu (TU/e) *A Reference Architecture for Adaptive Hypermedia Applications*
14 Wieke de Vries (UU) *Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems*
15 Rik Eshuis (UT) *Semantics and Verification of UML Activity Diagrams for Workflow Modelling*
16 Pieter van Langen (VU) *The Anatomy of Design: Foundations, Models and Applications*
17 Stefan Manegold (UvA) *Understanding, Modeling, and Improving Main-Memory Data-Base Performance*

2003

1 Heiner Stuckenschmidt (VU) *Ontology-Based Information Sharing in Weakly Structured Environments*
2 Jan Broersen (VU) *Modal Action Logics for Reasoning About Reactive Systems*
3 Martijn Schuemie (TUD) *Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy*
4 Petkovic (UT) *Content-Based Video Retrieval Supported by Database Technology*
5 Jos Lehmann (UvA) *Causation in Artificial Intelligence and Law – A Modelling Approach*
6 Boris van Schooten (UT) *Development and Specification of Virtual Environments*
7 Machiel Jansen (UvA) *Formal Explorations of Knowledge Intensive Tasks*
8 Yong-Ping Ran (UM) *Repair-Based Scheduling*
9 Rens Kortmann (UM) *The Resolution of Visually Guided Behaviour*
10 Andreas Lincke (UT) *Electronic Business Negotiation: Some Experimental Studies on the Interaction between Medium, Innovation Context and Cult*
11 Simon Keizer (UT) *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks*
12 Roeland Ordelman (UT) *Dutch Speech Recognition in Multimedia Information Retrieval*
13 Jeroen Donkers (UM) *Nosce Hostem – Searching with Opponent Models*
14 Stijn Hoppenbrouwers (KUN) *Freezing Language: Conceptualisation Processes across ICT-Supported Organisations*
15 Mathijs de Weerdt (TUD) *Plan Merging in Multi-Agent Systems*
16 Menzo Windhouwer (CWI) *Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouse*
17 David Jansen (UT) *Extensions of Statecharts with Probability, Time, and Stochastic Timing*
18 Levente Kocsis (UM) *Learning Search Decisions*

2004

1 Virginia Dignum (UU) *A Model for Organizational Interaction: Based on Agents, Founded in Logic*

2 Lai Xu (UvT) *Monitoring Multi-Party Contracts for E-Business*
3 Perry Groot (VU) *A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving*
4 Chris van Aart (UvA) *Organizational Principles for Multi-Agent Architectures*
5 Viara Popova (EUR) *Knowledge Discovery and Monotonicity*
6 Bart-Jan Hommes (TUD) *The Evaluation of Business Process Modeling Techniques*
7 Elise Boltjes (UM) *Voorbeeld$_{IG}$ Onderwijs; Voorbeeldgestuurd Onderwijs, een Opstap naar Abstract Denken, vooral voor Meisjes*
8 Joop Verbeek (UM) *Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale Politiële Gegevensuitwisseling en Digitale Expertise*
9 Martin Caminada (VU) *For the Sake of the Argument; Explorations into Argument-Based Reasoning*
10 Suzanne Kabel (UvA) *Knowledge-rich Indexing of Learning Objects*
11 Michel Klein (VU) *Change Management for Distributed Ontologies*
12 The Duy Bui (UT) *Creating Emotions and Facial Expressions for Embodied Agents*
13 Wojciech Jamroga (UT) *Using Multiple Models of Reality: On Agents who Know how to Play*
14 Paul Harrenstein (UU) *Logic in Conflict. Logical Explorations in Strategic Equilibrium*
15 Arno Knobbe (UU) *Multi-Relational Data Mining*
16 Federico Divina (VU) *Hybrid Genetic Relational Search for Inductive Learning*
17 Mark Winands (UM) *Informed Search in Complex Games*
18 Vania Bessa Machado (UvA) *Supporting the Construction of Qualitative Knowledge Models*
19 Thijs Westerveld (UT) *Using generative probabilistic models for multimedia retrieval*
20 Madelon Evers (Nyenrode) *Learning from Design: facilitating multidisciplinary design teams*

2005

1 Floor Verdenius (UvA) *Methodological Aspects of Designing Induction-Based Applications*
2 Erik van der Werf (UM) *AI techniques for the game of Go*
3 Franc Grootjen (RUN) *A Pragmatic Approach to the Conceptualisation of Language*
4 Nirvana Meratnia (UT) *Towards Database Support for Moving Object data*
5 Gabriel Infante-Lopez (UvA) *Two-Level Probabilistic Grammars for Natural Language Parsing*
6 Pieter Spronck (UM) *Adaptive Game AI*
7 Flavius Frasincar (TU/e) *Hypermedia Presentation Generation for Semantic Web Information Systems*
8 Richard Vdovjak (TU/e) *A Model-Driven Approach for Building Distributed Ontology-Based Web Applications*
9 Jeen Broekstra (VU) *Storage, Querying and Inferencing for Semantic Web Languages*
10 Anders Bouwer (UvA) *Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments*
11 Elth Ogston (VU) *Agent Based Matchmaking and Clustering - A Decentralized Approach to Search*

12 Csaba Boer (EUR) *Distributed Simulation in Industry*

13 Fred Hamburg (UL) *Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen*

14 Borys Omelayenko (VU) *Web-Service Configuration on the Semantic Web; Exploring How Semantics Meets Pragmatics*

15 Tibor Bosse (VU) *Analysis of the Dynamics of Cognitive Processes*

16 Joris Graaumans (UU) *Usability of XML Query Languages*

17 Boris Shishkov (TUD) *Software Specification Based on Re-usable Business Components*

18 Danielle Sent (UU) *Test-Selection Strategies for Probabilistic Networks*

19 Michel van Dartel (UM) *Situated Representation*

20 Cristina Coteanu (UL) *Cyber Consumer Law, State of the Art and Perspectives*

21 Wijnand Derks (UT) *Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics*

2006

1 Samuil Angelov (TU/e) *Foundations of B2B Electronic Contracting*

2 Cristina Chisalita (VU) *Contextual Issues in the Design and Use of Information Technology in Organizations*

3 Noor Christoph (UvA) *The Role of Metacognitive Skills in Learning to Solve Problems*

4 Marta Sabou (VU) *Building Web Service Ontologies*

5 Cees Pierik (UU) *Validation Techniques for Object-Oriented Proof Outlines*

6 Ziv Baida (VU) *Software-Aided Service Bundling - Intelligent Methods and Tools for Graphical Service Modeling*

7 Marko Smiljanic (UT) *XML Schema Matching - Balancing Efficiency and Effectiveness by means of Clustering*

8 Eelco Herder (UT) *Forward, Back and Home Again - Analyzing User Behavior on the Web*

9 Mohamed Wahdan (UM) *Automatic Formulation of the Auditor's Opinion*

10 Ronny Siebes (VU) *Semantic Routing in Peer-to-Peer Systems*

11 Joeri van Ruth (UT) *Flattening Queries over Nested Data Types*

12 Bert Bongers (VU) *Interactivation - Towards an E-cology of People, our Technological Environment, and the Arts*

13 Henk-Jan Lebbink (UU) *Dialogue and Decision Games for Information Exchanging Agents*

14 Johan Hoorn (VU) *Software Requirements: Update, Upgrade, Redesign - Towards a Theory of Requirements Change*

15 Rainer Malik (UU) *CONAN: Text Mining in the Biomedical Domain*

16 Carsten Riggelsen (UU) *Approximation Methods for Efficient Learning of Bayesian Networks*

17 Stacey Nagata (UU) *User Assistance for Multitasking with Interruptions on a Mobile Device*

18 Valentin Zhizhkun (UvA) *Graph Transformation for Natural Language Processing*

19 Birna van Riemsdijk (UU) *Cognitive Agent Programming: A Semantic Approach*

20 Marina Velikova (UvT) *Monotone Models for Prediction in Data Mining*

21 Bas van Gils (RUN) *Aptness on the Web*

22 Paul de Vrieze (RUN) *Fundaments of Adaptive Personalisation*

23 Ion Juvina (UU) *Development of Cognitive Model for Navigating on the Web*

24 Laura Hollink (VU) *Semantic Annotation for Retrieval of Visual Resources*

25 Madalina Drugan (UU) *Conditional Log-Likelihood MDL and Evolutionary MCMC*

26 Vojkan Mihajlovic (UT) *Score Region Algebra: A Flexible Framework for Structured Information Retrieval*

27 Stefano Bocconi (CWI) *Vox Populi: Generating Video Documentaries from Semantically Annotated Media Repositories*

28 Borkur Sigurbjornsson (UvA) *Focused Information Access using XML Element Retrieval*

2007

1 Kees Leune (UvT) *Access Control and Service-Oriented Architectures*

2 Wouter Teepe (RUG) *Reconciling Information Exchange and Confidentiality: A Formal Approach*

3 Peter Mika (VU) *Social Networks and the Semantic Web*

4 Jurriaan van Diggelen (UU) *Achieving Semantic Interoperability in Multi-Agent Systems: A Dialogue-Based Approach*

5 Bart Schermer (UL) *Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-Enabled Surveillance*

6 Gilad Mishne (UvA) *Applied Text Analytics for Blogs*

7 Natasa Jovanovic' (UT) *To Whom It May Concern - Addressee Identification in Face-to-Face Meetings*

8 Mark Hoogendoorn (VU) *Modeling of Change in Multi-Agent Organizations*

9 David Mobach (VU) *Agent-Based Mediated Service Negotiation*

10 Huib Aldewereld (UU) *Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols*

11 Natalia Stash (TU/e) *Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System*

12 Marcel van Gerven (RUN) *Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty*

13 Rutger Rienks (UT) *Meetings in Smart Environments; Implications of Progressing Technology*

14 Niek Bergboer (UM) *Context-Based Image Analysis*

15 Joyca Lacroix (UM) *NIM: A Situated Computational Memory Model*