Visual Identification, Categorization and Recollection

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

Martin Wiesmann

M.SC., University of Bern

Date of birth
17. 1. 1976

citizen of
Horgen (ZH)

Examiner and co-examiners

Prof. Dr. Richard Hahnloser
Prof. Dr. Alumit Ishai
PD Dr. Daniel Kiper

2009
Visual Identification, Categorization and Recollection

Summary

In this PhD dissertation I am describing three experiments on category learning and recollection memory in humans. In the first experiment I investigated the processes of identification and categorization of visual patterns. It is currently unknown whether identification and categorization of visual stimuli are supported by separate or identical neural circuits. Moreover, the dynamics of these processes are poorly understood: Does identification precede categorization, or vice-versa? To answer these questions, I measured how fast visual patterns are identified and categorized. In an identification task, subjects were initially trained to recognize a given pattern. In subsequent sessions, they had to indicate whether each of a series of patterns was identical to the learned pattern or not. In the categorization task, subjects trained with similar patterns defining an arbitrary category. They then had to indicate whether subsequently presented patterns belonged to the category or not. I measured accuracy and reaction times in both tasks. Performance improved during the first 3 to 4 sessions, then stabilized. While accuracy in both tasks showed similar dynamics with practice, the reaction times behaved differently: In the first session, the average reaction time for categorization was 716±193 ms, and 550±78 ms for identification. With training, categorization reaction times decreased more rapidly than in the identification task. Both types of reaction times then converged around 467 ms (500±81 ms categorization; 434±24 ms recognition). These results suggest that identification and categorization are two distinct processes.

The aim of the second experiment was to test whether the number of categories has an effect on the performance in a categorization task with rule-based and information-integration category structures. Glass patterns were divided into two different category structures. The rule-based categories can be learned via some explicit reasoning process, in this study the category boundaries were along one dimension, namely the distance between dots of a dot-pair. In the information-integration conditions, the form of the Glass patterns changed gradually from circular to radial, and this stimulus space was
divided into the different categories. This kind of category structure is hard to verbalize and accuracy is maximized only if information from two or more dimensions is integrated at some predecisional stage. These two category structures were divided into two or four categories, respectively. During the task, subjects indicated to which category the presented stimulus belonged. Subjects conducted seven sessions in each condition. The performance, averaged across all subjects and sessions, was virtually identical in the information-integration task for the two- and the four-category condition, whereas the performance in the rule-based four-category condition was significantly lower than in the two-category condition. The response latencies showed a similar pattern of results. While the reaction times of the two conditions in the information-integration task were not significantly different from each other, the response latencies in the four-category condition of the rule-based task were significantly higher than in the two-category condition, thus ruling out possible speed-accuracy trade-off explanations. These results support the hypothesis that rule-based and information-integration category learning is mediated by distinct neural substrates.

In the third study, event-related fMRI was used to investigate whether recollection- and familiarity-based memory judgments are modulated by the degree of visual similarity between old and new art paintings. Subjects performed a flower detection task, followed by a Remember/Know/New surprise memory test. The old paintings were randomly presented with new paintings, which were either visually similar or visually different. Consistent with the prediction, subjects were significantly faster and more accurate to reject new, visually different paintings than new, visually similar ones. The proportion of false alarms, namely remember and know responses to new paintings, was significantly reduced with decreased visual similarity. The retrieval task evoked activation in multiple visual, parietal and prefrontal regions, within which remember judgments elicited stronger activation than know judgments. New, visually different paintings evoked weaker activation than new, visually similar items in the intraparietal sulcus. Contrasting recollection with familiarity revealed activation predominantly within the precuneus, where the BOLD response elicited by recollection peaked significantly earlier than the BOLD response evoked by familiarity judgments. These findings suggest that successful
memory retrieval of pictures is mediated by activation in a distributed cortical network, where memory strength is manifested by differential hemodynamic profiles. Recollection- and familiarity-based memory decisions may therefore reflect strong memories and weak memories, respectively.
Visuelle Identifizierung, Kategorisierung und Erinnerungsvermögen

Zusammenfassung


Das Ziel der zweiten Studie war, den Effekt der Anzahl der Kategorien auf die Leistung in einer Kategorisierungsaufgabe mit entweder regel-basierter oder informations-

CURRICULUM VITAE

Martin Wiesmann, M.Sc.
Institute of Neuroradiology
University of Zurich
Winterthurerstrasse 190
8057 Zurich, Switzerland
Phone: +41-44-635-3445 Fax: +41-44-635-3449
Email: martin@ini.phys.ethz.ch

ACADEMIC POSITIONS
2008-present  Research Fellow, Institute of Neuroradiology, University of Zurich.
2004-2004  Research Fellow, Forschungsanstalt Agroscope Reckenholz-Tänikon ART, Zurich
2003-2003  Research Fellow, Space Sciences, University of Bern

EDUCATION
2004-present  Ph.D., Neuroinformatics, Institute of Neuroinformatics, ETH Zurich.
1997-2003  M.Sc., Physics, Space Sciences, University of Bern.

PUBLICATIONS

Peer-Reviewed Article


Other Publications


Abstracts


Acknowledgment

I would like to thank my supervisor, Dr. Daniel Kiper, for supporting me over the years and giving me the opportunity to do this PhD thesis. And I want to especially thank my other supervisor, Prof. Dr. Alumit Ishai, for her great support during the last two years, and for helping me with the manuscript, and also for giving me the opportunity to do fMRI-experiments. Furthermore, I would like to thank Dr. Armin Heinecke, for his support on the program BrainVoyager.
# Category learning

4.1 Neurobiology of category learning .................................. 38
  4.1.1 Neuroimaging experiments .................................... 41
4.2 Categorization and Identification ................................. 45
  4.2.1 Evidences for multiple memory systems .................... 45
4.3 Models of categorization ........................................... 52
4.4 Category learning tasks ............................................. 66

# Recollection and Familiarity

5.1 fMRI studies ......................................................... 74

# Outline of the experimental work

6 Outline of the experimental work .................................. 76

## Learning in pattern identification and categorization tasks

II Learning in pattern identification and categorization tasks 79

# Introduction

7 Introduction ............................................................. 79

# Methods

8 Methods ................................................................. 82
  8.1 Participants ...................................................... 82
  8.2 Stimuli ............................................................ 82
  8.3 Procedure ....................................................... 83
  8.4 Equipment ....................................................... 87
  8.5 Data analysis .................................................... 87

# Results

9 Results ................................................................. 88

# Conclusions

10 Conclusions ........................................................ 97

## The effect of category number on rule-based and information-integration category learning

III The effect of category number on rule-based and information-integration category learning 101
11 Introduction
  11.1 Overview of the study

12 Methods
  12.1 Participants
  12.2 Stimuli
  12.3 Procedure
  12.4 Equipment

13 Results
  13.1 Accuracy
  13.2 Reaction times

14 Discussion

IV Recollection- and Familiarity-based Decisions Reflect Memory Strength

15 Introduction

16 Behavioral pilot study
  16.1 Methods
  16.2 Results

17 Imaging study
  17.1 Methods
  17.1.1 Subjects
  17.1.2 Stimuli and tasks
  17.1.3 Data acquisition
  17.1.4 Data analysis
  17.2 Results
  17.2.1 Study phase
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Structure of the eye</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>The structure of the retina</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Sensitivity of rods and cones</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>Density of cones and rods in the retina</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>The visual pathway</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>Memory systems</td>
<td>31</td>
</tr>
<tr>
<td>7</td>
<td>Recognition and Categorization</td>
<td>38</td>
</tr>
<tr>
<td>8</td>
<td>Examples of stimuli from [1]</td>
<td>43</td>
</tr>
<tr>
<td>9</td>
<td>Behavioural results from [2]</td>
<td>47</td>
</tr>
<tr>
<td>10</td>
<td>The classic GT model</td>
<td>57</td>
</tr>
<tr>
<td>11</td>
<td>A GRT process</td>
<td>58</td>
</tr>
<tr>
<td>12</td>
<td>Example of category structure for GCM</td>
<td>61</td>
</tr>
<tr>
<td>13</td>
<td>Sensitivity parameter</td>
<td>65</td>
</tr>
<tr>
<td>14</td>
<td>The dot pattern categorization task</td>
<td>68</td>
</tr>
<tr>
<td>15</td>
<td>Example of category structure</td>
<td>72</td>
</tr>
<tr>
<td>16</td>
<td>Stimuli</td>
<td>84</td>
</tr>
<tr>
<td>17</td>
<td>Task</td>
<td>85</td>
</tr>
<tr>
<td>18</td>
<td>Control experiment</td>
<td>88</td>
</tr>
<tr>
<td>19</td>
<td>Results of categorization task</td>
<td>89</td>
</tr>
<tr>
<td>20</td>
<td>Results of identification task</td>
<td>90</td>
</tr>
<tr>
<td>21</td>
<td>Performance as a function of the distortion level</td>
<td>92</td>
</tr>
<tr>
<td>22</td>
<td>Performance as a function of time</td>
<td>93</td>
</tr>
<tr>
<td>23</td>
<td>Reaction times as a function of time</td>
<td>94</td>
</tr>
<tr>
<td>24</td>
<td>Reaction time as a function of performance</td>
<td>96</td>
</tr>
<tr>
<td>25</td>
<td>Indexes</td>
<td>98</td>
</tr>
<tr>
<td>26</td>
<td>Stimulus space from Maddox’s experiment [3]</td>
<td>104</td>
</tr>
<tr>
<td>27</td>
<td>Results from Maddox’s experiment [3]</td>
<td>105</td>
</tr>
<tr>
<td>28</td>
<td>Examples of stimuli</td>
<td>107</td>
</tr>
<tr>
<td>29</td>
<td>Performances and reaction times</td>
<td>112</td>
</tr>
</tbody>
</table>
30 Proportion correct averaged over all sessions . . . . . . . . . . 114
31 Reaction time averaged over all sessions . . . . . . . . . . . . . 115
32 Stimuli used in the experiment by Yago and Ishai [4] . . . . . 124
33 Activation in parietal cortex during memory retrieval (from [4]) 125
34 Behavioral data from the pilot study . . . . . . . . . . . . . . 127
35 Behavioral data from the encoding session . . . . . . . . . . . . 132
36 Activation during encoding . . . . . . . . . . . . . . . . . . . . 133
37 Behavioral data from the retrieval session . . . . . . . . . . . . 134
38 Activation as a function of task performance . . . . . . . . . . 136
39 Visual similarity effect . . . . . . . . . . . . . . . . . . . . . . 137
40 Activation of Remember judgments . . . . . . . . . . . . . . . 138
41 Main effect during retrieval . . . . . . . . . . . . . . . . . . . . 139
42 Recollection vs Familiarity . . . . . . . . . . . . . . . . . . . . 140
43 Activation in the medial temporal lobe . . . . . . . . . . . . . . 141
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>Anterior cingulate cortex</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>BA</td>
<td>Brodmann area</td>
</tr>
<tr>
<td>BOLD</td>
<td>Blood-oxygen-level dependent</td>
</tr>
<tr>
<td>CFE</td>
<td>Categorical fluency effect</td>
</tr>
<tr>
<td>COVIS</td>
<td>Competition between verbal and implicit systems</td>
</tr>
<tr>
<td>DNMS</td>
<td>Delayed non-match to sample</td>
</tr>
<tr>
<td>DOC</td>
<td>Dorsal occipital cortex</td>
</tr>
<tr>
<td>EBA</td>
<td>Exastriate body part area</td>
</tr>
<tr>
<td>ERP</td>
<td>Event-related potentials</td>
</tr>
<tr>
<td>FFA</td>
<td>Fusiform face area</td>
</tr>
<tr>
<td>FG</td>
<td>Fusiform gyrus</td>
</tr>
<tr>
<td>fMRI</td>
<td>Functional magnetic resonance imaging</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full-width-half-maximum</td>
</tr>
<tr>
<td>GCM</td>
<td>Generalized context model</td>
</tr>
<tr>
<td>GRT</td>
<td>General recognition theory</td>
</tr>
<tr>
<td>HIP</td>
<td>Hippocampus</td>
</tr>
<tr>
<td>IFG</td>
<td>Inferior frontal gyrus</td>
</tr>
<tr>
<td>IOG</td>
<td>Inferior occipital gyrus</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>IPS</td>
<td>Intraparietal sulcus</td>
</tr>
<tr>
<td>IT</td>
<td>Inferior temporal gyrus</td>
</tr>
<tr>
<td>LGN</td>
<td>Lateral geniculate nuclei</td>
</tr>
<tr>
<td>MDS</td>
<td>Multi-dimensional-scaling</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MTL</td>
<td>Medial temporal lobe</td>
</tr>
<tr>
<td>OFA</td>
<td>Occipital face area</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PFC</td>
<td>Prefrontal cortex</td>
</tr>
<tr>
<td>PHC</td>
<td>Parahippocampal cortex</td>
</tr>
<tr>
<td>PPA</td>
<td>Parahippocampal place area</td>
</tr>
<tr>
<td>PRS</td>
<td>Perceptual representation memory system</td>
</tr>
<tr>
<td>RF</td>
<td>Receptive field</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristics</td>
</tr>
<tr>
<td>ROI</td>
<td>Region-of-interest</td>
</tr>
<tr>
<td>SC</td>
<td>Superior culliculus</td>
</tr>
<tr>
<td>SCM</td>
<td>Similarity-choice model</td>
</tr>
<tr>
<td>SE</td>
<td>Standard error</td>
</tr>
<tr>
<td>SPL</td>
<td>Superior parietal lobule</td>
</tr>
<tr>
<td>TE</td>
<td>Inferotemporal cortex</td>
</tr>
</tbody>
</table>
V1 - Primary visual cortex
V2 - Visual area 2, or prestriate cortex
V3 - Third visual complex
V4 - Visual area 4
WCST - Wisconsin Card sorting test
Part I

Introduction

2 The Central Visual System

Although our visual system provides us with a coherent picture of the world around us, this picture has multiple facets. Objects we see have shape and color. They have a position in space, and sometimes they move. All these different information are extracted and integrated by neurons in the visual pathways, so that we perceive meaningful objects.

2.1 The structure of the eye

The eye is an organ specialized for the detection, localization and analysis of light.

Figure 1 shows the coarse structure of the eye. The pupil is the opening that allows light to enter the eye and reach the retina; it appears dark because of the light-absorbing pigments in the retina. The pupil is surrounded by the iris, whose pigmentation provides what we call the eye’s color. The iris contains two muscles that can vary the size of the pupil; one makes it smaller when it contracts, the other makes it larger. The light then reaches the retina.

The light is absorbed and detected by the photoreceptors. This information is propagated to bipolar cells and then further to ganglion cells. The ganglion cells fire action potentials in response to light, and these impulses propagate down the optic nerve to the rest of the brain. Besides the cells in this direct path from photoreceptor to brain, retinal processing is influenced by two additional cell types. Horizontal cells receive input from the photoreceptors and project neurites laterally to influence surrounding bipolar cells and photoreceptors. Amacrine cells receive input from bipolar cells
and project laterally to influence surrounding ganglion cells, bipolar cells, and other amacrine cells. The retina has a laminar organization: Cells are organized in layers, that are seemingly inside-out; light must pass through the ganglion cells and bipolar cells before it reaches the photoreceptors. Because the retinal cells above the photoreceptors are relatively transparent, image distortion is minimal, as light passes through them. One reason why this arrangement is advantageous is that the pigmented epithelium that lies below the photoreceptors plays a critical role in the maintenance of the photoreceptors and photopigments. The pigmented epithelium also absorbs any light that passes entirely through the retina, thus minimizing the reflection of light within the eye that would blur the image.

The organization of the cell layers of the retina is shown in Figure 2. The first layer is the ganglion cell layer, which contains the cell bodies of the ganglion cells. Next is the inner nuclear layer, which contains the cell bodies of the bipolar cells, the horizontal and amacrine cells. The next layer is the outer nuclear layer, which contains the cell bodies of the photoreceptors. Finally, the layer of photoreceptors outer segments contains the light-sensitive outer segments of the photoreceptors. The outer segments are embedded in the pigmented epithelium.

### 2.1.1 Photoreceptor structure

The conversion of the light signal into neural signals occurs in the photoreceptors at the back of the retina. Every photoreceptor has four regions: an outer segment, an inner segment, a cell body, and a synaptic terminal. The light is absorbed by the photopigments in the disk membranes of the outer segment, thereby triggers changes in the photoreceptor membrane potential.

There are two different types of photoreceptor cells: The rod photoreceptors have a long, cylindrical outer segment, containing many disks; and the cone photoreceptors have a shorter, tapering outer segment with fewer membranous disks. The higher number of disks and higher photopigment...
Figure 1: Structure of the eye

Figure 2: The structure of the retina
concentration in rods makes them over 1000 times more sensitive to light than cones. Indeed, under nighttime lighting only rods contribute to vision. On the other hand, under daylight, cones are the source for vision.

Another difference between rods and cones is that all rods contain the same photopigment, and the peak of their absorbance is at a wavelength of 498 nm. Whereas there are three types of cones, each containing a different pigment. The variations among pigments make the different cones sensitive to different wavelengths of light (see Figure 3). Thus, only the cones, not the rods, are responsible for our ability to perceive color.

The distribution of cones and rods vary from the fovea to the retinal periphery (see Figure 4). Near the fovea, the density of cones is much higher than that of the rods, whereas towards the retinal periphery the density of the cones decreases rapidly and the ratio of rods to cones increases. In the periphery there is also a higher ratio of photoreceptors to ganglion cells. The combined effect of this arrangement is that the peripheral retina is more sensitive to weak light, because rods are specialized for low light, and there are more photoreceptors feeding information to each ganglion cell.

These characteristics enable the peripheral retina to detect faint light, but the resolution in daylight is poor, because good visual acuity requires a low ratio of photoreceptors to ganglion cells and daytime vision requires cones. The region of retina most highly specialized for high-resolution vision is the fovea. Right at the fovea, there are only cones and no rods.

2.1.2 Bipolar cells

Bipolar cells can be categorized into two classes, the ON and OFF bipolar cell. An ON bipolar cell depolarizes, when it gets input from a photoreceptor, while the OFF bipolar cell hyperpolarizes. Each bipolar cell receives direct synaptic input from a cluster of photoreceptors. The number of photoreceptors in this cluster ranges from one at the center of the fovea to thousands in the peripheral retina. Additionally, bipolar cells also are connected via
Figure 3: Sensitivity of rods and cones

Figure 4: Density of cones and rods in the retina
horizontal cells to a circumscribed ring of photoreceptors that surrounds this central cluster.

The receptive field of a bipolar cell is the area of retina that changes the cell’s membrane potential, when stimulated with light. The arrangement of inputs to bipolar cells mentioned above leads to a so-called center-surround receptive field. This receptive field consists of two parts: a circular area of retina providing direct photoreceptor input, called the receptive field center, and a surrounding area of retina providing input via horizontal cells, called the receptive field surround. Bipolar receptive field diameters range from a fraction of a degree in the central retina to several degrees in the peripheral retina. The response of a bipolar cell’s membrane potential to light in the receptive field center is opposite to that of light in the surround. For example, an ON bipolar cell depolarizes when the center of the receptive field is stimulated, and hyperpolarizes when the surround is illuminated [5]. The center-surround receptive field is passed on from bipolar cells to ganglion cells via synapses.

2.1.3 Ganglion cells

Most ganglion cells in the mammalian retina have a center-surround receptive field with either an ON or an OFF center. They can be further categorized based on their appearance and connectivity. Two major types of ganglion cells in the human retina have been identified: large M-type ganglion cells and smaller P-type ganglion cells. P cells constitute about 90% of the ganglion cell population, M cells constitute about 5%, and the remaining 5% is made up of a variety of nonM-nonP ganglion cell types that are less well characterized. M cells have larger receptive fields than P cells, they are more sensitive to low-contrast stimuli, and they conduct action potentials more rapidly in the optic nerve.

Another important distinction between ganglion cell types is that some P cells are sensitive to differences in the wavelength of light. These color-
sensitive cells are called color-opponent cells, because a response to one wavelength in the receptive field center is canceled by showing another wavelength in the receptive field surround. Two types of opponency are found: red versus green and blue versus yellow [6]. The information from the ganglion cells is transmitted through the optic nerve to the brain.

2.2 Visual pathways

The neural pathway that leaves the eye, begins with the optic nerve (see figure 5). The optic nerves from both eyes combine to form the optic chiasm, which lies at the base of the brain. Axons originating in the nasal retinas cross from one side to the other, whereas axons from the temporal retina stay on the original side of the brain. The left visual hemifield from both eyes is so projected to the right side of the brain through the right optic nerve and the right visual hemifield is projected to the left side of the brain through the left optic nerve.

A small proportion of optic tract axons from synaptic connections with cells in the hypothalamus or continue past the thalamus to innervate the midbrain. These direct projections to part of the hypothalamus play an important role in synchronizing a variety of biological rhythms, including sleep and wakefulness, with the daily dark-light cycle. Direct projections to part of the midbrain, called the pretectum, control the size of the pupil and certain types of eye movements. And a small part of the ganglion cells in the retina project to a part of the midbrain called the superior colliculus (SC). The SC is involved in orienting the eyes in response to new stimuli in the visual periphery. It controls eye and head movements to bring a stimulus onto the fovea.

The major target of the optic nerves, however, are the right and the left lateral geniculate nuclei (LGN), which are located in the dorsal thalamus. And most neurons from the LGN project an axon to primary visual cortex via the optic radiation. The right LGN receives information about the left
visual hemifield from both eyes. At the LGN, input from the two eyes are kept separate, in different layers. Surprisingly, the receptive fields of LGN cells are almost identical to those of the ganglion cells of the retina that feed them. This is surprising, because the main source of synaptic input to the LGN is not the retina, but the primary visual cortex. Thus, one might reasonably expect that this feedback pathway would significantly alter the qualities of the visual responses recorded in the LGN. so far, however, a role for this massive input has not been clearly identified. The LGN also receives synaptic input from neurons in the brain stem whose activity is related to alertness and attentiveness [7, 8].

The LGN has a single major target: the primary visual cortex, which is also called the striate cortex or V1. From V1 on, two distinct pathways go to higher cortical areas.

2.2.1 Striate cortex

The major target of LGN cells is the primary visual cortex (V1), which is also called the striate cortex. (The term striate refers to the fact that area V1 has an unusually dense stripe of myelinated axons running parallel to the surface that appears white in unstained sections.)

Retinotopy The projection starting in the retina and extending to LGN and V1 illustrates a general organizational feature of the central visual system called retinotopy. Retinotopy is an organization whereby neighboring cells in the retina project information to neighboring cells in their target structures, in this case the LGN and the striate cortex. That means that the two-dimensional surface of the retina is mapped onto the two-dimensional surface of the subsequent structures.

Because visual space is not sampled uniformly by the cells in the retina, the mapping of the visual field onto a retinotopically organized structure is often distorted. This is due to the fact that the density of cones and ganglion
Figure 5: The visual pathway
cells is not homogeneous in the retina.

**Layers in striate cortex**  The neocortex in general, and striate cortex in particular, has neuronal cell bodies arranged into six layers. Starting at the white matter, which contains the cortical input and output fibers, the cell layers are named by Roman numerals VI, V, IV, III, II and layer I. Layer I has almost no neurons and consists almost entirely of axons and dendrites of cells in other layers. The full thickness of the striate cortex from white matter to pia is about 2 mm. Layer IV is further divided into three sublayers IVA, IVB and IVC.

**Receptive fields**  Most neurons in layer IVC in V1, which is the main target of LGN cells, have similar receptive fields to cells in LGN and the ganglion cells in the retina. This means they are generally small monocular center-surround receptive fields. Outside layer IVC, new receptive field characteristics, not observed in the retina or LGN, are found.

**Orientation selectivity**  Hubel and Wiesel [9] found that many neurons in V1 respond best to an elongated bar of light moving across their receptive fields. And the orientation of the bar is critical. The greatest response from the cell is given to a bar with a particular orientation; bars perpendicular to that orientation elicit much weaker responses. Neurons which have this type of receptive field are said to exhibit orientation selectivity. Most of the V1 neurons outside layer IVC are orientation selective.

Hubel and Wiesel [10] investigated whether orientation selectivity of nearby neurons is related. They used a microelectrode and penetrated the striate cortex radially, meaning perpendicular to the surface, from one layer to the next. They found that the preferred orientation remains the same for all the selective neurons encountered from layer II down through layer VI. Hubel and Wiesel called such a radial column of cells an orientation column.
As an electrode passes tangentially, parallel to the surface, through the
cortex in a single layer, the preferred orientation progressively shifts. Hubel
and Wiesel found that a complete 180° shift in preferred orientation can be
seen when traversing about 1 mm, on average, in layer III.

The detection of stimulus orientation is thought to be one of the most
important functions of striate cortex. These orientation-selective neurons are
specialized for the analysis of object shape.

**Direction selectivity**  A subset of orientation-selective cells exhibits
also direction selectivity. They respond best when a bar of light at the
optimal orientation moves perpendicular to the orientation in one direction
but not in the opposite direction [11, 12]. Direction-selective neurons are
thought to be specialized for the analysis of object motion.

**Simple and Complex receptive fields**  Many orientation-selective
neurons have a receptive field elongated along a particular axis, with an ON-
center or OFF-center region flanked on one or both sides by an antagonistic
surround. This arrangement of ON and OFF areas is also seen in retinal
and LGN receptive fields. Striatal neurons receive a converging input from
three or more LGN cells with receptive fields that are aligned along one axis.
Hubel and Wiesel called neurons of this type simple cells. This segregation of
ON and OFF regions is a defining property of simple cells, and it is because
of this receptive field structure that they are orientation selective.

Complex cells, on the other hand, do not have distinct ON and OFF
regions, they elicit ON and OFF responses to stimuli throughout the receptive
field. Hubel and Wiesel proposed that complex cells are constructed from
the input of several similarly oriented simple cells.

**Binocularity**  Each neuron in layer IVC receives input from a layer in
LGN representing either eye. These monocular neurons from either eye are
also clustered in V1 rather than randomly intermixed. Layers superficial to
IVC are binocular, responding to light in either eye. They are said to have binocular receptive fields, meaning that they actually have two receptive fields, one in the ipsilateral and one in the contralateral eye. Binocular receptive fields are essential to form a single image of the world around us from inputs of both eyes. Retinotopy is preserved because the two receptive fields of a binocular neuron are precisely placed on the retinas such that they are looking at the same point in space.

2.2.2 Extrastriate cortex

Striate cortex is called V1, for ‘visual area one’, because it is the first cortical area to receive information from the LGN. Beyond V1 lie other distinct areas of cortex, each of which contains a representation of the visual world. The contributions to vision of these extrastriate areas are still being debated. However, there is an emerging picture that there are two major cortical streams of visual processing, one projecting dorsally from striate cortex toward the parietal lobe and the other projecting ventrally toward the temporal lobe.

The ventral stream is thought to be involved in the perception of the visual world and the recognition of objects, while the dorsal stream appears to analyze visual motion and the visual control of action.

2.2.3 The ventral visual pathway

The importance of the inferotemporal cortex (TE) for visual object recognition has been known for many years as a result of lesion studies in the monkey. Ungerleider and Mishkin [13] delineated the functional dichotomy between the serial cortical pathway leading from the primary visual cortex (V1) to TE and the pathway leading from V1 to the parietal cortex. The former was named the ventral visual pathway, and the latter the dorsal visual pathway.
The ventral visual pathway goes from V1, through V2, V4, posterior TE, to TE. The ‘step-by-step’ projections are the strongest, but there are also connections that jump one step, such as those from V2 to posterior TE and those from V4 to TE. From experiments with monkeys, it is known that a lesion in one of the stages along the ventral visual pathway disrupts the ability to recognize objects [14].

The functional magnetic resonance imaging (fMRI) studies of Sereno et al. [15], Tootell et al. [16] and DeYoe et al. [17] in humans have demonstrated that occipital regions surrounding V1 are arranged in retinotopically organized areas that can be activated by simple stimuli. In accordance with the retinotopic arrangements in the monkey brain, they designated these regions as the human homologues of V2, V3, V3a and V4. The retinotopic organization becomes coarser along the pathway, and no sign of retinotopic organization has been reported in TE. The TE can be activated by complex stimuli.

Among the projections sites from TE, the perirhinal cortex (Brodmann Area (BA) 35 and 36) and the upper bank of the anterior superior temporal sulcus are dominantly visual. In this respect, TE does not represent the final stage of the ventral visual pathway.

Along the ventral visual pathway, the features and identity of an object is detected.

2.2.4 The dorsal visual pathway

A series of PET studies by Haxby and Ungerleider’s group [18, 19] has demonstrated the distinction between the object identification pathway (ventral visual pathway) and the object location pathway, which is called the dorsal visual pathway.

In their studies, the same stimuli composed of one sample face and two ‘choice’ faces were used in two different tasks. In the face-identification task, the subjects had to indicate which of the two choice faces matched the sam-
ple face. In the location task, the subjects had to match the position of the face relative to the side of the display box marked by a double bar. When compared to the activation produced by a sensory-motor control task using scrambled images, the face-identification task and the location-task evoked activation over large posterior brain regions, with an extensive overlap centered on the occipital lobe. When compared to each other, however, specific activations were split into the ventral and dorsal regions. The activation specific to the face-identification task was distributed along the ventral surface of the occipital cortex and the posterior part of the temporal cortex. Whereas the location task elicited activation along the surface of the dorsal cortex. Haxby and Ungerleider concluded that along the dorsal visual pathway, the location of an object is determined and encoded.

On the basis of new neuropsychological observations [20, 21], Goodale and Milner re-interpreted the function of the dorsal stream, in that it mediates the visual control of skilled actions, such as grasping, that are directed at the seen objects. They examined a patient with a profound disorder in the perception of some object qualities, like the object’s spatial location, its form, orientation and size. This patient nevertheless was accurate in guiding her hand and finger to the very object whose qualities she fails to perceive. They concluded that the neural substrates for the visual perception of object qualities such as shape, orientation and size are distinct from those underlying the use of those qualities in the control of manual skills.

2.3 Mechanisms of visual object recognition

Visual object recognition is a key function of the primate brain and it involves the ventral visual pathway. It tolerates considerable changes in images, such as those caused by variable illumination and different viewing angles and articulations of the object; it is also capable of automatic generalizations. The neural mechanisms of visual object recognition have been investigated both in behavioral studies of brain-damaged patients and monkey. The de-
velopment of non-invasive measurement techniques for examining the human brain has provided new and powerful tools, which have increased the speed of exploration [22].

Bruner [23] assumed that perceptual experience is the end product of a categorization process. Whatever is perceived achieves its meaning from a class of percepts with which it is grouped.

Two main mechanisms of visual object recognition have been identified, categorization and identification. Categorization means that many objects get the same label, such as all tools belong to the same category 'tool', whereas in identification, a label is reserved for a particular object, such as a hammer.

It is still debated whether these two processes are mediated by the same neural substrate. In this work (see Part 2), I investigated the dynamics of the learning of a categorization and an identification task. The aim was to find evidence for or against a dual-process system.

Another dualism in visual object recognition emerges when studying Recollection and Familiarity. When a person is confronted with a stimulus, which he encountered previously and he remembers having seen it before, as well as recollects contextual information about the previous encounter, such as location or circumstances, we call this Recollection. In contrast, Familiarity means that the person has a feeling of having seen the stimulus, but can’t recollect any contextual information.

It is unclear yet whether Recollection and Familiarity are two distinct processes, mediated by different neural systems or whether the difference is quantitative and Recollection reflects a stronger memory trace. I address this issue in Part 4, by means of an fMRI experiment.

2.3.1 Effects of Size and Position

The effects of size changes on recognition have been examined in experiments in which subjects distinguish stimuli as being the same or different,
disregarding changes in size. Recognition performances in these tasks are typically assessed by measuring response latencies. It was observed that the performance varies as a function of the size ratio between the two stimuli. Increasing size discrepancy between the two shapes resulted in longer latencies [24, 25]. In another study [26], in which the viewing distance and the size of novel objects were manipulated, it was found that the perceived, and not the retinal, size of the objects determines the size ratio effect.

One of the key properties of neurons, in terms of position dependence, is the size of their receptive fields (RF). In V1, foveal RFs are typically small (~1° of visual angle), consistent with a position-specific representation, but RF size increases as one moves along the ventral visual pathway [27]. In early studies of anterior IT [28, 29], it was found that the receptive fields are large (>20° of visual angle), and their object-selectivity is preserved within the RF [30]. This indicates that, despite the position specificity of early visual areas, IT cells might be independent of translations of stimuli to different positions. However, more recent studies have also found smaller RFs. Op de Beeck and Vogels [31] reported a range of RF sizes from 2.8 to 26°, with a mean size of 10°, and large variability in response within RFs. Thus, although there are some neurons with large RFs, there is a wide distribution of RF size. Such heterogeneity in RF size makes it difficult to predict the degree of behavioral position dependence from the responses of single neurons. Whereas IT neurons with small RFs suggest a position specificity, cells with large RFs could support position-independent performance. However, it is unlikely that object representations arise directly from the responses of individual cells but rather from a population-level response across an ensemble of IT neurons [32]. Thus to understand the position-dependency of object recognition, we need to understand how responses of an ensemble of neurons is integrated. Hung [33] used a linear classifier to predict the object category after translation of 4° from single-neuron data from IT-cells in macaque monkeys. The activity of small neuronal populations (~100 randomly selected cells) over very short
time intervals (as small as 12.5 ms) contained unexpectedly accurate and robust information about both object identity and category. This information generalized over a range of object positions and scales, even for novel objects. Coarse information about position and scale could also be read out from the same population.

2.3.2 Object-selective regions

Object-selective regions respond strongly when subjects view pictures of objects but not when they view textures, noise or highly scrambled pictures. These regions comprise a large constellation of areas in both the ventral and dorsal visual pathways that lie anterior and lateral to early retinotopic cortex. The parcellation of these areas is difficult because they are largely non-retinotopic [34] and different types of objects activate slightly different regions [35].

Kanwisher and colleagues [36, 37] have suggested that ventral temporal cortex contains a limited number of modules specialized for the recognition of special categories such as faces (fusiform face area; FFA; [36]), places (parahippocampal place area; PPA [37]) and body parts (extrastriate body part area; EBA [38]).

However, Ishai and colleagues [39, 40] argue that the representation of objects is not restricted to a region that responds maximally to that object, but rather is distributed across a broader expanse of cortex. In their functional MRI study, subjects performed passive viewing and delayed match-to-sample tasks, with faces, houses, chairs and scrambled pictures. Three distinct regions of ventral temporal cortex that responded preferentially to one of the three stimulus categories were found. However, they found that each category also evoked significant responses in the regions that responded maximally to other stimuli. Moreover, each category was associated with its own differential pattern of response across ventral temporal cortex. They concluded that the representation of an object is not restricted to a region
that responds maximally to that object, but rather that object-responsive areas are within a distributed cortical network and are overlapping.

2.3.3 Inferotemporal cortex

A large number of studies confirmed initial findings that IT is the last exclusively visual area in the ventral pathway. More than 85% of the neurons in this area are excited or inhibited by different simple or complex visual stimuli [41]. The observed properties of IT cells change significantly as one moves from the most posterior part, where cells have similar properties to those observed in V4, to the most anterior part, where neurons rarely respond to such simple stimuli. Among the changing characteristics are topography, receptive field size, and stimulus selectivity [42].

There is growing evidence that responses of single cells in the monkey (IT) correlate well with overall perception. Ito et al [30] presented several shapes to monkeys and found a group of cells which responded to particular shapes over 64-fold size difference. Sary et al [43] reported similar results. The shape-selectivity of IT cells in awake monkeys does not vary with the size and position of a stimulus, and also does not vary with the visual cue used to define the shape.

Logothetis’s group [44, 45] conducted a series of experiments using the effect of binocular rivalry. When different images are presented to the left and right eye, a subject does not perceive a combination of the two images, but rather each image alternately. The experimenters presented a stimulus that evoked a strong response from a recorded cell to one eye, and a stimulus, which doesn’t activate the recorded cell to the other eye. The monkey had been trained to make differential responses according to two predetermined stimulus categories. Faces were used in the experiments with IT cells. Then they compared the responses of the recorded cells with the monkey’s reported perception. They found a correlation in 18% of the cells in V1 between their responses and the monkey’s perception. This proportion increased to 25%
in V4 [44], and nearly all the inferotemporal cells responded consistently with the monkey’s reported perception [45]. These results suggest that the conscious perception of objects is better correlated with cell activities in IT than in earlier visual areas.

2.3.4 Face perception

Face perception is a highly developed visual skill in primates. Converging evidence from neuropsychology, neurophysiology, and cognitive development indicates that face perception may be mediated by a specialized neural system in the brain [46]. Face perception has consistently activated a region in the lateral fusiform gyrus [19, 36], also called the ‘fusiform face area’. However, other studies showed that this region also responds to other categories, such as houses, chairs and tools, thus ruling out its status as a face ‘module’ [47, 40, 39, 48].

Ishai and colleagues [49, 48] conducted a series of experiments to identify areas, which are activated when subjects look at faces. They found a network of face-responsive regions which included the inferior occipital gyrus, fusiform gyrus, superior temporal sulcus, hippocampus, amygdala, inferior frontal gyrus and orbitofrontal cortex. They could also show that while all face stimuli evoked activation within all regions, compared with neutral faces, viewing famous and emotional faces resulted in larger spatial extents of activation and higher amplitudes of the BOLD signal.

Prosopagnosia Lesions in the inferomedial occipito-temporal region of the right hemisphere have been shown to disrupt the recognition of familiar faces [50]. The lesions typically involve the fusiform and lingual gyri or their interconnections and are caused by strokes of the right posterior cerebral artery. Prosopagnosia, as this disorder was called by Bodamer [51], is considered a specific agnosia, which renders human patients incapable of recognizing the faces of familiar or famous persons but spares their ability to recognize com-
mon objects. However, prosopagnosic patients can recognize individuals by their voices or sometimes by their movements [52]. They can also recognize a face as the object ‘face’ and name and point to its parts [53, 52]. Furthermore, the complexity of faces does not seem to be the cause of this disorder, since the patients are able to recognize complex visual stimuli [54, 55].

More recent studies have shown that some prosopagnosic patients exhibit normal patterns of activation in the fusiform face area (FFA), despite their inability to recognize faces [56], suggesting that activation in this region is not sufficient for face recognition.

Rossion and colleagues reported a patient with bilateral occipito-temporal lesions, resulting in a deficit restricted to face perception. They observed normal activation of the right FFA in response to faces in the patient, despite the absence of any feedforward inputs from right occipital face area (OFA), located in a damaged area of cortex. They conclude that the FFA and OFA in the right hemisphere and their integration are necessary for normal face processing [57].

**Deficits in the Recognition of Facial Expression** Another dissociation in the recognition of objects is that observed between the identification of faces and the recognition of facial expression. Facial expressions, such as smiles and frowns, are complex transformations that have to be ignored in order to recognize the same face under different emotional states. At the same time, however, facial expressions have a number of different meanings that are essential for social interactions. There’s evidence that these facial expressions are processed independently of face identification. Tranel and colleagues [58] have reported prosopagnosic patients, who fail to recognize a familiar faces, but at the same time are able to identify facial expression, gender and age of presented face stimuli.

On the other hand, Adolphs et al. [59] reported a case of a 30-year old woman with a confined amygdala lesion. She suffered from Urbach-
Wiethe disease, which led to almost complete bilateral destruction of the amygdala, while sparing her hippocampus and all neocortical structures. The patient was tested in rating facial expressions, such as happiness, surprise, fear, anger, disgust, sadness, as well as neutral faces, and was found to have a much lower performance at this task than age-matched controls. Her ability to identify individual faces, however, was completely preserved.

2.4 Summary

In this chapter I have outlined the organization of the sensory pathway from eye to thalamus to cortex. Vision involves the perception of several different properties of object features - color, form, movement - and these properties are processed in parallel by different cells and regions of the visual system. As we move along the visual pathway, neurons respond to more and more complex stimuli.

In the following chapters, I will describe the memory and category learning systems. Specifically, I will explain how visual objects are represented and how categorical representations of visual objects are memorized and retrieved.

3 Memory

3.1 Multiple memory systems

There is growing consensus that human memory is mediated by multiple qualitatively distinct systems. Two major memory systems have been proposed: the declarative and the non-declarative memory system [60].

3.1.1 Declarative memory system

Declarative memories are those accessible to conscious awareness. Typically this includes working memory, episodic memory and often also semantic
memory. Declarative memory presumably would mediate any learning strategy that the subject could accurately describe. The declarative memory is also called the explicit memory, since it consists of information that is explicitly stored and retrieved.

Working memory is the ability to maintain and manipulate limited amounts of information during brief periods of cognitive activity. It is heavily used in reasoning and problem solving, and because of this, it is often associated with a wide variety of cognitive tasks. Because working memory is effective only for brief time intervals, it cannot store a lasting category representation, but it could be the primary mediating memory system in tasks where the categories are learned quickly.

Episodic memory refers to memories of specific past events or episodes in our personal history. Episodic memory provides a context-rich representation of an event that might include information from all sensory modalities.

By contrast, semantic memory, or the memory for facts, is context-poor and typically involves only one modality.
3.1.2 Non-declarative memory systems

Non-declarative memory systems, also known as implicit memory systems, are not consciously accessible. These include two different memory systems. First the procedural memory, which is the memory of skills that are learned though practice. Traditionally these have been motor skills, such as those used when playing football or a musical instrument. There are several signatures of procedural learning that make it qualitatively different from learning that is mediated by declarative memory systems. There typically is little conscious recollection or even awareness of the details of procedural memories. And procedural learning is slow and incremental and it requires immediate and consistent feedback.

And secondly the perceptual representation memory system (PRS), which was proposed to account for various repetition-priming effects in which exposure to a degraded stimulus improves performance on subsequent presentation of that stimulus, even when subjects have no conscious recollection of the degraded stimulus, and even in patients with amnesia. Cognitive neuroscience theories postulate that the learning that underlies the PRS occurs within visual cortex. The idea is that repeated presentation of the same stimulus during some relatively brief time interval enhance the firing of the visual cortical unit that is maximally sensitive to that stimulus.

3.1.3 Explicit and Implicit Category Knowledge

A fundamental dissociation among memory types is seen in the memory function of amnesic patients who have damage to the medial temporal lobe (MTL). These patients have impaired declarative memory for facts and events (conscious, explicit memory), but exhibit intact nondeclarative (nonconscious, implicit memory) memory [61]. Intact nondeclarative memory in amnesic patients implies the existence of memory systems outside MTL, but the study of amnesic patients does not tell us how these memory systems operate, what brain areas support them, or how they interact with intact
declarative memory in healthy people.

When healthy participants are given tasks that depend on nondeclarative memory, they generally appear to be unaware that they have acquired new information or a new skill and participants usually attribute their improved performance to guessing, instinct, or gut feeling. The term *implicit learning* was first used to introduce this idea of learning without awareness [62] and inspired a great deal of research and debate about whether memory could truly occur outside awareness.

Reber et al. [63] investigated the dissociation of implicit and explicit category knowledge with the dot-pattern paradigm using fMRI. One group of subjects (IC) learned the category in the traditional incidental fashion by simply observing the patterns during the study phase. They were asked to identify the central dot in each pattern. For this group, no mention was made of the existence of the underlying category structure. A second group of participants (EC) learned the identical category by viewing the same study items after being told that the study items came from the same category and were instructed to attempt to determine the category.

In the subsequent fMRI test phase, participants viewed patterns from the target category and from an unfamiliar category, and were asked to indicate whether it came from the same category as the study items for each patterns.

Both groups exhibited knowledge of the category during the categorization test by correctly endorsing novel category members more often than the noncategorical patterns. The noncategorical patterns were generated from an unfamiliar prototype with the same relative structure as the target category, making it impossible to succeed on the test without being influenced by the previously studied patterns.

For the IC group, reduced activation was observed in the posterior occipital cortex for the categorical patterns compared with the noncategorical patterns. This effect replicates previous reports (see above, [64, 1]) that a categorical fluency effect (CFE) occurs after learning the visual category.
The CFE reflects a change in visual processing of dot patterns such that categorical patterns are processed more fluently and thus evoke less activity than noncategorical patterns. With the described study, this finding is extended to the case of a short (5-item) study phase (previous studies of the same and other authors used a 40-item study phase).

In the EC group, the same stimulus-based contrast identified several regions in which increased activity was observed for the categorical patterns. Although the pattern of increased activity for the categorical patterns is qualitatively similar to the increases in activity in prefrontal and posterior cortical areas observed during recognition of previously presented dot patterns [64], the specific areas of increased activity differ, suggesting that conscious categorization is not driven solely by recognition memory.

These results suggest that two separate sets of processes occur depending on whether the knowledge of the category was acquired incidentally or intentionally. Categorization after intention learning resulted in a network of increased activity for categorical patterns without evidence for decreases. Categorization following incidental learning resulted in decreased activity for categorical patterns in the posterior occipital cortex without evidence for increases.

3.2 Hippocampus and Memory

There is growing evidence that the medial temporal lobe, and especially the hippocampus, has an essential role in declarative memory processes [65]. Damage to the hippocampus and nearby parts has been shown to result in amnesia in humans.

In monkeys, damage to the hippocampus or to some of its connections, such as the fornix, produces deficits in learning about where objects are and where responses must be made [66]. For example, macaques and humans with damage to the hippocampus or fornix are impaired in object-place memory tasks in which not only the objects seen, but where they were seen, must be
Monkeys studied by Kluever and Bucy [69] following temporal lobectomy, the bilateral removal of the temporal lobes, had a peculiar way of interacting with their environment, in addition to a host of other abnormalities. The monkeys explored their room by placing objects in their mouths. If an object was edible, they would eat it, if it wasn’t they would drop it. However, they would repeatedly go back to the same inedible objects, put them in their mouth, then toss them aside. This problem with object recognition is probably related to memory function in the temporal lobe. The effects of temporal lobectomy make a strong case that one or more structures in the medial temporal lobe are essential for the formation of declarative memories.

Because the macaque monkey brain is similar in many ways to the human brain, macaques are frequently studied to investigate amnesia. In a commonly used task, called delayed non-match to sample (DNMS), the monkey faces a table that has several small wells in its surface. It first sees the table with one object on it covering a well. The monkey is trained to displace the object so that it can grab a food reward in the well under the object. After the monkey gets the food, a screen prevents it from seeing the table for some period of time (the delay interval). Finally, the monkey gets to see the table again, but now there are two objects on it: One is the same as before, and another is new. The monkey’s task is to displace the new object (the non-matching object) in order to get another food reward in a well below it. Normal monkeys are relatively good at this task. With delays between the two stimulus presentations of anywhere from a few seconds to ten minutes, the monkey correctly displaces the non-matching object in about 90% of the trials. Memory required in the DNMS task has been called recognition memory because it involves the ability to judge whether a stimulus has been seen before.

Experiments demonstrated that severe deficits on the DNMS task result from bilateral medial temporal lesions in macaque monkeys [70, 71]. Perfor-
mance was close to normal if the delay between the sample stimulus and the two test stimuli was short, a few seconds. This is very important because it indicates that the monkey’s perception was still intact after the lesioning and it remembered the DNMS procedure. But when the delay was increased from a few seconds to a few minutes, the monkey made increasingly more mistakes choosing the non-matching stimulus. With the lesion, the animal was no longer as good at remembering what the sample stimulus was in order to choose the other object. This behavior suggests that it forgot the sample stimulus if the delay was too long. Importantly, this deficit was also observed if the monkey was allowed to touch but not see the object. That means that the deficit in recognition memory produced by the lesion was not modality specific. It appears that the medial temporal lobe structures are critical for the consolidation of memory [72].

In the 1950’s, Scovillo and Milner [73] described the severe amnesia that followed bilateral surgical removal of the medial temporal lobe (patient H.M.). This important case demonstrated that memory is a distinct cerebral function, dissociable from other perceptual and cognitive abilities. This patient appeared to have a complete loss of memory for events that followed the bilateral medial temporal lobe removal, together with a partial retrograde amnesia for the three years leading up to this operation. But early memories are seemingly normal and there is no impairment of personality or general intelligence.

4 Category learning

Is the plant edible or poisonous? Is the person a friend or a foe? Was the sound made by a predator or by the wind? All organisms assign objects and events in the environment into separate classes or categories. If they did not, they would die, and their species would become extinct. Therefore, categorization is among the most important decision tasks performed by
organisms.

Living organisms have a remarkable capacity to achieve adaptive responses in diverse situations. In this respect, two basic response patterns can be distinguished: we may select a specific response for a particular object or, alternatively, the same action is performed for a number of different stimuli that share behaviorally salient features. The former case refers to identification or recognition, which requires one-to-one stimulus-response mapping (Figure 7).

For example, humans have a repertoire of behavior when encountering with dogs in general, but some actions remain specific for our own dog. General behavior is based on categorization while selecting a specific action towards a particular object is based on recognition.

The essence of categorization is transfer to novel stimuli, after training, observers are able to use category-specific responses for stimuli that they have never seen before. This abstraction process makes our behavior economical, diminishing the need of separate response learning for each and every individual object in our environment, a doubtlessly impossible venture. However, category learning is not a unitary process. According to the structure of training, at least three types of paradigms can be distinguished. First, subjects may learn categories by the observation of category members, or exemplars. Second, corrective feedback can be provided after category-decisions, leading to a gradually improving performance from chance level to a full ability to classify novel items. Feedback is a basic form of reward, giving value to objects of our environment and related responses. Third, getting acquainted with abstract rules, humans are able to achieve a remarkably good categorization performance, even without seeing any exemplar [74].

For the most part, early theories of category learning did not specify an underlying memory system, at least not in the language used by memory theorists. However, virtually all theories have assumed a single learning system. A growing body of recent evidence, however, suggests that category
During recognition, a particular stimulus /S/ is associated with a specific response /R/. During categorization, a group of similar but individually different stimuli (circles with different internal textures) is associated with a single response [74].

Considering these very different approaches, we may assume that category learning is not homogeneous at the neurobiological level, and hence the discovery of neuronal correlates is not straightforward [75].

4.1 Neurobiology of category learning

Most of the research on category learning in non-humans has involved either single-cell recordings in monkeys or lesion-studies in rats. The three most widely studied brain areas have been the inferotemporal cortex (IT), the prefrontal cortex (PFC), and the basal ganglia [76].

**Inferotemporal Cortex (IT)**  The IT is the terminal sensory association region on the ventral (or “what”) pathway out of the primary visual cortex [13]. Many studies have implicated IT in the high-level representation and
processing of visual objects [77].

A variety of reported category-specific agnosias, resulted by lesions in IT, have led the interest to IT as a possible neural locus of category learning. The most widely known of such deficits, which occur with human faces, and is called prosopagnosia, are associated with lesions to the fusiform gyrus (FG) in IT.

A lesion in IT resulting in a category-specific agnosia does not necessarily imply that category learning occurs in IT. It is well known that neighboring cells in IT tend to fire to similar stimuli [78]. Thus, damage to a region within IT is likely to lead to perception deficits within a class of similar stimuli.

In fact, there is now strong evidence that IT is not a site of category learning. Rolls et al. [79] recorded IT-cells of monkeys while they were watching different visual objects (e.g., food, tools). Associations of selected visual stimuli with a food reward or an aversive taste did not affect the magnitude of the responses of the neurons to the stimuli either during the learning or after the period of learning.

More recent studies have found similar null results [80, 81, 82]. In these studies, monkeys were trained to classify visual objects into one of two categories (e.g., tree versus non-tree, two categories of arbitrary complex shapes). Single-cell recordings showed no change of the firing rate properties of IT cells with learning.

This evidence suggests that IT does not mediate the learning of new categories. However, IT is crucial to the categorization process, because it encodes the highest level of representation of visual stimuli. Thus, other candidates for the neural locus of category learning could be structures receiving direct projections from IT. There are three regions: the PFC, the medial temporal lobe (i.e., the hippocampal system), and the basal ganglia.

**Prefrontal Cortex (PFC)** Strong evidence that PFC plays an important role for category learning has been found using the Wisconsin Card Sorting
Test (WCST) with patients with frontal lobe lesions. The WCST is a classic neuropsychological assessment that requires participants to learn a series of simple categories [83].

Milner [84] showed that patients with frontal lesions are impaired in the WCST. It has been suggested that deficits on the WCST are classic symptoms of frontal lobe pathology [85]. It could be verified in more recent lesion studies in rats that the PFC plays a critical role in WCST [86].

Single-unit recording studies in the PFC have demonstrated two different phenomena that are particularly relevant to category learning. First one set of studies reported neurons in the lateral PFC that show category-specific responding. For example, in an experiment by Freedman et al. [87, 88, 80], monkeys learned to categorize computer-generated images as dogs or cats. These images were created by morphing the prototypes of each category in varying degrees. After training the animals to a high level of accuracy, single-unit recordings were collected in the lateral PFC region receiving direct input from IT. They found that many neurons responded category-specific, that is, the response to members of the category was very similar for one group of cells, while they showed little or no response to members of the other category. They also reported that the firing properties of these cells were better predicted by category membership than by visual similarity.

A second finding was that the PFC plays an important role in learning and applying categorization rules. For example, Asaad et al. [89] trained monkeys to classify visual objects according to one rule (e.g. spatial) or another (e.g. associative). The rule to be applied to the current stimulus, is signaled to the monkey at the beginning of each trial by a cue. They then collected single-cell recordings in the PFC. Many cells were observed to respond rule-specific, that is, they fire during application of one of the rules, but not during the other, regardless of which stimulus is shown. Many lesion studies have verified the important role played by the PFC in rule learning and use [90].
**Basal Ganglia** Many lesion studies in rats and monkeys showed that the tail of the caudate nucleus is both necessary and sufficient for visual discrimination learning. It was shown that animals with lesions at the tail of the caudate nucleus are impaired in learning visual discrimination that require one response to one stimulus and a different response to some other stimulus [91, 92].

The sufficiency of the caudate nucleus for visual discrimination learning was shown in a series of studies by Gaffan and colleagues that lesioned all pathways out of visual cortex except into the tail of the caudate. Eacott and Gaffan [93] lesioned projections into the PFC, and projections to the hippocampus and amygdala were lesioned by Gaffan and Harrison [94]. None of these lesions affected visual discrimination learning. Many researchers have hypothesized that the primary role of the neostriatum is in learning to associate an appropriate response with each stimulus [95, 96].

### 4.1.1 Neuroimaging experiments

**fMRI** In an experiment by Reber and colleagues [64], subjects studied dot patterns that were distortions of an underlying prototype and then made yes-no category judgments about new dot patterns. These subjects showed an increase in BOLD response for categorical decisions compared to noncategorical decisions bilaterally in the superior frontal gyrus and the right inferior frontal gyrus. Learning the category resulted in changes in visual processing for the dot-pattern stimuli so that subsequent processing of categorical stimuli was more fluent than processing of noncategorical stimuli, i.e. less activity was observed in early visual areas for categorical patterns compared with noncategorical patterns. Aizenstein et al. [97] also compared activity for categorical and noncategorical patterns, but did so during acquisition of the category. Reduced activity was once again observed for the categorical patterns compared with noncategorical patterns when the category was learned implicitly.
Together, these studies provide evidence that the neural basis of this visual category learning task is a form of perceptual learning that occurs outside awareness.

In an additional study [1], subjects performed a categorization and a recognition task with dot patterns as described by Knowlton and Squire [98]. Examples of stimuli are illustrated in figure 8. In this study, the authors could replicate their previous results [64], in that the posterior occipital cortex showed a decreased activation for categorical patterns relative to noncategorical patterns. For the recognition memory task, a number of areas exhibited increased activity in response to the old patterns as compared with the new patterns. In particular, increased activity was observed in right occipital cortex. The difference in brain activity in posterior visual cortex between the categorization task and the recognition memory task is particularly striking. During categorization, a sizable decrease in activity was observed in occipital cortex that was similar to decreases reported previously in association with perceptual priming tasks [99, 100]. In contrast, during recognition, an increase in activity occurred in this region. The specific occipital areas exhibiting these changes were not identical in the two tasks. With a different analysis of the data Reber et al. [101] found subregions of the posterior occipital cortex which showed an increase in activity during recognition and a decrease in categorization for categorical patterns compared with unrelated dot patterns.

These previous studies compared activity in early visual areas during categorization judgments for categorical and noncategorical dot patterns to observe the fluency effect. This comparison does not identify changes in activity in areas that are active during the categorization process, since both conditions involve performing categorization. To identify brain areas involved in performing categorization, Reber et al. [63] compared activity of categorization to those of a control task. In this study they used the same categorization and recognition tasks as in the previous studies, additionally,
Figure 8: Examples of stimuli from [1]
The study items were high distortions of a prototype pattern. The test items, illustrated from left to right, were presentations of the training prototype, low and high distortions of the training prototype, and a random dot pattern.

Subjects performed a control task. During the control task, subjects saw dot patterns consisting of four to seven dots, and were asked to count the dots and report ’yes’ or ’no’ whether there was an odd number of dots shown.

Comparing categorization with counting revealed activation in 15 regions identified with a two-stage analysis. Six areas were identified that exhibited increased activity during categorization compared with counting and also exhibited increased activity during recognition. These areas likely reflect brain areas involved with processing the dot patterns and making decisions about categorization or recognition memory.

During categorization, additional activity was observed in nine areas that did not exhibit increased activity during recognition. In the frontal lobes, these areas included bilateral inferior prefrontal cortex, although the in-
creased activity on the right side was of a much greater extent than the left, as well as medial prefrontal cortex, and left frontopolar cortex. In addition, a left inferior parietal area exhibited increased activity, possibly the lateral match to the larger right inferior parietal cortical increase observed during both categorization and recognition. Two areas in left temporal cortex (middle and superior temporal gyri) also exhibited increased activation.

As in categorization, increased activity in the recognition task compared to the counting task was observed bilaterally in the inferior and middle occipital gyri (BA 18), in right inferior parietal cortex and in right ventral occipito-temporal cortex. Although the locations of these observed increases are similar, the extent of the BA 18 activations during recognition was considerably larger in recognition.

In addition, during recognition extensive increases in activity were observed throughout the precuneus bilaterally and the posterior cingulate. Two areas of increased activity were observed in the frontal lobes, in the right medial frontal gyrus and right dorsolateral prefrontal cortex.

A double subtraction, to compare activity associated with recognition and categorization revealed more activity during recognition than categorization in seven areas. These include two regions in the medial aspect of the precuneus, two occipital cortical areas, two areas in the cerebellum and a region of the posterior, right hippocampus. No areas were identified for which categorization produced significantly greater activation than recognition.

The comparison between brain activity observed during recognition and categorization identifies a number of brain areas commonly associated with memory retrieval that exhibited more activity during recognition than categorization. Although the region exhibiting increased activity in the right posterior hippocampus for recognition is small, the role of the medial temporal lobe (including the hippocampus) in recognition memory is well-established. Increased activity in the hippocampus during retrieval has been reported in recent fMRI studies [102, 103]. Several studies of memory retrieval have
reported increased activity in the precuneus [104] suggesting that this area plays an important role in conscious retrieval processes.

4.2 Categorization and Identification

4.2.1 Evidences for multiple memory systems

Ashby and Lee [105] studied the relation between the identification, similarity judgments, and categorization of multidimensional perceptual stimuli. They compared the data with predictions from several different versions of the general recognition theory (GRT), from versions of the multi-dimensional-scaling (MDS) choice model, and from the biased-choice model. In one application, two subjects first identified a set of confusable stimuli and then made judgments of their pairwise similarity. The stimulus set consisted of nine semicircular figures, constructed by factorially combining three sizes (i.e. radii) and three orientations.

The second application was to Nosofsky's [106] identification-categorization experiment. The most general GRT substantially outperformed the biased-choice model and any version of the MDS-choice model. This result is especially significant because in this application, GRT has 35 fewer parameters than the biased-choice model. The model that best predicted the similarity judgments was a GRT model with one free attention parameter.

Larry Squire and Barbara Knowlton conducted a series of experiments [107, 2] in which they studied the categorization and identification abilities of amnesic patients. In the earlier study amnesic patients and healthy control subjects studied 40 dot-patterns, which all belonged to the same category. In a subsequent test, subjects had to classify novel stimuli according to whether they did or did not belong to the learned category. The amnesic patients performed in this task as good as the controls. In contrast, they performed much worse in an identification task in which they had to recognize a single stimulus after it was presented 40 times in succession.
In the study of 1995 [2] they describe a profoundly amnesic patient (E.P.), who cannot learn and remember specific instances, i.e. he has no detectable declarative memory. Nevertheless E.P. performed normal in the categorization task, but was unable to recognize a single stimulus in the identification task (Figure 9).

In a similar experiment [108] amnesic patients, including E.P., and healthy controls had to categorize cartoon animals. These animals consisted of 9 features, which could take either one of two values (e.g. striped body or spotted body). Subjects studied small distortions of a prototype cartoon animal during a training phase, and immediately afterwards categorized novel cartoon animals according to whether they belong to the category. Subsequently subjects had to name the features of the animals on the according values. Amnesic patients performed as well as the controls in the categorization task, but scored significantly worse in the naming task, E.P. wasn’t able to name any of the features. In the categorization, two of their amnesics actually categorized stimuli opposite to the way they should have. That is, they mistakenly called the prototype and low distortions nonmembers and called the antiprototype and high distortions members. Reed et al. suggested that amnesics had a spared implicit category learning system that partitioned members and nonmembers but that perhaps declarative memory was needed to remember which partition corresponded to the items they had previously been exposed to.

Flanery and colleagues [109] conducted a series of experiments to investigate the process of category learning by using the paradigms from the previously described studies. In the first experiment, subjects were trained either on 40 times the prototype (40P condition) or 40 high distortions of a prototype (40H condition) or were not trained at all, and then performed a recognition or a categorization task, respectively. Flanery could replicate Squire’s and Knowlton’s [2] results, in that the performance of the untrained subjects was comparable to the performance of the amnesic patients. When
Figure 9: Behavioural results from [2].
(A) Classification 1. Classification of 84 novel dot patterns after studying 40 different training patterns that were distortions of a prototype dot pattern. Control subjects (n=4, open bars) and E.P. (closed bars) performed similarly, endorsing test items as a function of how closely they resembled the prototype of the training category. In each panel, brackets for the control subjects for each item type (4 prototypes, 20 low distortions, 20 high distortions and 40 random dot patterns) show the SE of their four scores, each of which was an average from six tests. For E.P., the brackets for each item type show the SE of his performance on six tests. (B) Overall percentage correct scores for classification 1 (A) and a parallel test of recognition memory (recognition 1). E.P. performed as well as control subjects at classification. In contrast, he scored at chance and much worse than controls (49.1% vs. 95.2% correct) when trying to recognize a single dot pattern 5 min after it had been presented 40 times consecutively. The recognition test consisted of 8 repetitions of the study pattern intermixed with 76 random dot patterns [2].
subjects performed a categorization task in the 40P condition, their performance was significantly higher than in the 40H condition. These results suggest that the information presented during initial exposure can have a significant effect on the categorization performance. To examine this relationship more detailed, they induced amnesia in normal subjects by delaying the test by several weeks. During the test phase subjects were tested either on the category they learned or a different one, without being told. Interestingly, in the 40H condition the performance was virtually identical in both test conditions, same or different category. Whereas in the 40P condition, the performance was higher for the same category as in the 40H conditions, but was at chance-level for a different category. Subjects appear to be making categorization decisions based on what they acquired during earlier phases of the experiment. A good performance can be achieved in the absence of any prior exposure to the category members. And it shows that very good categorization performance can be achieved when people are tested on items that are different from what they had actually studied. But this seems to only occur when subjects have been initially exposed to a very diffuse category structure consisting of high distortions that are not very similar to one another. When subjects have been exposed to a clear category structure through repetition of a single prototype, they attempt to categorize items based on that acquired category knowledge, not on information presented during categorization test.

Flanery also addresses the experiments by Reed et al. [108], especially the question if those two amnesic patients, who responded the opposite way as they should have, remembered the learned category unconsciously or if they learned the category anew while performing the categorization task. To address this issue, they replicated the paradigm of Reed et al., but had three different conditions during the test phase. For a third of the subjects, the test started immediately after the training and the categories were the same as in the training. In this condition, the subjects performed good. In the second
condition, the test was delayed by one week to induce mild amnesia in the subjects. In this condition, subject’s performance in the categorization was comparable to the performance in the first condition. But the performance of the recall was significantly lower in this second condition. And the third condition included also a delay of one week, but additionally the category was redefined by the experimenter, so subjects performed a categorization task with an untrained category. Even so, the performance in the categorization was the same as in the other two conditions, and the performance of the recall lay between the other conditions.

Palmeri and Flanery [110] found evidence that normal subjects can acquire information about a category in the absence of prior study and even in opposition of prior study. Classification decisions made during a categorization test may not be based solely on information acquired during a study task, but may also be based on information acquired during the test itself.

They conclude that care must be taken in selecting items for a categorization test so as not to provide additional information about the category being tested or so as not to change the information about the categories that may have been previously acquired that may have been previously acquired. A preferable way to test individuals in a neutral manner might be to sample all possible test stimuli from a uniform distribution.

An important addition to the discussion of these observed dissociations between categorization and identification would be an experiment showing a double dissociation, i.e. patients showing a deficit in the identification task, but not in the categorization task (amnesic patients), and patients showing the opposite, normal behavior in the identification task and a deficit in the categorization task. Knowlton and colleagues conducted an experiment addressing this issue [111]. Amnesic patients and non-demented patients with Parkinson’s disease were given a probabilistic classification task in which they learned which of two outcomes would occur on each trial, given the particular combination of cues that appeared. Amnesic patients exhibited
normal learning of the task but had severely impaired declarative memory for the training episode. In contrast, patients with Parkinson’s disease failed to learn the probabilistic classification task, despite having intact memory for the training episode.

Furthermore patients with Huntington’s disease were severely impaired in a probabilistic classification task (see section 3.4 'Category learning tasks'; [112], which is considered to be independent of declarative memory.

Alternatively, researcher have proposed that the reported dissociations can also be attributed to a single memory system. The results of the study by Knowlton and Squire [98] can also be predicted by the exemplar-memory-model [113], which implies a single memory system. With this model the dissociations between categorization and recognition can be interpreted in terms of allowances for parameter differences across groups. In one experiment, a parameter change in memory sensitivity was induced by testing categorization and recognition at varying delays; the results closely matched the ones observed by Knowlton and Squire [98] for normal and amnesic participants. To achieve this, Nosofsky [114] adjusted a parameter, which reflects memory sensitivity (see 'Exemplar model' in section 3.3). A high value was used for normal subjects, and for amnesics the value was lower, so that the reported performances could be replicated. The exemplar model also yielded good quantitative predictions of the categorization-recognition dissociation.

When profound amnesia is induced in normal subjects by eliminating the study session, these subjects perform as good as control subjects in the categorization task, but are at chance-level in the recognition task [110]. These findings suggest that subjects are able to discover which clusters of patterns are likely to be members of the category simply because many members are similar to one another and all nonmembers are dissimilar from one another. Successfully performing the categorization task may require only the use of working memory which is known to be spared in amnesia.

When amnesic patients perform a categorization task with two categories,
their performance is marginally worse than that of normal subjects [115]. These results could be predicted with the exemplar-model.

The important dissociation between the two groups are the observed correct categorization probabilities for the various types of items. The dot-patterns were divided into groups according to their similarity to the prototype.

However, other evidences that these two processes rely on separate neural networks have been found in other studies.

In a study by Grill-Spector & Kanwisher [116], in which natural images were used, the performances of a detection task and a categorization task were similar over several stimuli exposure times, whereas the performance of an identification task was significantly lower. In a second experiment they compared performances of a detection task and a categorization task, identification task resp., on a trial-by-trial basis with exposure times of 17 and 33 ms. A stimulus appeared on the screen, immediately followed by a mask, then a second stimulus was shown, also followed by a mask, one stimulus contained an object, the other was a random dot-pattern. Subjects first had to indicate in which interval the object was, then categorize or identify the object in a two-alternative forced choice. A successful detection leads to a successful categorization and vice versa, but no such correlation could be seen between detection and identification.

Psychophysical experiments with monkeys and human observers have shown that, in some cases, categorization of objects can be achieved extremely fast, making it unlikely that categorization follows identification [117, 116].

In another test of the multiple-systems hypothesis, subjects performed either a simple or a complex categorization task, while performing a simultaneous numerical Stroop task [118]. In the simple categorization tasks, each set of contrasting categories was separated by a unidimensional explicit rule,
whereas the complex tasks required integrating information from three stimulus dimensions and resulted in implicit rules that were difficult verbalize. In the concurrent task condition, the participants had to perform a numerical analogue of the Stroop task along with learning the categorization rules. The concurrent task required the participant to remember which of two numbers was physically larger and which was numerically larger. The concurrent Stroop task dramatically impaired learning of the simple explicit rules, but did not significantly delay learning of the complex implicit rules.

4.3 Models of categorization

During the past decades, various scientists tried to establish models in order to explain human performance in various category learning tasks. Despite this extensive research, there is no conclusion about the superiority of any of these approaches. Instead, it is likely that they reflect processes that work in a cooperative, supplementary or alternative way rather than in an exclusive manner.

On each trial of a categorization task, a subject is shown one of \( n \) stimuli and is asked to assign it to one of \( m \) experimenter defined categories, where \( m < n \).

*Prototype models* assume that when asked to assign a stimulus to one of several categories, the subject responds with the category possessing the most similar prototype [119].

*Exemplar models* assume that subjects perform some sort of global match between the representation of the presented stimulus and the memory representation of every exemplar of each category and then choose a response on the basis of these similarity computations [120, 106].

*Decision bound models*, which are derived from general recognition theory (GRT; [121]), assume that the subject learns to assign responses to regions of the perceptual space. On each trial, the subject determines in which region
the stimulus representation falls and then emits the associated response. The
decision bounds are the partitions between competing response regions.

In spite of the apparently large differences between these models, they all
make three kinds of assumptions: 1) representation assumptions, 2) retrieval
assumptions, and 3) response selection assumptions.

**Representation assumptions**  Prototype, exemplar and decision bound
models all assume that the percept associated with any single exposure to
a stimulus can be represented as a point in a multidimensional space. A
second representation assumption currently make only by the GRT decision
bound models is that repeated presentation of the same stimulus does not
always lead to the same perceptual effect, that is, a fundamental property of
all biological systems is noise.

**Retrieval assumptions**  GRT assumes that subjects attempt to solve
the computational problem of maximizing categorization accuracy. The op-
timal solution to this problem is to divide the stimulus space into response
regions. On each trial, the optimal classifier determines in which region the
stimulus is located and the emits the associated response. The partition be-
tween two competing response regions is called the decision bound. GRT
assumes that subjects attempt to respond optimally, but because of inherent
limitations in their information processing capabilities, they fail in this goal.
Five such limitations are identified: 1) perceptual noise, 2) selection of a
suboptimal decision bound, 3) variability in the memory of this bound, 4)
response bias, 5) variability in the memory of the response criterion.

Prototype models assume that $P(R_a|S_i)$, the probability that Stimulus
$S_i$ elicits Response $R_a$, is determined by the similarity of stimulus $S_i$ to each
of the category prototypes. Similarity is assumed to be a monotonically de-
creasing function of the psychological distance between point representations
of the stimulus and the prototype. A number of different versions of the
model can be constructed depending on how distance and the function that
relates distance and similarity are defined.

Exemplar models assume that categorization is based on a match of the stimulus representation to the representation of all exemplars of every category. Perhaps the most widely known exemplar model is the context model of Medin and Schaffer ([120]). As originally proposed, the context model assumed that the stimulus representation was binary valued on each perceptual dimension. Nosofsky ([106]) generalized the context model to continuous valued stimulus dimensions and he allowed for possible response biases.

**Response selection assumptions** Two kinds of response selection assumptions are popular. The first is that given stimulus $S_i$, the probability of responding $R_a$ is computed from the similarity-choice model (SCM; [122, 123]). Specifically, in a categorization task with two categories, A and B, the SCM predicts that the conditional probability of responding A on trials when stimulus $S_i$ was presented is equal to

$$P(R_a|S_i) = \frac{b_a * n_{ia}}{b_a * n_{ia} + (1 - b_a) * n_{ib}}$$ (1)

where $b_a$ is the response bias toward category A and $n_{ix}$ is the similarity of stimulus $S_i$ to category x.

The SCM has been incorporated into the most widely known exemplar models, namely the context model ([120]; although in this model $b_a = 0.5$) and the generalized context model (GCM; [106]). In this case, $n_{ix}$ is equal to the sum of the similarities of stimulus $S_i$ to all exemplars in category x.

The second kind of response selection assumption is that response selection is a deterministic process. Let $h(x_i)$ be some function of the perceptual coordinates with the property that category A is favored when $h(x_i)$ is negative and category B is favored when $h(x_i)$ is positive. For example, in exemplar or prototype models $h(x_i)$ might equal $n_{ib} - n_{ia}$. Then the deterministic decision rule is to

Respond $R_a$ if $h(x_i) < d$; otherwise respond $R_b$.
Note that $d$ is a bias parameter. When $d > 0$, response $R_a$ is favored and when $d < 0$, response $R_b$ is favored.

With perceptual noise, the perceived value of the stimulus is $x_{pi} = x_i + e_{pi}$ and with criterial noise the subject determines the referent to be $d + ec$, where $ec$ is a random variable with mean 0. Under these conditions, the rule becomes

Respond $R_a$ if $h(x_{pi}) < d + ec$; otherwise respond $R_b$.

Note that this rule is no longer deterministic in the sense that the same response is always given to the same stimulus. However, it is deterministic in the sense that response $R_a$ is given with probability 1 if $h(x_{pi}) < d + ex$. In most experiments, both perceptual and criterial noise are expected.

**General recognition theory (GRT)**

**Boundary model** General recognition theory (GRT), which was first introduced by Ashby and Townsend ([121]), is a multivariate extension of signal detection theory [124]. It assumes that, on any given trial, the perceptual effect of a stimulus can be represented as a point in a multidimensional space. However, like signal detection theory, GRT assumes that repeated presentations of the same stimulus do not always lead to the same perceptual effect. Thus, over trials, the perceptual effects of a stimulus are represented by a multivariate probability distribution [105].

GRT assumes that on any single trial, the percept elicited by stimulus $i$ can be represented by the vector $x'_i = [x_{1i}, x_{2i}, ..., x_{ni}]$, where $x_{ij}$ is the perceived value of stimulus $i$ on dimension $j$. Because of stimulus and neural noise, $x_i$ is assumed to be a random vector that varies across trials. In most applications of GRT, the distribution of $x_i$ is assumed to be multivariate normal.

GRT assumes that to select a response, the subject partitions the perceptual space into two (or more) regions and associates a different response with each region. On each trial, the subject determines which region the percept
is in and then emits the associated response. The decision bound is the set of all points that separate the regions. In the case where there are two response alternatives, it is possible to define a function $y_i = h(x_i)$ with the property that

$$y_i = h(x_i) \begin{cases} > 0 & \text{for all } x_i \text{ in one response region} \\ = 0 & \text{for all } x_i \text{ on the decision bound} \\ < 0 & \text{for all } x_i \text{ in the other response region} \end{cases} \quad (2)$$

This static version of GRT is illustrated in Figure 10.

Ashby [126] proposed the process interpretation of GRT described in Figure 11 (illustrated for the special case in which the number of stimulus dimensions, $n$, equals two). In this interpretation, there is a different sensory channel tuned to each stimulus component or dimension. For example, when the stimuli are lines that vary in length and orientation, channel 1 would be a size (i.e., spatial frequency) sensitive channel, whereas channel 2 would be orientation sensitive. The crossing lines on the left allow for the possibility that channel 2, for example, might respond to the channel 1 component. Ashby showed that whether or not these lines cross is closely related to whether the stimulus components are perceived integrally (i.e., crossed lines) or separably (i.e., no crossing). The crossed lines in the middle of Figure 11 allow for possible lateral interactions between the channels (i.e., lateral inhibition or excitation). Ashby showed that this more central type of perceptual interaction is closely associated with the phenomenon of perceptual dependence.

The output of the sensory channels is the percept $x$, which serves as input to the decision process. The decision process is assumed to compute the discriminant value $y = h(x)$. For mathematical equivalence with FRT, the output of each sensory channel is a single numerical value on each trial, as is the output of the decision process.
Figure 10: The classic GT model

Figure 10a shows the perceptual representation of stimulus $i$, which in this example is assumed to vary on two perceptual dimensions $x_1$ and $x_2$. The ellipse is a contour of equal likelihood from the multivariate perceptual distribution that is the experiment-level perceptual representation, and the point is an example of the perceptual representation from a single trial. Figure 10b illustrates the decision process. See text for more details [125].
Figure 11: A GRT process
A process interpretation of GRT for the special case in which there are only two perceptual dimensions (and two stimulus dimensions) [125].

As a consequence, the decision rule: 'Respond A to all points on one side of the bound and B to all points on the other side', is equivalent to the rule:

Respond A if \( y_i > 0 \), and respond B if \( y_i < 0 \).

Because the function \( y_i = h(x_i) \) discriminates between stimuli on either side of the decision bound, it is often called a discriminant function.

The probability of responding \( R_j \) on trials in which stimulus \( S_i \) was presented is equal to the probability that a random sample from the perceptual distribution associated with stimulus \( S_i \) falls in the response region associated with response \( R_j \). Formally,

\[
P(R_j | S_i) = \int \int f_i dxdy
\]

where \( f_i(x, y) \) is the bivariate normal probability density function for stimulus \( S_i \), and R signifies the \( R_j \) response region.

**Context theory**  A very similar model to the MDS-choice model has been proposed by Medin and Schaffer [120] for categorization, which is known
as the context theory. According to the theory, the probability that stimulus $S_i$ is classified in Category $C_j$, $P(R_j|S_i)$, is given by

$$P(R_j|S_i) = \frac{b_j \sum_{j \in C_j} n_{ij}}{\sum_{k=1}^{m} (b_k \sum_{k \in C_k} n_{ik})}$$  \hspace{1cm} (4)$$

where the parameter $b_j$ represents the bias for making category response $R_j$. As before, the symbol $n_{ij}$ denotes the similarity between stimuli $S_i$ and $S_j$. The index $j \in C_j$ is intended to read all $j$ such that $S_j$ is a member of $C_j$.

As is evident, the context model response rule (Equation 4) bears a striking structural resemblance to the choice model for stimulus identification (Equation 3). Indeed, the two response rules can be linked in a simple way. The one-to-one mapping of stimuli onto responses in identification is transformed into a many-to-one mapping of stimuli onto responses in categorization. A natural starting hypothesis for a quantitative model relating the two paradigms was proposed by Shepard, Hovland, and Jenkins [127] and Shepard and Chang [128]: To predict categorization performance from identification performance, one should simply cumulate over all stimulus-response cells in the identification matrix that would map onto a given stimulus-response cell in the categorization matrix. Stated another way, all inter item confusions in the identification paradigm that are within-class confusions would result in correct categorization responses. Only between-class confusions would result in categorization errors. This hypothesis was also referred to as the mapping hypothesis [106].

The mapping hypothesis formalizes the idea that the principles of stimulus generalization underlying identification will also underlie categorization. The response rule of Medin and Schaffer’s context model (Equation 3) arise essentially by combining the mapping hypothesis with the assumption that the choice model accurately characterizes performance in identification paradigms. The only difference is that the identification response bias parameters in Equation 3 are replaced by categorization response bias
Exemplar model  The exemplar model is the generalized context model (GCM). The GCM generalizes the original version of the context model and integrates it with classic theories and ideas proposed in the areas of choice and similarity. The GCM uses a multidimensional scaling (MDS) approach to modeling similarity. According to the model, exemplars are represented as points in a multidimensional representational space, and similarity between exemplars is a decreasing function of their distance in the space. Selective attention processes systematically modify the structure of the space in which the exemplars are embedded. An important working hypothesis is that, with experience in a given task, observers often learn to distribute their attention over psychological dimensions in a manner that tends to optimize performance.

These ideas are illustrated schematically in Figure 12. In the top panel, there are two categories, X and O, defined by five exemplars each. The exemplars reside in a two-dimensional space. Exemplars X2 and O4 are close together in the space, and so are highly similar to one another, whereas Exemplars X and O2 are far away, and so are dissimilar. Suppose that an observer needs to classify item i (illustrated in the space). According to the model, the observer sums the similarity of item i to all the X exemplars and to all the O exemplars. The classification decision is based on the relative magnitude of these sums.

Specifically, in an experiment involving multiple categories and in which there are no response biases, the probability that item i is classified into Category J is given by

$$P(J|i) = \frac{(\sum_{j \in J} S_{ij})^\gamma}{\sum_K (\sum_{k \in K} s_{ik})^\gamma}$$

(5)

where $s_{ij}$ denotes the similarity of item i to exemplar j and the index $j \in J$ denotes that the sum is over all exemplars j belonging to Category J.
Figure 12: Example of category structure for GCM
Schematic category structure for illustrating the workings of the generalized context model. X, exemplars of Category X; O, exemplars of Category O. Panel A: Category structure with equal attention to both dimensions. Panel B: Category structure with selective attention to the horizontal dimension [129].
The parameter $\gamma$, first introduced into the GCM response rule by Ashby and Maddox [130], is a response-scaling parameter. When $\gamma = 1$, the observer responds by 'probability matching' to the relative summed similarities, whereas when $\gamma$ grows greater than one, the observer responds more deterministically with the category that yields the largest summed similarity. A direct process interpretation for the emergence of the summed similarity computation and the $\gamma$ response-scaling parameter was developed by Nosofsky and Palmeri [131] in terms of their exemplar-based random walk model of categorization.

A critical component assumption in the GCM is that similarity between exemplars is not an invariant relation but a highly context-dependent one. In particular, it is assumed that selective attention processes modify psychological similarity relations among exemplars, usually in a manner that tends to optimize performance for any given task. This selective attention process can have a dramatic influence on the classification predictions that are made by the model.

For example, in the top panel of Figure 12, item $i$ is roughly equally similar to the exemplars of Category X and Category O, and so would be classified in each category with roughly equal probability. However, the horizontal dimension is far more relevant than is the vertical dimension for discriminating the categories. An experienced observer would presumably learn this aspect of the category structure and would attend selectively to this relevant dimension.

This process is represented in the model in terms of a set of selective attention weights that stretch the space along attended relevant dimensions and shrink it along unattended irrelevant ones, as illustrated in the bottom panel of Figure 12. Note from the illustration that this selective attention process would tend to optimize performance because it would separate further in psychological space the two categories that need to be discriminated. In addition, note that whereas item $i$ was roughly equally similar to the X and O exemplars when equal attention was given to the dimensions (top panel),
it takes on far greater similarity to the exemplars of Category X following selective attention to the relevant dimension (bottom panel).

These ideas are represented formally in the model as follows. Assume that the exemplars reside in an M-dimensional psychological space, and let $x_{im}$ denote the value of exemplar $i$ on psychological dimension $m$. (These psychological coordinate values for the exemplars are often derived by conducting a variety of similarity-scaling studies in which MDS solutions for the exemplars are derived: for a review, see [132]) The distance between exemplars $i$ and $j$ is computed by using the weighted Minkowski power-model formula,

$$d_{ij} = (\sum_{m} w_{m} |x_{im} - x_{jm}|^r)^{\frac{1}{r}} \quad (6)$$

In this equation, the value $r$ defines the distance metric of the space. Common values are $r = 1$, which defines a city block distance metric; and $r = 2$, which defines a Euclidean distance metric. The city block metric is typically assumed when modeling distances among highly separable dimension stimuli, whereas the Euclidean metric is used to model distances among integral-dimension stimuli. The parameters $w_{m}$ are the attention weights. The $w_{m}$ parameters model the degree of attention that an observer gives to each dimension in making psychological distance judgments among exemplars. As illustrated previously, a geometric interpretation for the attention weights is that of stretching and shrinking the psychological space along its dimensions.

The similarity between exemplars $i$ and $j$ ($s_{ij}$) is assumed to be a nonlinearly decreasing function of their distance ($d_{ij}$),

$$s_{ij} = \exp(-c \times d_{ij}^p) \quad (7)$$

where $c$ is an overall scaling or sensitivity parameter and the value $p$ defines the precise form of the similarity gradient. Common values of the
generalization gradient are \( p = 1 \), which defines an exponential similarity gradient, and \( p = 2 \), which defines a Gaussian similarity gradient [106]. The exponential model is favored in situations in which observers learn to classify fairly discriminable, nonconfusable stimuli.

The sensitivity parameter \( c \) in Equation 7 determines the rate at which similarity declines with distance. Its role is illustrated in Figure 13. The top panel shows an example in which the value of \( c \) is relatively high. In this case, the generalization gradient relating similarity to distance is steep, so that exemplars that are even a moderate distance apart in the space are judged as very dissimilar. By contrast, as shown in the bottom panel, when \( c \) is small, the generalization gradient is shallow, and exemplars that are separated by large distances in the space may still be judged as similar.

Finally, the \( s_{ij} \) values computed from Equations 6 and 5 are substituted into Equation 5 to yield the classification predictions made by the GCM.

In summary, the GCM is defined by the system of Equations 5-7. Its free parameters include the response-scaling parameter \( \gamma \), the sensitivity parameter \( c \), and the attention weights \( w_m \).

Prototype model Prototype models use a similar multidimensional representational space as seen in the case of exemplar models. The crucial difference between prototype and exemplar approaches is that in the prototype model categories and individual exemplars have separate representations. When observers are shown exemplars during training they extract the central tendency or summary representation of these items, called the prototype. For recognition, a stimulus is compared with individually stored exemplars, while for categorization, a stimulus is compared with the prototype in order to decide whether or not it is a member of the category.

Luce-Shepard Choice Model (Identification) A widely and successfully used model of identification processes is the similarity choice model developed by Shepard and Luce [123, 122]. According to the model, the
Figure 13: Sensitivity parameter
Illustration of exponential gradients relating similarity to distance. Top panel: high value of sensitivity. Bottom panel: low value of sensitivity [129].
probability that stimulus $i$ leads to response $j$ in an identification experiment, $P(R_j|S_i)$, is given by

$$P(R_j|S_i) = \frac{b_j \ast n_{ij}}{\sum_{k=1}^{n} b_k \ast n_{ik}}$$

(8)

where $0 \leq b_j \leq 1$, $\sum b_j = 1$, $n_{ij} = n_{ji}$ and $n_{ii} = 1$. The $b_j$ parameters are interpreted as response bias parameters and the $n_{ij}$ parameters as similarity measures on the stimuli $S_i$ and $S_j$.

In Shepard’s [123] original formulation of the model, the similarity parameters were given an explicit interpretation in terms of distances in a psychological space. He assumed that

$$n_{ij} = f(d_{ij})$$

(9)

where $f$ is some monotonically decreasing function and where the $d_{ij}$’s are distances that satisfy the metric axioms. To reduce the number of parameters to be estimated, Shepard suggested that the stimuli be represented as points in a low-dimensional psychological space. The $d_{ij}$’s could then be derived by computing the distances between the points in space. The configuration of points that achieved the best account of the identification data would then be taken as the multidimensional scaling (MDS) solution for the stimulus set.

This model is often referred to as the MDS-choice model.

### 4.4 Category learning tasks

There is a large number of category learning tasks and many of them are specifically related to different theoretical models. In this section, most common tasks will be described.

**The dot pattern task** The prototype stimulus consists of a certain number of dots randomly placed in an area. Category exemplars are gen-
erated by various degrees of displacement of the dots. If the degree of displacement is high, exemplars will be less similar to the prototype (high distortions) than when the degree of displacement is smaller (low distortions) (Figure 14). In the training phase, subjects are exposed with distortions of a prototype, with or without explicit instructions to learn to category. In the testing phase, new distortions (low and high), the prototype, and category-independent random dot patterns are sequentially presented, and subjects are asked to decide whether or not a stimulus is a member of the trained category. Interestingly, categorization performance is best for the prototype, followed by low and high distortions, respectively. This pattern of performance is observable despite the fact that subjects did not see the prototype in the training phase [133].

In the recognition version of the test, subjects view a prototype during the training phase, and are asked to memorize it. During the test phase, the memorized dot pattern (target) must be selected from distractors that are low and high distortions of the target. The dot pattern task is the cornerstone of prototype theories.

**Artificial grammar learning** This test has several versions. In one version, subjects are presented with letter-strings, which are generated according to a specific rule that determine the order of letters. For example, strings QA TA and GFBF are different exemplars, because they consist of different letters. However, they belong to the same category, because they have been generated according to the same abstract rule: the second and fourth letters are always the same. In contrast, ATAQ is very similar to QATA in its surface characteristics (letters), albeit they have been generated according to different abstract rules. Similarly to the dot pattern task, participants are not directly instructed to learn the rule, they passively view category exemplars [134].

In the test phase, decisions are made about new letter strings; the task is
The basic pattern (prototype) consists of nine dots. Category exemplars are distorted versions of the prototype, whereas random patterns have a completely different structure. While the prototype (P) is localized in the center of the representational space (right side of the figure), the exemplars have a more peripheral position. Low distortions (LD) are more similar to the prototype than the high distortions (HD), and hence LDs are closer to the prototype in the representational space [74].
to determine whether the new stimuli belong in the group of 'grammatical' items, depending on whether they were generated according to the rule of the training phase. In different paradigms, feedback is used or subjects are explicitly instructed to direct their attention to the underlying rule. In the case of feedback learning of artificial grammar, boundary models may be preferred [121].

**Set-shifting and stimulus-reward learning**  In these tasks, stimuli vary along a number of dimensions such as orientation, shape and color. Subjects first learn the relevant dimensions for decisions (e.g. 'select the red stimulus'), by using feedback in each trial. During intradimensional shifting, new stimuli are presented but the same dimensions remain relevant. This phase of discrimination learning represents the classic criterion of categorization: generalization of knowledge for never-seen exemplars. During extradimensional shifting, the context changes and a previously irrelevant dimension must be used for decisions (e.g. shape instead of color). Therefore, following the altered feedback, observers should override their previous strategy and shift their attention to the new dimension. This process is an essential component of the Wisconsin Card Sorting Test (WCST), which is often mentioned as a classic rule-based category learning task [135]. In the WCST, the boundary between categories is unambiguous, and cues such as color, form and orientation determine category-membership in an all-or-none fashion. Therefore, the WCST is mediated by the explicit system that is based on easily verbalized rules.

Reversal learning is an ideal approach to investigate the stimulus-reward learning component of visual discriminations. In the first phase of this task, feedback is used to establish stimulus-response associations (for example, rewarding feedback is given when key A is pressed for a red stimuli and when key B is pressed for a green stimuli) [135]. In the reversal phase of the task, the contingency changes and subjects are requested to adopt the new strategy.
Gambling tasks provide a more sophisticated approach to investigate stimulus-reward patterns in category learning. The Iowa Gambling Task [136, 137] includes categorical decisions (selecting from different decks of cards), which are associated with different reward-punishment profiles. Advantageous decisions require the avoidance of categories associated with high reward but even higher punishment (overall loss) and the preference of categories associated with smaller rewards but minimal penalties (overall gain). In the 'Gamble' task, the larger reward is associated with the least likely outcome in a probabilistic decision-making situation. Risk-taking behavior is characterized by the selection of larger immediate reward despite its less advantageous long-term consequences.

Probabilistic classification learning In this test, subjects are asked to decide whether geometrical forms (cues) predict rain or sunshine (weather prediction task). Each cue is associated with rain and sunshine with a certain probability. After each trial, including a decision about weather outcome, subjects receive feedback. In this way, during the first 50 trials there is a gradual increase in performance. After completing the test, subjects are asked questions about the cues and other details of the test in order to evaluate their explicit knowledge [138, 111]. Probabilistic classification learning is best related to boundary models. Because of the probabilistic nature of the task, category rules are hard to verbalize, and hence this task may be mediated by the implicit system.

Rule-based and information-integration paradigms These tasks are also motivated by the boundary models of category learning in which stimulus-response and stimulus-reward associations are important components. Consider a two-dimensional representational space including the length of vertical and horizontal line-segments of stimuli (Figure 15a). This representational space can be divided into two parts for two corresponding cat-
egories according to linear rules. In the dimensional integration condition, the length of both vertical and horizontal segments must be taken into consideration (Figure 15b), whereas in the selective attention condition only one dimension is relevant for categorization (Figure 15c). Finally, the representational space can also be divided according to non-linear rules (Figure 15d).

Category rules related to linear conditions are regularly easy to verbalize, while non-linear rules are sometimes impossible to word. The first case is a rule-based mode of category learning (as seen in the case of WCST), whereas in the latter case information from many exemplars must be accumulated for categorical decisions in an implicit-procedural manner. However, it must be noted that many information-integration tasks have linear category boundaries. Rule-based tasks are those in which the best possible performance can be achieved via a rule that is easy to describe verbally.

5 Recollection and Familiarity

There is increasing evidence from cognitive, neuropsychological and neuroimaging studies that recognition memory performance reflects two distinct memory processes or types of memory, often referred to as recollection and familiarity. Familiarity is defined as recognizing a person, or a visual object, as familiar, but not being able to recollect who the person is or where they were previously encountered. Whereas in recollection, one remembers having encountered the person, or the object, previously and contextual information about the previous encounter is recalled [139]. In the existing recognition memory literature, we can observe mainly four empirical dissociations between recollection and familiarity that have been used to support for the claim that recognition memory involves more than a single type of memory.

First, recollection is more severely disrupted than familiarity by certain brain lesions, indicating that these two processes are mediated by different brain regions. For example, amnesic patients exhibit significantly greater memory
Figure 15: Example of category structure
(a) An example for experimental stimuli consisting of two line segments.
(b) Dimensional integration condition. The two-dimensional representational space (length of the horizontal and vertical segments) is divided into two categories. The length of both segments is relevant for categorization. c) Selective attention condition. Only the length of vertical segment is relevant for categorization. (d) Non-linear condition. The representational space is divided by non-linear rules that are difficult to word [74].
impairments on associative than item recognition tests, indicating that the regions damaged in amnesia are more important for the former than latter type of recognition judgment [140]. In associative recognition tasks, subjects are asked to recollect specific information about the study event, such as determining when or where an item was previously studied. In an item recognition test, the participant has to distinguish recently studied items from non-studied items.

Second, studies investigating the processing speed of these two processes have indicated that familiarity is faster than recollection. Hintzman and Caulton [141] conducted a study with healthy human subjects to investigate the processing speed in an associative and an item recognition task under speeded test conditions. Subjects were found to be able to make accurate discriminations that can be based on familiarity (i.e. the item recognition task) faster than they can make discriminations that require them to recollect specific information about the study event (associative recognition task). Other studies have shown that, when the subject is allowed to take more time for the decision, the probability of accepting a new item that is either similar to a studied item, or is from an inappropriate study list, first increases then decreases [142, 143]. These results suggest that a fast familiarity process leads new items, which are similar to studied items, to be incorrectly accepted, and only with additional retrieval time are subjects able to recollect the information that allows them to reject those items.

Third, the analysis of recognition confidence responses indicates that recollection and familiarity can produce distinct receiver operating characteristics (ROCs). When one plots the hit rates against false alarm rates as a function of response confidence (i.e., an ROC), the observed empirical functions change their shapes across conditions, such that they require at least two functionally independent memory parameters to describe them, indicating that at least two separate memory components are needed to account for recognition performance [144, 145]. The ROCs in item recognition tests
are curvilinear, as long as performance is above chance level. Whereas in associative recognition tests the ROCs are quite distinct in the sense that they are often much more linear in shape [144, 146].

Fourth, neuroimaging and electrophysiological studies found distinct neural correlates of recollection and familiarity. For example, event related potentials (ERPs) recorded on the scalp during recognition tests indicate that items that are remembered or that are associated with accurate memory for some detail of the study event are related to ERP that exhibit distinct temporal and spatial scalp distributions from those related to items that are recognized on the basis of familiarity in the absence of recollection [147]. Although these results do not indicate which brain regions support recognition performance, they do indicate that there are at least two separate brain processes involved.

### 5.1 fMRI studies

During recognition memory retrieval, hippocampal and parahippocampal regions are related to recollection, but not familiarity. For example, in recognition memory tests of drawings, bilateral hippocampal and parahippocampal regions were related to associative recognition, but not to item recognition [148]. Moreover, in a Remember / Know recognition test with words, Remember responses activated left hippocampal and parahippocampal regions, but not Know responses [149]. Both recollection and familiarity activate prefrontal regions during retrieval, but the two processes seem to rely on partially distinct sub regions. Henson [149] found that left anterior prefrontal regions are more active for items that are remembered than for those that are known.

In recognition retrieval tasks, activation was found also in the parietal lobe [150, 149]. It appears to be related primarily to recollection. For example, several studies have found greater parietal activation for remember responses than know responses [149, 102], suggesting that this region may be more involved in recollection than familiarity.
Studies of Recollection and Familiarity in amnesic patients have shown that these patients are impaired in both processes, but generally more so in recollection. Selective hippocampal damage appears to disrupt recollection, but not familiarity. In studies that have included amnesic patients with damage that extends beyond the hippocampus, such as the surrounding temporal lobe, recollection is sometimes found to be more disrupted than familiarity, but other studies with similar groups of patients reported comparable recollection and familiarity deficits [151, 152].

Squire [153] showed that extensive temporal lobe damage can reduce both recollection and familiarity memory judgments to chance level. There is some evidence that the discrepancies observed in these groups can be caused in part by the different methods that have been used to match performance in amnesic and control subjects [154]. However, many studies have found that recollection is disrupted to a greater extent than familiarity in amnesic patients [140, 155]. These amnesic patients, however, had a selective hippocampal or fornix damage. Aggleton and Shaw [156] found that in the latter patients familiarity performance is in the normal range, whereas the performance for recollection was significantly impaired. The observation that disproportionate recollection deficits are seen in hippocampal patients suggests that the hippocampus is particularly important for recollection. In contrast, the finding that the proportional familiarity performance decreases are sometimes found in patients with more extensive temporal lobe lesions suggests that the regions surrounding the hippocampus may be important for familiarity.

Further evidence for this spatial dissociation was found in Remember / Know experiments [157, 158]. In these studies, patients with extensive temporal lobe damage were included, and it was found that recollection was severely disrupted, whereas familiarity was impaired to a lesser extent. One recent study contrasted Remember / Know judgments in patients with extensive temporal lobe lesions with patients who had a more selective hip-
pocampal lesion. The study found that patients with extensive temporal lobe lesions had deficits in recollection and familiarity, whereas the patients with more selective hippocampal lesions had selective deficits in recollection [159].

6 Outline of the experimental work

In this PhD dissertation I describe three experiments about identification, categorization and recall memory. In the first experiment (Part 2) the dissociation between visual pattern categorization and identification were investigated. Whether identification and categorization of visual stimuli are supported by separate or identical neural circuits remains a matter of debate. Similarly, the dynamics of these processes are poorly understood: Does identification precede categorization, or vice-versa? To answer these questions, I measured how fast visual patterns are identified and categorized. The hypothesis in this study was that a dissociation in the learning rates of the two tasks would support the theory that identification and categorization are mediated by distinct neural circuits. In the identification task, subjects were initially trained to recognize a given pattern. In subsequent sessions, they had to indicate whether each of a series of patterns was identical to the learned pattern or not. In the categorization task, subjects trained with similar patterns defining an arbitrary category. They then had to indicate whether subsequently presented patterns belonged to the category or not. I measured accuracy and reaction times in both tasks. Performance improved during the first 3 to 4 sessions, then stabilized. While accuracy in both tasks showed similar dynamics with practice, the reaction times behaved differently: In the first session, the average reaction time for categorization was $716 \pm 193$ ms, and $550 \pm 78$ ms for identification. With training, categorization reaction times decreased more rapidly than in the identification task. Both types of reaction times then converged around $467$ ms ($500 \pm 81$ ms catego-
rization; 434±24 ms recognition). These results suggest that identification and categorization are two independent processes.

In the second experiment (Part 3), I examined different category learning systems. There is strong evidence that rule-based category learning and information-integration category learning are mediated by different categorization systems that are associated with distinct, but partially overlapping, neurobiological substrates. Rule-based category learning tasks are those in which the category structure can be learned through some explicit reasoning process. Information-integration category learning tasks, on the other hand, are those in which accuracy is maximized only if information from two or more stimulus components is integrated at some predecisional stage. The aim of this study was to test whether different numbers of categories have a different impact on rule-based than on information-integration category learning. A dissociation in the effect of the number of categories in the two category learning conditions would support the hypothesis that there are multiple category learning systems. Subjects learned in this experiment to distinguish two, or four respectively, similar dot-pattern categories. The categories were separated in two different ways: In the information-integration condition the category boundaries were difficult to describe verbally, and could not be distinguished by a reasoning process. While in the rule-based condition the boundaries of the categories was along one stimulus dimension, and could be reasoned out. I found that the category number affects rule-based, but not information-integration category learning. These results suggest that rule-based and information-integration are mediated by two distinct category learning systems.

The third experiment (Part 4) investigated the two memory processes, recollection and familiarity. Event-related fMRI was used to investigate whether recollection- and familiarity-based memory judgments are modulated by the degree of visual similarity between old and new art paintings. We tested in this study whether recollection and familiarity judgments about
complex pictures are two separate mechanisms or whether they reflect memory strength. We hypothesized if recollection and familiarity judgments about complex pictures reflect strong memories and weak memories, respectively, correctly remembered items would be associated with stronger neural activation than known items. Subjects performed a flower detection task, followed by a Remember/Know/New surprise memory test. The old paintings were randomly presented with new paintings, which were either visually similar or visually different. Subjects were significantly faster and more accurate to reject new, visually different paintings than new, visually similar ones. The proportion of false alarms, namely remember and know responses to new paintings, was significantly reduced with decreased visual similarity. The retrieval task evoked activation in multiple visual, parietal and prefrontal regions, within which remember judgments elicited stronger activation than know judgments. New, visually different paintings evoked weaker activation than new, visually similar items in the intraparietal sulcus. Contrasting recollection with familiarity revealed activation predominantly within the precuneus, where the BOLD response elicited by recollection peaked significantly earlier than the BOLD response evoked by familiarity judgments. These findings suggest that successful memory retrieval of pictures is mediated by activation in a distributed cortical network, where memory strength is manifested by differential hemodynamic profiles. Recollection- and familiarity-based memory decisions may therefore reflect strong memories and weak memories, respectively.
Part II

Learning in pattern identification and categorization tasks

7 Introduction

Living organisms have a remarkable capacity to achieve adaptive responses in diverse situations. In this respect, two basic response patterns can be distinguished: we may select a specific response for a particular object or, alternatively, the same action is performed for a number of different stimuli that share behaviorally salient features. The former case refers to identification or recognition, which requires one-to-one stimulus-response mapping, whereas the latter refers to categorization, characterized by many-to-one stimulus-response mapping [74]. For example, we have a repertoire of behavior when encounter with dogs in general, but some actions remain specific for our own dog. General behavior is based on categorization, while selecting a specific action towards a particular object is based on identification [160].

It is so far still debated whether categorization and identification are two different processes. Dissociations between categorization and identification of visual stimuli have been reported by several researchers. Most current models of object recognition rely on "bottom-up" processing [161, 162, 163, 164]. Bottom-up processing means that object features are first analyzed locally and independently, and then combined into a unified, coherent percept. Some of these models imply that an object is first identified, and subsequently categorized into one of many perceptual categories.

There are, however, reasons to question the bottom-up approach, and the implicit assumption that categorization follows or is equivalent to identification. Psychophysical experiments with monkeys and human observers
have shown that, in some cases, categorization of objects can be achieved extremely fast, making it unlikely that categorization follows identification [117, 116, 165]. The performances of a detection task and a categorization task were similar over several stimuli exposure times, whereas the performance of an identification task was significantly lower [116]. In a second experiment Grill-Spector and Kanwisher compared performances of a detection task and a categorization task, and identification task respectively, on a trial-by-trial basis with exposure times of 17 and 33 ms. In this experiment, a natural stimulus appeared on the screen, immediately followed by a mask, then a second stimulus was shown, also followed by a mask, one stimulus contained an object, the other was a random dot-pattern. Subjects first had to indicate in which interval the object was, then categorize or identify the object in a two-alternative forced choice. A successful detection leads to a successful categorization and vice versa, but no such correlation could be seen between detection and identification.

In a study by Thorpe [166], monkeys and humans had to categorize natural visual stimuli, the exposure time was only 30 ms. Nevertheless performance of monkeys was 90% and that of humans 94%. The measured reaction times for monkeys was 250-300 ms, and for humans 250-450 ms.

Theoretical works also suggest that objects might be categorized before they are identified [167, 168].

Nosofsky proposed an alternative explanation for this dissociation in his exemplar model [160, 162, 113], which implies that categorization and identification rely on the same neural network. In this model the ability to discriminate a prototype, or a category respectively, from a distractor is represented in terms of a sensitivity parameter. The dissociation in the performance of amnesic patients and controls in the identification task can be explained by applying a different value to the sensitivity parameter. Furthermore in a task where the subjects had to categorize dot-patterns into two categories, the performance of the amnesic patients was lower than that of the con-
trols, which was predicted by the model [161]. Similarly, it was observed that healthy subjects perform worse in an identification task compared to a categorization task, when a delay of several days was introduced between training and test [169, 115], consistent with the model.

However, evidence that these two processes rely on separate neural networks has been found in other studies. It has been observed that categorization and identification activate different neural systems. Reber and colleagues compared the brain activities of a categorization and an identification task [1]. In the categorization task, subjects learned a category of dot-patterns, which consisted of distortions of a prototype pattern and, in a subsequent test, judged new dot-patterns whether they belong to the learned category. In the identification task, subjects viewed five dot-patterns eight times each and had to judge whether dot patterns presented during the test have been studied earlier. They found different patterns of brain activity in visual processing areas for the two tasks. During the categorization task, the familiar stimuli were associated with decreased activity in posterior occipital cortex, whereas during the identification task, the familiar stimuli were associated with increased activity in this area.

Furthermore, studies with amnesic patients show a dissociation between categorization and identification. In a categorization task with dot-patterns, the performance of amnesic patients and healthy participants are similar, but differ significantly in an identification task [2]. Amnesic patients were hardly able to recognize a specific dot-pattern, but could categorize dot-patterns as well as normal controls. Similar results were found by Reed [108] in a study with cartoon-animals with nine binary features as stimuli.

Category learning appears to be independent of declarative (explicit) memory for training instances and independent of the brain structures essential for declarative memory that are damaged in amnesia. Knowledge about categories can be acquired implicitly by accumulating information from multiple examples [98]. Using healthy subjects in various states of simulated
amnesia, researchers [109, 170] could show that subjects with simulated amnesia (delay between training and test) performed as well as normal subjects in the categorization task, but their performance decreased in the identification task.

By studying how quickly a new identification and categorization task is learned, we can work out the dependence, or independence, of these two tasks. If these two tasks rely on the same neural network, we expect that the learning rate will be the same, whereas if the learning rates diverge, it would be an indication that there is a dissociation between the two tasks.

In this study I tested whether identification and categorization have the same time-course of learning. Namely does identification precede categorization or vice versa? And whether they improve at the same rate or not. I wanted to investigate this question by examining the dynamics of the two tasks, using completely artificial stimuli in a psychophysical experiment.

8 Methods

8.1 Participants

Ten healthy persons (6 men, 4 women, mean age was 31 ± 4 years) with normal or corrected to normal visual acuity participated in this study.

8.2 Stimuli

I used dot-patterns as stimuli, which were introduced by Posner [133]. Posner originally used 9 dots for his patterns, whereas in this study 30 dots were used instead to increase the difficulties of the tasks. A dot-pattern prototype is generated by placing the dots randomly in an area of $6\circ \times 6\circ$ of visual angle. Distortions of it were created by shifting the dots by a certain distance in a random direction. The shift distance of each dot is a random value with a normal-distributed probability function with mean $x_0$ and $\sigma = 2$ Pixels
($\approx 6'$ of visual angle). $x_0$ is measured in Pixels and can have any value, it is defined as the distortion level. The boundary of a pattern-category is set by the experimenter to a shift distance of 6 Pixels ($\approx 18'$ of visual angle). Eight distortion levels have been defined for this study: The prototype, three low-distortions ($x_0 = 2, 4$ and 5 Pixels), which are within the category, three high-distortion ($x_0 = 7, 8, 10$ Pixels) outside of the category and random dot-patterns. Four examples of the stimulus, a prototype, a low-distortion (distortion level 4), a high-distortion (distortion level 8) and a random dot-pattern, are shown in figure 16. During the task, the stimuli were shifted about $1^\circ$ in a random direction, to avoid that subjects concentrate on a sub-pattern. The visual angle of the stimulus during a session was $6^\circ$ and the size of each dot was $12'$. The dots were displayed white on a black background with high luminance. The subject’s head was kept in a chin rest, to assure that the viewing distance remains constant at 50 cm.

8.3 Procedure

Each subject performed a categorization and an identification task and, as a control experiment, a reaction time task.

Reaction time In a control experiment I wanted to rule out that subjects are faster pressing one button than the other. In 150 trials the subject was asked to press the appropriate button as fast as possible when a circle appeared on the screen. The circle may appear either in the upper or in the lower part of the screen, and each button is assigned to one of the positions. After 75 trials the button-assignment was reversed. This reaction time experiment was conducted twice with each subject, once before the first session of the categorization, or the identification task respectively, and once after the last session.
Figure 16: Stimuli
Four examples of the dot-pattern stimuli used in this experiment. A: Prototype; B: Low-distortion of the prototype (distortion-level 4); C: High-distortion of the prototype (distortion-level 8); D: Random dot-pattern.
Each trial started with a fixation point, which remained for 1 to 2 seconds, then a pattern was displayed for 150 ms, followed by a mask. The screen stayed then blank until the next trial starts. The subject reported whether the pattern belonged to the studied category (categorization task), or whether the pattern is identical with the prototype (identification task) respectively, by pressing a button. He then received an auditory feedback if his response was incorrect.
Categorization task  The first session of the categorization task was a training session, in which subjects saw 40 low-distortions for 5 seconds each. These stimuli were generated from one prototype, which was unique for each subject. The displayed low-distortions build up a category and the subject was asked to learn it. Subsequently subjects went through ten task-sessions. Each task-session consisted of 3 blocks of 100 trials, these 100 trials were identical for each block in a given session and contained 5 times the prototype, 45 low-distortions, 36 high-distortions and 14 random patterns. The order of the trials was randomized for each block. Between each sessions was a break of 1 to 3 days. The sequence of a trial is schematically shown in Figure 17. In each trial a fixation point first appeared on the screen for 1 to 2 seconds (randomized). Then the pattern is shown for 150 ms, followed by a mask for another 150 ms, after which the screen stayed blank until the next trial. The subject was asked to indicate whether the current pattern is a member of the learned category or not as fast and accurate as possible, by pressing a button. An auditory feedback was provided if his answer was wrong. Accuracy and reaction times were measured.

Identification task  In the training session of the identification task, the subject saw a prototype 40 times 5 seconds and was asked to memorize it. As in the categorization task subjects performed ten sessions, which consisted also of 3 block of 100 identical trials in randomized order. In each block were 23 times the prototype, 17 low-distortions, 36 high-distortions and 14 random dot-patterns. The sequence of one trial is the same as in the categorization task (see Figure 17). The subject had to answer the question if the current pattern is identical with the one from the training session as fast and accurate as possible.
8.4 Equipment

I used a VSG graphics card (Cambridge Research Systems, UK) with a screen (Sony, GDM-F520) with a refresh rate of 120 Hz. Connected to the graphics card was a response box (CB6, Cambridge Research Systems, UK). The stimulus generation and the tasks were programmed on Matlab using the VSG-Software toolbox. The subjects placed their heads on a chin rest to ensure that the distance to the screen was constantly 50 cm.

8.5 Data analysis

From the behavioral data, the proportion of correct responses was calculated, and compared the performances in the two task with t-tests, on a session-by-session basis. Additionally a three-way analysis of variance (ANOVA) was computed with the performance and the reaction time of correct responses. The three variables were session number (1 to 10), task type (categorization or identification) and distortion level (eight different levels).

The patterns were shifted during the task slightly in a random direction, to prevent subjects concentrating on a sub-pattern. To see if this shift had any effect on the reaction times or accuracy, I calculated an index as following:

\[
index = \frac{Performance_{top} - Performance_{bottom}}{Performance_{top} + Performance_{bottom}} \tag{10}
\]

where \(Performance_{top}\) means the performance when the pattern appeared in the upper part of the screen, and \(Performance_{bottom}\) when the pattern was presented on the lower half of the screen. This index is positive if subjects performed better when the pattern was on the upper part of the screen, negative if patterns in the lower part of the screen exhibited a higher performance, and zero if both performances are equal. Similarly, the index for the difference in performance in the right-left direction, as well as for the reaction times in both direction (top/bottom and right/left) was calculated.
In the control experiment, subjects had to press a button as fast as possible, depending on where a circle appeared, they had to press one of two buttons. The subject’s mean reaction time was 338±49 ms. The reaction times for the two buttons were not significantly different (Button A: 338±25 ms; Button B: 325±37 ms; unpaired t-test, p<0.05). Furthermore, when the assignment of the buttons were reversed, the reaction times were virtually identical, also compared to the reaction times of the original assignment (Button A:348±69 ms; Button B: 343±57 ms; unpaired t-test, p<0.05; see Figure 18).

Figure 19 shows the proportion of yes-responses during the categorization
Figure 19: Results of categorization task
Mean proportions of Yes-responses, averaged across all subjects, during the categorization task
Figure 20: Results of identification task
Mean proportions of Yes-responses, averaged across all subjects, during the identification task.
task for each distortion level. The error bar indicate standard error of the mean and the red line shows chance-level, which is at 50%. The asterisks signify a significant difference to chance-level (t-test, p<0.05). This graph shows that the subjects were able to discriminate category-members from non-members, proportion of yes-responses are significantly above chance-level for the low-distortions (distortion levels 2, 4 and 5) and the prototype. And subjects responded less with 'yes' to high-distortions (distortion-levels 7, 8 and 10) and the random dot-pattern. However, patterns of the distortion-level 7, elicited yes-responses in half the trials, indicating that this distortion-level is ambiguous, that is, it is close to the category boundary, and thus subjects are not sure whether it belongs to the category or not.

Similarly, Figure 20 shows the proportion of yes-responses during the identification task for each distortion level. Here, all distortion levels exhibited responses significantly different from chance-level. However, only the proportion of yes-responses for the prototype is above chance-level, all others are below. This is the expected result, since subjects were only supposed the respond with 'yes' to the prototype. The proportion of yes-responses decreases with increasing distortion level, indicating that subjects made less errors the less the pattern resembled the prototype.

Another way to display this behavior, is to plot the proportion of correct responses as a function of the distortion-level (Figure 21). In this figure, one can see that the performance is lowest in the two tasks around the category boundary. For the identification task, the boundary is at the prototype and for the categorization task at the distortion-level 6. The performance is lowest for patterns with a distortion level near the category boundary, and increases with higher and lower distortion levels.

In the next step, the change of the performance over the different sessions was analyzed. Figure 22 displays the performances of the two tasks as a function of the session number. Subjects' accuracy increased during the first 4 to 5 sessions then stays constant at about 73% in the categorization task
Figure 21: Performance as a function of the distortion level
Mean performance, averaged across all subjects, of the categorization (black) and the identification (red) task as a function of the distortion level
Figure 22: Performance as a function of time
Mean performance, averaged across all subjects, of the categorization (black) and the identification (red) task, as a function of session number
Figure 23: Reaction times as a function of time
Mean response latencies, averaged across all subjects, of the categorization (black) and the identification (red) task, as a function of session number
and at 87% in the identification task. Changes of performances over time were identical in the two tasks. Further analysis of the performances of the blocks within sessions (not shown here) revealed a high increase in accuracy within the first two sessions, and a much lower increase after that.

The reaction time as a function of time is shown in Figure 23. Initially identification precedes categorization. The difference in the first session is significant and the reaction times were 776 ± 63 ms in the categorization task and 608 ± 58 ms in the identification task. However, the difference of the reaction times is not significant after eight sessions, the reaction times converge at about 500ms. A three-way ANOVA showed that the three main effects are significant for both the performance and the reaction times (p < 0.01), indicating that performance and reaction time are modulated by the task type, the distortion-level and the number of session. It also revealed a significant interaction between the task and the distortion-level for both performance and reaction time (p < 0.0001). This means that performance and reaction times are different for a given distortion-level for the two tasks. The performance data showed additionally a significant interaction between distortion-level and session number (p < 0.01), implying that the performance developed differently for the different distortion-levels. Finally, a strong interaction between task and session for the reaction time data (p < 0.0001) was found, meaning that the reaction times develop differently in the two tasks over sessions.

Since the performances of the two tasks are significantly different from each other throughout the experiment, the reaction times as a function of the performance is displayed, as a control for the possibility that the performance modulates the reaction times more than the task. Performances and reaction times of all the distortion-levels of all blocks from all subjects and sessions were calculated separately, and reaction times of trials exhibiting a similar performance were pooled together. The results of these calculations are shown in Figure 24. The reaction times in the categorization task are
Figure 24: Reaction time as a function of performance

Mean response latencies, averaged across all subjects, of categorization (black) and the identification (red) task, as a function of performance.

constantly higher than the ones in the identification task. However, the difference of these reaction times is not dependent on the performance, ruling out the possibility that the different developments of the reaction times is due to the different difficulties.

As mentioned in the 'Data analysis' section, an index was calculated to control if the shift of the dot-pattern around the center of the screen had an impact on performance or reaction time. Figure 25 presents these indexes for the performance (left) and the reaction time (right). On the y-axis, the index of the comparison between stimuli presented in the upper half of the screen and the lower half of the screen is displayed. The x-axis represents the comparison between the right and the left half of the screen. If the index of
a given data point on the y-axis is above zero, the subject performed better, or his reaction time was higher respectively, for dot-patterns presented in the upper half of the screen, else his performance and reaction times were higher for patterns presented on the lower half of the screen. The meaning of the index on the x-axis is analogue: positive, the subject had a higher performance, or reaction times respectively, for patterns presented on the right side of the screen. The black stars represent the indexes of single subjects for the categorization task, and the blue crosses represent indexes of single subjects for the identification task. The big red and green crosses represent the mean and the standard deviation of the two tasks across all subjects. All four mean indexes are not significantly different from zero, indicating that the shift of the stimulus around the center of the screen had no impact on performance or reaction time.

10 Conclusions

I used a psychophysical approach to assess the relationship between identification and categorization of visual patterns. Accuracy and reaction times in categorization and identification tasks were measured. Ten subjects participated in this study and performed 10 sessions in each task. The behavioral results show that the subjects did learn the category and the prototype, respectively. In the categorization task low-level distortions exhibited a significantly higher proportion of yes-responses than chance, whereas high-level distortions display a lower proportion of yes-responses. In the identification task only the prototype evoked a higher proportion of yes-responses than chance. Although performances of the two tasks are different, they equally increase over time. Performances increase with practice over the first five sessions then reach a plateau with a constant difference between categorization and identification. The reaction times of the two tasks are initially different, identification precedes categorization. With practice subjects become
Figure 25: Indexes

Mean index values of the categorization (big red cross) and the identification (big green cross) task for performance (left) and reaction times (right). The extent of these crosses show standard deviation of the mean. The black pluses, and blue crosses show the index values of single subjects. Details are described in the section ’Data analysis’.
faster in both tasks, however, improvement in categorization is better than in the identification task. After about eight sessions the difference of the reaction times is not significant anymore, they converge around 500ms. A 3-way ANOVA also shows that the reaction times develop differently with practice, whereas the performances develop parallel. This can be seen in the highly significant interaction between tasks and sessions. I would expect the reaction times of the categorization task to decrease further to the level of those of the identification task, or even below. Showing that the reaction times of the categorization task would indeed decrease further than the reaction times of the identification task would require an extended version of this experiment. This reveals some practical problems, since subjects would have to come every two to three days for about twenty sessions, which only few subjects would be able to perform. A non-significant interaction between tasks and sessions would imply that the reaction times would develop parallel in the two tasks over the sessions, which would lead to the conclusion that these processes are mediated by the same neural substrate.

Measured reaction times of a categorization task in other studies [117, 116, 165, 166] were short, so that categorization is as fast as identification or even faster, whereas in this study an initial longer latency for categorization was found. I used completely artificial categories, which the subjects have never seen before, and in the mentioned literature they used natural stimuli, such as faces, animals or houses. These natural stimuli have been learned extensively by the subjects throughout their life, so that they were already experts in detecting these categories.

One might also argue that the difference in the reaction times is due to the different difficulties. Reaction times of all distortion levels were sorted, which exhibited the same performance in the two tasks and looked at the reaction times as a function of performance (Figure 26). The difference of the reaction times of the two tasks is independent of performance.

The difference in the development of the reaction times found in this study
cannot be accounted to the different difficulties of the tasks, because the performances develop equally over time and the reaction times are not dependent on performance. One might conclude from these results that identification and categorization are two distinct processes. However, these conclusions are derived mainly from the significant interaction of a 3-way ANOVA between task type and session number, and the independence of the reaction times from the performance. The different development of the reaction times is not readily visible in the figure. If we this interaction between task type and session number would not be significant, one might easily conclude from these results that identification and categorization are one and the same process.

Alternatively, a possible conclusion from these results might be that identification and categorization are mediated by the same neural network, but have different learning rate, and thus would show identical behaviour as observed in this experiment. In this conclusion identification would be a special case for categorization, a category with only one element.

Future imaging studies will have to reveal the neural network(s) for these two processes.
Part III

The effect of category number on rule-based and information-integration category learning

11 Introduction

An increasing number of studies suggest that the learning of different types of category structures is mediated by different categorization and memory systems [171, 172, 173]. Most multiple systems theorists agree that one system is explicit and one is implicit. For example, the model Competition between Verbal and Implicit Systems (COVIS) assumes an explicit, hypothesis-testing system, and an implicit, procedural-learning system [174]. The explicit system is assumed to dominate the learning of rule-based categorization learning tasks, whereas the implicit system dominates learning of information-integration learning tasks.

In rule-based category learning tasks, category structures can be learned via some explicit reasoning process [174]. The optimal rule, i.e. the rule that maximizes performance, is usually easy to describe verbally. In many applications, only one stimulus dimension is relevant, and the subject’s task is to discover this relevant dimension and then to assign the stimulus to the correct category. Sometimes two dimensions are relevant and the subject has to learn a decision criterion along each dimension and then assign the dimensional values to a category. The decisions along each dimension is then integrated to determine category membership. Since a decision is first made about the value of each dimension, this integration is post-decisional and generates the response.
In information-integration tasks, on the other hand, accuracy is maximized only if information from two or more dimensions is integrated at some predecisional stage [175]. In most cases, the optimal rule in information-integration tasks is difficult or even impossible to describe verbally.

Empirical evidence for multiple category learning systems comes from several different studies. For example, Ashby, Queller and Berretty [176] conducted an experiment in which subjects had to learn categories unsupervised, i.e. without a trial-by-trial feedback. Subjects were able to learn simple uni-dimensional rules, whereas no learning effect was seen with more complex two-dimensional rules. It is well known that with supervision, subjects can learn complex rules that are difficult or impossible to verbalize [177, 178]. However in the absence of a trial-by-trial feedback, subjects almost always use uni-dimensional rules [179, 176].

11.1 Overview of the study

In the last chapter I investigated the processes categorization and identification. In this chapter I examined different kinds of categorization processes. Does the brain process different category structures differently? Or is there a universal categorization process? The goal of the present study was to test whether different numbers of categories have the same impact on rule-based and information-integration category learning. Different effects in the two conditions would be a support for the multiple system theory. There is only one study, which systematically manipulated the numbers of categories while holding the nature of the category structures fixed [3].

Maddox et al. used line segments with varying length and orientation. They divided the stimulus space in two or four categories. In the rule-based case the category boundaries were parallel to the dimensions (see Figure 27), so that the rule is easily described verbally. Short lines belong to one category, long lines to the other. In the information-integration condition, the category boundary was oblique to the dimensions, so that the rule is not so easy to
describe. They found that performance in the information-integration task was independent of the number of categories, whereas in the rule-based task, the performance decreased with increased number of categories (see Figure 28).

In this study I used Glass patterns. In the information-integration condition I gradually change the form of the Glass pattern from radial to circular and divide this stimulus space into two or four categories. These category rules are hard to describe verbally.

In the rule-based condition, I used only circular Glass patterns, but varied the dot-separation continuously. These category rules are easy to verbalize, namely that small dot-separations are assigned to category A and large dot-separations belong to category B, in the two-category case.

I also asked the subjects to respond as fast as possible, so that meaningful reaction times could be measured. Additionally I let the subjects perform the task several times with an interval of about two days, to investigate whether the changes over time of the performances and reaction times are similar in the four tasks or different. Differences in the development of performances and reaction times would be further evidence for a multiple category learning system.

12 Methods

12.1 Participants

Twenty healthy students (four women, mean age was 27 years) from the University of Zurich participated in this study. All subjects had normal or corrected-to-normal vision. The subjects were randomly assigned to one of the four groups.
Figure 26: Stimulus space from Maddox’s experiment [3]

Scatter plot in the length and orientation space for the categorization stimuli from each of the four conditions. For the two-category conditions, the filled squares denote stimuli from Category A, and the open triangles denote stimuli from Category B. For the four-category conditions, the open squares denote stimuli from Category A, the filled squares denote stimuli from Category B, the open triangles denote stimuli from Category C, and the filled triangles denote stimuli from Category D. The dashed lines in each plot denote the optimal decision bound(s). Reprinted with permission.
Figure 27: Results from Maddox’s experiment [3]
Proportion correct for the two- and four-category/rule-based and information-integration category structures averaged over blocks (standard error bars included). RB = rule-based; II = information-integration. Reprinted with permission.
12.2 Stimuli

I used Glass patterns [180] as stimuli. A Glass pattern is generated by placing dots in random locations in a certain area. This dot-pattern is the transformed, by translation, rotation, expansion or similar operations, and superimposed on itself. Since all dot-pairs are now aligned to a certain pattern, a perception of a form arises. In the case of a rotational operation, a circular pattern is perceived. In practice, the Glass patterns are created by calculating the orientation of the dot-pairs, which is a function of their location.

I generated two sets of stimuli, one for the rule-based task and one for the information-integration task. The latter set consisted of Glass patterns with a constant dot-separation and orientations of the dot-pairs that smoothly manipulated the perception of the pattern from radial over spiral to circular. I divided then the continuum of patterns into two or four categories. In the upper part of Figure 29, an example from each of the categories is shown. In the case of two categories the first two and the last two belong to the same category.

The set of stimuli for the rule-based task consisted of circular-shaped Glass patterns with variable dot-separations. The dot-separation varied smoothly between the two extremes. This continuum is the divided into two or four categories. The lower part of Figure 29 shows an example from each of the categories. In the case of two categories the first two and the last two belong to the same category.

The size of the pattern was 10° of visual angle, and the dot-sizes were one pixel (≈3'). In each pattern there were 180 dot-pairs. The dot-separation in the information-integration task was 24’ of visual angle and in the rule-based task, it varied between 8’ and 40’.
Figure 28: Examples of stimuli
12.3 Procedure

Two tasks were defined, the information-integration task and the rule-based task, with either two or four categories. In the information-integration task, subjects categorized Glass patterns according to their form. In the two-category case the categories were either radial or circular, whereas in the four-category case, subjects also had to distinguish the intermediate types, namely the clockwise spiral and the anti-clockwise spiral form.

During the rule-based task, subjects categorized the Glass pattern according to the dot-separation. In the case of the task with two categories, one category consisted of Glass patterns with short dot-separations and one consisted of Glass patterns with large dot-separations. The range of dot-separations was divided into four categories in the four-category case, and subjects had to distinguish the four categories.

The procedure was the same for all four groups. Subjects first went through a training session and in seven subsequent sessions performed the assigned task.

During the training, subjects were presented with examples from each of the categories they were assigned to. First the category number was displayed for five seconds, then six examples of the corresponding category were subsequently presented for five seconds each. Having presented all assigned categories, this procedure was repeated three times, so that subjects saw 18 examples of each category. This repetition is to ensure that the subject has a basic knowledge about the category.

During the subsequent seven task sessions, subjects categorized the Glass patterns into the learned category. Each session consisted of 100 trials, 25 stimuli for each category. The order of the trials was randomized.

In each trial a fixation point first appeared on the screen for one to two seconds (randomized). Then the pattern is shown for 200 ms, after which the screen stayed blank until the next trial. The subject was asked to indicate to which category the current pattern belongs to as fast and accurate as possible.
Subjects gave their responses by pressing one of two, or four respectively, buttons. In case of an incorrect response, an auditory feedback was given. The task sessions followed each other in an interval of two days on average.

Reaction times and accuracy were measured.

12.4 Equipment

I used a VisaGe system (Cambridge Research Systems, UK) with a screen (Sony, GDM-F520) with a refresh rate of 120 Hz. Connected to the VisaGe system was a response box (CB6, Cambridge Systems, UK). The stimulus generation and the tasks were programmed on Matlab 7.1 using the CRS-software toolbox. Subjects placed their heads on a chin rest to ensure that the distance to the screen was constantly 50 cm.

13 Results

13.1 Accuracy

All subjects were able to discriminate the two and four categories. Their performance was significantly above chance-level.

Figure 29A shows the accuracy of the information-integration task with two and four categories as a function of the session number. The error bar in this and subsequent graphs indicate standard error. Initially, the performance in this task was lower (86 %) in the case of four categories than in the case of two categories (78 %), although this difference was not significant. The proportion of correct responses changed significantly with practice, the difference in performance of the first and the last session was significant (p<0.05) in both cases. The two performance curves were virtually identical.

The accuracy results for the rule-based task are shown in Figure 29B. Apparent in this graph is that the performance of the four-category condition
was constantly and significantly lower than in the two-category condition. The performance in the two-category condition showed a slight increase with practice. Whereas the difference between the first and the last session was not significant, the difference between the second and the last session was (p<0.05). Also, the performance of the last four sessions of the two-category case was not significantly different from the performance of the last sessions of the information-integration task. The proportion of correct responses in the four-category condition was in all sessions significantly lower than in the two-category condition, and it didn’t increase with practice. The performance of the first and the last session, but also of the second and last session, was not significantly different.

A three-way analysis of variance (ANOVA) was conducted on the performance data, with the variables: task type (information-integration or rule-based), category number (two or four) and session (seven sessions). All three main effects were significant (F(1,111)=196 for the task type; F(1,111)=179 for the category number; F(6,111)=7 for the session number; p<0.0001 for all three main effects). The task type and category number interaction was significant (F(1,111)=148, p<0.0001), whereas the other two interactions and the three-way interaction were not (F(6,111)<1.8, p>0.8).

The most important finding from the preceding analysis was the task type by category number interaction, which is shown in figure 31. It shows the performance of all four groups, averaged over all sessions and subjects. To determine the locus of this interaction a number of t-tests on the performance data collapsed across sessions have been conducted. First, T-tests comparing two versus four categories performance were conducted separately for the rule-based and information-integration conditions. In the information-integration condition the two performances were statistically equivalent (90 % for two-category and 88 % for the four-category condition), whereas in the rule-based condition the performance for two categories was significantly better (88 %) than for four categories (63 %, p<0.0001).
Second, T-tests comparing rule-based versus information-integration category performances were conducted separately for the two- and four-category cases. In the two-category condition the performance in rule-based and the information-integration task were not significantly different. However, in the four-category case the performance in information-integration was significantly higher (p<0.0001).

13.2 Reaction times

The reaction times were also analyzed. The data is shown in Figure 29C for the information-integration cases and in Figure 29D for the rule-based cases. In the analysis, as well as in the graph, only reaction times of correct responses were considered.

In the information-integration condition the reaction times of the two-category case stayed constant. Although there seemed to be a slight trend towards faster response times, the difference between the first and the last session was not significant. The response times in the four-category case were generally slightly higher than those of the two-category case. During the first session, the reaction times in the two-category case were higher than in subsequent sessions.

Reaction times of the two- and four-category conditions in the rule-based task were over all sessions significantly different (p<0.0001). In the two-category condition, the response times decreased with increasing practice, the difference in reaction times of the first and the last session was significant (p<0.05). Whereas in the four-category condition, the reaction times stayed constant over all sessions, first and last session were statistically equivalent.

As for the behavioral data, a three-way ANOVA on the reaction times was conducted. The three variables were again: task type, category number and session. All three main effects were significant (F(1,111)=5, p<0.05 for the task type; F(1,111)=110, p<0.0001 for the category number; F(6,111)=3.6, p<0.05 for the session number). The task type and category number inter-
Figure 29: Performances and reaction times
action is significant ($F(1,111) = 157, p<0.0001$), whereas the other two interactions and the three-way interaction are not ($F(6,111) < 2, p>0.2$).

Figure 32 shows the reaction times of all four groups, averaged across all sessions and subjects. A series of t-tests on this data was performed. First, t-test comparing two versus four categories reaction times separately for the rule-based and information-integration conditions were conducted. In the information-integration condition the reaction times of the four-category condition were slightly higher (698 ms for the two-category and 926 ms for the four-category condition, $p<0.01$), whereas in the rule-based condition the reaction times for two categories were significantly better (692 ms) than for four categories (1058 ms, $p<0.0001$).

Second, t-tests comparing rule-based versus information-integration category reaction times separately for the two- and four-category cases were calculated. In the two-category condition the reaction times in rule-based and the information-integration task were not significantly different. In the four-category case the reaction times in information-integration were slightly lower, but the difference was not significant.

14 Discussion

The results from this experiment indicate an interaction between the number of categories and the nature of the category structure, rule-based or information-integration, on category learning. While no performance difference was observed across information-integration two-category and four-category conditions, there was a large performance difference observed across the rule-based two-category and four-category conditions. The performance in the rule-based two-category condition was about 25 % better than in the rule-based four-category condition (see Figure 30).

This experiment revealed a similar pattern of results as W.T. Maddox et al. observed in their study [3]. They reported no accuracy difference in
Figure 30: Proportion correct averaged over all sessions
Figure 31: Reaction time averaged over all sessions
the two information-integration conditions, but a significant lower accuracy for the rule-based four-category condition compared to the rule-based two-category condition. While they used line segments with varying length and orientation, Glass patterns have been used in this study. The Glass patterns have the advantage that the stimuli always have the same physical properties, the density of dots stays constant across stimuli. Thus ruling out that physical properties were the cause of this pattern of result. And since the two two-category conditions exhibited the same performance, the difference in accuracy in the rule-based conditions is not based on huge difference in difficulty of the tasks. The category structures, information-integration and rule-based, are processed differently in the brain.

The reaction time data generally supported the conclusions derived from the performance data, and ruled out possible speed-accuracy trade-off explanations. The response times for the two information-integration conditions and the rule-based two-category condition were not significantly different from each other, while the response times for the rule-based four-category condition were significantly higher. Thus, response times increased as performance decreased. This increase in reaction times in the rule-based four-category condition mimicked the performance results.

Overall, these findings suggest that for the rule-based categories increasing category number places a greater demand on executive processes and working memory, and such demands result in performance decreases that can be seen in both accuracy rates and reaction times.

Generally, subjects don’t improve their performance with practice. In the information-integration two-category condition, the performance in the first session is 86 % and in the last session 91 %, this difference is not significant. A similar small and not significant improvement can be observed during the first three sessions in the four-category condition. Similarly, a small and insignificant increase during the first three sessions of the rule-based two-category condition can be observed, while the performance in the four-
category condition stays mainly constant at about 63%. Subjects don’t reach perfect performance even in the two-category conditions, because stimuli near to the category boundary are perceived ambiguously. However, improvement of task behavior in the reaction time data (Figure 29 C+D) was observed, where a general trend towards faster responses with increasing practice was observed.

Several studies from literature provide further evidence that rule-based and information-integration category structures are mediated by different neural circuits. In particular, Ashby and his colleagues have proposed that with rule-based structures, category learning is mediated by a circuit that includes the anterior cingulate, the prefrontal cortex, and the head of the caudate nucleus, whereas in information-integration tasks, learning is mediated largely within the tail of the caudate nucleus for visual stimuli [174, 171, 181, 182].

There are also studies showing that the nature and timing of the feedback was critical for efficient information-integration category learning. For example Ashby et al. [176] showed that information-integration structures with two categories cannot be learned without trial-by-trial feedback, whereas unidimensional rule-based structures with two categories can. Similarly, Ashby et al. [183] and Maddox et al. [184] showed that timing of the feedback was important for information-integration, but not for rule-based category learning. Ashby and Maddox investigated the effects of two different kinds of categorization training. In observational training, observers are presented with a category label and then shown an exemplar from that category. In feedback training, they are shown an exemplar, asked to assign it to a category, and then given feedback about accuracy of their response. Rule-based categorization learning is not affected by the type of training, whereas in information-integration categories, the performance was significantly lower when no feedback was presented. Furthermore, responses were less accurate when the feedback was delayed in the information-integration category.
learning, but not in rule-based category learning.

Other studies suggested that rule-based category learning, but not information-integration category learning, is adversely affected by tasks known to recruit frontal structures, such as the Stroop task [118] and working memory [184].

These results suggest that the mechanism that mediates learning-related changes in the cortex are qualitatively different, and such differences indicate that category number may have different effects on rule-based and information-integration tasks.

The rule-based category learning system is under conscious control and has full access to working memory and executive attention, and accurate rule-based category learning depends upon good use of these processes. As the complexity of the rules increases, the demand on working memory and attention resources increases, and performance decreases. The information-integration system, on the other hand, is not under conscious control, and instead learns through a gradual incremental learning process.

This present study gives further evidence that distinct category-learning processes exist. The results indicate that rule-based and information-integration category learning is mediated by separate neural systems.

In chapter 2 I showed that identification and categorization are two distinct processes. The experiment described in chapter 3 revealed that there are distinguishable categorization processes in the brain. In the next chapter I describe an experiment in which subjects had to perform a memory task, they had to identify previously seen paintings. We tested whether recollection and familiarity judgments were influenced by the degree of visual similarity between old and new paintings, and whether recollection and familiarity are mediated by the same identification process.
Part IV

Recollection- and
Familiarity-based Decisions
Reflect Memory Strength

15 Introduction

Results from cognitive, neuropsychological, and neuroimaging studies of human memory suggest that recognition memory performance reflects two distinct memory processes or types of memory, often referred to as recollection and familiarity [139]. Recollection is defined as a retrieval of qualitative information about a study event, such as where or when an item was encountered. Whereas familiarity is a feeling of having encountered the item, but no recollection of where or when the item was encountered occurs. Human neuroimaging studies have indicated that prefrontal, parietal and medial temporal cortices are involved in recognition memory of prior episodes, but the functional roles that these regions play in recognition are widely debated [185].

The dissociation between recollection and familiarity have been first proposed by studies in which amnesic patients showed a more severely disrupted recollection than familiarity. For example, amnesic patients exhibit significantly greater memory impairments on associative than item recognition tests, indicating that the regions damaged in amnesia are more important for the former than latter type of recognition judgments [140, 186]. Moreover, receiver operating characteristic (ROC) studies in amnesics have indicated that only one process (i.e., familiarity) is needed to account for their recognition performance [187], as expected if they exhibited a severe deficit in
recollection.

Knowlton and Squire [188] found that amnesics with damage to the hippocampal formation or diencephalon showed reduced levels of both Remember and Know judgments, whereas Schacter et al. [189] found that amnesics showed reduced levels of Remember but not Know judgments.

Further dissociation have been identified by previous neuroimaging studies, they found regions related to recollection and familiarity, either by contrasting recognition responses accompanied by reports of conscious remembering to those accompanied by reports of familiarity in the absence of recollection [190] or by contrasting recognized items for which subjects can correctly recognize where or when the item was studied (i.e., accurate source recognition) to those for which they fail to recollect source information [191, 192]. Two fMRI studies contrasted the activity elicited by test words subjected to remember/know judgments [149, 102]. In both cases greater activity was found in left lateral parietal cortex (Brodmann Areas 39/40) for old items endorsed as remembered rather than known. A second finding, reported only by Eldridge [102], was of greater activity in the hippocampus and adjacent MTL cortex for items endorsed as remembered than for items classified as known or new. In a third study, subjects had to recognize new and previously studied words, and additionally give their confidence of their response [193]. Higher activation for correct recognition with low confidence than correct recognition with high confidence was found in the right dorsolateral prefrontal region (inferior frontal sulcus, BA 9/46).

Dissociations in activation between familiar and novel items have been reported in numerous studies (for a review see [194]). In the study of Ishai and Yago [195], caucasian subjects had to memorize asian faces and four days later performed a memory retrieval task in the MR scanner. The familiar faces were recognized faster and more accurately, and elicited stronger activation in face-selective regions in the visual cortex, parietal and prefrontal regions, and the hippocampus. Tulving and colleagues conducted a study [196] in
which subjects, while undergoing PET scanning, viewed previously studied pictures in one type of scan, and previously not seen pictures in another. They observed activation for old items in the left and right frontal areas and posterior regions, whereas novelty activation was found in the right limbic regions, and bilaterally in temporal and parietal regions.

Cansino and colleagues found greater activation for recollected pictures than for familiar pictures in the right hippocampal formation and the left prefrontal cortex [197]. In a series of fMRI experiments, Henson and colleagues found reduced activation to old items compared to new items in anterior medial temporal lobe (MTL) encompassing perirhinal cortex and the anterior part of Brodmann area 35 [198].

Several event-related fMRI studies have explored whether parietal activation varies according to whether recognition is accompanied by recollection or familiarity. Wheeler et al.,[199] found that both recollection and familiarity were associated with common processes, but recollection additionally recruits regions specific to retrieved content. In a memory retrieval task with words, they found higher activation for correct remember and know responses compared to new responses in the left parietal cortex, near intraparietal sulcus (IPS). Whereas the left inferior temporal cortex selectively associates with remembering.

In a study by Yago and Ishai [4], subjects were instructed to learn and memorize paintings, four days later they performed an old-new recognition memory task in the MR scanner. Old paintings were randomly presented with new exemplars. The new exemplars were either visually similar to the prototypes, ambiguous, i.e. somewhat similar, or dissimilar (Figure 32). In attention-related regions in the parietal cortex, namely intraparietal sulcus (IPS) and superior parietal lobule (SPL), decreased activation with decreased visual similarity between the prototypes and the novel exemplars was observed (Figure 33). In prefrontal areas, old paintings evoked stronger activation than the new ones, regardless of their visual similarity. Whereas
in the hippocampus and precuneus, stronger responses were evoked by the new, visually different paintings. They concluded that recognition memory is mediated by classification of new items as a match or a mismatch, based on their degree of visual similarity to old items. This old-new task, however, did not address the issue of memory processes and the extent to which the observed behavioral and neural responses were due to recollection- or familiarity-based memory decisions.

A number of models have been proposed that assume that recognition memory judgments can be based on two distinct forms of more. Several dual-process models have been developed with the primary purpose of accounting for the memory deficits observed in patients with focal brain lesions. A number of models assume that medial temporal lobe damage (e.g., hippocampus and the surrounding temporal lobe including the parahippocampal gyrus) results in a form of amnesia that disrupts recollection but leaves familiarity intact. For example, Huppert and Piercy [200, 201] argue that healthy subjects can base recognition memory judgments on assessments of item familiarity or on the basis of retrieving contextual information about study events, whereas amnesics are limited primarily to assessments of familiarity. In more recent models, it is suggested that the hippocampal region is critical for recollection, whereas surrounding structures in the medial and inferior temporal lobe (e.g. parahippocampal gyrus) are important for familiarity [202].

Jacoby and colleagues [192, 203] have argued that recognition memory judgments can be based on the assessment of processing fluency (i.e. familiarity) or on the recovery of the context and elaboration given to an item when it was initially studied (i.e. recollection). A fundamental distinction between the two processes is assumed to be that recollection reflects an analytic, consciously controlled process, whereas familiarity is a relatively automatic process.

Yonelinas and colleagues [204, 187, 205] have argued that recollection
and familiarity differ in terms of the type of information that they provide and in the extent to which each process influences recognition confidence. Familiarity is assumed to reflect the assessment of 'quantitative' memory strength information in a manner similar to that described by signal detection theory. In contrast, recollection reflects a threshold retrieval process whereby 'qualitative' information about a previous event is retrieved. That is, subjects may retrieve many different aspects of a study event (e.g., where or when the event occurred), but, for some items, subjects are unable to retrieve any accurate qualitative information about the study event (i.e., some items fall below the recollective threshold). For these items subjects are expected to rely on assessments of familiarity.

To test whether recollection and familiarity judgments were influenced by the degree of visual similarity between old and new pictures, we used event-related fMRI with portraits, landscapes and abstract compositions by painters with a unique style. Based on the previous studies by Yago and Ishai [4, 195], we predicted that subjects would correctly reject more new paintings that were visually different from the old ones, and would make more false alarms to new, visually similar paintings. Furthermore, we hypothesized that if recollection and familiarity judgments about complex pictures reflect strong memories and weak memories, respectively, correctly remembered items would be associated with stronger neural activation than known items [206].

16 Behavioral pilot study

16.1 Methods

Fourteen (7 females, mean age 25 years) healthy, right-handed subjects with normal or corrected to normal vision participated in the behavioral pilot study. Stimuli were presented on a monitor using Presentation (www.neurobs.com, version 10.3). During the encoding session, subjects were instructed to learn
Figure 32: Stimuli used in the experiment by Yago and Ishai [4]
Subjects memorized prototypes of portraits, landscapes and abstract paintings taken from six painters (shown from top to bottom are examples from Modigliani, Van Gogh and Miro). Four days later, in the memory retrieval session, the familiar prototypes were mixed with new exemplars that were visually similar, ambiguous or dissimilar, and subjects indicated whether they had seen each picture before. The degree of visual similarity between the prototypes and the new exemplars was assessed in a separate behavioral pilot (see Methods in [4]).
Figure 33: Activation in parietal cortex during memory retrieval (from [4])
Coronal sections, taken from two individuals, illustrating stronger activation evoked by the prototypes and similar exemplars than by ambiguous and dissimilar exemplars in the IPS (top; p < 0.0001) and the SPL (bottom; p < 0.005). Mean parameter estimates for all stimulus types were averaged across all 6 runs (portraits, landscapes and abstract paintings) and all subjects.
and memorize paintings (portraits by Modigliani and Renoir, landscapes by Pissarro and Van Gogh, abstract paintings by Kandinsky and Miro, Figure 32). Each run included 30 old pictures from one category, 15 from each painter, and the order of the paintings as well as the order of the three runs was randomized across subjects. Each picture was presented in the center of the screen for 5 s. Three days later, subjects performed a memory retrieval task. The 15 familiar paintings from each painter were presented randomly with 45 new exemplars. The new pictures were either visually similar to the old ones, ambiguous, or dissimilar. (Figure 32) Each painting was presented for 3 s and subjects pressed a response button on the keyboard to indicate whether they remember the painting, know it (i.e. the painting looks familiar) or whether it is new.

16.2 Results

Figure 34 shows the results of the behavioral pilot study. On the left, the percentages of remember, know and new responses for the different stimulus types (old, new similar, new ambiguous and new different) are plotted. In accordance with our prediction, the proportion of false alarms, namely remember and know responses to new paintings, was reduced with decreased similarity. Moreover, most false alarms were due to feelings of familiarity. Subjects were faster and more accurate to reject new, visually different items than new visually similar ones. The reaction times (Figure 34, right) of new responses decrease with decreased similarity. Know responses were associated with longer latencies than remember or new responses, suggesting that subjects hesitated to make a familiarity-judgment. Latencies for remember responses remained virtually identical for all stimulus types.
Figure 34: Behavioral data from the pilot study
Mean responses, averaged across all subjects, to old, new similar, new ambiguous and new different paintings (left) and mean reaction times for the different responses (right). The error bars represent the standard error and asterisks show significant differences (p < 0.05).
17 Imaging study

17.1 Methods

17.1.1 Subjects

Twenty-one healthy, right-handed subjects with normal or corrected to normal vision participated in the study. The 21 subjects (11 females) ranged in age from 20 to 29 years (mean 25 years). All subjects gave informed written consent for the procedure in accordance with protocols approved by the University Hospital Zurich and were paid for their participation. All subjects were unfamiliar with the stimuli and reported visiting art museums once a year or less.

17.1.2 Stimuli and tasks

Stimuli were displayed using Presentation (www.neurobs.com, version 10.3) and were projected with a magnetically shielded LCD video projector onto a translucent screen placed at the feet of the subject. During the encoding session, subjects were instructed to indicate whether they see any flowers in each painting (portraits by Modigliani and Renoir, landscapes by Pissarro and Van Gogh, abstract paintings by Kandinsky and Miro, Figure 32). Each run included 20 old paintings from one category, 10 from each painter, and the order of the paintings as well as the order of the three runs was randomized across subjects. Each picture was presented in the center of the subject's field of view for 3 s.

After a 10-min delay, subjects performed a surprise memory retrieval task. In an event-related design, the 20 familiar paintings from each category were randomly presented with the same number of new exemplars in three runs, one for each category. The new pictures were either visually similar to the old

---

ones or visually dissimilar. Each picture was presented for 3 s and subjects had to make a Remember / Know / New decision by pressing one of three buttons. Subjects were instructed to respond 'Remember' if they were sure that the painting was presented during the flower detection task, 'Know' if the picture seemed familiar, but they were not sure if it was presented during the study phase and 'New' if the picture was not presented in the flower detection task.

In each run five paintings alternated with 15 s of fixation, which were used as a visual baseline.

The degree of visual similarity between the new exemplars and the old ones was assessed in a separate pilot study [4]. In this study, seven subjects rated the new pictures as visually similar, ambiguous, or dissimilar using a scale (0 = Dissimilar, 1 = Somewhat similar, 2 = Similar, 3 = Very similar). The mean ratings were: similar portraits 2.29 ± 0.5; similar landscapes: 2.2 ± 0.5; similar abstract paintings: 2.1 ± 0.53; ambiguous portraits: 1.1 ± 0.65; ambiguous landscapes: 1.4 ± 0.8; ambiguous abstract paintings: 1.3 ± 0.6; dissimilar portraits: 0.27 ± 0.4; dissimilar landscapes: 0.47 ± 0.46; dissimilar abstract paintings: 0.26 ± 0.23.

17.1.3 Data acquisition

Data were collected using a 3T Phillips Intera whole body MR scanner (Phillips Medical Systems, Best, The Netherlands). Changes in blood-oxygenation level-dependent MRI signal were measured by using sensitivity encoded gradient-echo echoplanar sequence (SENSE, [207]) with 35 axial slices, TR = 3 s, TE = 35 ms, flip angle = 82°, field of view = 220 mm, acquisition matrix = 80 x 80, reconstructed voxel size = 1.72 x 1.72 x 4 mm, and SENSE acceleration factor R = 2. High-resolution spoiled gradient recalled echo structural images were collected in the same session for all the subjects between study and test phase (180 axial slices, TR = 20ms, TE = 2.3ms, field of view = 220 mm, acquisition matrix = 224x 224, reconstructed voxel size = 0.9 x 0.9 x 0.75
These high-resolution anatomical images provided detailed anatomical information for the region-of-interest (ROI) analysis and were used for 3D normalization to the Talairach and Tournoux atlas [208].

17.1.4 Data analysis

Accuracies and reaction times were computed for each subject, object category (portraits, landscapes, abstract paintings), stimulus type (Old (O), and the new, similar (S) and dissimilar (D) exemplars), and response type (remember, know, new). Pair-wise t-tests were used to compare between the proportions of the different responses, remember, know, new, and the reaction times of the responses.

Functional MRI data were analyzed in BrainVoyager QX Version 1.8 (Brain Innovation, Maastricht, The Netherlands). All volumes were realigned to the first volume, corrected for motion artefacts and spatially smoothed using a 5-mm full-width-half-maximum (FWHM) Gaussian filter. Face- and object-selective responses were analyzed using multiple regression [209] with the contrast paintings versus fixation. Based on this contrast, a set of ROIs was anatomically defined for each subject with clusters that showed a significant difference in activation (P < 0.01, uncorrected). These regions included bilateral face- and object-selective regions in the dorsal occipital cortex (DOC), inferior occipital gyrus (IOG), fusiform gyrus (FG), intraparietal sulcus (IPS), inferior frontal gyrus (IFG), anterior cingulate cortex (ACC), and the insula. To compare activation evoked by remember responses with activation evoked by know or new responses, a general linear model was constructed with the three response types (remember, know, new) as predictors. Based on these contrasts, a set of ROIs was anatomically defined for each subject with clusters that showed a significant effect (P < 0.05, uncorrected) which included the precuneus and the hippocampus. In each subject and each ROI, the mean parameter estimates were calculated separately for each stimulus type (O, S, D) and response (yes / no during the study phase and
remember / know / new during the test phase) in all correct trials.

17.2 Results

17.2.1 Study phase

Behavior In this phase, subjects were presented portraits, landscapes and abstract paintings and indicated whether each painting contained any flowers. In this flower detection task, most paintings did not include any flowers (Figure 35, left). The proportion of paintings including flowers was highest for the landscapes and lowest for abstract paintings. Surprisingly in more than 20% of the abstract paintings did the subjects detect flowers, although these paintings don’t depict any concrete items. Subjects were fastest to decide whether a painting contains any flowers for the portraits and significantly slower for abstract paintings ($P < 0.001$; Figure 35, right). While the latencies for yes- and no-responses for portraits and abstract paintings were the same within the category, the reaction times for no-responses for the landscapes is significantly higher than the reaction times for yes-responses ($P < 0.05$), suggesting that subjects were expecting landscapes to include flowers.

Imaging Data The main effect, namely all paintings versus fixation baseline, is shown in Figure 36. Activation was found in multiple regions of the visual, parietal and prefrontal cortex, including DOC (mean Talairach coordinates: $-28, -82, 12; 33, -82, 12$), IOG ($-24, -80, -3; -21, -80, -5$), FG ($-28, -53, -12; 28, -53, -10$), IPS ($-32, -51, 47; 29, -51, 44$), IFG ($-44, 3, 33; 43, 4, 34$) and ACC ($8, 7, 53$). In all these regions, activation evoked by yes-responses is virtually identical with the activation elicited by no-responses (Figure 36, bottom).
Figure 35: Behavioral data from the encoding session.
Mean responses (right) and reaction times (right) as a function of the category of the paintings.
Figure 36: Activation during encoding.
Activation evoked by portraits, landscapes and abstract paintings during the study phase. Top: coronal sections, taken from a representative subject, illustrating activation in DOC, IOG, FG, IPS, IFG and ACC. Bottom: Mean parameter estimates, averaged across all subjects and both hemispheres.
17.2.2 Test phase

Behavior  Ten minutes after performing the flower detection task, subjects performed a surprise memory test. In this test phase the paintings from the study phase were randomly presented with new paintings that were either visually similar or different from the old ones and subjects made Remember / Know / New decisions. The behavioral results of the pilot study were replicated during the memory retrieval task in the MR scanner (Figure 37). Again, the proportion of false alarms, namely Remember and Know responses to new paintings, was reduced with decreased similarity. Moreover, most
false alarms were due to feelings of familiarity. Subjects were faster and more accurate to reject new, visually different items than new visually similar ones (P < 0.00001). “Know” responses were associated with longer latencies than “Remember” or “New” responses.

**Imaging Data**  The main effect of paintings, namely responses evoked by all paintings as compared with the fixation baseline, revealed activation in a distributed cortical network that included regions in the visual, parietal and prefrontal cortex. In the visual cortex, bilateral activation was found in the DOC (mean Talairach coordinates: -27, -81, 11; 32, -81, 11), IOG (-29, -77, -1; 28, -77, -6) and FG (-28, -55, -15; 33, -55, -14). In the parietal cortex, activation was found in the IPS (-23, -54, 38; 27, -54, 39), and in the prefrontal cortex activation was revealed in the IFG (-44, 3, 33; 43, 4, 34), insula (-42, 8, 5; 39, 8, 2) and the ACC (2, 15, 40). As activation in the left and right hemispheres was very similar, the parameter estimates were averaged across both hemispheres.

In all the analyzed regions, the remember responses evoked stronger activation than know or new responses, except in the ACC, where the activation for the three different responses were virtually identical (Figure 41).

Contrasting recollection, namely correctly remembered items, with familiarity, namely correct know judgments, reveals predominantly activation in the precuneus, not only is the activation evoked by recollection higher than familiarity (p < 0.01), but also the hemodynamic response elicited by recollection peaks earlier (5.6 s) than the response evoked by familiarity (10.3 s, p < 0.004; Figure 41, left). A similar pattern of activation was revealed when contrasting correctly remembered pictures with correct rejections of new pictures. Activation in precuneus was stronger and peaked significantly earlier during remember (5.7 s) than new (10.2, p < 0.002; Figure 41, center) judgments. Whereas contrasting correct know responses with correct rejection of new paintings revealed similar hemodynamic response profiles and the peak
Figure 38: Activation as a function of task performance
Top: A coronal section, taken from a representative subject, illustrates activation in the FG, IPS, IFG and ACC. Bottom: Mean parameter estimates for correct responses to old items (HT); incorrect responses to old items (MS); correct responses to new items (CR) and incorrect responses to new items (FA). Parameter estimates were averaged across all subjects and both hemispheres.
Figure 39: Visual similarity effect

Top: A coronal section, taken from a representative subject, illustrates activation in the FG, IPS, IFG and ACC. Bottom: Mean parameter estimates for correctly recognized old and new items, as a function of their visual similarity. Parameter estimates were averaged across all subjects and both hemispheres.
Figure 40: Activation of Remember judgments.
Activation evoked by Remember judgments for old and new, visually similar paintings. Top: Coronal section, taken from a representative subject, illustrating activation in FG and IPS. Bottom: Mean parameter estimates, averaged across all subjects and both hemispheres.
Figure 41: Main effect during retrieval
Activation evoked by portraits, landscapes and abstract paintings during the test phase. Top: coronal sections, taken from a representative subject, illustrating activation in DOC, IOG, FG, IPS, IFG and ACC. Bottom: Mean parameter estimates, averaged across all subjects and both hemispheres.
Figure 42: Recollection vs Familiarity
Top: Group statistical maps illustrating significant activation in the pre-cuneus during the following contrasts: correct remember vs correct know (left); correct remember vs correct new (middle); and correct know vs correct new (right). Bottom: Analysis of the BOLD signal indicated that correctly remembered paintings evoked responses with higher amplitude and a significantly earlier peak than both known and new items. Data were averaged across all 21 subjects.
Figure 43: Activation in the medial temporal lobe.
Top: Group statistical map illustrating activation in the parahippocampal cortex (PHC) and the hippocampus (HIP). Bottom: Mean parameter estimates for correct Remember / Know / New responses, averaged across all subjects.
of the BOLD signal (11.1 s and 10.2 s, respectively; Figure 41, right) was not statistically significant.

Higher activation for correct remember judgments than correct know judgments in the medial temporal lobe was observed, namely in the parahippocampal cortex (PHC) and the hippocampus (Figure 43). In both regions, correctly remembered paintings evoked significant stronger activation than correctly known pictures (p < 0.01 in PHC and p < 0.02 in the hippocampus). The difference between remembered and new paintings and between known and new items, however, was not statistically significant.

Consistent with the previous study [4], in the IPS, a region implicated in many attention tasks, decreased activation with decreased visual similarity between the old and the novel exemplars was observed (Figure 39). In the IPS, the old portraits, landscapes and abstract paintings elicited stronger activation as compared to the novel, similar (did not reach significance) and dissimilar exemplars (p < 0.01) in both hemispheres. Moreover, the new, similar exemplars evoked stronger activation than the dissimilar exemplars, although this difference did not reach significance. In the IPS, old paintings evoked stronger activation than new exemplars, regardless of the response made by the subject (not shown). Furthermore, Remember responses to old paintings evoked stronger activation in the FG and the IPS than Remember responses to new, visually similar paintings (Figure 40).

Hits, namely remember or know responses to old paintings, evoked stronger activation than Misses (p < 0.05), namely new responses to old pictures, correct rejections (p < 0.01), namely new responses to new items, and false alarms (p < 0.001) in the regions FG, IPS and IFG (Figure 38).

18 Discussion

We investigated the neural correlates that mediates recollection and familiarity. During the study phase, subjects performed an attention demanding
flower detection task. The behavioral data showed that response latencies were significantly shorter for the representational paintings, i.e., portraits and landscapes, in which familiar objects were clearly depicted. In contrast, the visual search for flowers in abstract compositions took much longer, in line with our previous studies of object indeterminacy in art paintings [210]. During the study phase activation in distributed cortical network, that includes visual, parietal and prefrontal cortices was found. The activation evoked by paintings including flowers were virtually identical to the activation evoked by paintings without flowers, implying that subjects paid equally attention to both types of stimuli.

The behavioral data of the surprise Remember / Know / New memory retrieval shows that most of the old paintings were correctly recognized and that responses to the new exemplars depended on their visual similarity to the old ones. The proportion of false alarms, namely remember and know responses to new paintings, was significantly reduced with decreased visual similarity. Subjects also responded faster and more accurately to the visually different paintings. Longer latencies for correct responses were associated with the novel, similar paintings, probably due to their resemblance to the old pictures. Consistent with previous findings (e.g.[211]), correct and incorrect know responses were associated with significantly longer latencies, suggesting that subjects hesitated before deciding that a picture looked familiar but they could not recollect additional information about its prior experience.

The recognition memory task evoked activation within a distributed cortical network that included similar visual, parietal and prefrontal regions that were activated during the study phase. Within DOC, IOG, FG, IPS and IFG, correctly remembered old items elicited stronger activation than both correctly known and new items. In some of these regions, greater activation for remembered than known words has been previously found [149], suggesting that recollection elicits enhanced activation than familiarity regardless of stimulus format.
Contrasting recollection, namely correct Remember judgments, with familiarity, namely correct Know judgments, revealed a significant cluster of activation in the precuneus, a region implicated in many episodic memory retrieval (e.g., [212, 213, 185, 4]) and visual imagery studies [214, 215]. Within the precuneus, correctly remembered paintings not only elicited stronger activation than known items, but a latency shift of the BOLD response was observed, with a significantly earlier peak during recollection than familiarity-based judgments. Within the temporal lobe, activation in the hippocampus and PHC was stronger during recollection than during familiarity-based memory decisions. Our findings provide empirical evidence in support of a recent perspective according to which the Remember / Know procedure separates strong memories from weak memories [216, 217]. It is of interest that within the PHC and the hippocampus, activation evoked by remembered paintings was not statistically significant from activation evoked by new items. Our fMRI findings are consistent with a previous report, in which activation in the hippocampus was recorded in epileptic patients using depth electrodes. Some hippocampal neurons increased their firing rate in response to old pictures, whereas other neurons signaled novelty by increased firing rate in response to new pictures [218]. Taken collectively, our findings that recollection evoked stronger activation than both familiarity and novelty suggest that the Remember / Know / New experimental procedure reflects memory strength and not independent memory processes.

Activation in parietal cortex during the memory test revealed stronger responses to the old, correctly remembered paintings than to the new items, consistent with previous ERP and fMRI studies (e.g., [219, 220, 221]). Furthermore, activation within the IPS was reduced with decreased similarity between the new paintings and the old ones. The IPS, traditionally considered a region of the dorsal frontoparietal attention network, was implicated in many cognitive studies of attention, particularly in target detection tasks [222, 223], and the segmentation of old from new items [224]. Numer-
ous recognition memory studies have further shown that posterior parietal cortex does not merely detect old items but, rather, mediates higher order cognitive processes associated with memory retrieval [225, 226, 213]. Taken collectively, our previous [4, 195] and current findings suggest not only that the parietal cortex mediates the old/new effect, but also processes the degree of visual similarity between old and new items. These findings support the ‘mnemonic accumulator’ hypothesis, according to which recognition memory decisions are based on the integration of sensory signals [227].

Models of recognition memory assume that recollection and familiarity are independent processes during retrieval [139]. Evidence for such neuroanatomical dissociation came from studies in which subjects were instructed to rate their memory confidence [185] or to indicate the amount of recollected information [228]. Future studies will determine the extent to which the dissociation of the neural correlates of recollection and familiarity could be generalized across various experimental paradigms. Although remember and know responses are exclusive, recollected items are also familiar ones. It is therefore highly likely that retrieval of mnemonic information about complex pictures is modulated by activation within a distributed neural system, where memory strength modulates the neural response. The redundant relationship between recollection and familiarity has been corroborated by neuropsychological [229] and electrophysiological [230] studies. Our current findings suggest that the same cortical structures are activated during recollection- and familiarity-based judgments and are consistent with models of a continuum of mnemonic information on which the subject establishes a criterion [231, 232, 233]. This criterion, as shown in our current and previous studies [4, 195], depends on the degree of visual similarity between old and new items. A distributed neural system for recognition memory is consistent with recent neuroanatomical findings in amnesic patients, indicating that the ability to recollect remote memories depends not only on the medial temporal lobe but on widely distributed neocortical areas in the occipital, parietal
and prefrontal lobes [234, 235]. Furthermore, a distributed memory network is not only physiologically and ecologically plausible, but also confirms with previous fMRI studies showing that the representation of objects and faces in the human brain is not modular, but rather distributed across a wide expanse of cortex [39, 40, 48]. Within this distributed neural system, the memory strength is manifested by differential BOLD responses during recollection and familiarity judgments. Future studies will determine the extent to which these two memory processes are manifested by differential patterns of effective connectivity among regions.

In summary, these results show that recollection- and familiarity-based memory decisions depend on the degree of visual similarity between old and new items. Furthermore, recognition memory of complex pictures is mediated by activation within a distributed cortical network, where remembered and known items evoke differential BOLD responses that reflect their memory strength.
Part V

Discussion

In this PhD dissertation I described three studies in which category-learning and the memory processes of recollection and familiarity were investigated. In the first study, the relationship between identification and categorization of visual dot patterns were studied. A psychophysical paradigm was utilized to measure performance and reaction time changes with practice. The behavioral data showed that the subjects learned the category and the prototype, respectively. Subjects improved their performance with practice over the first five sessions in both tasks and, in subsequent sessions, the proportion of correct responses stayed constant. The performance in the categorization task was higher than in the identification task in all sessions, probably due to a different level of difficulty in the two tasks. However, a 3-way ANOVA revealed no significant interaction between task type and session number, implying that the performances of the two tasks changed with the same rate over time. Initially, the reaction times in the two tasks were significantly different from each other. Response latencies in the categorization task were longer than in the identification task. With practice, subjects responded faster to the target stimuli in both tasks. However, the response latencies of the categorization task decreased faster with practice than the response latencies of the identification task. The difference in reaction times became insignificant after eight sessions, suggesting that the reaction times of the two tasks converged. This difference in the development was also revealed by a 3-way ANOVA that showed a significant interaction between task type and session.

Because the difference in the reaction times in the two tasks was independent of the performance, we can rule out that the difference in the development was due to the different difficulty level. Since this difference
in the development of the response latencies cannot be attributed to task difficulty, we can argue that the learning rates in identification and categorization diverge, thus suggesting that the two processes are distinct and are likely mediated by separate neural substrates. Measured reaction times during categorization tasks in other studies [116, 117, 165, 166] were very short, as fast as or even faster than reaction times of an identification task, whereas response latencies of the categorization task in this study were initially significantly longer than reaction times in the identification task. This inconsistency can be explained by the nature of the stimuli used in the different studies. Whereas, natural stimuli, such as faces, animals and houses, were used in those studies, in our study, artificial, never-seen before stimuli and categories were used. Natural stimuli and categories have been seen and learned by the subjects throughout their lives, subjects were already experts in categorizing and identifying natural categories, and could not improve their performance any further.

Future studies will determine whether longer training in these two tasks would result in shorter response latencies in the categorization task than in the identification task, or whether the reaction times converge and stay constantly at the same level. Reber and colleagues found dissociations in brain activity between categorization and identification tasks with dot-patterns [64]. However, no studies have compared brain activity between newly learned categories and categories in which subjects were experts at. To summarize, identification and categorization of visual dot patterns show different learning rates. The response latencies in the categorization task decreased faster than in the identification task. These results suggest that identification and categorization are two distinct processes. However, whether they are mediated by the same neural network, or by different ones, will have to be shown in future imaging studies.

The aim of the second experiment was to test whether the number of categories has an effect on the performance in a categorization task with
rule-based and information-integration category structures. Subjects learned two or four categories of Glass patterns. During the task, subjects indicated to which category the presented stimulus belonged.

The Glass patterns were divided into two different category structures. The rule-based categories can be learned via some explicit reasoning process, in this study the category boundaries were along one dimension, namely the distance between dots of a dot-pair. The set of stimuli for the rule-based task consisted of circular-shaped Glass patterns with variable dot-separations. The dot-separation varied smoothly between two extremes. This continuum is then divided into two or four categories. In the information-integration conditions, the form of the Glass patterns changed gradually from circular over spiral to radial, while the dot-separation was held constant. This kind of category structure is hard to verbalize and accuracy is maximized only if information from two or more dimensions is integrated at some predecisional stage. This continuum of patterns was divided into two or four categories. Subjects were asked to learn the categories assigned to them. Subjects conducted seven sessions in each condition. The performance, averaged across all subjects and sessions, was virtually identical in the information-integration task for the two- and the four-category condition, whereas the performance in the rule-based four-category condition was significantly lower than in the two-category condition. The response latencies showed a similar pattern of results. While the reaction times of the two conditions in the information-integration task were not significantly different from each other, the response latencies in the four-category condition of the rule-based task were significantly higher than in the two-category condition, thus ruling out possible speed-accuracy trade-off explanations. These results confirm findings from a study by Maddox et al., and support the hypothesis that rule-based and information-integration category learning are mediated by distinct neural substrates.

Converging evidence indicates that the learning of rule-based and information-
integration category structures might be mediated by different neural circuits, and that literature provides further reason to expect that the experimental manipulations studied here have different effects on these two types of tasks [175, 171, 181, 182, 236]. A direct association between procedural learning and information-integration category learning was shown by Maddox et al. [237]. The information-integration category learning was negatively affected by changes in the response location associated with unique category labels. Other studies suggested that rule-based category learning, but not information-integration category learning, is negatively affected, when subjects simultaneously performed tasks known to recruit frontal structures, such as the Stroop task [118] and working memory tasks [184].

The rule-based category learning system is under conscious control and has full access to working memory and executive attention and accurate rule-based learning depends upon good use of these processes. As the complexity of the rule increases, the demand on working memory and executive attention resources increases, and performance is expected to decrease when subjects use a rule-based strategy. When the optimal rule is simple, as in the two-category condition, learning requires only a small amount of working memory and executive attention resources. Under these conditions, subjects used nearly the optimal rule, and thus reached high accuracy. As the complexity of the optimal rule increases, the working memory and executive attentional resource required is much greater. Under these conditions overall performance is significantly lower. As the complexity increases further, performance continues to decline significantly, and some subjects start to apply information-integration strategies [184].

Ashby and Waldron [182] proposed a neurobiologically plausible model of information-integration category learning with visual stimuli, according to which visual stimuli are represented perceptually in higher-level visual areas, such as the inferotemporal cortex (IT). In primates, all of the extrastriate visual cortex projects directly to the tail of the caudate nucleus, with about
10'000 visual cortical cells converging on each caudate cell [238]. It is assumed that a low-resolution map of the perceptual space is represented among the striatal units. Cells in the tail of the caudate then project to prefrontal and premotor cortex. Ashby and Waldron hypothesized that, through a procedural learning process, each caudate unit learns to associate a category label with a large group of visual cortical cells (i.e. all that project to it). Each subject was informed about the number of categories prior to each condition, and across conditions the complexity of the categories was held constant (i.e. the number of stimulus clusters was constant). Given the proposed neurobiology of this information-integration learning system, there is no a priori reason to expect a performance difference between information-integration category learning with two versus four categories.

To summarize, category number affected rule-based category learning but not information-integration category learning. The performance in information-integration category learning task was not affected by the number of categories, whereas in rule-based category learning task, the four-category condition elicited a significantly lower performance than the two-category condition. The notion of multiple-category systems is supported by these results.

In the third study, the memory processes recollection and familiarity were investigated. To test whether recollection and familiarity judgments were influenced by the degree of visual similarity between old and new pictures, we used event-related fMRI with portraits, landscapes and abstract compositions by painters with a unique style. Based on a previous study [4], we predicted that subjects would correctly reject more new paintings that were visually different from the old ones, and would make more false alarms to new, visually similar paintings. Furthermore, we hypothesized that if recollection and familiarity judgments about complex pictures reflect strong memories and weak memories, respectively, correctly remembered items would be associated with stronger neural activation than known items. During the study phase subjects were presented with paintings and performed an attention
demanding flower detection task. Activation was found in a distributed cortical network, that includes visual, parietal and prefrontal areas. Activation evoked by paintings that include flowers was not significantly different from activation evoked by paintings without flowers. This suggests that subjects paid equal attention to both types of paintings. Interestingly, shorter response latencies for paintings with flowers predicted subsequent Remember judgments during the memory test, suggesting that fast identification and detection of the flower targets facilitated later recollection of these paintings. Stronger activation in the FG for paintings with flowers predicted subsequent Remember and Know judgments, whereas reduced activation for paintings without flowers predicted subsequent Know judgments. Ten minutes after the study phase, subjects performed a surprise Remember / Know / New memory test. Old paintings, i.e. paintings shown during the study phase, were intermixed with new ones that were either visually similar or dissimilar to the old ones. Most of the old paintings were correctly recognized by the subject and performance on new paintings depended on the visual similarity to the old ones. The proportion of false alarms was significantly lower for new, similar paintings than for new, dissimilar paintings. Subjects also responded faster and more accurately to the visually different paintings. Longer latencies for correct responses were associated with the novel, similar paintings, probably due to their resemblance to the old pictures. Consistent with previous findings (e.g. [211]), correct and incorrect know responses were associated with significantly longer latencies, suggesting that subjects hesitated before deciding that a picture looked familiar but they could not recollect additional information about its prior experience.

The recognition memory task evoked activation within a distributed cortical network that included similar visual, parietal and prefrontal regions that were activated during the study phase. Within DOC, IOG, FG, IPS and IFG, correctly remembered old items elicited stronger activation than both correctly known and new items. Contrasting recollection, i.e. correct Remember
judgments, with familiarity, namely correct Know judgments, revealed a significant cluster of activation in the precuneus, a region implicated in many episodic memory retrieval (e.g., [212, 213, 185, 4]) and visual imagery studies [214, 239, 215]. Within the precuneus, correctly remembered paintings not only elicited stronger activation than known items, but a latency shift of the BOLD response was observed, with a significantly earlier peak during recollection than familiarity-based judgments. Within the temporal lobe, activation in the hippocampus and PHC was stronger during recollection than during familiarity-based memory decisions. Our findings provide empirical evidence in support of a recent perspective according to which the Remember / Know procedure separates strong memories from weak memories [216, 217].

Most dual-process models assume that recollection and familiarity operate independently during retrieval [240]. The empirical evidence consistently shows that the two processes are independent under most standard test conditions. However, there are test conditions under which this assumption of independence does not appear to hold [241]. In order to determine whether recollection and familiarity operate independently, it is necessary to assess whether they can be functionally dissociated. That is, if one process operates independently of the other, it should be possible to find variables that influence one process without influencing the other. Empirical findings indicate that recollection and familiarity can be doubly dissociated by retrieval manipulations, verifying that the two processes operate independent at retrieval. For example, response speeding and dividing attention during time of test reduce recollection but leave familiarity unaffected [139]. Future studies will determine the extent to which the dissociation of the neural correlates of recollection and familiarity could be generalized across various stimuli (word vs. picture) and experimental paradigms, and differentiate the subprocesses that are involved in recollection and familiarity.

Our current findings suggest that the same cortical structures are activated during recollection- and familiarity-based judgments and are consistent
with models of a continuum of mnemonic information on which the subject establishes a criterion [231, 232, 233]. This criterion, as shown in our current and previous studies [4, 195], depends on the degree of visual similarity between old and new items.

In summary, our results show that recollection- and familiarity-based memory decisions depend on the degree of visual similarity between old and new items. Recognition memory of complex pictures is mediated by activation within a distributed cortical network, where remembered and known items evoke differential BOLD responses that reflect their memory strength.

In this thesis I described three experiments, in the first one I've tested whether identification and categorization are mediated by the same process. The second experiment followed the question whether there exists more than one categorization process in the brain. And the third experiment investigated the identification processes recollection and familiarity. These experiments revealed that identification and categorization are two distinct processes. Furthermore, the brain processes at least two category structures differently, the categorization of rule-based and the information-integration category structures are mediated by distinct neural substrates, supporting the notion of multiple-category systems. Whereas recollection and familiarity, as two identification processes, are mediated by the same neural network, and the difference is the strong, or weak respectively, memory strength, which is reflected by differential BOLD responses.
References


[148] Yonelinas, A., Hopfinger, J., Buonocore, M., Kroll, N., and Baynes, K. Hippocampal, parahippocampal and occipital-temporal contributions


Recollection- and familiarity-based decisions reflect memory strength

Martin Wiesmann and Alumit Ishai

1. Institute of Neuroinformatics, University of ETH Zurich, Zurich, Switzerland
2. Institute of Neuroradiology, University of Zurich, Zurich, Switzerland

INTRODUCTION

Functional brain imaging studies have shown that medial temporal, parietal and prefrontal cortices are involved in recognition memory of prior episodes (Rugg and Wilding, 2000; Rugg and Yonelinas, 2003). The functional role that these regions play in memory retrieval, however, is still debated. Specifically, it is unclear whether recollection, the retrieval of specific information about a past experience, and familiarity, a sense that an event has been previously experienced (Tulving, 1985), are mediated by dissociated neural systems or separate strong memories from weak memories. Some studies suggest that separate cortical networks (Yonelinas et al., 2005) and differential activation in parietal cortex (Vilberg and Rugg, 2007) mediate these two distinct memory processes, whereas other studies suggest that recollection and familiarity reflect differences in the strength of a common memory trace (Donaldson, 1996; Dunn, 2004; Gonsalves et al., 2005; Squire et al., 2007; Wixted, 2007).

In most episodic retrieval studies written words have been used as stimuli (e.g., Cabeza et al., 2001; Henson et al., 1999) and the neural correlates of retrieving pictures from memory are not fully understood. We have recently conducted a series of event-related fMRI studies to investigate the neural mechanisms of retrieving pictorial information from long-term memory and the effects of visual similarity between old and new pictures. Subjects memorized paintings (Yago and Ishai, 2006) or unfamiliar Asian faces (Ishai and Yago, 2006) and 4 days later performed an old-new recognition memory task in the MR scanner. The old pictures were presented with new ones that were visually similar, somewhat similar or visually different from the old paintings. Consistent with our hypothesis, subjects were slower and less accurate to reject new, visually similar paintings. We found activation in a distributed cortical network that included face- and object-selective regions in the visual cortex, as well as prefrontal areas where the old paintings evoked stronger activation than the new ones, regardless of their visual similarity. Moreover, activation elicited by new paintings in parietal cortex was reduced with decreased similarity to the old items, whereas in the hippocampus and precuneus, stronger responses were evoked by the new, visually different paintings. We concluded that recognition memory is mediated by classification of new items as a match or a mismatch, based on their degree of visual similarity to old items (Ishai and Yago, 2006; Yago and Ishai, 2008). Our old-new task, however, did not address the issue of memory processes and the extent to which the observed behavioral and neural responses were due to recollection- or familiarity-based memory decisions.

To test whether recollection and familiarity judgments were influenced by the degree of visual similarity between old and new pictures, we used event-related fMRI with portraits, landscapes and abstract compositions by painters with a unique style (Figure 1). Based on our previous study, we predicted that subjects would correctly reject more new paintings that were visually different from the old ones, and would make more false alarms to new, visually similar paintings. Furthermore, we hypothesized that if recollection and familiarity judgments about complex pictures...
were randomly presented with 60 new paintings. The new paintings were subjects performed a surprise memory test, in which the 60 old paintings alternated with 15-s fixation epochs.

series collected during the study phase, paintings (4 epochs of 15 s each) any flowers by pressing one of two buttons. In each of the three time series collected in the test phase, paintings (8 epochs of 15 s each) alternated with 15-s fixation epochs. In both study and test, each run included one category of paintings (i.e., portraits, landscapes or abstract compositions).

New memory retrieval test in which the paintings from the study phase (“old”) were randomly presented with new paintings that were either visually similar or visually different.

Figure 1. Stimuli and tasks. During study, subjects were presented with portraits, landscapes and abstract paintings (shown from top to bottom are examples from Renoir, Pissaro and Kandinsky) and performed a flower detection task. Ten minutes later, subjects performed a surprise Remember/Know/New memory retrieval test in which the paintings from the study phase (“old”) were randomly presented with new paintings that were either visually similar or visually different.

reflect strong memories and weak memories, respectively, correctly remembered items would be associated with stronger neural activation than known items.

MATERIALS AND METHODS

Subjects

Twenty-one normal, right-handed subjects (11 females, mean age 25 years) with normal or corrected to normal vision participated in the study. All subjects gave informed written consent for the procedure in accordance with protocols approved by the University Hospital of Zurich. The subjects, students from the University of Zurich, had no formal art education and reported visiting art museums once a year or less. Post-scan questionnaires revealed that all subjects were unfamiliar with the paintings and had not seen them prior to the experiment.

Stimuli and tasks

Stimuli were displayed using Presentation (www.neurobs.com, version 9.13) and were projected with a magnetically shielded LCD video projector onto a translucent screen placed at the feet of the subject. During the study phase, subjects performed a flower detection task on 20 portraits by Modigliani and Renoir, 20 landscapes by Pissarro and Van Gogh and 20 abstract compositions by Kandinsky and Miro (Figure 1). Each painting was presented for 3 s and subjects had to indicate whether it includes any flowers by pressing one of two buttons. In each of the three time series collected during the study phase, paintings (4 epochs of 15 s each) alternated with 15-s fixation epochs. We then collected the anatomical images and after this 10-min scan, subjects performed a surprise memory test, in which the 60 old paintings were randomly presented with 60 new paintings. The new paintings were either visually similar or visually different from the old paintings. The degree of visual similarity between the new paintings and the old ones was assessed in a separate behavioral pilot (for details see Yago and Ishai, 2006). Each painting was presented for 3 s and subjects had to make a Remember/Know/New decision by pressing one of three buttons. Subjects were instructed to respond “Remember” if they were absolutely convinced that the picture was presented during the flower detection task; “Know” if the picture seemed familiar; and “New” if the picture was not presented in the flower detection task. In each of the three time series collected in the test phase, paintings (8 epochs of 15 s each) alternated with 15-s fixation epochs.

Data acquisition

Data were collected using a 3T Philips Intera whole body MR scanner (Philips Medical Systems, Best, The Netherlands). Changes in blood-oxygenation level-dependent MRI signal were measured by using sensitivity encoded gradient-echo echoplanar sequence (SENSE, Pruessmann et al., 1999) with 35 axial slices, TR = 3 s, TE = 35 ms, flip angle = 82°, field of view = 220 mm, acquisition matrix = 80 × 80, reconstructed voxel size = 1.72 mm × 1.72 mm × 4 mm and SENSE acceleration factor R = 2.

High-resolution spoiled gradient recalled echo structural images were collected in the same session for all the subjects (180 axial slices, TR = 20 ms, TE = 2.3 ms, field of view = 220 mm, acquisition matrix = 224 × 224, reconstructed voxel size = 0.9 mm × 0.9 mm × 0.75 mm). These high-resolution anatomical images were collected after the study phase and provided detailed anatomical information for the region-of-interest (ROI) analysis and for 3D normalization to the Talairach and Tournoux (1998) atlas.

Data analysis

Accuracies and reaction times were computed for each subject, category of paintings (portraits, landscapes, abstract paintings) and response type (Yes/No during the flower detection task; Remember/Know/New during the memory retrieval test). ANOVA was used to compare the various conditions.

Functional MRI data were analyzed in BrainVoyager QX Version 1.8 (Brain Innovation, Maastricht, The Netherlands). All volumes were realigned to the first volume, corrected for motion artefacts and spatially smoothed using a 5-mm FWHM Gaussian filter. The main effects during the study and test were analyzed using multiple regression (Friston et al., 1995). Based on the contrast of paintings vs. fixation, a set of ROIs was anatomically defined for each subject with clusters that showed a significant effect (p < 0.0001, uncorrected). These regions included the inferior occipital gyrus (IOG), fusiform gyrus (FG), dorsal occipital cortex (DOC), intraparietal sulcus (IPS), inferior frontal gyrus (IFG), insula and the anterior cingulate cortex (ACC). The contrasts of Remember vs. Know and Remember vs. New further revealed significant activation in the precuneus and in two medial temporal lobe structures, the parahippocampal cortex (PHC) and the hippocampus. In each subject and each ROI, the mean parameter estimates were calculated separately for each response type (Yes/No during flower detection task; Remember/Know/New during memory test) and were used for between-subjects random-effects analyses.

Finally, we tested whether reaction times and fMRI activation during the study phase could predict subsequent behavioral and neural responses to the old paintings during the test phase. Thus, responses during the flower detection task were sorted based on subsequent Remember and Know judgments subjects made during the retrieval test.

RESULTS

Study phase

Behavioral data. In this phase subjects were presented with portraits, landscapes and abstract paintings and indicated whether each painting contained any flowers. The behavioral data collected while subjects
performed the task in the scanner are shown in Figure S1. As most paintings did not include flowers, the proportion of No responses was higher than the proportion of Yes responses ($t(124) = 10.9, p < 0.00001$) for portraits, landscapes and abstract paintings. Furthermore, subjects made significantly faster Yes responses than No responses ($t(120) = 3.3, p < 0.001$) for portraits, landscapes and abstract paintings. Interestingly, subjects reported seeing flowers in more than 20% of the abstract paintings, which do not depict natural objects, but rather use purely visual forms of line, color and shape. Moreover, it took subjects significantly longer to decide whether these abstract compositions contained flowers. Thus, the differences in response latencies between Yes abstract responses and No responses to portraits and landscapes were significant ($t(36) = 4.2, p < 0.001$ and $t(36) = 2.9, p < 0.01$, respectively), as well as the difference in response latencies between No responses to abstract paintings and No responses to portraits ($t(40) = 4.4, p < 0.001$).

**Imaging data.** The main effect, namely responses evoked by all paintings as compared with the fixation baseline, revealed activation within a distributed cortical network that included multiple visual, parietal and prefrontal regions (Figure S2). Significant activation was found in the LOC (mean Talairach coordinates: $−28, −82, 12; 33, −82, 12$), IOG ($−24, −80, −3, 21, −80, −5$), FG ($−28, −53, −12; 28, −53, −10$), IPS ($−32, −51, 47; 29, −51, 44$), IFG ($−44, 3, 33; 43, 4, 34$) and ACC ($8, 7, 53$). Within all regions, activation evoked by Yes and No responses during the flower detection task was virtually identical, ruling out differential effects of attention during task performance.

We then tested whether behavioral and neural responses during study could predict subsequent memory performance during test. We found that shorter response latencies for paintings with flowers predicted subsequent Remember judgments during the memory test. In terms of the neural response, we found that Yes responses during the flower detection task resulted in similar activation in the FG during subsequent Remember and Know judgments (Figure S3).

**Test phase**

**Behavioral data.** Ten minutes after performing the flower detection task, subjects performed a surprise memory task. In this test phase the paintings from the study phase were randomly presented with new paintings that were either visually similar or visually different from the old ones (see Figure 1) and subjects made Remember/Know/New decisions. The behavioral data collected while subjects performed the task in the scanner are shown in Figure 2. The proportion of remember responses to the old items was significantly higher than both know and new responses ($t(61) = 5.8, p < 0.000001$ in both comparisons). Consistent with our prediction, the proportion of false alarms, namely remember and know responses to new items, significantly decreased with decreased visual similarity between the old and the new items [remember new similar vs. remember new different, $t(61) = 5.7, p < 0.000001$; know new similar vs. know new different, $t(61) = 10.5, p < 0.000001$]. In terms of response latencies, know decisions took significantly longer than both remember and new responses [know vs. remember, $t(187) = 9.0, p < 0.000001$; know vs new, $t(187) = 8.2, p < 0.000001$ for all paintings]. Finally, consistent with our prediction, subjects responded significantly faster to the new, visually different than to the new, visually similar paintings ($t(61) = 4.9, p < 0.000001$).

**Imaging data.** The main effect, namely responses evoked by all paintings as compared with the fixation baseline, revealed activation within a distributed cortical network that included multiple visual, parietal and prefrontal regions (Figure 3). Significant activation was found in the LOC (mean Talairach coordinates: $−27, −81, 11; 32, −81, 11$), IOG ($−29, −77, −1, 28, −77, −6$), FG ($−28, −55, −15; 33, −55, −14$) IPS ($−23, −54, 38, 27, −54, 39$), IFG ($−44, 3, 33; 43, 4, 34$) and ACC ($2, 15, 40$). Within the LOC, IOG, FG, IPS and IFG activation elicited by correctly remembered paintings was significantly higher than activation evoked by correctly know and new judgments ($p < 0.000001$ in both comparisons).

We then analyzed the activation evoked by correct and incorrect responses in all ROIs (Figure 4). Within the IPS, hits, namely correct responses to old items, evoked stronger activation than misses ($p < 0.01$), correct rejection of new paintings ($p < 0.01$) and false alarms ($p > 0.001$). To test our hypothesis about visual similarity, we compared correct responses to old paintings with correct responses to new, visually similar and new, visually different items (Figure 6). Consistent with our previous study (Yago and Ishai, 2006), we found that within the IPS, new, visually different paintings evoked less activation than new, visually similar ones ($p < 0.01$). To further understand the effect of visual similarity on memory decisions, we compared remember responses to old items (correct responses) with remember responses to new, visually similar items (false alarms). Within the FG, the difference between these responses was not significant, however in the IPS, remember responses to old items evoked significantly stronger activation than remember responses to new, visually similar paintings.
Memory Retrieval: Main Effect

Figure 3. Activation evoked by portraits, landscapes and abstract paintings during the memory retrieval test. Top: coronal sections, taken from a representative subject, illustrating activation in DOC, IOG, FG, precuneus, IPS, IFG, insula and ACC. Bottom: Mean parameter estimates, averaged across all subjects and both hemispheres, for correct Remember/Know/New judgments.

Figure 4. Task performance and visual similarity. Top: A coronal section, taken from a representative subject, illustrates activation in the FG and IPS. Bottom: Left: Mean parameter estimates for correct responses to old items (HT); incorrect responses to old items (MS); Correct responses to new items (CR) and incorrect responses to new items (FA). Right: Mean parameter estimates for correctly recognized old and new items, as a function of their visual similarity. Parameter estimates were averaged across all subjects and both hemispheres.
similar items \( (p < 0.02) \), consistent with previous reports about the role of the IPS in mediating the old/new effect.

We then contrasted the memory responses. Interestingly, comparing correct remember judgments with correct know judgments revealed activation predominantly in the precuneus (Figure 5). Analysis of the BOLD response within this region indicated not only stronger responses to remembered items, but also a BOLD latency shift, with a significantly earlier peak for recollection \( (5.6 \text{ s}) \) than familiarity \( (10.3 \text{ s}; t(40) = 3.1, p < 0.004) \). Comparing correctly remembered pictures with correct rejections of new pictures further revealed a similar pattern of activation, albeit with a lower threshold. Activation in the precuneus was stronger and peaked significantly earlier during remember \( (5.7 \text{ s}) \) than new \( (10.2 \text{ s}; t(40) = 3.37, p < 0.002) \) judgments. Finally, contrasting correct know responses with correct rejection of new items revealed similar hemodynamic response profiles and the peak of the BOLD response \( (11.1 \text{ and } 10.2 \text{ s, respectively}) \) was not statistically significant \( [t(40) = 0.54, p = 0.59] \).

Comparing correct remember judgments with correct know judgments also revealed activation in the medial temporal lobe, namely in the PHC and the hippocampus (Figure 6). In both regions, correctly remembered paintings evoked significantly stronger activation than correctly known items \( [t(40) = 2.75, p < 0.009 \text{ in the PHC and } t(40) = 2.47, p < 0.018 \text{ in the hippocampus}] \). The difference between remembered and new paintings, however, was not statistically significant \( [t(40) = 1.55, p = 0.13 \text{ in the PHC and } t(40) = 1.16, p = 0.25 \text{ in the hippocampus}] \).

**DISCUSSION**

We investigated the neural correlates that mediate recognition memory of portraits, landscapes and abstract paintings. During the study phase, subjects performed an attention demanding flower detection task. The behavioral data showed that response latencies were significantly shorter for the representational paintings, i.e., portraits and landscapes, in which familiar objects were clearly depicted. In contrast, the visual
search for flowers in abstract compositions took much longer, in line with our previous studies of object indeterminacy in art paintings (Ishai et al., 2007). The flower detection task evoked activation within a distributed cortical network, in which paintings with and without flowers evoked virtually identical responses, ruling out differential effects of attention during encoding. Interestingly, shorter response latencies for paintings with flowers predicted subsequent Remember judgments during the memory test, suggesting that fast identification and detection of the flower targets facilitated later recollection of information about these paintings. Stronger activation in the FG for paintings with flowers predicted subsequent Remember and Know judgments, whereas reduced activation for paintings without flowers predicted subsequent Know judgments.

The surprise Remember/Know/New memory retrieval test revealed that most of the old paintings were correctly recognized and that responses to the new items depended on their visual similarity to the old ones. Consistent with our hypothesis, the proportion of false alarms, namely remember and know responses to new paintings, was significantly reduced with decreased visual similarity. Consistent with previous findings (e.g., Dewhurst and Conway, 1994), correct and incorrect know responses were associated with significantly longer latencies, suggesting that subjects hesitated before deciding that a picture looked familiar but they could not recollect additional information about its prior experience.

The recognition memory task evoked activation within a distributed cortical network that included similar visual, parietal and prefrontal regions to those activated during the study phase. Within DCC, IGS, FG, IPS and IFG, correctly remembered old items elicited stronger activation than both correctly known and new items. In some of these regions, greater activation for remembered than known words has been previously found (Henson et al., 1999), suggesting that recollection elicits enhanced activation than familiarity regardless of stimulus format. When remember judgments were directly contrasted with know judgments, a significant cluster of activation was found in the precuneus, a region implicated in many episodic memory retrieval (e.g., Fletcher et al., 1995; Shannon and Buckner, 2004; Yago and Ishai, 2006; Yonelinas et al., 2005) and visual imagery (Ishai et al., 2000a, 2002; Mechelli et al., 2004) studies. Within the precuneus, correctly remembered paintings not only elicited stronger activation than known items, but a latency shift of the BOLD response was observed, with a significantly earlier peak during recollection than familiarity-based judgments. Within the temporal lobe, activation in the hippocampus and PHC was stronger during recollection than during familiarity-based memory decisions. Our findings provide empirical evidence in support of a recent perspective according to which the Remember/Know procedure separates strong memories from weak memories (Squire et al., 2007; Wixted, 2007). It is of interest that within the PHC and the hippocampus, activation evoked by remembered paintings was not statistically significant from activation evoked by new items. Our fMRI findings are consistent with a previous report, in which activation in the hippocampus was recorded in epileptic patients using depth electrodes. Some hippocampal neurons increased their firing rate in response to old pictures, whereas other neurons signaled novelty by increased firing rate in response to new pictures (Ruthschauser et al., 2006). Taken collectively, our findings that recollection evoked stronger activation than familiarity within multiple regions suggest that the memory decisions reflect memory strength and not independent memory processes.

Activation in parietal cortex during the memory test revealed stronger responses to the old, correctly remembered paintings than to the new items, consistent with previous ERP and fMRI studies (e.g., Curran and Cleary, 2003; Kahn et al., 2004; Wilding, 2000). Furthermore, activation within the IPS was reduced with decreased similarity between the new paintings and the old ones. The IPS, traditionally considered a region of the dorsal frontoparietal attention network, was implicated in many cognitive studies of attention, particularly in target detection tasks (Corbetta et al., 2000; Kincaide et al., 2005; Shulman et al., 2001) and the segmentation of old from new items (Pollmann et al., 2003). Numerous recognition memory studies have further shown that posterior parietal cortex does not merely detect old items but, rather, mediates higher order cognitive processes associated with memory retrieval (Konishi et al., 2000; Shannon and Buckner, 2004; Wheeler and Buckner, 2003). Taken collectively, our previous (Ishai and Yago, 2006; Yago and Ishai, 2006) and current findings suggest not only that the parietal cortex mediates the old/new effect, but also processes the degree of visual similarity between old and new items. These findings support the “mnemonic accumulator” hypothesis, according to which recognition memory decisions are based on the integration of sensory signals (Wagner et al., 2005).

Models of recognition memory assume that recollection and familiarity are independent processes during retrieval (Yonelinas, 2002). Evidence for such neuroanatomical dissociation came from studies in which subjects were instructed to rate their memory confidence (Yonelinas et al., 2005) or to indicate the amount of recollected information (Vilberg and Rugg, 2007). Future studies will determine the extent to which the dissociation of the neural correlates of recollection and familiarity could be generalized across various experimental paradigms. Although remember and know responses are exclusive, recollected items are also familiar ones. It is therefore highly likely that retrieval of mnemonic information about complex pictures is modulated by activation within a distributed neural system, where memory strength modulates the neural response. The redundant relationship between recollection and familiarity has been corroborated by neuropsychological (Knowlton, 1998) and electrophysiological (Nolte and Paller, 2004) studies. Our current findings suggest that the same cortical structures are activated during recollection- and familiarity-based judgments and are consistent with models of a continuum of mnemonic information on which the subject establishes a criterion (Donaldson, 1996; Dunn, 2004; Gonsalves et al., 2005). This criterion, as shown in our current and previous studies (Ishai and Yago, 2006; Yago and Ishai, 2006), depends on the degree of visual similarity between old and new items.

A distributed neural system for recognition memory is consistent with recent neuroanatomical findings in amnesic patients, indicating that the ability to recollect remote memories depends not only on the medial temporal lobe but on widely distributed neocortical areas in the occipital, parietal and prefrontal lobes (Bayley et al., 2005; Squire and Bayley, 2007). Furthermore, a distributed memory network is not only physiologically and ecologically plausible, but also confirms with previous fMRI studies, showing that the representation of objects and faces in the human brain is not modular, but rather distributed across a wide expanse of cortex (Haxby et al., 2001; Ishai et al., 1999, 2000b). Within this distributed neural system, the memory strength is manifested by differential BOLD responses during recollection and familiarity judgments. Future studies will determine the extent to which these two memory processes are manifested by differential patterns of effective connectivity among regions.

In summary, our results show that recollection- and familiarity-based memory decisions depend on the degree of visual similarity between old and new items. Furthermore, recognition memory of complex pictures is mediated by activation within a distributed cortical network, where remembered and known items evoke differential BOLD responses that reflect their memory strength.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGEMENTS

We thank Edward Wilding and Andrew Yonelinas for reading the manuscript. This study was supported by the Swiss National Science Foundation grant 3200B0-105278 and by the Swiss National Center for Competence in Research: Neural Plasticity and Repair.

SUPPLEMENTARY MATERIAL

See Figures S1–S3.
Recollection, familiarity and memory strength

Figure S1. Behavioral data during the flower detection task. Mean responses (left) and reaction times (right), averaged across 21 subjects. Error bars indicate standard error of the mean (SEM).

Figure S2. Activation evoked by portraits, landscapes and abstract paintings during the flower detection task. Top: coronal sections, taken from a representative subject, illustrating activation in DOC, IOG, FG, IPS, IFG and ACC. Bottom: mean parameter estimates averaged across all subjects and both hemispheres. Yes-flower and No-flower responses evoked virtually identical activation within all ROIs. Error bars indicate standard error of the mean (SEM).


REFERENCES


Recollection, familiarity and memory strength