
Aus dem Institut für Medizinische Psychologie der
Ludwig-Maximilians-Universität München

Vorstand: Prof. Dr. Ernst Pöppel

Untersuchungen zur Kontextabhängigkeit der visuellen Objekterkennung

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von

Markus Andreas Gschwind
aus
Basel

2007

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter: Prof. Dr. rer. nat. habil. Dr. med. habil. Ingo Rentschler
Prof. Dr. med. Dr. h.c. Thomas Brandt

Mitberichterstatter: Prof. Dr. Ing. Dr. med. habil. Florian Kolb
Prof. Dr. med. Volker Klauß

Mitbetreuung: PD. Dr. rer. nat. Dr. med. habil. Hans Brettel

Dekan: Prof. Dr. med. Dietrich Reinhardt

Mündliche Prüfung: 21. Juni 2007

Danksagung

Ich möchte mich bei all jenen, die diese Arbeit ermöglicht oder unterstützt haben, herzlich bedanken. Ganz besonderen Dank schulde ich Herrn *Prof. Dr. Ingo Rentschler*, der die Anregung zu diesem Themenkreis gab, dessen große, immer fächerübergreifende Perspektive die Diskussionen bestimmte und der mit außerordentlicher Umsicht, Hilfsbereitschaft und persönlichem Einsatz alles unternahm, was einem Doktoranden das Leben erleichtern konnte.

In der Arbeitsgruppe danke ich besonders Herrn *PD Dr. Hans Brettel*, Paris, der mit viel Einsatz die komplizierte Programmierung des Experimentier-Computers betreute, die Gedankengänge dieses Projektes kontinuierlich begleitete und auch aus der Pariser Ferne immer für Fragen zur Verfügung stand. Meinem Vorgänger *Erol Osman* danke ich für die Vorarbeiten bei der Aufstellung des Experimentes, seine tatkräftige Hilfe bei Datenauswertungsproblemen und für vielfältige fachliche Unterstützung, Herrn *PD Dr. Martin Jüttner*, Birmingham, für sein waches begleitendes Mitdenken und sicheres Urteil. Ohne die unermüdliche Hilfe meiner Kollegen *Markus Hofbauer* und *Ulrich Nuding* wäre die Erstellung der MATLAB-Skripten für die Datenauswertung nie geglückt. Und schließlich danke ich *Mme Dr. Sophie Schwartz*, Genève, und meiner Kollegin *Stéphanie Duhoux*, Genève, für eine sehr gelungene Kooperation, ihre Unterstützung und für maßgeblich weiterführende kritische Fragen zu meinen Experimenten.

Besonders danke ich auch Frau *Dr. Isolde von Bülow* für die große Unterstützung durch das DFG-Graduiertenkolleg 267 „Sensorische Interaktion in biologischen und technischen Systemen“, das nicht nur mein Leben für ein halbes Jahr ermöglichte, sondern auch die Bezahlung der Probanden übernahm und mir durch zahlreiche wissenschaftliche Unternehmungen viele Anregungen gab. Ebenso unterstützten das Projekt die Stiftung für Bildung und Behindertenförderung, Stuttgart, und das Generation Research Project des Humanwissenschaftlichen Zentrums der Universität München.

Ich danke auch meinen Probanden, von denen einige wahrhaft unerschütterlichen Optimismus zeigten und trotz großer Mühen bis zum Ende des Experiments durchhielten.

Ein besonderer Dank geht an Herrn *Prof. Dr. med. Dr. h. c. Johannes W. Rohen*, Erlangen, für sein Interesse an dieser Arbeit und wertvolle Ratschläge.

Nicht zuletzt möchte ich mich bei allen Mitgliedern des Instituts für Medizinische Psychologie bedanken für die freundliche Aufnahme und die große Hilfsbereitschaft bei all meinen Belangen. Mein Dank gilt auch seinem Vorstand, Herrn *Prof. Dr. Ernst Pöppel*, dafür, dass er meine Arbeit an diesem Institut in einer wahrhaft interdisziplinären Atmosphäre ermöglicht hat.

Inhalt

1	Einleitung.....	10
2	Das Wahrnehmen von Objekten	13
2.1	Neurophysiologie visueller Objektwahrnehmung	13
2.1.1	Frühe Verarbeitung visueller Information	13
2.1.2	Auftrennung der Bildmerkmale im visuellen Assoziationskortex.....	16
2.1.3	Sehen als Analyse versus Sehen durch Schlussfolgern	20
2.1.4	Multimodale Objektrepräsentation: Haptische und visuelle Wahrnehmung	23
2.2	Interne Objektrepräsentation und kategorielles Objektwissen	25
2.2.1	Repräsentation und Bindungsproblem.....	25
2.2.2	Kategorie und Klasse	27
3	Theorien der Objekterkennung.....	31
3.1	Was das Objekt zusammenhält - die Gestalttheorien	31
3.2	Techniken im maschinellen Bilderkennen.....	33
3.2.1	Strukturmerkmale im Bild – Indexing primitives.....	34
3.2.2	Segmentierung und Merkmale.....	35
3.3	Psychologische Theorien der Objekterkennung	36
3.3.1	Marrs algorithmischer Ansatz.....	36
3.3.2	Biedermans Objekterkennung durch Teilkörper.....	37
3.3.3	„View-dependent“ Theorien	39
3.3.4	Zusammenfassung: Die Objektstruktur in der Repräsentation.....	41
3.3.5	Fazit: Verschiedene Probleme im Objekterkennen	43
4	Experimentelle Untersuchung der Objekterkennung unter Berücksichtigung des Kontextes.....	45
4.1	Erkennungsleistung versus Repräsentation	45
4.2	Messung der Erkennungsleistung – die Klassifikation.....	46
4.3	Rolle der Bildinformation.....	48
4.4	Berücksichtigung des Kontextes.....	48
4.4.1	Das Vorwissen über die Objekte	49
4.4.2	Relative Konfiguration der Testobjekte.....	51
4.4.3	Die Objektumgebung.....	51
5	Ziele der vorliegenden Untersuchung	53
5.1	Experiment I: Multimodale Repräsentation der 3D-Objektstruktur	54
5.2	Experiment II: Bildmerkmale und Lernbarkeit der 3D-Objektstruktur.....	55

6	Methoden	56
6.1	Probanden	56
6.1.1	Homogene experimentelle Gruppen	56
6.2	Stimuli.....	58
6.2.1	Die 3D-Testobjekte und die Konstruktion der Objektansichten.....	58
6.2.2	Objektansichten für rein-strukturbasierte Klassifikation.....	61
6.2.3	Variation der Stimuli für Experiment II	65
6.2.4	Experimentelle Präsentation der Ansichten.....	66
6.3	Objektvorwissen: Experimentelle Bedingungen	68
6.4	Psychophysische Methode: Das Klassifikations-Paradigma	70
6.4.1	Überwachtes Kategorienlernen (supervised category learning)	70
6.4.2	Generalisierungstest.....	72
6.4.3	Introspektionsberichte der Probanden	72
6.4.4	Praktische Rekonstruktion der Objekte im Modell.....	72
6.5	Datenauswertung	73
6.5.1	Variablen.....	73
6.5.2	Dynamik des Lernens	76
6.5.3	Statistische Methoden.....	76
7	Ergebnisse	78
7.1	Experiment I	78
7.1.1	Lernen	78
7.1.2	Generalisierung.....	80
7.1.3	Objektkonstruktion	83
7.1.4	Zusammenfassung der Ergebnisse.....	84
7.2	Experiment II	85
7.2.1	Lernen	85
7.2.2	Generalisierung.....	87
7.2.3	Objektkonstruktion	88
7.2.4	Zusammenfassung der Ergebnisse.....	89
7.3	Vergleich von Experiment I und Experiment II: Kugeln versus Würfel... 90	
7.3.1	Lernen	90
7.3.2	Generalisierung.....	92
7.3.3	Dynamik des überwachten Lernens.....	94
7.3.4	Zusammenfassung des Vergleiches zwischen Experiment I und Experiment II.....	96
8	Diskussion	97
8.1	Experiment I	97
8.2	Experiment II	98
8.3	Schlussfolgerung.....	100
8.4	Kritik und Ausblick	103

9	Zusammenfassung	106
10	Literaturverzeichnis	108
11	Anhang	115
11.1	Anweisung an den Probanden.....	115
11.2	Introspektion der Probanden aus Experiment I.....	116
11.2.1	Probanden der Kontrollgruppe.....	116
11.2.2	Probanden der Priming-Gruppe vision	117
11.2.3	Probanden der Priming-Gruppe motor	118
11.3	Neben-Experiment „Kugel-Inform“	119
11.3.1	Information zur Chiralität	119
11.3.2	Introspektion eines Probanden (Kugel-Inform).....	119
11.4	Plausibilität des Maßes „Relapse“	120
11.5	Beispiele zu den Lernkurven aus dem überwachten Kategorienlernen...	121
11.6	Poster	124
	Curriculum Vitae.....	125

1 Einleitung

Ein Chirurg schaut vor einer anstehenden Operation auf das Röntgenbild, um sich einen Eindruck von den zu operierenden Strukturen im Körperinneren zu verschaffen. Je besser er sich dabei vorstellen kann, was ihn erwartet, desto gezielter kann er vorgehen. Jeder erfahrene Chirurg muss in der Lage sein, sich aufgrund solcher schwarz-weiß schattierter, zweidimensionaler Abbildungen eine Vorstellung von den dreidimensionalen Gegebenheiten im Körper zu machen. Er könnte auch nur mit einem Auge auf die Röntgenaufnahme schauen – diese Fähigkeit ist nicht an das stereoskopische Sehen gebunden. Sie ist eine Frage der Erfahrung und des Sehenlernens – des Bildverstehens. Der Chirurg muss das Röntgenbild „richtig sehen“.

Das „richtige Sehen“ bedeutungsvoller dreidimensionaler Strukturen aus zweidimensionalen Abbildungen soll in der vorliegenden Arbeit näher untersucht werden. Es ist eine Tätigkeit, die wir alle täglich leisten. Jedoch auch das Auge selbst funktioniert auf der Basis einer zweidimensionalen Abbildung der dreidimensionalen Umwelt: Es gilt daher zu verstehen, wie aus dem flächenhaften Netzhautbild die notwendigen Informationen gewonnen werden, um sich in einer dreidimensionalen Umwelt adäquat verhalten zu können. Dies ist eine der wichtigsten Fragen in der Erforschung der visuellen Objekterkennung.

Das Erkennen von Objekten ist eine eigenständige Gehirnleistung. Aus der Neuropsychologie sind spezifische Störungen der visuellen Objekterkennung bekannt, die visuellen Agnosien (Grüsser & Landis, 1991). An ihnen wird deutlich, worin diese Gehirnleistung besteht. Bei der *apperzeptiven Agnosie* werden Bildmerkmale wie Flächen, Kanten oder Farben, nicht zu einem sinnvollen Ganzen integriert. Einzelne Details können zwar benannt werden, aber das Objekt wird mit visuell ähnlichen Objekten verwechselt. Bei der *assoziativen Agnosie* ist die visuelle Integration intakt, der Zugang zum Objektwissen aber gestört. Die Zuordnung zu Oberbegriffen wie „Tier“ oder „Möbel“ gelingt dem Patienten wohl, nicht aber die genaue Identifizierung und das Benennen des Objektes. Bei den visuellen Agnosien ist das Phänomen der Doppeldissoziation zu beobachten, das heißt, zwei beschriebene Hirnleistungen können jeweils unabhängig voneinander ausfallen. Dies gilt gemäß der neuropsychologischen *Modularitätshypothese* als Nachweis für eine jeweils eigene neuronale Grundlage dieser Leistungen. So werden für die höhere Hirnleistung der Objekterkennung verschiedene Teilsysteme im visuellen System verantwortlich gemacht.

Mathematisch gesehen kann eine zweidimensionale Form als die Projektion einer unendlichen Mannigfaltigkeit verschiedener möglicher dreidimensionaler Formen interpretiert werden. Aus diesem Grund hat der inverse Prozess der Ableitung einer dreidimensionalen Objektstruktur aus dem zweidimensionalen Netzhaut-

bild während des Sehvorgangs, das so genannte „inverse Problem“ (Pizlo, 2001), für sich genommen keine eindeutige Lösung. Zu seiner Klärung sind immer Zusatzinformationen erforderlich.

Biologisch relevantes Wahrnehmen funktioniert nie in einer ausschließlich visuellen Welt; beim Sehvorgang wird eine Vielzahl von außerhalb der Netzhaut liegenden, zusätzlichen Faktoren mitintegriert. Aus unseren eigenen Bewegungen, aus unserem Wissen, unserer Erfahrung und gelenkt durch unsere Aufmerksamkeit fließen die notwendigen Zusatzinformationen mit ein. All diese Faktoren bilden den Kontext des Sehvorgangs und beeinflussen über das Wirkungsgefüge der visuellen Wahrnehmung unser Erleben der Umwelt (Albright, 1995). Ohne den Kontext mit-einzubeziehen ist visuelles Erkennen nicht zu verstehen.

Die kognitionspsychologische Forschung zur Objekterkennung konzentrierte sich in der Vergangenheit auf die Leistungen des visuellen Systems. Dabei stand insbesondere die Frage nach der Natur der internen Repräsentation der dreidimensionalen Objekte im Zentrum der Diskussion (Tarr & Vuong, 2002). Von einigen Forschern wurde eine Abhängigkeit der Repräsentation von den bisher bekannten Objektansichten beschrieben und daraus auf eine piktorale, zweidimensionale Art der Repräsentation geschlossen. Das Verständnis der dreidimensionalen Struktur des Objektes soll gemäß dieser Auffassung über eine Verkettung multipler zweidimensionaler Repräsentationen erreicht werden (Bülhoff & Edelman, 1992; Tarr, 1995). Demgegenüber zeigten andere Autoren, dass man bei vielen Objekten schon auf den ersten Blick sagen kann, welche dreidimensionale Struktur sie haben. Deswegen müsse die Objektrepräsentation als solche struktureller und nicht piktoraler Natur sein, und somit weitgehend stabil gegenüber den Veränderungen der Objektansicht, die durch ungewohnte Blickrichtungen herbeigeführt werden (Biederman, 1987; Biederman & Bar, 1999).

Beide Auffassungen aber konzentrieren sich ausschließlich auf das visuelle System und berücksichtigen die Kontextbedingtheit der Wahrnehmungsleistung nicht. Sie setzen die gemessene Erkennungsleistung mit der visuellen Repräsentation gleich (Liu, Kersten & Knill, 1999) und bieten damit wenig Kontrolle über die anderen am Erkennungsvorgang beteiligten Elemente. So bleiben die folgenden Kernfragen offen:

- Wie werden die Objektrepräsentationen während des Lernens aufgebaut?
- Welche Vorerfahrung haben die Probanden mit den Testobjekten und wie fließt diese in die Erkennungsleistung mit ein?
- Welche Rolle spielt die Bildinformation für die Erkennungsleistung?

Die vorliegende Arbeit stellt zwei Verhaltensexperimente zur menschlichen 3D-Objekterkennung dar, die auf diese Fragen Bezug nehmen. Um die hier verwendete experimentelle Herangehensweise verständlich zu machen, wird in den einleitenden Kapiteln 2 bis 4 Hintergrundwissen dazu dargestellt.

Kapitel 2 stellt die Grundlagen der Objektwahrnehmung dar: Die Strategien der neuronalen Verarbeitung visueller Information einerseits und die psychologischen Befunde zur Struktur des repräsentierten Objektwissens andererseits.

Kapitel 3 beschäftigt sich mit den wichtigsten Theorien über die Funktion der visuellen Objekterkennung, in deren Kontext auch die hier berichteten Experimente stehen: kognitionspsychologische und maschinelle Konzepte konkurrieren und befruchten sich gegenseitig.

Kapitel 4 diskutiert die Bedingungen der experimentellen Untersuchung der visuellen Objekterkennung unter Berücksichtigung des Kontextes.

Daraus ergeben sich Motivation und Ziele der vorliegenden Arbeit, wie sie in *Kapitel 5* beschrieben sind.

2 Das Wahrnehmen von Objekten

Visuelles Wahrnehmen spielt sich, wie jede kognitive Leistung, in zwei Bereichen ab. Es setzt einerseits eine neuronale Grundlage der visuellen Verarbeitung, andererseits das Vorhandensein intern repräsentierter Information voraus. Im folgenden Kapitel werden zuerst die grundlegenden Strategien neuronaler Verarbeitung der visuellen Information dargestellt und danach wird auf die Struktur der internen Repräsentation eingegangen. Trotz beachtlicher Forschungsergebnisse auf beiden Gebieten ist jedoch eine Korrespondenz physiologischer und psychologischer Befunde nur sehr bedingt gegeben. Objekterkennung bewegt sich deshalb immer in zwei Beschreibungsebenen.

2.1 Neurophysiologie visueller Objektwahrnehmung

2.1.1 Frühe Verarbeitung visueller Information

Retina. Die Verarbeitung der visuellen Reize beginnt bereits in den Zellschichten der Retina: Schon die Bipolarzellen in den Mittelschichten fassen für die Verarbeitung essentielle Information in so genannte rezeptive Felder (RF) zusammen, die jeweils aus einem Zentrum und einer Randzone bestehen. Die On-Bipolarzellen werden durch eine Licht-Reizung ihres Zentrums erregt und durch eine Licht-Reizung der Randzone gehemmt, die Off-Bipolarzellen verhalten sich genau umgekehrt.

Die Organisation der Retina in sich gegenseitig überlappende rezeptive Felder bewirkt, dass schon auf dieser Ebene nicht homogene, monochrome Flächen detektiert werden, sondern Grenzen, Kanten und Kontraste. Durch konvergente Signalweiterleitung und laterale Inhibition wird die Kontrastinformation noch zusätzlich verstärkt.

Corpus geniculatum laterale (CGL). Das CGL erhält seine Projektionen nicht nur von der Retina, sondern auch aus anderen Thalamuskernen, vom Hirnstamm und besonders vom Kortex. Dabei ist der Informationsfluss vom Kortex hinab zum CGL (top-down) erstaunlicherweise wesentlich stärker als der vom CGL hinauf zum Kortex (Sherman & Koch, 1986). Das CGL des Thalamus ist die erste Umschaltstation für 90% der Fasern des Sehnervs. Die restlichen 10% ziehen als retino-colliculärer Trakt zum Colliculus superior im Mesencephalon, einer Struktur, die an der Steuerung der Augenbewegungen beteiligt ist.

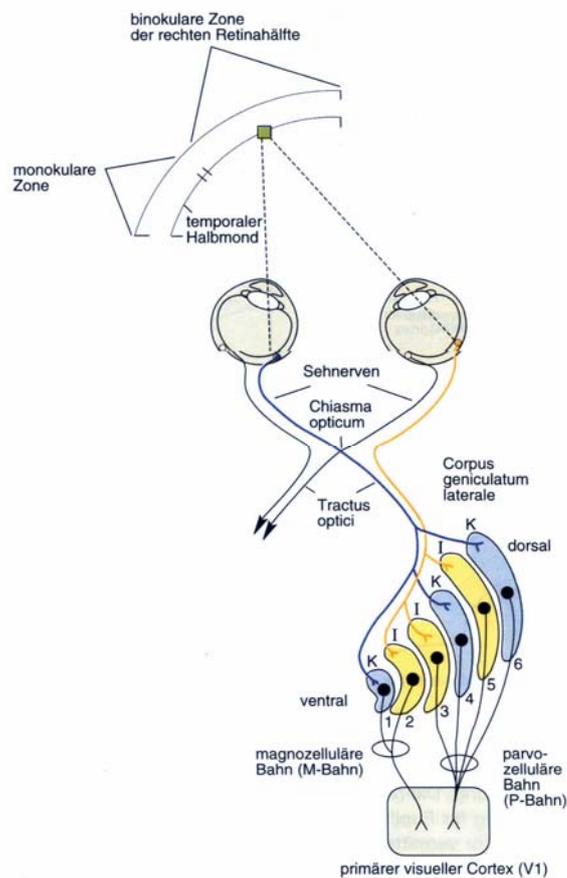


Abb. 1: Verlauf der Sehbahn von der Retina über das Chiasma opticum zum Colliculus superior. I = ipsilaterale Fasern, K = kontralaterale Fasern (aus Kandel, 1996).

Vom CGL wird die visuelle Information über den Tractus opticus zum primären visuellen Kortex geleitet. Der Tractus opticus führt erstaunlicherweise nur noch 40% der Fasern, die von der Retina zum CGL führen und ist von allen Projektionen zum CGL hin oder vom CGL weg die faserärmste. Die Projektionen vom Kortex zum CGL hinab sind weit zahlreicher.

Wie Schiller, Logothetis & Charles (1990) zeigen konnten, wird das visuelle Eingangssignal schon auf der Ebene des CGL aufgetrennt. Dabei wird Retinotopie beibehalten; das heißt, dass Impulse aus benachbarten rezeptiven Feldern der Retina auch im CGL benachbart ankommen. Die sechs Schichten des CGL empfangen ihre Eingangssignale jeweils entweder aus magnozellularen oder aus parvozellulären Ganglienzellen des ipsilateralen oder des kontralateralen Auges (**Abb. 1**).

Die Auftrennung erfolgt nach visuellen Merkmalen:

- *nach Bewegung:* M-Ganglienzellen, Schichten 1 und 2
- *nach Farbe, Textur, Muster und räumlicher Tiefe:* P-Ganglienzellen, Schichten 3 bis 6).

Primäre Sehrinde (V1). In der primären Sehrinde werden die visuellen Impulse nach Merkmalen des Ortes, der Länge, der Orientierung, der Ortsfrequenz und der Bewegung analysiert. Durch die Arbeiten von Hubel und Wiesel (Hubel, 1995) wurden bahnbrechende Erkenntnisse über den visuellen Kortex des Säugetiers zu Tage gefördert. Es gelang ihnen, mit Mikroelektroden aus dem Katzenhirn einzelne Neuronenaktivitäten abzuleiten, während sie das Tier auf verschiedene Muster schauen ließen. Sie entdeckten dabei drei verschiedene Arten von Zelltypen im primären visuellen Kortex:

- Einfache kortikale Zellen (*simple cells*) reagieren am stärksten auf Linien an einem bestimmten Ort und in einer bestimmten Orientierung (senkrecht, waagrecht oder diagonal).
- Komplexe Zellen (*complex cells*) antworten am stärksten auf Linien mit einer bestimmten Bewegungsrichtung. Sie empfangen Eingangssignale von mehreren über die Retina verteilten „simple cells“ und reagieren, wenn diese Signale einen bestimmten zeitlichen Abstand haben.
- „Endinhibierte Zellen“ (*endstopped* oder *hypercomplex cells*) antworten am stärksten auf Linien einer bestimmten Länge und Bewegungsrichtung. Hier liegt eine höhergradige Verschaltung aus einfachen und komplexen Zellen zugrunde, die diese raumzeitliche Reizverteilung detektiert.

Hubel und Wiesel haben die Neuronen in V1 auch als *Merkmalsdetektoren* bezeichnet. Das orientierungsselektive Antwortverhalten dieser Zellarten wird alleine durch den Aufbau und die Verteilung von erregenden und hemmenden Bereichen im rezeptiven Feld der Zelle erreicht. Das Konzept der Äquivalenz des Neurons mit dem Stimulus, der es am stärksten aktiviert, bildet seither eine der Grundlagen in der Beschreibung der neuronalen Repräsentation eines gesehenen Gegenstandes. Die neuronale Antwort ist auf diese Weise eine quantitativ beschreibbare Größe, die direkt aus dem Bild vorhergesagt werden kann.

V1 ist wie das CGL ebenfalls retinotop organisiert. Den rezeptiven Feldern aus der Fovea wird allerdings in V1 etwa fünfmal so viel Raum eingeräumt wie denjenigen aus der Retinaperipherie (*foveale Vergrößerung; kortikaler Vergrößerungsfaktor*). Hubel und Wiesel beschrieben auch die vertikale Organisation der primären Sehrinde in Säulen (*columns*), und zwar nach drei unterschiedlichen Reizeigenschaften: Zellen des gleichen Ortes auf der Retina (Positionssäulen), Zellen gleicher Merkmalsdetektion (Orientierungssäulen) und Zellen, die jeweils optimal auf eines der beiden Augen ansprechen (Augendominanzsäulen). Diese Säulen sind alle etwa 1 mm dick. Ein Kubikmillimeter-Block des Kortex bildet so, indem er jeweils eine Säule jeder Sorte zusammenfasst, ein Verarbeitungsmodul für Position und Orientierung. Eine solche so genannte *Hyperkolumne* bedient jeweils einen kleinen Teil der linken oder der rechten Netzhaut (**Abb. 2**).

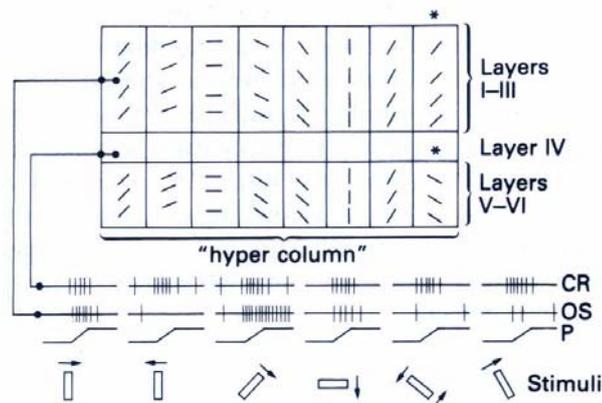


Abb. 2: Die Hyperkolumnen erstrecken sich durch die Schichten I bis VI des primären visuellen Kortex. Die Antworten einer orientierungs-selektiven Zelle aus Schicht II (OS) und einer orientierungsunempfindlichen Zelle aus Schicht IV (CR) sind dargestellt. Die Orientierungssensitivität verändert sich systematisch zwischen den einzelnen Kolumnen. In der mit * bezeichneten Kolumne antworten die Zellen der Ebene I-III anders als jene der Ebenen V-VI (aus Grüsser & Landis, 1991).

2.1.2 Auftrennung der Bildmerkmale im visuellen Assoziationskortex

Von der primären Sehrinde, die etwa 10% der ganzen Hirnoberfläche bedeckt, projizieren etwa 300 verschiedene Wege zu etwa 30 mit der Weiterverarbeitung befassten Zentren (Desimone & Ungerleider, 1989; Felleman & Van Essen, 1991). Dies sind nicht nur Projektionen der visuellen Verarbeitung in die visuellen Assoziationskortex, sondern auch Bahnen zu motorischen Zentren und zu Zentren anderer Sinnesmodalitäten. Dabei gilt wie beim CGL auch hier, dass nicht nur von „unten nach oben“ projiziert wird, sondern in demselben Maße auch von oben nach unten.

Ventraler und dorsaler Pfad. Ein Meilenstein auf dem Weg zur Kenntnis der visuellen Verarbeitung wurde durch Mishkin, Ungerleider & Macko (1983) gesetzt, indem sie die Existenz von zwei getrennten Verarbeitungsbahnen vom primären visuellen Kortex in den visuellen Assoziationskortex nachweisen konnten: dem so genannten „Was-Pfad“ (*ventral stream*) in den Temporallappen und dem „Wo-Pfad“ (*dorsal stream*) in den Parietallappen. Der ventrale Pfad ist für die Objekterkennung entscheidend, der dorsale Pfad für die Objektlokalisierung (**Abb. 3**).

Dass diese Bahnen nicht als gänzlich getrennt betrachtet werden können, forderten allerdings Milner & Goodale (1995), die in Ergänzung zu dem Konzept des „Was“ und „Wo“ auch die Beteiligung des dorsalen Pfades an der Eigenbewegungssteuerung, der Manipulation eines Objektes („Wie“) aufzeigten: Erst die Verbindung beider Pfade lässt eine sinnvolle Manipulation eines Objektes zu. Aufgrund der Tatsache, dass auch die motorische Manipulation bekannten optischen Illusionen unterliegt, ist diese Aufteilung jedoch nicht unumstritten (Franz, 2001).

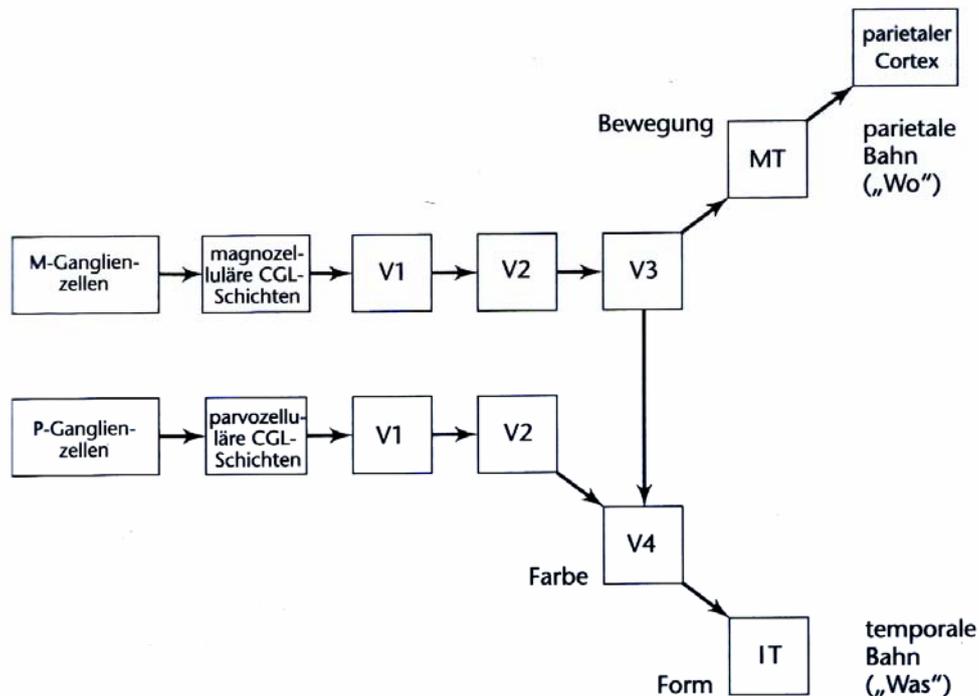


Abb. 3: Vereinfachtes Schema der beiden visuellen Verarbeitungsbahnen zum mediotemporalen Kortex (MT) und inferotemporalen Kortex (IT). Dazu bestehen allerdings zusätzlich zahlreiche Querverbindungen und Feedbackschleifen (aus Goldstein, 2002).

Die Komplexität der Signale, die zu einer Neuronendepolarisation führen, nimmt mit aufsteigender Verarbeitung zu. Die Anzahl der Komplexzellen ist in der sekundären Sehrinde (V2) höher als in der primären (V1) und die Anzahl der Hyperkomplexzellen in V4 größer als in V2 (Tanaka, 1997). Gleichzeitig nimmt auch die räumliche Größe der rezeptiven Felder stark zu, und die retinotopie Organisation geht verloren.

Der inferiore Temporkortex (IT). Die kompliziertesten der bisher bekannten Reiz-Reaktionsmuster wurden bei Primaten im Inferotemporkortex (IT) gefunden, der sich aufsteigend in einen posterioren (TEO) und einen anterioren Teil (TE) unterteilt.

In den 90er Jahren brachten die Arbeiten von Tanaka und Mitarbeitern eine Reihe von bemerkenswerten Ergebnissen zum Verständnis des Inferotemporkortex und der Art und Weise, wie dort Objekte repräsentiert werden (Tanaka, Saito, Fukuda & Moriya, 1991; Fujita, Tanaka, Ito & Cheng, 1992; Tanaka, 1996, 1997, 1997). Diese Forscher bestimmten „kritische Merkmale“ für einzelne Neuronen – dies sind Merkmale, die für eine Reizantwort notwendig sind und nicht weiter minimiert werden können, ohne die Zellantwort zum Erliegen zu bringen. Nachdem sich diese als einfache Grundformen herausgestellt hatten, war es möglich, zu zeigen, dass die Zellen des TE selektiv auf komplette, wenn auch primitive Objektformen reagieren (**Abb. 4**). Tanaka und Mitarbeiter unterschieden *primäre* Zellen, die am stärksten auf

einfache Formen antworten, wie Ellipsen, Punkte oder Rechtecke, und *elaborierte* Zellen, die spezifisch auf Formen, Farben und Textur in Kombination antworten. Beispielsweise war die Antwort einer Zelle am stärksten auf eine vereinfachte Form eines Apfels (ausgefüllter Kreis mit dünnem Stiel), jedoch nur schwach, wenn der Stiel dick war oder fehlte, oder der Kreis fehlte, oder an seiner Stelle ein Stern oder ein Rechteck am Stiel hing. Ebenfalls fand Tanaka Zellen, die ansichtsselektiv nur bei einem aufrecht dargebotenen Gesicht feuerten und nachließen, wenn das Gesicht gekippt oder um 180° gedreht war (Tanaka et al., 1991).

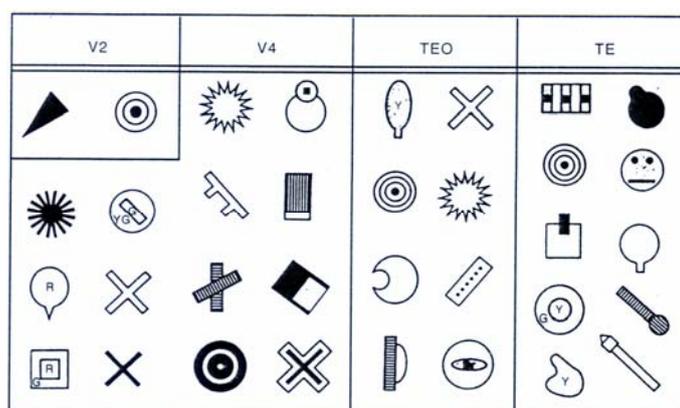


Abb. 4 Beispiele für komplexere „kritische Reizmerkmale“ von Neuronen in V2, V4, dem posterioren Inferotemporalkortex (TEO) und dem anterioren Inferotemporalkortex (TE), wie sie von Tanaka und Mitarbeitern verwendet wurden (aus Zigmond, Bloom, Landis, Roberts & Squire, 1999).

Diese für gesichtsähnliche Reize selektiven Zellen wurden im oberen Sulcus temporalis und im TE gefunden. Ihre Zellantwort hängt entscheidend davon ab, ob die Gesichtsmerkmale (Mund, dunkle Augen, kreisförmige Begrenzung) *in Kombination* vorliegen und ein vollständiges Gesichtsschema ergeben. Die Selektivität ist gestört, wenn diese Ordnung nicht vorliegt oder das Bild kontrastinvertiert ist.

Auch im TE zeigen benachbarte Zellen eine ähnliche Merkmalsselektivität, so dass in Analogie zu Hubel und Wiesel's Entdeckung ebenfalls von einer *columnaren* Organisation des TE gesprochen werden kann (Fujita et al., 1992).

Optical Imaging erlaubt, das nach dem Sauerstoffverbrauch desoxygenierte Hämoglobin auf der experimentell offengelegten Hirnoberfläche des Affen mittels Laserlicht direkt darzustellen. Mit dieser modernen Methode gelang es Tanaka (1997) sogar, eine Verschiebung der neuronalen Aktivität bei Drehung des dargebotenen Gesichtsmusters in Echtzeit nachzuweisen.

Sind bisher nur größenselektive, ortsselektive und ansichtsselektive Neurone erwähnt worden, so muss nun auch auf die Existenz von größeninvarianten, ortsinvarianten und ansichtsinvarianten Neuronen hingewiesen werden (Tanaka et al., 1991). Dies sind Neurone, die unabhängig von der Größe, dem Ort oder der Position des Objektes antworten. Im inferioren Temporalkortex gibt es also eine sensorische Codierung der Objekte. Man unterscheidet dabei *Einzelzellkodierung*: je ein Neuron reagiert selektiv auf einen Stimulus, und *Populations- oder Ensemblekodierung*: die

Kodierung ist auf ein ganzes Neuronennetzwerk verteilt. Das Muster des Feuerns unterscheidet die einzelnen Objekte voneinander.

Der laterale okzipitale Komplex (LOC). Diese okzipito-temporale Region wird als das Analogon des Menschen zum IT des Affen angesehen. Der LOC liegt am lateralen Rand des Gyrus fusiformis. Er ist ein weitgehend nicht-retinotopes Gebiet, in das sowohl Fasern des kontralateralen als auch des ipsilateralen Gesichtsfeldes projizieren. Wie unten aufgeführte Arbeiten zeigen, gilt der LOC als essentielles Areal für die Objekterkennung.

Der LOC wird stärker durch zusammenhängende Objektansichten und Objektfragmente aktiviert als durch bloße Texturen oder zufällig kombinierte Objektmerkmale. Dabei spielt es keine Rolle, ob die gesehenen Objekte vertraut oder neu sind (Grill-Spector, Kourtzi & Kanwisher, 2001). Neuere fMRI-Studien zeigen außerdem, dass die Aktivierung im LOC weitgehend unempfindlich ist gegenüber einem Wechsel von Größe, Position oder Konturmerkmalen der Objekte (Grill-Spector et al., 2001; Vuilleumier, Henson, Driver & Dolan, 2002).

Bisher ist nicht klar, ob es mehrere verschiedene Module gibt, die auf die Erkennung verschiedener Kategorien spezialisiert sind. Jedenfalls wurde eine Region identifiziert, die selektiv auf Gesichter reagiert: das fusiforme Gesichtsareal (fusiform face area, FFA) (Puce, Allison, Gore & McCarthy, 1995). Es gibt andererseits auch Anzeichen dafür, dass eine Aktivierung dieses fusiformen Gesichtsareals (FFA) nichts anderes als ein hohes Maß an Expertise bezüglich irgendeiner Kategorie widerspiegelt (Gauthier, Skudlarski, Gore & Anderson, 2000). Moore & Engel (2001) fanden, dass die Aktivität im LOC nach Präsentation von 3D-Körpern gegenüber der von 2D-Formen anstieg.

Der LOC spielt also eine wichtige Rolle in der Objekterkennung. Das Anschauen sowohl von unbekanntem als auch von bekannten Objekten aktiviert den LOC (Grill-Spector et al., 2001), zum Objekt gehörende *auditorische* Signale aktivieren den LOC jedoch *nicht* (Amedi, Jacobson, Hendler, Malach & Zohary, 2002). Diese Tatsache spricht dafür, dass der laterale okzipitale Komplex eine Art allgemeines System des Wahrnehmens der Objektform ist.

Der Präfrontale Kortex (PFC). Der Frontallappen kann als die Zielregion für die zwei bilderkennenden Verarbeitungswege angesehen werden, die beide im Okzipitallappen beginnen: den der räumlichen Zusammenhänge (dorsaler Pfad) und den der Objekterkennung (ventraler Pfad).

Die Funktion des präfrontalen Kortex ist die Auswahl des Verhaltens, abhängig vom Kontext und dem Vorwissen (Sturm, Herrmann & Wallesch, 2000). Er ist verbunden mit den meisten sensorischen, motorischen und untergeordneten Strukturen des Neokortex. Er integriert also sensorische Information aus verschiedenen Modalitäten mit der Information über den inneren Zustand des Organismus – Wachheit, Antrieb und Motivation – wie auch den emotionalen Gehalt eines Sinnesin-

gangs und seine anregende Bedeutung. So befindet sich der präfrontale Kortex in der Position, die kognitiven Prozesse zu kontrollieren, so dass die richtigen Handlungen zur rechten Zeit und am rechten Ort erfolgen können (Miller, 2000; Fuster, 2001). Die Rückprojektionen zu den sensorischen Arealen erlauben es dem PFC, einen „top-down“-Einfluss auf eine Vielzahl anderer Gehirnprozesse zu nehmen.

Die Neurone des PFC können ihre Aktivität über mehrere Sekunden aufrechterhalten. Sie können auf diese Weise „stehende Verbindungen“ zwischen verschiedenen, zeitlich voneinander getrennten Ereignissen bilden, die auch dann aufrechterhalten werden, wenn dazwischen andere, störende Ereignisse auftreten. Diese aktive Leistung unterscheidet das Arbeitsgedächtnis (working memory) von einer einfachen kurzfristigen Pufferung. Im IT dagegen wird aufrechterhaltene Aktivität viel leichter unterbrochen (Miller, 2000).

Läsionen des Frontalhirns haben neuropsychologisch einen Verlust des Integrationsvermögens in den sozialen Kontext zur Folge. Daraus ergibt sich meist auch ein unangemessenes Verhalten (Fuster, 2001).

2.1.3 Sehen als Analyse versus Sehen durch Schlussfolgern

Bis hier wurde die Verarbeitung eines Bildes als analytischer Prozess aufgefasst: Das Bild gelangt als neuronale Information über das CGL und die primäre Sehrinde V1 zu den Assoziationskortexen und wird dort in verschiedene Elemente aufgespalten. Die Assoziationskortexen sind alle hierarchisch miteinander verbunden, je höher man die Verarbeitungsstufen hinaufsteigt, desto komplizierter werden die Filteroperationen der jeweiligen Neuronen.

Die Neuronen ihrerseits sind charakterisiert durch ihre rezeptiven Felder, deren Beschaffenheit davon abhängt, in welcher Form die eingehenden Impulse aus den tiefer liegenden Schichten zusammengeführt werden. Dieses Modell gilt auch noch für die Neuronen im IT. Auf dieser hohen Ebene in der Verarbeitungshierarchie besitzen sie – aufgrund der konvergenten Verschaltung der Projektionen aus V4 und TEO – typischerweise sehr große rezeptive Felder und sind spezifisch für komplexe Stimuli.

So wird das visuelle System als ein Feedforward-Netzwerk betrachtet, durch das immer kompliziertere Eigenschaften des visuellen Eingangssignales extrahiert werden. Dieser Analyseprozess endet damit, dass schließlich die Identität eines gesehenen Objektes und seine Lage im Raum an die Zentren für Entscheidungsprozesse im Frontalhirn, wie den präfrontalen Kortex (PFC), weitergeleitet werden.

Zwischen allen Verarbeitungszentren innerhalb dieser vertikal-hierarchischen Struktur gibt es immer auch horizontale Verbindungen. Ihre Aufgabe ist es, die rein von unten nach oben (bottom-up) definierten Eigenschaften der rezeptiven Felder zu modulieren, abhängig von ihrer unmittelbaren Umgebung. Weiter sprechen Auf-

merksamkeitseffekte für einen gewissen Grad an Kontrolle von oben nach unten (top-down).

Diese funktionale Interpretation des visuellen Systems aus der Anordnung der kortikalen Zentren, die durch die Zahl ihrer Konnektivitäten (Felleman & Van Essen, 1991) definiert ist, berücksichtigt aber eine wichtige anatomische Tatsache überhaupt nicht: Die feedforward-Verbindungen sind, wie oben bereits erwähnt, nur ein kleiner Teil aller Verbindungen. Nur 5% aller erregenden Synapsen eines durchschnittlichen Neurons in V1 empfangen aufsteigende Signale aus dem CGL (Peters, Payne & Budd, 1994; Young, 2000). Alle anderen Verbindungen stammen aus V1 selbst oder aus den *höher liegenden* kortikalen und subkortikalen Zentren.

Dass in einem höchst optimierten System, wie der visuellen Verarbeitung, so viele Ressourcen an horizontale und an top-down Verbindungen vergeben werden, rechtfertigt die Annahme, dass diese Verbindungen viel mehr Funktionen haben müssen als nur die Neuronen zu modulieren.

Deshalb haben mehrere Autoren angenommen, dass das Sehen weitgehend ein Prozess des Schlussfolgerns (*Inference*) sein könnte (Neisser, 1996; Binkofski, Buccino, Posse, Seitz, Rizzolatti & Freund, 1999; Young, 2000). Eine aus verschiedenen Quellen gespeiste Schlussfolgerung und nicht nur eine einfache Analyse ist in der Tat zwingend notwendig, um 3D-Objekte aufgrund ihrer 2D-Projektionen auf der Retina erkennen zu können, denn dies ist theoretisch gesehen ein „unter-bestimmtes“ („ill-posed“) Problem mit nicht eindeutiger Lösung (Pizlo, 2001)(vgl. auch Kap. 3.2).

Der schlussfolgernde Prozess des Sehens basiert deshalb weitgehend auf bereits verfügbaren Kenntnissen über die gesehenen Objekte. Innerhalb eines solchen Systems signalisieren die Neuronen die Wahrscheinlichkeit des Auftretens eines Merkmals in der äußeren Welt. Dieses Schlussfolgern stützt sich auf die statistische Struktur der visuellen Welt und ist immer abhängig vom Kontext des Auftretens wie auch dem Vorwissen über gelernte Kategorien (vgl. auch Kap. 2.2.2).

Auf die dominante Rolle der Feedbackverbindungen in der visuellen Verarbeitung weisen besonders die großen Unterschiede im zeitlichen Reaktionsverhalten der höheren kortikalen Zentren hin. So zeigen die hierarchisch tiefstgelegenen Zentren keineswegs immer die kürzesten Latenzen, und die höher gelegenen die längeren, wie man es bei einer einfachen Bottom-up Verarbeitung erwarten würde. Insbesondere die Neurone in den höheren kortikalen Arealen wie MT, MST und FEF (frontal eye field) zeigen nur sehr kurze Verzögerungen (Bullier, 2001). Darauf aufbauend wurde gezeigt, dass es verschiedene Bahnen der Verarbeitung geben muss, eine schnellere und eine langsamere.

Die schnelle Bahn wird von M-Neuronen als „magnozelluläre Bahn“ von der gesamten Retina gespeist. Sie führt vom CGL über den mediotemporalen Kortex zum parietalen Kortex und entspricht dem dorsalen Pfad. Sie arbeitet zwar sehr rasch, aber ihre örtliche Auflösung ist nur gering. Sie registriert eine Art unscharfes hell-dunkel Bild, das nun top-down als „active blackboard“ für die nachfolgende, verzögert aufsteigende Feinverarbeitung dient. Diese wird durch die zweite, langsa-

me Bahn geleistet und führt von den p-Zellen im Zentrum der Retina als „parvozelluläre Bahn“ zum inferotemporalen Kortex (IT) in den Temporallappen. Diese Bahn entspricht dem ventralen Pfad, arbeitet viel langsamer, leistet aber eine hohe örtliche Auflösung, wodurch ein konturen- und detailreiches Bild vermittelt wird (Bullier, 2001).

Darüber hinaus gibt es eine Anzahl Arbeiten, die aufzeigen, dass die Antwort der Neuronen in den visuellen Arealen auf gegebene visuelle Impulse nicht konstant ist, sondern je nach Kontext des Erkennungsvorgangs durch verschiedene Einflussfaktoren verändert wird.

Lee und Mitarbeiter zeigten, dass die Neuronen in V1 sensitiver werden für shape-from-shading Stimuli, wenn diese für das Verhalten benutzt werden sollen (Lee, Yang, Romero & Mumford, 2002). Die Forscher korrelierten neuronale Signale mit langer Latenz in V1 und V2 mit Verhaltensleistungen von Affen. Sie schließen daraus, dass die Sensitivität für die dreidimensionale Form über Neurone in V1 durch Feedbackverbindungen aus V2 und anderen extrastriären Arealen vermittelt wird (Lee et al., 2002).

In einer Diskriminationsaufgabe unter der Benutzung von Liniensegmenten fanden Li, Piech & Gilbert (2004), dass Neuronen in V1, abhängig von den Attributen der gelernten Formen, ihre funktionalen Eigenschaften wechseln können. Diese Eigenschaften hingen außerdem von der ausgeführten Wahrnehmungsaufgabe ab, so dass auf ein und denselben visuellen Stimulus sehr verschiedene Antwortmuster in V1 auftreten können.

Ähnliche Ergebnisse wurden auch für den IT erhalten, wo die Neuronen zuerst globale Eigenschaften kodieren (Affengesicht vs. Menschengesicht) und mit einer Verzögerung von ungefähr 50 ms detaillierte Eigenschaften, wie Identität oder Gesichtsausdruck (Sugase, Yamane, Ueno & Kawano, 1999). Auch nach diesem Experiment stellen die Autoren die Hypothese auf, dass die globale Information als „Schlagzeile“ agiere, um den Verarbeitungspfad in die höheren kortikalen Areale zu leiten und die Zielareale vorzubereiten auf die nachfolgende, detaillierte Information.

Die Neuronen im IT zeigen außerdem eine Sensibilisierung auf die zu erfüllende Kategorisierungsaufgabe: Sie antworten nicht selektiv auf einen bestimmten Stimulus oder eine inhaltliche Kategorie, sondern sie werden auf Merkmale abgestimmt, die für die jeweilige Aufgabe diagnostisch sind (Sigala, 2004).

All diese Ergebnisse unterstützen die Ansicht, dass auf der neurophysiologischen Ebene die Beschreibung einer reinen Feedforward-Verarbeitung, die auf hierarchisch geordneten Filtern aufsteigender Komplexität basiert, nicht ausreichend ist, um die essentiellen Eigenheiten des menschlichen Sehsystems zu erklären. Stattdessen muss auch auf der physiologischen Ebene ein adaptierendes Verhalten mitberücksichtigt werden, das durch den Kontext des Erkennungsvorganges bestimmt wird. Die Funktion der einzelnen visuellen Areale kann während des zeitlichen Verlaufs der visuellen Analyse ebenso wechseln wie auch bedingt durch das Vorwissen und die Umstände der visuellen Erkennung.

2.1.4 Multimodale Objektrepräsentation: Haptische und visuelle Wahrnehmung

Wurde klassischerweise jede der Sinnesmodalitäten einzeln betrachtet, unter Vernachlässigung der Querverbindungen zu anderen Modalitäten, so rückt nun in der Forschung vermehrt die Tatsache ins Bewusstsein, dass das visuelle System keineswegs isoliert arbeitet. Die Umwelt, in der wir leben, bietet sich nicht nur einer einzigen Sinnesmodalität an, sondern viele Informationen werden von mehreren Modalitäten gleichzeitig oder ergänzend vermittelt. Objekterkennung ist keineswegs nur ein visueller Vorgang. Wir können zum Beispiel sichtbare Objekte auch über unseren Tastsinn erfassen. Dieser beinhaltet allerdings verschiedene Sinnessysteme.

Nach Loomis & Lederman (1986) wird der Tastsinn in drei Systeme unterteilt, den kutanen Sinn, den kinästhetischen Sinn und den haptischen Sinn. Der kutane oder taktile Sinn vermittelt über die Hautrezeptoren Qualitäten wie Oberflächenbeschaffenheiten (rau/glatt und hart/weich) und Temperatur. Der kinästhetische Sinn liefert über die Rezeptoren in den Muskeln, Sehnen und Gelenken Information über die Stellung und Bewegung der eigenen Gliedmaßen. Der haptische Sinn benutzt beide Systeme, die Hautrezeptoren und die kinästhetischen Rezeptoren zur Erkundung von Objekten. Eine motorische Komponente ist dabei unentbehrlich, wie Lederman & Klatzky (1998) ihre grundlegenden Forschungen zusammenfassen. Das ertasten präziser Information und die Identifikation komplexer Objekte erfordern willentliche, explorative Bewegungen, die ihrerseits durch die Wahrnehmungsziele des Beobachters gelenkt werden. Dabei wird Handlung im Dienst von Wahrnehmung ausgeführt, abhängig von der Verarbeitung der kutanen und kinästhetischen Reize durch das haptische System¹ (Lederman & Klatzky, 1998).

Die Vernetztheit der Sinneseindrücke des visuellen und haptischen Systems wird zunehmend untersucht. Dieses Forschungsfeld profitiert zudem stark von den modernen Möglichkeiten der funktionellen Kernspintomographie.

In den peripheren sensorischen Arealen ist die Informationsverarbeitung strikt modalitätsspezifisch. Zangaladze und Mitarbeiter zeigten aber, dass bei einer Unterbrechung der Funktion des okzipitalen Kortex durch den Gebrauch transkranieller Magnetstimulation (TMS) nicht nur die visuelle, sondern auch die haptische Objekterkennung gestört ist, nicht aber die haptische Diskrimination im allgemeinen (Zangaladze, Epstein, Grafton & Sathian, 1999). Dies beweist die wichtige Rolle des visuellen Kortex auch für die haptische Objekterkennung.

Binkofski et al. (1999) fanden beim blinden Manipulieren von komplexen Objekten eine Aktivierung nicht nur des ventralen prämotorischen Kortex (vPMC)

¹ Es soll hier kurz erwähnt werden, dass der Begriff „Haptik“ in verschiedenen Disziplinen sehr unterschiedlich belegt ist: Im Kontext der Kognitionspsychologie wird Haptik meist als manuelles Formerkennen verwendet. Im Kontext der virtuellen Realität beschreibt er das künstliche Vermitteln einer Tastempfindung. Im technisch-industriellen Kontext beschäftigt sie sich mit dem Design von handgerechten Schaltern und Knöpfen.

und der sensorischen Areale S1 und S2, sondern auch des superioren Parietallappens. Sie vermuteten deshalb bei Primaten und Menschen die Existenz eines frontoparietalen Kreislaufs (fronto-parietal circuit). Diese Aktivierung trat übrigens nur bei der Manipulation von komplexen Objekten auf, nicht aber bei der Manipulation einer zum Vergleich verwendeten glatten, einfachen Kugel. Frühere Studien, die eher einfache Objekte benutzt hatten und nur kurz dauernde Bewegungen verlangten, fanden diesen Kreislauf nicht.

Vom LOC wurde anfangs angenommen, dass es ein Areal sei, das der rein visuellen Verarbeitung dient. Neuere fMRI-Studien haben allerdings nachgewiesen, dass das haptische Befühlen von Objekten ebenfalls Areale aktiviert, die an den LOC angrenzen und ihn sogar überlappen.

Amedi und Mitarbeiter (Amedi, Malach, Hendler, Peled & Zohary, 2001) verglichen die Aktivierungsunterschiede beim Anschauen von bloßer Textur und von Objekten (die klassische Methode, um den LOC mittels fMRI nachzuweisen) mit derjenigen beim haptischen Befühlen von Textur und von Objekten. Sie fanden eine somatosensorische Aktivierung in der okzipito-temporalen Region, die beim Befühlen der Objekte stärker war als beim Befühlen der Textur. Die meisten aktivierten Voxels lagen innerhalb des LOC, die anderen daran angrenzend. In einem weiteren Experiment gelang es diesen Autoren zu zeigen, dass diese multimodale Überlagerung in zwei Punkten spezifisch war: (1) Diejenigen Voxels, die sowohl durch taktile als auch durch visuelle Stimuli aktiviert wurden, reagierten eher auf greifbare Objekte als auf Gesichter oder Häuser. (2) Sinnvolle akustische Stimuli, die aber nicht zum Objekt gehörten, bewirkten hingegen keine Aktivierung dieses Areals (Amedi et al., 2001).

Eine weitere fMRI-Studie untersuchte die Interaktion zwischen der visuellen und der haptischen Sinnesmodalität, indem sie neuartige Knetobjekte verwendete (James, Humphrey, Gati, Servos, Menon & Goodale, 2002). Eine rein haptische Objektexploration (mit verbundenen Augen) bewirkte hier einerseits eine Aktivierung in den bekannten somatosensorischen Gebieten, aber auch in okzipitalen Regionen. Die Aktivierungen während einer rein visuellen und einer rein haptischen Vorerkundung überlagerten sich dabei im mittleren okzipitalen Areal des LOC. Wurden die Objekte nach der visuellen Vorerkundung (Priming) angeschaut, konnte eine stärkere Aktivierung in medialen und lateralen okzipitalen Regionen des LOC gemessen werden. Das gleiche galt aber auch, wenn die Objekte nur haptisch vorerkundet wurden. Deshalb vermuten die Autoren eine Mitnutzung der visuellen Objektrepräsentation im ventralen Pfad durch die haptische Objektwahrnehmung.

2.2 Interne Objektrepräsentation und kategorielles Objektwissen

Die im Kapitel 2.1 dargestellten neurobiologischen Forschungsergebnisse zur Wahrnehmung eines Objektes zeigen ein dominierendes Prinzip in der Wahrnehmungsverarbeitung: die Dissoziation. Alle Qualitäten des wahrgenommenen Objektes werden einzeln verarbeitet: die Farbe, die Form, die Bewegung, der Ort, und auch die Bedeutung. Für jede dieser Objekteigenschaften gibt es ein zuständiges Hirnareal. Und doch sehen wir einen Hund auf der Wiese und nehmen nicht „Fell“, „braun“, „grün“, „springend“, „Tier“, „schnell“, „rechts“ als einzelne Komponenten wahr.

Wie ist es möglich, dass wir trotz aller dissoziativen Verarbeitung ein einziges kohärentes Objekt wahrnehmen? Diese Frage wird gegenwärtig als das so genannte „Bindungsproblem“ diskutiert. Sie ist wiederum eng verbunden mit der Frage, was die „interne“ Repräsentation eines wahrgenommenen Gegenstandes ist.

Wie diese interne Repräsentation auf psychologischer Ebene als kategorielles Objektwissen strukturiert ist, wird im zweiten Teil dieses Kapitels unter „Kategorie und Klasse“ dargestellt.

2.2.1 Repräsentation und Bindungsproblem

Die „interne“ oder „mentale Repräsentation“ eines Gegenstandes ist die Form seines Auftretens im Bewusstsein. Die Frage nach dem *Format* dieser mentalen Repräsentation ist allerdings die gemeinsame Frage aller Disziplinen, die unter dem Namen „Kognitionswissenschaften“ zusammengefasst werden. Entsprechend vielgestaltig sind die möglichen Antworten.

Scheerer (1993) unterscheidet grob vier verschiedene Verwendungen des Begriffes „Repräsentation“. (1) Repräsentation als „inhaltsvoller mentaler Zustand“, (2) „Re-Präsentation“ (wiederholte Abbildung), (3) „strukturerhaltende Darstellung“ oder (4) „Stellvertretung“ des Wahrgenommenen.

Unter (3) „strukturerhaltende Darstellung“ fallen einerseits die bildliche oder analoge Darstellung und andererseits auch die nichtbildliche Darstellung nach Regeln (propositionale Darstellung). Beide Verwendungsformen sind geistesgeschichtlich alt: Vorstellung wurde schon immer mit inneren Bildern gleichgesetzt; andererseits verwendet auch schon Leibniz „Repräsentation“ als „strukturerhaltende, isomorphe ‚Abbildung‘ im mathematischen Sinn“ (Scheerer, 1993).

In der so genannten *imagery*-Debatte stehen sich noch heute Piktoralisten und Propositionalisten (mit den Hauptvertretern Stephen Kosslyn und Zenon Pylyshyn) gegenüber. Es geht dabei um die Entscheidung, welche der beiden Positionen zur Repräsentation bildlicher Vorstellungen eher experimentell bestätigt werden kann: (1) Vorstellungen sind reaktivierte, aber analoge sensorische Sinnesbilder, die direkt inspiziert und interpretiert werden können (Kosslyn) oder (2) Vorstellungen sind strukturell-symbolische Beschreibungen modalitätsspezifischer Information. Diese

wird analog interpretiert, ist aber als solche eine propositionale Wahrnehmung (Pylyshyn) (Jüttner, 2003).

Das Bindungsproblem. Steht die „mentale“ oder „interne Repräsentation“ eines Gegenstandes für die Form seines Auftretens im Bewusstsein, so beschreibt die „neuronal Repräsentation“ dagegen die damit korrelierende Form der neuronalen Aktivierungsverteilung. Die Untersuchungen der neuronalen Repräsentation basieren hauptsächlich auf der Analyse von Aktivitätsmustern *einzelner* Neurone, die eines nach dem anderen abgeleitet werden (Singer, 1999). So ist es nur natürlich, dass zwar für bestimmte Stimuli spezifisch reagierende Neurone gefunden wurden, aber es ist nicht viel darüber bekannt, in welcher Weise die vielen betroffenen Neuronen zeitlich und örtlich miteinander interagieren, und welche Rolle diese Interaktion für die neuronale Repräsentation überhaupt spielt. Hier könnte sich eine mögliche Lösung des Bindungsproblems zeigen: Nicht die Neuronen selbst repräsentieren die Wahrnehmungsinhalte, sondern die Muster ihrer Synchronisation (Singer, 1999).

Der Erkennungsprozess als Abgleich von Eingangsreiz und Repräsentation. Für die *Objekterkennung* schließt sich hier ein weiteres ungelöstes Problem an: Die mentale Repräsentation eines Objektes wird nicht nur für den Aufbau eines inneren Objektbildes verwendet, sondern ist ebenso für das Erkennen des Objektes von essentieller Bedeutung. Beim Erkennungsvorgang wird das Sinneseingangssignal, das als neuronales Erregungsmuster vorliegt, mit der gespeicherten Objektrepräsentation abgeglichen (sog. *Matching-Prozess*): Stimmen beide überein, wird das Objekt erkannt.

Dies bedeutet gleichzeitig, dass die Repräsentation grundsätzlich dieselben Objektmerkmale verwenden muss, wie sie auch für die Objekterkennung durch die Sinnesorgane geliefert werden. Für den Vorgang dieses Abgleichs gibt es grundsätzlich zwei verschiedene Möglichkeiten:

- (1) Aus den Eingangssignalen wird ein Objektmodell synthetisiert und dieses mit der gespeicherten Objektrepräsentation abgeglichen.
- (2) Aus der gespeicherten Objektrepräsentation werden mögliche Eingangssignale generiert und diese dann mit den tatsächlichen Eingangssignalen verglichen.

Im Bereich des maschinellen Sehens sind diese zwei Möglichkeiten als (1) *inverse modelling* und als (2) *feedforward modelling* bekannt.

Beim Verständnis des Erkennungsvorganges als eines Abgleichprozesses zwischen dem neuronalen Erregungsmuster und der Objektrepräsentation sind zwei Verhältnisse von Interesse: das Verhältnis der Eingangsdaten zum gespeicherten Modell und das Verhältnis von „neuronaler“ Objektrepräsentation zur „mentalen“ Objektrepräsentation.

2.2.2 *Kategorie und Klasse*

Der Begriff der Kategorie hat sich in seiner Bedeutung verändert. Aristoteles gilt als der Urheber einer Kategorienlehre, die die objektive Realität in allgemeinsten Zügen systematisch formuliert. Aus der Analyse des Sprechens über die Dinge entwickelte er insgesamt zehn verschiedene Kategorien (*categorias*, gr. Aussage), die die ontologischen Grundprinzipien auf allgemeinsten Ebene beschreiben. Sie sind die "höchsten" und "ersten" Begriffe, in denen sich fundamentale Bestimmungen der Dinge ausdrücken. Neben der wichtigsten, der subjekterzeugenden „Substanz“ nennt er ihre sog. Akzidentien: Qualität, Quantität, Relation, Ort, Zeit, Zustand, Lage, Wirken (aktiv) und Leiden (passiv). Die Aristotelischen Kategorien bestimmten grundlegend alles weitere Philosophieren über Kant und Hegel bis ins 20. Jh. Sie standen aber, wegen ihrer Eigenschaft als die Grundfesten möglichen Denkens überhaupt, über die Jahrhunderte hinweg nie selbst als Gegenstände einer Theorie zur Diskussion.

Erst Ludwig Wittgenstein (2003) rüttelte mit seiner Überzeugung, dass die Grenzen der Sprache auch die Grenzen der beschreibbaren Welt markieren, zum ersten Mal an der Klarheit der Trennlinien zwischen den einzelnen Begriffen, Arten und Klassen. Mit seinem Vorgehen, die „Bedeutung des Wortes aus dessen Gebrauch in der Sprache“ zu definieren, kann er als der Begründer der modernen kognitiven Kategorisierungstheorien gelten. Die Kategorien sind nun nicht mehr die grundlegenden Prinzipien des Denkens im Aristotelischen Sinn, sondern werden synonym zu „Klasse“ verwendet. „Kategorisieren“ heißt klassifizieren, einordnen des Wahrnehmungsobjektes in einen Bedeutungszusammenhang.

Die kognitiven Kategorisierungstheorien nahmen ihre entscheidende Entwicklung dann hauptsächlich in den Arbeiten der amerikanischen Psychologin Eleanor Rosch (1975; 1976). Ihr Verdienst bestand im Schaffen einer konsequenten Methodik, die eine Reihe von überraschenden Effekten offen legte.

Ausgangspunkt ihrer Kritik an den „klassischen Kategorisierungen“ war das Argument, dass, wenn eine Klasse nur durch gemeinsame Eigenschaften ihrer Mitglieder bestimmt sei, es nicht sein konnte, dass einzelne Mitglieder typischere Exemplare dieser Klasse sind als andere. Ihr zweiter Kritikpunkt war, dass die klassische Einteilung nur durch Eigenschaften der Klassenmitglieder selbst bestimmt sei und dabei die Besonderheiten des kategorisierenden Subjekts selbst außer Acht gelassen werde: die menschliche Neurophysiologie, die menschlichen Fähigkeiten zur Wahrnehmung, zur Bildung mentaler Bilder, zum Lernen, Erinnern und Organisieren des Gelernten und zur Kommunikation.

Mit ihren Arbeiten schuf Rosch eine allgemeine Perspektive auf ein Problem, das neben Psychologen gleichermaßen auch Linguisten, Anthropologen und Philosophen interessierte. Ihre Theorien sind seither unter dem Namen „Prototypen-Theorie“ bekannt geworden und umfassen vor allem zwei Themenkreise: den „Basic-level Effekt“ und den „Prototypen Effekt“.

Der Basic-level Effekt. Die erste Beschreibung dieses Effektes ist Roger Brown zu verdanken (Brown, 1958). Er hatte beobachtet, dass es eine „erste Ebene“ gibt, in der Kinder lernen, die Dinge ihrer Umwelt zu benennen. Diese erste Ebene ist weder die allgemeinste noch die speziellste, sondern wird charakterisiert durch unverkennbare Merkmale und kürzere und häufiger benutzte Namen. Brown nannte diese Kategorisierungsebene „die natürliche“ und sie wurde später der „Basic-level“ oder auch „Entry-level“ genannt:

Wirbeltier - Säugetier – *Hund* – Dackel
Möbel – *Tisch* – Gartentisch
Pflanze – *Baum* – Buche.

Auch die Arbeiten von Rosch betonten, dass die mittlere Ebene der taxonomischen Hierarchie psychologisch die wichtigste sei, im Gegensatz zur traditionellen Theorie, die alle Klassen als gleichwertig behandelte.

Die mittlere Ebene (der Basic-level) ist nach Rosch et al. (1976) definiert als

- die höchste Ebene, auf der die Kategorienmitglieder eine ähnlich wahrgenommene Gesamtgestalt haben
- die höchste Ebene, auf der eine einzige Vorstellung die Kategorie voll repräsentieren kann,
- die höchste Ebene, auf der man mit ähnlichen Handlungen mit allen Kategorienmitgliedern interagieren kann,
- die Ebene, auf der die Kategorie am schnellsten klassifiziert wird,
- die Ebene mit den gebräuchlichsten Bezeichnungen für die Kategorienmitglieder,
- die Ebene, die zuerst von Kindern gebraucht und verstanden wird,
- die Ebene, die ohne weitere Erklärung ohne Kontext gebraucht werden kann (Bsp. er ist vom Baum gefallen),
- die Ebene, auf der das meiste Wissen enthalten ist.

Die Taxonomie ist also nicht nur ein Mittel, die uns umgebenden Erscheinungen zu systematisieren, sondern die Mitte der taxonomischen Hierarchie erfüllt gleichzeitig so grundlegende Aufgaben wie Gestaltwahrnehmung und Vorstellungsfähigkeit, Wissensrepräsentation und Merkfähigkeit, Kommunikation und Bewegungsinteraktion.

Basic-Level-Kategorien bestehen nie abstrakt, sondern ihre Strukturen wachsen vielmehr im kulturellen Gebrauch. Sie werden jeweils davon bestimmt, wie das Verhältnis vom Teil zum Ganzen wahrgenommen wird.

Der Prototypen-Effekt. Rosch konnte experimentell nachweisen, dass Klassen in der Regel typische, „beste“ Exemplare besitzen, was sie methodisch auf die folgende Weise erreichte:

- Sie ließ ihre Probanden einige Unterarten einer Klasse auf einer Skala danach bewerten, wie gut sie als Beispiel für die Klasse dienten.

- Die Reaktionszeit auf die Frage, ob ein bestimmtes Beispiel einer Klasse angehöre, wurde gemessen.
- Das freie Auflisten von Beispielen für bestimmte Klassen zeigte signifikante Häufungen bei „typischen“ Exemplaren.
- Asymmetrie nach Ähnlichkeit: Die Ähnlichkeit von weniger typischen Beispielen zu typischen innerhalb einer Klasse wurde als größer angesehen als die Ähnlichkeit von typischen zu weniger typischen.
- Asymmetrie in der Generalisierung: Neue Aussagen über ein Klassenmitglied wurden eher vom untypischen zum typischen hin übertragen als umgekehrt.
- Rosch gelang es, empirisch zu demonstrieren, dass die Wittgenstein'sche „Familienähnlichkeit“ nicht nur eine philosophische a-priori Spekulation war: Die wahrgenommene „Familienähnlichkeit“ zwischen Klassenmitgliedern einer Klasse korrelierte mit der numerischen Häufigkeit, wie „beste“ Beispiele dieser Klasse eingeordnet wurden.

Ludwig Wittgenstein hatte in seinen „Philosophischen Untersuchungen“ anhand der Beispiele der Begriffe Sprache, Spiel und Sprachspiel aufgezeigt, dass bestimmte Kategorien von Dingen mit einer taxonomischen Klassifikation (hierarchischen Systematik) nicht hinreichend erfasst werden können (Wittgenstein, 2003). Es gibt nach ihm keine allgemeinen Merkmale, die für alle Sprachen, Spiele und Sprachspiele gleichermaßen gelten. Es gibt zwar einige Spiele mit gemeinsamen Merkmalen, die aber mit anderen wiederum überhaupt keine Gemeinsamkeiten aufweisen. Diese Form von lockerer Gemeinsamkeit nannte er Familienähnlichkeit (*family resemblance, cluster definition*). Die Mitglieder einer Klasse sind wie Mitglieder einer Familie, deren Ähnlichkeit sich durch den statistisch verteilten Anteil an einer Gesamtheit von Merkmalen bestimmt.

Allerdings sind Prototypenkonzept und Wittgenstein'sche Familienähnlichkeit nicht vollständig vereinbar. Unter den Mitgliedern einer Familie gibt es kein „bestes Beispiel“, keine typischeren und weniger typischen Exemplare. Ein Prototyp einer Klasse wäre darin ein Spezialfall.

Durch Roschs Arbeiten war die Existenz von Prototypen auch experimentell belegt. Barsalou (1983) zeigte sogar, dass Prototypen-Effekte nicht nur für wichtige Dinge unserer Kultur gelten, sondern auch für die von ihm so benannten „ad-hoc Kategorien“ wie zum Beispiel „was man zum Geburtstag schenkt“ oder „was man aus dem Haus rettet, wenn es brennt“. Dieses Phänomen ist umso erstaunlicher, als diese spontan entworfenen Kategorien, für die es kein klares Wort gibt, eigentlich keinerlei Präexistenz haben sollten.

Wie waren diese Effekte nun zu interpretieren? Zuerst nahm Rosch an, die Prototypen entstünden auf der Wahrnehmungsebene durch hervorspringende Merkmale, gute Erinnerbarkeit und Verallgemeinerungstauglichkeit der Exemplare. Insbesondere wurde aber die Frage diskutiert, inwieweit der Prototypen-Effekt auch die Kategorie charakterisiere, in der Form, wie sie mental repräsentiert ist. Rosch bejahte zunächst diese Hypothese, indem sie von den Annahmen der informationsverarbeitenden Psychologie ausging (Lakoff, 1987).

Später nahm Rosch aber von einer eins-zu-eins Entsprechung von Prototyp und Repräsentation wieder Abstand. Prototypen bestimmen zwar die Gedächtnisstrukturen, sind aber nicht hinreichend für das Umgrenzen der Repräsentation. Es müsse darüber hinaus angenommen werden, dass hinter dem Prototyp eine nicht näher bestimmbare Quelle liegt, die erst möglich macht, dass man bei Kategorien mit klarer Begrenzung auch vom Prototyp stark abweichende Exemplare (beispielsweise ein Pinguin in der Kategorie „Vogel“) noch zweifelsfrei als innerhalb der Kategorie liegend bestimmen kann.

Die Überzeugungskraft der Prototypen-Effekte führte dazu, dass die Prototypen-Theorie in den kognitiven Theorien zu Repräsentation, Gedächtnis und Lernen einen festen Platz eingenommen hat, obwohl sie bisher durch kein Modell hinreichend bestimmt werden konnte.

Generalisierung. Sind die verschiedenen Kategorien gelernt, dann ist für ein erfolgreiches Verhalten von entscheidender Wichtigkeit, ob das kategoriale Wissen auch generalisierbar ist. Generalisierung bedeutet die Fähigkeit, Wissen, das von bestimmten Exemplaren einer Kategorie erlernt ist, auf andere ähnliche Exemplare anzuwenden (Squire, 1992). Dabei ist immer auch eine Diskrimination erforderlich, denn Exemplare, die zwar ähnlich sind, aber nicht zur gefragten Kategorie gehören, müssen von solchen unterschieden werden, die zur Kategorie gehören, auch wenn jene unähnlicher sind als diese. Die Eigenschaft der Kategorie muss also für eine erfolgreiche Generalisierung als invariant erkannt werden. Die Generalisierungsfähigkeit gilt deshalb als ein Beweis für die Existenz einer präzisen mentalen Repräsentation und somit einer erfolgreichen Objekterkennung.

3 Theorien der Objekterkennung

Die moderne psychologische Forschung zur Objekterkennung ist wesentlich durch die Erkenntnisse inspiriert worden, die aus den Bemühungen um eine maschinelle Bilderkennung geflossen sind. Aus diesem Grund werden hier, nach einer Beschreibung der Gestaltphänomene, auch die Prinzipien maschineller Bilderkennung kurz behandelt. Im Anschluss daran werden die wichtigsten psychologischen Theorien zur Objekterkennung dargestellt, in deren Kontext auch die hier berichteten Experimente stehen.

3.1 Was das Objekt zusammenhält - die Gestalttheorien

Die vorherrschende psychologische Schule des ausgehenden 19. Jh. war der Assoziationismus. Er postulierte in Anlehnung an John Locke und Thomas Hobbes einen vollkommenen Parallelismus zwischen Reiz und Empfindung, zwischen physiologischen und psychologischen Abläufen (Assoziation). Das Erleben baute sich danach additiv aus elementaren Sinnesempfindungen auf („Empfindungsatomismus“).

Die Gestaltpsychologie entstand als Gegenreaktion zu dieser mechanistischen Anschauung. Max Wertheimer verhalf ihr 1912 zu grundlegender Evidenz, indem er das Phänomen der Scheinbewegung beschrieb, wie sie zwischen zwei abwechselnd blinkenden, benachbarten Streifen auftritt. Es sieht dann so aus, als ob sich ein Streifen hin- und herbewegt. Die wahrgenommene Bewegung des Lichtstreifens könne aber keinerlei physiologische Entsprechung haben, da es dort keine Lichtreize gibt. Wertheimer folgerte daraus, dass Wahrnehmung allgemein dem Euklid'schen Satz folgen müsse, „Das Ganze ist mehr als seine Teile“. Auch das Phänomen der Scheinkonturen wurde anschließend oft als bestätigendes Argument für dieses Prinzip verwendet (*Abb. 5*).

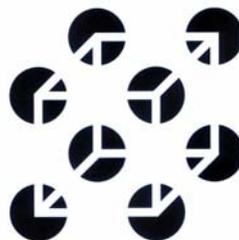


Abb. 5 Ein berühmtes Beispiel für Scheinkonturen. Der Würfel besteht nicht realiter, die durchgehend wahrgenommenen Kanten werden ergänzt (aus Goldstein, 2002).

Wertheimer beschreibt in seiner Schrift „Untersuchungen zur Lehre von der Gestalt“ (Wertheimer, 1922; Wertheimer, 1923) die wichtigsten Faktoren, die eine Wahrnehmung zur Gruppierung von Einheiten benutzt. Er nennt sie Gestaltfaktoren. Die wichtigsten seien hier kurz aufgeführt:

- *Prägnanz oder gute Gestalt (Einfachheit)*: Jedes Reizmuster wird so gesehen, dass die resultierende Struktur die einfachstmögliche ist.
- *Ähnlichkeit*: Ähnliche Dinge erscheinen in zusammengehörenden Gruppen geordnet.
- *Gestaltgerechte Linienfortsetzung*: Punkte, die als gerade oder sanft geschwungene Linien gesehen werden, wenn man sie verbindet, werden als zusammengehörig wahrgenommen. Linien werden tendenziell so gesehen, als folgten sie dem einfachsten Weg.
- *Nähe*: Dinge, die sich nahe beieinander befinden, erscheinen als zusammengehörig.
- *Gemeinsames Schicksal*: Dinge, die sich in gleicher Geschwindigkeit in die gleiche Richtung bewegen, erscheinen als zusammengehörig.
- *Vertrautheit*: Dinge bilden mit größerer Wahrscheinlichkeit Gruppen, wenn die Gruppen vertraut erscheinen oder etwas bedeuten.

Parallel mit der so beschriebenen Gruppierung der Elemente geht auch das Unterscheiden einer Figur vom Hintergrund einher. Die *Figur-Grund-Trennung* war für die Gestaltpsychologie als Grundlage zur Objekterkennung von entscheidender Bedeutung. Sie wurde anhand von Kippfiguren untersucht: zweideutige Bilder, auf denen Figur und Hintergrund gegenseitig ihre Rolle vertauschen können. Die Figur wirkt dinghafter und kommt schneller ins Gedächtnis als der Hintergrund. Sie wird als vor dem Hintergrund stehend wahrgenommen, der seinerseits als ungeformtes Material gesehen wird und die Konturen, die die Figur vom Hintergrund trennen, scheinen zur Figur zu gehören (**Abb. 6**).



Abb. 6 Version einer Kippfigur. Es können wahlweise eine Vase oder zwei Gesichter gesehen werden (aus Goldstein, 2002).

Obwohl die Erkenntnisse der Gestaltpsychologie beinahe 100 Jahre alt sind, gelten sie auch heute noch als wertvoll. Sie zeigen bereits die komplexe Struktur der Informationsverarbeitung auf und weisen auf die Beteiligung von präkognitiven und automatischen Prozessen hin.

Die Gestaltfaktoren und Gestaltgesetze werden heute als eine Sammlung von Phänomenen beim Bilderkennen aufgefasst. Die Voraussetzung für diese Phänomene liegt in der Tatsache begründet, dass die uns umgebende Welt nicht regellos und zufällig ist, sondern vielmehr physikalischen, biologischen und sozialen Gesetzen gehorcht, auf die sich das visuelle System aus Erfahrung verlassen kann. Trotzdem ist die Frage bisher ungelöst, wie während des Erkennungsvorgangs entschieden wird, welche dieser Regeln bei der Auswertung eines Bildes jeweils zur Anwendung kommt.

3.2 Techniken im maschinellen Bilderkennen

Wesentliche Anregungen zur Beantwortung der Frage, wie das visuelle System funktionieren könnte, kommen aus dem Bereich des maschinellen Sehens (*computer vision*). In diesem Fachbereich kann die Tauglichkeit eines Funktionsmodells direkt an seiner Leistung überprüft werden, weil hier algorithmisch durchgeführt werden muss, was andernorts nur beschreibende Theorie bleibt (Caelli, Johnston & Robison, 1993). Jedoch verfolgen diese Bestrebungen nicht unbedingt die Frage, wie das biologische visuelle System arbeitet, sondern sie entstehen aus den Bedingungen ingenieurtechnischer Anwendungen und verstehen sich als reine Möglichkeit einer tatsächlich funktionierenden effizienten Lösung (Caelli et al., 1993). Dies sollte in der Diskussion immer beachtet werden.

Die überraschend großen Schwierigkeiten, die es beim maschinellen Erkennen eines einfachen Bildes zu überwinden gilt, sind vor allem auf das Problem zurückzuführen, dass aus einem zweidimensionalen Bild dreidimensionale Strukturen

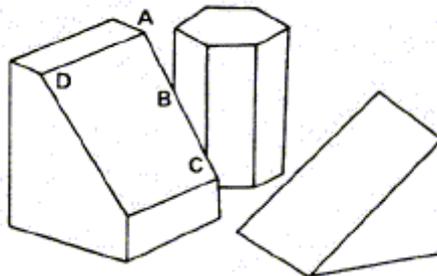


Abb. 7: Beispiel für die zu lösenden Schwierigkeiten beim maschinellen Bilderkennen: Die Punkte A, B, C und D sind alles T-Stücke. Während A und D Ecken eines Körpers bilden, kommen B und C zustande, weil ein dahinter liegender Körper verdeckt wird (aus Ballard & Brown, 1982).

erkannt werden müssen. Dazu muss der analysierende Algorithmus zum Beispiel entscheiden, ob drei Linien, die in einem Schnittpunkt zusammentreffen, zum selben Gegenstand gehören (Ecke) oder von der Überlagerung unterschiedlicher Gegenstände, die hintereinander liegen, herrühren. Auch muss der Algorithmus Hypothesen anstellen, ob ein Gegenstand an seiner Begrenzungskontur tatsächlich endet oder ob er vielleicht durch etwas anderes verdeckt ist und welche Form sein verdeckter Teil haben könnte (*Abb. 7*).

3.2.1 Strukturmerkmale im Bild – Indexing primitives

Die Elemente, die verwendet werden, um eine Struktur in ein Bild zu bringen, das heißt, verschiedene Objekte darin zu unterscheiden, sind die so genannten *Indexing Primitives*. Sie sind diejenigen Bildstrukturen, die verwendet werden, um aus einer Liste bekannter Objekte das gesuchte herauszugreifen. Sie liefern Indices in die Datenbank der gelernten Objekte. Die Art und Weise, wie ihre Extraktion aus einem gegebenen Bild organisiert ist, ist gleichzeitig das bestimmende Kriterium für die Leistung eines Modells zur Bildanalyse.

Die *Indexing Primitives* können verschiedene Komplexitätsgrade aufweisen, von einfachen Punkten (x/y-Koordinaten) über 3D-Punkte (x/y/z-Koordinaten), 2D-Konturen (Linien in der Fläche), 3D-Konturen (Linien im Raum), 2D-Konturgruppen, bis zu 3D-Oberflächen und 3D-Volumenformen (volumetric primitives)(Dickinson, 1993).

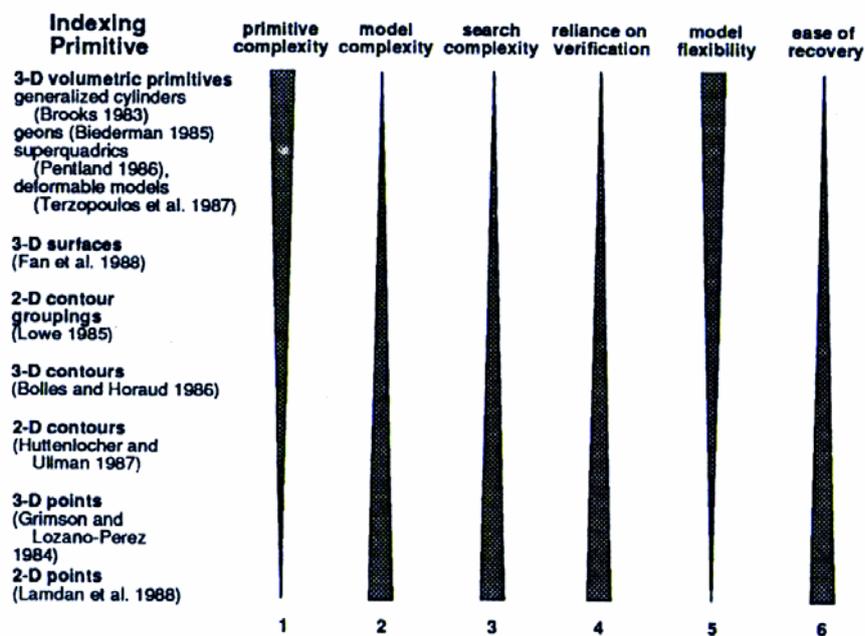


Abb. 8: Übersicht über die Auswirkungen der Auswahl von Indexing Primitives aus verschiedenen Komplexitätsstufen. (aus Dickinson, 1993).

Abhängig vom Komplexitätsgrad der indexing primitives gestalten sich die anderen Modellierungskriterien einfach oder komplex. **Abb. 8** zeigt die wechselweisen Beziehungen zwischen diesen Kriterien: Wenn die indexing primitives komplexer werden (1), nehmen ihre für das Objekt benötigte Anzahl (2) und damit auch die Anzahl ihrer hypothetisch möglichen Kombinationen (3) ab. Ebenso ist ein komplexes indexing primitive weniger zweideutig, es muss sich daher weniger auf eine zusätzliche top-down-Verifikation stützen (4). Damit ist es aber auch flexibler und eher unempfindlich gegen kleine Formveränderungen des Objekts (5). Schließlich ist ein komplexes – also dreidimensionales – indexing primitive nur sehr schwer aus einem zweidimensionalen Bild extrahierbar (6).

3.2.2 Segmentierung und Merkmale

In der Beschreibung eines gegebenen Bildes bedienen sich maschinelle Algorithmen der Technik der Segmentierung (recognition-by-parts). Die Bildoberfläche wird in Teile (Parts) aufgeteilt, so dass die komplexe geometrische Beschreibung eines Objektes zu kleineren handlichen Portionen reduziert wird. Diese Segmentierung ist jedoch nicht möglich ohne zusätzliches Bildwissen (externe Eingabe), und sie bedeutet auch nicht, dass das gesamte Bild gleichermaßen Berücksichtigung finden muss. Es reicht beispielsweise, wenn Kanten und Ecken lokalisiert werden, ohne zu definieren, was „Nicht-Kanten“ und „Nicht-Ecken“ sind (Caelli et al., 1993). Es ergibt sich hiermit bereits eine Hierarchie innerhalb des Bildes. Der Verlauf der Grenzen zwischen den Teilen, die durch die Segmentierung entstehen, wird dabei ausschließlich aufgrund von algorithmischen Gesichtspunkten bestimmt (z.B. Minima einer Kurve). In den Teilen können verschiedene Arten von Merkmalen (*attributes*) extrahiert werden:

- **morphologische Merkmale**, die aus dem gesamten Objekt abgeleitet werden, wie beispielsweise die Anzahl der durch die Segmentierung entstandenen Teile, die totale Fläche und der gesamte Umfang;
- **unäre (komponentenspezifische) Merkmale** wie die Größe einer einzelnen Fläche, ihre Spannweite oder ihre Helligkeit;
- **binäre (relationale) Merkmale**, wie die Länge der gemeinsamen Grenze, die Entfernungs- oder Winkelbeziehungen zwischen den Teilen.

Für die technische Objekterkennung entwickelte recognition-by-part-Mustererkennungs-Systeme wie beispielsweise EBS (Jain & Hoffman, 1988) oder CLARET arbeiten auf dieser Grundlage. Nach der Zerlegung des gegebenen Objektes in geeignete Komponenten (Teile, *parts*) wird jede Komponente durch einen Vektor aus unären Merkmalen, und jedes Paar benachbarter Komponenten durch einen Vektor aus binären Merkmalen beschrieben.

Innerhalb der so definierten Merkmalsräume werden mittels erlernter Regeln (neuronales Netz) Felder definiert, die probabilistisch als Zugehörigkeit zu einer bestimmten Objektklasse interpretiert werden.

In allen Ansätzen stellt die Formulierung (Codierung) der 3D-Objektstruktur aus einer 2D-Darstellung (z.B. auch die Retina) das größte Problem dar. Probleme ergeben sich dabei auf vier verschiedenen Stufen:

- (1) Wie wird segmentiert: Nach Intensitäten, Kanten oder Krümmungen?
- (2) Welche Merkmale werden extrahiert (externe Eingabe)?
- (3) Wie wird der Merkmalsraum in Bereiche ähnlicher Merkmalszusammensetzung unterteilt?
- (4) Wie wird aus der erfolgten Unterteilung des Merkmalsraumes eine (probabilistische) Klassifikationsentscheidung getroffen?

Für die psychophysische Analyse der Objekterkennung wird aus diesen Ausführungen erst deutlich, wie viele Teilschritte überhaupt notwendig sind, um einen solch komplizierten Vorgang wie Objekterkennung in Einzelschritte zu zerlegen, die der sequentiell arbeitende Algorithmus nacheinander abarbeiten muss.

3.3 Psychologische Theorien der Objekterkennung

3.3.1 Marrs algorithmischer Ansatz

Als einer der Pioniere des algorithmischen Ansatzes in der Wahrnehmungsforschung gilt David Marr (Marr, 1982). Seine Analyse des Wahrnehmungsvorgangs beruhte auf einem stufenweisen Vorgehen: 1. Formulieren des Wahrnehmungsproblems, 2. Suchen der möglichen (algorithmischen) Lösungen dafür und 3. die Implementierung dieser Lösungen im Wahrnehmungssystem. Damit hat Marr allgemeine Regeln zur Wahrnehmungsforschung aufgestellt.

Marrs algorithmisches Modell der Objekterkennung folgt vom 2D Bild eines Gegenstandes zur dreidimensionalen Beschreibung ebenfalls einem stufenweisen Vorgehen: Als erstes werden Kanten und elementare Merkmale in einer primären Rohskizze (*primal sketch*) festgehalten (extrahiert). Als zweites wird die Rohskizze zu einer zweieinhalbdimensionalen Skizze erweitert, indem die Merkmale, die in Größe und Orientierung ähnlich sind, nach Gestaltgesetzen gruppiert werden. Diese zweieinhalbdimensionale Skizze, die die Flächen und ihre Anordnungen beschreibt, wird zuletzt zur dreidimensionalen Repräsentation erweitert, die unsere Wahrnehmung bestimmt.

Der Erkennungsvorgang ist dann abgeschlossen, wenn diese Wahrnehmung mit einer gespeicherten Repräsentation des Objektes abgeglichen worden ist. Für den Ansatz von Marr & Nishihara (1978) sind für die Repräsentation zwei Punkte wesentlich: (1) Das Vorhandensein eines objektzentrierten Koordinatensystems und (2) darauf aufbauend eine hierarchisch modulare Objektbeschreibung in volumetrischen Primitiva („generalised cones“), die sich nach den aus dem Bild bestimmbaren Längsachsen ausrichten (**Abb. 9**).

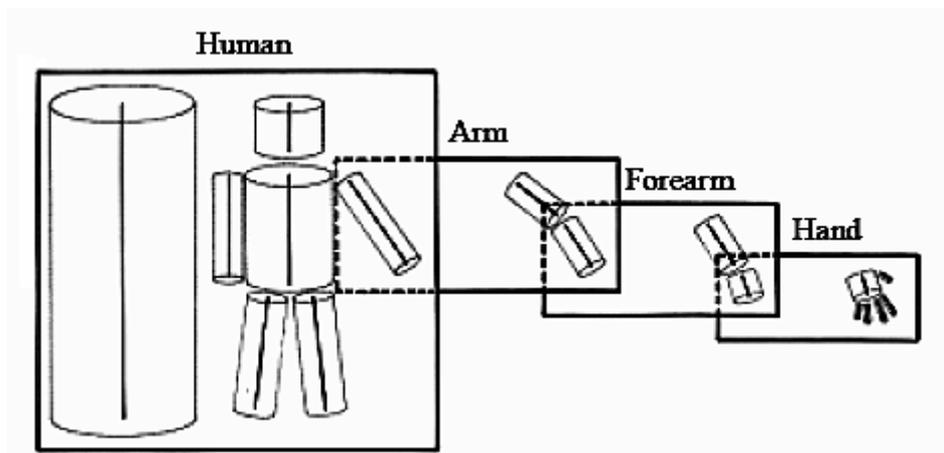


Abb. 9 Das Modell von Marr & Nishihara (1978) ist ein hierarchisch strukturiertes 3D-Modell. Auf jeder Stufe der Auflösung wird das vorliegende Objekt durch „generalized cones“ approximiert, die über die Lage ihrer Hauptachsen zueinander in Beziehung stehen und so die Gesamtstruktur bewahren.

Die Stärken von Marrs System liegen in der konsequent vollzogenen formalen Beschreibung und der daraus sich theoretisch ergebenden, vollständigen Berechenbarkeit. Auch die Anwendung eines Bottom-up-Prinzips ist konsequent vollzogen: Keine zusätzliche Information muss in das System hinein gegeben werden. Dies ist aber aus heutiger Perspektive gesehen auch wiederum die Schwäche dieses Ansatzes. Neuroanatomie und Neurophysiologie haben überzeugend eine große Anzahl Rückkopplungsverbindungen nachgewiesen (vgl. Kap. 2.1.3), so dass ein reiner „bottom-up“-Ansatz biologisch nicht ausreichend sein kann.

Ebenfalls kritisch zu beurteilen ist die Tauglichkeit des Systems in Fragen der Kategorisierung und Generalisierung – wann genau gehören „generalized cones“ der gleichen Klasse an? Außerdem gestaltet sich auch die praktische Umsetzung schwierig: Die Berechnung der „cones“ aus einer gegebenen Ansicht ist längst nicht immer möglich. Trotzdem war Marrs System konzeptionell sehr einflussreich.

3.3.2 Biedermans Objekterkennung durch Teilkörper

Ein auf Marr aufbauender Ansatz ist das von Irving Biedermann vorgelegte Konzept der „Recognition-by-Components“ (RBC) (Biederman, 1987). Wie Marr geht auch Biederman von einer Beschreibung durch volumetrische Elementareinheiten aus, die aus dem Bild aufgrund der Konturen von Ecken, Kanten und Grenzen bestimmt werden. Diese volumetrischen Einheiten sind bei ihm durch eine Ansammlung von 36 festen Volumenformen gegeben. Die so genannten „geons“ sind geometrische Körper wie Quader, Kegel, Zylinder, gebogener Kegel, stumpfer Kegel etc. Sie werden von Biedermann als Analoga zu den Phonemen in der Sprachforschung verstanden, von denen eine relativ geringe Anzahl ausreicht, die komplette Lautbildung aller Sprachen abzudecken.

Die Konturen der Geons werden jeweils aus Schlüsselmerkmalen im Bild bestimmt, den so genannten „nonaccidental properties“ (NAPs). Das sind allgemeine Gesetze des Transfers einer zweidimensionalen Darstellung auf ein dreidimensionales Volumen: „collinearity“ (eine Gerade bleibt eine Gerade), „curvilinearity“ (eine Kurve bleibt eine Kurve), „symmetry“ (symmetrisch bleibt symmetrisch), „parallel edges“ (parallele Kanten bleiben parallel) und „coterminating edges“ (zugleich endende Kanten bleiben bestehen). Diese nonaccidental properties sind dabei weitgehend invariant gegenüber Blickrichtungsveränderungen. So gehen die 2D-Konturen im Bild direkt in eine 3D-Geon-Beschreibung über (**Abb. 10**).

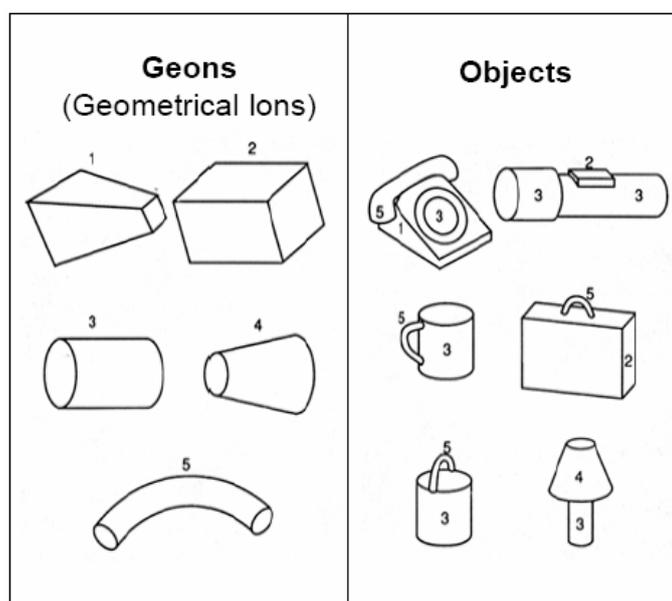


Abb. 10: Beispiele für einige Geons und Objekte, die aus ihnen zusammengesetzt werden können. Durch relational unterschiedliche Konstruktion können verschiedene Objekte entstehen (aus Biederman, 1985).

Das Ziel seines Ansatzes ist nach Biedermann die perzeptive Erstzuordnung, die initiale Kategorisierung für sich allein dargebotener Objekte (sog. basic-level categorisation). Für dieses Konzept sprechen folgende experimentelle Befunde (Biederman, 1987):

- (1) Objekte, die nur partiell dargeboten werden, können auch unter kurzen Darbietungszeiten gut erkannt werden.
- (2) Experimente mit reduzierter Konturinformation zeigen die Bedeutung von aussagekräftigen Konturbestandteilen wie Ecken und Kanten für das Erkennen auf.

Damit Recognition-by-components (RBC) wirksam sein kann, müssen allerdings drei Voraussetzungen gelten (Biederman & Gerhardstein, 1993): (1) Das zu erkennende Objekt muss in einzelne Geons unterteilbar sein, so dass die Geons und deren Relationen überhaupt extrahiert werden können. (2) Die zu unterscheidenden Objekte müssen verschiedene Geons in sich tragen. (3) Über die verschiedenen Blickrichtungen müssen die identischen „nonaccidental properties“ sichtbar sein.

Bis heute ist allerdings die Frage unbeantwortet geblieben, wie die Geons selbst gelernt, und wie die konstituierenden Geons eines neuen unbekanntes Objektes gefunden werden. Daraus ergibt sich, dass das RBC-Konzept nur für bereits gelernte Objekte effizient funktioniert. Außerdem fehlt bislang ein Existenznachweis für Geons überhaupt, und ihre mit 36 angegebene Anzahl erscheint willkürlich.

3.3.3 „View-dependent“ Theorien

Die deutlichste Infragestellung von strukturbeschreibenden Theorien erfolgte aufgrund von Experimenten, die zeigten, dass die von Marr und Biederman behauptete stabile Erkennungsleistung über verschiedene Blickwinkel (Invarianz) nicht in jedem Fall gegeben war. Dazu zählen die psychophysischen Untersuchungen (Rock, Di Vita & Barbeito, 1981; Bülthoff & Edelman, 1992) sowie die neurobiologischen Experimente (Logothetis, Pauls, Bülthoff & Poggio, 1994).

Es handelt sich dabei um durchwegs ähnlich aufgebaute Experimente, die folgenden Sachverhalt zeigten: Wenn Probanden unvertraute (novel) Objekte nur aus einzelnen kontrollierten Blickrichtungen kennen lernen und danach die Objekte unter weiter davon entfernt liegenden Blickrichtungen erkennen sollen, steigen die benötigten Antwortzeiten und die Fehlerraten mit zunehmendem Abstand von der ursprünglich gelernten Ansicht des Objektes linear an. Aufgrund dieser Abhängigkeit der Erkennungsleistung vom Rotationswinkel (view-dependency) wurde behauptet, dass das Objekt vom Beobachter innerhalb eines egozentrischen Referenzrahmens mental repräsentiert werde. Das heißt, dass die Objekte in der Gestalt im Gedächtnis auftreten, wie sie gesehen wurden, nämlich ansichtenbasiert oder piktoral.

Ihren Anfang nahmen diese Experimente in den Arbeiten von Shepard, Metzler und Cooper (Shepard & Cooper, 1982). Diese Autoren hatten festgestellt, dass die Zeit, die benötigt wird, um zu entscheiden, ob zwei aus verschiedenen Blickwinkeln dargebotene Objekte identisch oder spiegelsymmetrisch sind, linear von der Rotationswinkeldifferenz zwischen den Objekten abhängt. Der Vorgang, der diesem Phänomen zugrunde liegt, wurde als „mentale Rotation“ bezeichnet, weil die Probanden zum Zweck des Vergleichs mental das eine Objekt in die Position des anderen rotieren.

Tarr & Pinker (1989) fanden, dass diese Zeitverzögerung immer, und nicht nur bei der Unterscheidung von spiegelsymmetrischen Objektpaaren, auftritt, dass sie aber mit zunehmender Vertrautheit mit den Objekten stark abnimmt. Des Weiteren stellten sie fest, dass das Erkennen der spiegelsymmetrischen Gegenstücke von bereits bekannten Objekten jeweils mit einer konstanten Zeitverzögerung geschieht, unabhängig von deren Ausrichtung. Diese Ergebnisse deuteten sie dahingehend, dass die Objektrepräsentation konkret oder bildhaft in dem Sinn sein müsse, dass sie spezifisch ist für die lokale Anordnung der Objektteile im visuellen Feld unter einer bestimmten Blickperspektive. Die Invarianz gegenüber verschiedenen Objektansich-

ten entstehe darüber hinaus durch mehrere (multiple) 2D-Ansichten, die zur Repräsentation zusammengehängt werden.

Eine weitere Unterstützung findet diese Theorie durch Forschungen zur Frage der Existenz einer so genannt „kanonischen Ansicht“: Es wurde gefunden, dass Probanden übereinstimmend ein oder zwei Ansichten eines Objektes subjektiv als „besser als alle anderen“ bewerteten (Palmer, Rosch & Chase, 1981). Auch wenn die Objekte in repräsentativer Darstellung gezeichnet werden sollten, wurden übereinstimmend immer ähnliche Blickperspektiven gewählt.

Dieses Phänomen wurde ursprünglich rein deskriptiv und qualitativ definiert, dann aber von Blanz, Tarr & Bülthoff (1999) systematisch untersucht. Die Wahl des Blickwinkels zur kanonischen Ansicht hängt nach diesen Autoren ab von der:

- (1) **Objektgeometrie**: Typische Merkmale sind möglichst gut sichtbar präsentiert, was meistens in einer Ansicht von schräg vorne resultiert (kleiner als 45°).
- (2) **Aufgabe**: Abhängig von der geforderten Kategorisierungsebene (basic-level oder subordinate level (vgl. Kap. 2.2.2)), werden unterschiedliche Ansichten ausgewählt.
- (3) **Erfahrung** (Vertrautheit und Funktionalität): Die Objekte werden in derjenigen Lage ausgewählt, in der sie in der Alltagserfahrung auftreten. Ein Flugzeug beispielsweise von unten, ein Teekessel, so dass er gut ergriffen werden kann. Ebenso ist die kanonische Ansicht eines Objekts diejenige, in der das Objekt spontan mental vorgestellt wird.

Für das Benennen einer kanonischen Objektansicht treten insgesamt die niedrigsten Antwortzeiten und Fehlerraten auf. Beide steigen aber mit zunehmend abweichendem Ansichtswinkel linear an. Auch diese Tatsache kann von den „view-dependent“-Theorien problemlos integriert werden. Sie wird als Beweis für die Existenz einer durch multiple piktorale Ansichten gebildeten Objektrepräsentation angesehen (Bülthoff, Edelman & Tarr, 1995).

Rotation-for-handedness-Hypothese. Erwähnt werden muss hier noch die Rotation-for-handedness-Hypothese, die eine Ausarbeitung der obigen view-dependent-Theorien ist und heute von einigen Autoren vertreten wird (Tarr, 1995; Gauthier, Hayward, Tarr, Anderson, Skudlarski & Gore, 2002). Sie nimmt an, dass es sich bei der Objekterkennung um einen anderen Vorgang handelt als bei der Beurteilung der Chiralität eines gegebenen Objektes (linke oder rechte Version). Die menschliche 3D-Objekterkennung beruhe auf ansichtenbasierten und reflexionsinvarianten Repräsentationen. Die Händigkeit der Objekte werde dagegen durch mentale Rotation bestimmt.

3.3.4 Zusammenfassung: Die Objektstruktur in der Repräsentation

Wie in Kapitel 1 ausgeführt, wird die Einheit des wahrgenommenen Objektes in der visuellen Verarbeitung in seine verschiedenen Qualitäten aufgespalten: Form, Farbe, Bewegung, kategoriale Bedeutung. Alle diese Bestandteile werden in unterschiedlichen Hirnarealen verarbeitet.

Gewöhnlich erkennt man ein Objekt an seiner Form. Wenn beispielsweise ein Elefant blau dargestellt ist, erkennen wir ihn. Wenn aber bei unverändertem Elefanten-Grau seine typische Elefanten-Form verändert wird, ist er kein Elefant mehr. Die Form ist dominant in der Objektcharakterisierung und damit auch in der gespeicherten Repräsentation.

Die dargestellten Theorien der biologischen Objekterkennung widersprechen sich in wesentlichen Punkten, was die Kodierung der Objektstruktur betrifft: Marr und Biederman (Kap. 3.3.1 und Kap. 3.3.2) fordern eine modulare Rekonstruktion der Objektstruktur aus geometrischen Grundformen. Dadurch soll die Strukturinformation des Objekts und eine Stabilität des Objekterkennens über unterschiedliche Blickrichtungen hin aufrechterhalten werden.

Die „view-dependent“ Theorien (Kap. 3.3.3) halten dem die Abhängigkeit der Erkennungsleistung von den Bildveränderungen durch unterschiedliche Blickperspektiven entgegen. Ihre Vertreter machen sogar auf eine bestehende Asymmetrie aufmerksam: Man könne aus einer blickrichtungsinvarianten (*viewpoint-invariant*) Erkennungsleistung nicht direkt auch auf eine blickrichtungsinvariante Objektrepräsentation schließen, denn solche Ergebnisse könnten ebenso durch zwei andere Ursachen bedingt sein (Bülthoff et al., 1995):

- (1) durch ein *diagnostisches Merkmal*: Im Kontext der experimentell definierten Erkennungsaufgabe können die Probanden die Objekte ansichtsunabhängig an spezifischen diagnostischen Merkmalen (*diagnostic features*) auf dem Objekt erkennen
- (2) durch *multiple gespeicherte Ansichten*: Das Objekt ist den Probanden bereits so vertraut, dass schon viele verschiedene Ansichten gespeichert wurden (*multiple views*). Für jede mögliche Blickrichtung stünde dann bereits eine gespeicherte 2D-Repräsentation zur Verfügung, woraus sich die Erscheinung der Invarianz der Erkennungsleistung erst ergäbe.

Die verschiedenen Theorien bewegen sich somit auf einer Achse zwischen zwei Polen: einerseits der konsequenten Durchalgorithmisierung und andererseits der psychophysisch-experimentellen Evidenz. Die theoretische Durchalgorithmisierung, wie beispielsweise in Marrs Theorie, ist psychophysisch nicht belegt. Andererseits fehlt aber den psychophysisch gut belegten „view-dependent“-Theorien bisher das klare algorithmische Konzept.

So ist noch nicht befriedigend geklärt, wie die postulierte Ansammlung von vielen Ansichten („multiple views“) zur Repräsentation eines einzigen Objekts verbunden wird, und wie sich aus dieser Sammlung von 2D-Bildern die 3D-Objektstruktur rekonstruieren lässt. Das hier angeführte Stichwort heißt Interpolation

oder Normierung aufgrund der „Ähnlichkeit“. Dies war lange Zeit nur ein einfacher Abgleich mit Schablonen (Template-matching) auf der Basis von Pixelwerten. Inzwischen wurden unter der Verwendung neuronaler Netzwerke weitere, auch komplexere Theorien entwickelt (Riesenhuber & Poggio, 1999). Das Problem der Verbindung von verschiedenen piktoralen Ansichten besteht besonders dann, wenn sich diese sehr unähnlich sind und die Frage zuerst gelöst werden muss, welches Merkmal der einen Ansicht mit welchem Merkmal der anderen Ansicht überhaupt korrespondiert (Korrespondenz-Problem).

Nach einigen Jahren kontroverser Auseinandersetzungen wurde der Vorschlag gemacht, die Lösung der Unvereinbarkeit der beiden Ansätze darin zu suchen, dass die verschiedenen Konzeptionen verschiedene Gültigkeitsbereiche hätten: Danach sollte die perzeptive Erstzuordnung eines Objektes unter Rückgriff auf die 3D-Modelle des Biedermanschen RBC-Ansatzes geschehen, was der basic-level Kategorisierung nach E. Rosch entspricht. Die Feinverarbeitung, die subordinate-level Kategorisierung, sollte jedoch auf der Basis der pictoralen 2D-Repräsentation der „multiple-views“-Theorien geschehen (Bülthoff et al., 1995).

Grundsätzlich betrachtet beginnt die Suche nach der Art der Repräsentation der geometrischen Objektstruktur jedoch bereits beim einzelnen 2D-Bildmerkmal. Ein gegebenes Bildmuster als bloße Ansammlung isolierter Einzelmerkmale zu repräsentieren, wäre problematisch. Eine solche Zerlegung käme einer Atomisierung des Bildes gleich, mit dem gleichzeitigen Verlust der Information über die übergeordnete Gesamtstruktur. Dies stünde auch im Widerspruch zu den Gesetzen der Gestalttheorie.

Dass schon auf elementarer Ebene von Struktur gesprochen werden muss, zeigt folgendes Beispiel mit einfachsten Elementen (**Abb. 11**): Die Objekte 1 und 2 können als Kombinationen von Komponenten auf verschiedene Weise aufgefasst werden: 1 als zwei nebeneinander stehende Sanduhrgläser (1B) oder als zwei sich gegenseitig durchdringende Dreiecke (1C), auch die Deutungen 1D oder 1E sind möglich.

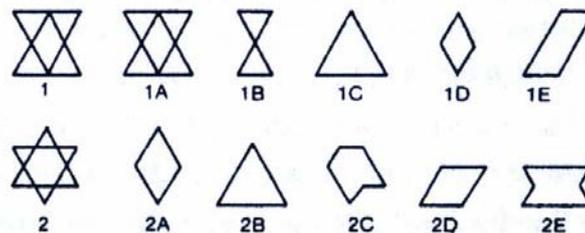


Abb. 11: Objektkomponenten und Objektstruktur. 1B – 1E sind mögliche Komponenten von Objekt 1, 2A – 2E mögliche Bestandteile von Objekt 2 (aus Jüttner, 2003).

Mithilfe dieser Demonstration (**Abb. 11**) ist es leicht einzusehen, dass es nur schwer möglich ist, das Auftreten von Struktur im Sinne einer übergeordneten Ordnung der einzelnen Komponenten (Indexing Primitives), auf einer bestimmten Komplexitäts-

ebene zu lokalisieren. Zwar können Indexing Primitives durchaus Relationen untereinander aufbauen, aber sie selbst bestehen – je nach den sie konstituierenden Komplexitätsstufen (vgl. *Abb. 8*) – selbst wiederum aus Relationen noch basalerer Elemente.

3.3.5 *Fazit: Verschiedene Probleme im Objekterkennen*

In der Diskussion zur Objekterkennung müssen deshalb mehrere verschiedene Problemstellungen auseinander gehalten werden:

- (1) ***Figur-Hintergrund-Trennung:*** Das Unterscheiden des Objekts von der Umgebung. Dies ist jedoch nur die Voraussetzung für das Erkennen. Die Mustermerkmale, die konstitutiv für individuelle Objekte sind, müssen als zusammengehörig gruppiert werden und die Objekte müssen voneinander getrennt werden, um überhaupt identifizierbar zu sein. Dabei muss diese Gruppierung der Bildelemente allerdings unbewusst ablaufen, da sie der Objektidentifikation vorausgeht. Sie muss möglich sein, bevor man überhaupt weiß, welche Objekte in einem Bild sind. Hier laufen die Automatismen ab, die in der Gestaltphänomenologie beschrieben werden.
- (2) ***Diagnostisches Merkmal:*** Das Wiedererkennen eines Objektes unter anderen ähnlichen Objekten anhand eines – oder mehrerer – dafür diagnostischen Merkmals. Diese Leistung ist nicht von vornherein mit der entry-level-Kategorisierung gleichzusetzen. Auch eine subordinate Kategorisierung kann auf dieser Ebene getroffen werden. Um beispielsweise eine blaue Billardkugel von einer roten zu unterscheiden, muss die unterschiedliche Farbe als Kriterium berücksichtigt werden. Dies ist ein diagnostisches Merkmal auf der subordinaten Ebene, das nichts mit der 3D-Struktur zu tun hat.

Zu diesem Sachverhalt gehört auch das Einordnen eines unbekanntes Objektes in eine bekannte Kategorie, wenn sie aufgrund von diagnostischen Merkmalen erfolgt. Bei einem sehr ungewöhnlichen Designerstuhl etwa erkennt man als diagnostisches Merkmal vielleicht eine Art Sitzfläche in einer Höhe, wie sie Stühle normalerweise aufweisen.

- (3) ***Strukturverstehen:*** Das komplette Verstehen der räumlichen Verhältnisse (3D-Struktur) des Objektes, so dass sie in Worten angemessen wiedergegeben werden oder aktiv reproduziert und weiterverwertet werden kann. Ein „Minimalbeispiel“ für das „Strukturverstehen“ ist die Unterscheidung von händigen oder chiralen Objekten. Diese können nicht auf visuelle Ähnlichkeit hin unterschieden werden, sondern nur unter Berücksichtigung ihrer relationalen Struktur.
- (4) ***Finden einer neuen Kategorie:*** Das Finden einer neuen, noch unbekanntes Kategorie aufgrund von Schlussfolgerungen aus der beobachteten Merkmalsverteilung in einer gegebenen Objektmenge. Diese Ebene der Objekterkennung wird im Alltag selten gefordert, aber sie ist ebenso möglich: Man denke an die Arbeit

eines mit der Klassifikation von Organismen befassten Biologen oder an das Problem einer angemessenen Taxonomie der Krankheiten.

Außerdem ist anzunehmen, dass es nicht *die eine* Repräsentation eines Objektes gibt, sondern dass die interne Objektrepräsentation je nach Art der geforderten Erkennungsaufgabe in unterschiedlicher Weise von den Merkmalen des Objektes Gebrauch macht.

4 Experimentelle Untersuchung der Objekterkennung unter Berücksichtigung des Kontextes

Der Fokus der psychologischen Arbeiten zur Objekterkennung, wie sie im vorangehenden Kapitel 3 dargestellt sind, liegt auf der Frage nach der Natur der visuellen Objektrepräsentation. Die Natur einer biologisch relevanten visuellen Repräsentation ist jedoch ohne den Kontext des Wahrnehmungsvorgangs nicht abschließend zu beschreiben. Dieses Kapitel diskutiert die Bedingungen einer experimentellen Untersuchung der Objekterkennung unter Berücksichtigung des Kontextes.

4.1 Erkennungsleistung versus Repräsentation

In der psychologischen Literatur zur 3D-Objekterkennung wurde in der Regel von den Erkennungsleistungen direkt auf die funktionalen Eigenschaften der internen Repräsentationen geschlossen (Liu et al., 1999). Besonders der Grad der Ansichten-Abhängigkeit (view-dependency) der Objektrepräsentationen ist ein kontrovers diskutiertes Problem (Tarr & Vuong, 2002). Die Abhängigkeit der Erkennungsleistung vom Blickwinkel der gezeigten Objektansichten wird dabei als Beweis für die Ansichten-Abhängigkeit der zugehörigen Objektrepräsentation angesehen.

Die experimentell ermittelten Leistungen beim Erkennen von Objekten können aber nicht mit der Repräsentation gleichgesetzt werden. Die Repräsentation ist experimentell nur indirekt zugänglich, wie folgendes Schema zeigt (**Abb. 12**). Die Eingangsdaten im Bildformat (Bild/Stimulus) müssen für den Erkennungsvorgang mit den gespeicherten Objektdarstellungen (Objektmodellen oder Objektrepräsentationen) verglichen werden. Falls die Eingangsdaten mit der Repräsentation hinreichend übereinstimmen, kann eine Zuordnung zur entsprechenden Objektkategorie erfolgen. In der gemessenen Erkennungsleistung zeigt sich jedoch nicht direkt die Objektrepräsentation, sondern nur, wie gut dieser Abgleich von Eingangsdaten und gespeicherter Repräsentation gelang. Es kann keine direkte Aussage über die Repräsentation gemacht werden.

Wie gut der Abgleich-Vorgang gelingt, wird durch jegliches Vorwissen, das außerhalb der Bildinformation zur Verfügung steht, wesentlich beeinflusst. Jede Art von außerbildlicher Information, die zu einem anderen Zeitpunkt oder an einem

anderen Ort gewonnen wurde, kommt dafür in Frage. Menschliches Objekterkennen ist ohne solches Kontextwissen nicht möglich (Albright, 1995).

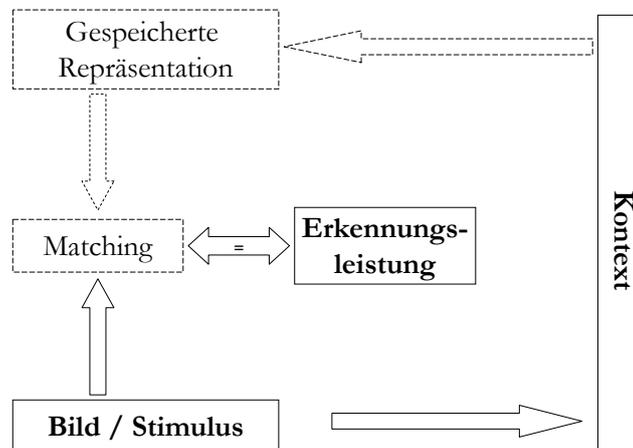


Abb. 12: Schema des Erkennungsvorgangs als Abgleich von Eingangsdaten und gespeicherter Repräsentation. Die experimentell direkt zugänglichen Größen sind: Stimulus, Erkennungsleistung und Kontext (im ausgezogenen Rahmen angegeben). Auf die Repräsentation kann nur indirekt geschlossen werden (in gestrichelten Rahmen). Der Abgleichsprozess (Matching) wird per Definition mit der Erkennungsleistung gleichgesetzt.

Wird der Erkennungsvorgang aus diesem Gesichtspunkt betrachtet, folgt daraus, dass bei der experimentellen Untersuchung der Objekterkennung nicht nur (1) die Erkennungsleistung sondern auch (2) die Bildinformation und (3) das Kontextwissen miteinbezogen werden müssen. Nur dann können verlässliche Aussagen über die Natur der internen Objektrepräsentation gemacht werden.

Die Messung der Erkennungsleistung wird in Kap. 4.2, die Rolle der Bildinformation in Kap. 4.3 und die Berücksichtigung des Kontextwissens in Kap. 4.4 diskutiert.

4.2 Messung der Erkennungsleistung – die Klassifikation

Das heutige Standardparadigma der experimentellen Messung von Erkennungsleistung ist die Reizunterscheidung bei zeitlich verzögerter Darbietung (Delayed Matching to Sample, DMS).

Delayed Matching to Sample (DMS). Bei der Reizunterscheidung mit zeitlich verzögerter Darbietung wird nach einem Bild A ein Bild B präsentiert und der Proband muss entscheiden, ob das auf Bild B dargestellte Objekt identisch mit dem auf Bild A dargestellten ist (*same/different task* mit Zeitverzögerung). Dieses Paradigma testet die *Objektdiskrimination* und gilt heute als die Standardmethode, da sie von der Mehrzahl der Forscher, die sich mit den Problemen der Objekterkennung befassen,

angewendet wird (Tarr & Pinker, 1989). Auch für die Experimente zur mentalen Rotation wird diese Methode verwendet (Shepard & Metzler, 1971; Tarr & Pinker, 1989; Gauthier et al., 2002; Parsons, 2003).

Beim *DMS Task* gibt es nur zwei verschiedene Klassen von Bildern, die auseinander gehalten werden müssen: „gleich“ und „verschieden“. Die Beurteilung beruht auf einem Ähnlichkeitsvergleich von zwei kurzfristig hintereinander gezeigten Bildern im Kurzzeitgedächtnis (short-term memory). Die Bilder müssen perzeptiv nur soweit analysiert werden, wie nötig ist, um ihre Unterschiedlichkeit festzustellen. In den meisten Experimenten wissen die Probanden schon im Voraus, worauf sie zu achten haben (in einer Aufgabe der Objekterkennung: „gleiches Objekt ja/nein“, für mentale Rotation: „spiegelverkehrt ja/nein“), und sie haben die Objekte bereits gut kennengelernt (Tarr, 1995; Gauthier et al., 2002). Deshalb stellt die Bewältigung des Bilderkennungs Vorgangs keine Schwierigkeit dar. Eine Beteiligung des Langzeitgedächtnisses wird vollständig vermieden, wenn bei jeder Darbietung neue Reizmuster verwendet werden.

Doch funktioniert biologisch relevante Objekterkennung nicht nur im Vergleich zweier gleichzeitig oder kurz hintereinander präsentierter Bilder. Eine wesentliche Rolle spielt darin – in Kapitel 2 dargestellt – die Zuordnung des Wahrnehmungsinhaltes zum System der Kategorien. Außerdem sollte diese Zuordnung auch dann funktionieren, wenn die Objekte schon länger nicht mehr gesehen wurden. Eine Beteiligung des Langzeitgedächtnisses ist für die menschliche Objekterkennung unentbehrlich.

Aus diesem Grund haben einige Autoren Interesse an einem Paradigma zur Objekterkennung entwickelt, das eine Klassifikationsleistung vom Probanden erfordert (Caelli, Rentschler & Scheidler, 1987; Rentschler, Jüttner & Caelli, 1994; Langguth, 1998; Rentschler & Jüttner, 2000; Jüttner, Langguth & Rentschler, 2004). Bei der Klassifikation handelt es sich, im Unterschied zur Diskrimination, um eine höhere Hirnleistung, mit der der sensorischen Erfahrung Bedeutung zugemessen werden kann (Edelman, 1987).

Klassifikation. Grundsätzlich gibt es zwei Möglichkeiten von Klassifikationsaufgaben (*classification* oder auch *categorisation*): (1) die Zuordnung von Elementen zu bereits bekannten Klassen und (2) die Konfrontation mit einem Kollektiv von Elementen bei unbekannter Klasseneinteilung. Bei beiden Formen ist die Unterscheidung von beliebig vielen Klassen während einer Aufgabe möglich. Sind die Kriterien für die verschiedenen Klassen dem Probanden zu Anfang völlig unbekannt, müssen sie innerhalb eines Lernvorgangs von ihm erst erarbeitet werden. In einem zweiten Abschnitt, der Generalisierung, wird dann getestet, wie gut der Proband auch weitere Testbilder diesen gelernten Klassen zuordnen kann (vgl. Kap. 2.2.2).

Die Klassifikation ist ein allgemein verwendbares Paradigma. Im Fall der Objekterkennung bilden die verschiedenen Objekte die Klassen, denen die verschiedenen Objektansichten zugeordnet werden müssen. Die Bilder müssen hier auf alle möglichen Merkmale hin rezipiert werden und das Bild einer hypothetisch zu fin-

denden, aber nicht unmittelbar anwesenden Klasse zugeteilt werden. Dabei bestimmt die Auffälligkeit der Klassenmerkmale im Gesamtkontext die Geschwindigkeit des Lernens. Auf diese Weise gelingt es, das Entstehen einer Objektrepräsentation im Experiment zu beobachten.

4.3 Rolle der Bildinformation

Liu, Kersten & Knill (1999) haben darauf aufmerksam gemacht, dass aus der Erkennungsleistung nicht direkt auf die Natur der Repräsentationen geschlossen werden kann. Die gemessenen Erkennungsleistungen der Beobachter werden von den (physikalischen) Objektredundanzen gegenüber der Änderung von Darbietungsbedingungen ebenso wie von den Eigenschaften der Repräsentationen beeinflusst (S. 603). Verlässliche Rückschlüsse auf interne Objektrepräsentationen setzen also voraus, dass die Objektredundanzen und die Eigenschaften der Repräsentationen separiert werden.

Das Trennen von Bildinformation und Repräsentation kann auf folgendem Weg gelingen: durch das Quantifizieren der Objektredundanzen und die Normierung der beobachteten Erkennungsleistungen im Hinblick auf diese Redundanzen. Im Rahmen einer Modellrechnung muss dafür der Einfluss der physikalisch messbaren Objekteigenschaften auf die Erkennungsleistung quantifiziert werden. Aus dem Vergleich der theoretischen Voraussagen mit der tatsächlich beobachteten Erkennungsleistung kann dann auf die Eigenschaften der Repräsentation geschlossen werden. Für eine solche Berechnung der Objektredundanz sind verschiedene Verfahren verwendet worden:

- (1) Ein Verfahren wurde von Liu et al. (1999) selbst verwendet, indem die Autoren eine „object complexity“ nach bestehenden Freiheitsgraden zur Objektrekonstruktion aus den Bildelementen definierten und danach die Stimulusinformation mittels eines modellierten „idealen Beobachters“ („ideal observer“) quantifizierten.
- (2) Ein anderes Verfahren benutzten Osman, Pearce, Juttner & Rentschler (2000; Osman, in Vorbereitung): Der von diesen Autoren verwendete Mustererkennungs-Algorithmus CLARET misst die Ähnlichkeitsabstände zwischen den einzelnen Muster-Klassen aufgrund eingegebener kritischer Merkmale im Merkmalsraum.

4.4 Berücksichtigung des Kontextes

Jegliches Objektwissen, das außerhalb des Objektes zur Verfügung steht, das zu einem anderen Zeitpunkt oder an einem anderen Ort gewonnen wurde, zählt zum „Kontextwissen“ (Albright, 1995). Bei der experimentellen Untersuchung der Objekterkennung fallen darunter im Wesentlichen drei verschiedene Aspekte:

- (1) **das Vorwissen über die Objekte:** Das Vorwissen, das der Proband von den Testobjekten haben kann, ist gegeben durch die Erfahrung mit den Testobjekten aus anderen Zusammenhängen.
- (2) **die relative Konfiguration der Objekte im Testsatz:** Die Unterschiede zwischen den Objekten zueinander bedingen ein bestimmtes Unterscheidungsverhalten.
- (3) **die Objektumgebung:** Die Umgebung, in der das Objekt auftritt, wird bestimmt durch den Hintergrund, die Kontraste, die kategoriale Umgebung („die Landschaft“, in der das Objekt auftritt).

4.4.1 Das Vorwissen über die Objekte

Es gibt prinzipiell verschiedene Ansatzpunkte, das Objektvorwissen zu kontrollieren:

- (1) Für das Experiment werden neuartige Objekte (*novel objects*) verwendet, mit denen der Proband bisher keine Erfahrung gemacht hat. Sie haben meist eine komplex abstrakte oder fantastische Form, um jede kategoriale Assoziation zu vermeiden.
- (2) Das Objektvorwissen wird direkt in das Experiment integriert in Form von kontrollierter Vor-Exposition bei konstantem anschließenden Paradigma, so dass unterschiedliches experimentelles Verhalten nur auf den Unterschied an Objektvorwissen zurückgeführt werden kann (*varierte Priming-Bedingungen*).

Unbekannte Objekte („Novel Objects“). Das Vorwissen über die zu lernenden Objektkategorien kann nur dann experimentell verlässlich kontrolliert werden, wenn die Versuchsperson mit den betreffenden Objekten nicht vertraut ist. Das hat verschiedene Autoren dazu veranlasst, neuartige Objekte (*novel objects*) zu erfinden, die in der natürlichen Umgebung der Probanden nicht vorkommen. Es sind meist abstrakte, komplizierte geometrische Figuren, die aufgebogenen Büroklammern („*Paperclips*“ oder *Drahtmodelle*) ähneln. Sie wurden sowohl in experimentellen Arbeiten (Rock et al., 1981; Tarr & Pinker, 1989; Bülhoff & Edelman, 1992; Logothetis et al., 1994; Bülhoff et al., 1995; Christou, Tjan & Bülhoff, 2003) wie auch in theoretischen Studien zur Objekterkennung (Poggio & Edelman, 1990; Logothetis et al., 1994) verwendet.

Neben ihrer Neuartigkeit erfüllen diese Objekte noch zwei weitere Kriterien: Sie bestehen sozusagen nur aus Form und, da sie beinahe keine Oberfläche aufweisen, werden unerwünschte Effekte, die auf Oberflächenmerkmale wie Farbe oder Textur zurückzuführen sein könnten, eliminiert. Eine andere Variante von „*novel objects*“, die dem Auge des Probanden mehr Flächen bieten, sind die gemüseartigen, amorphen Fantasie-Knollen („*Greebles*“ und „*Amöben*“, **Abb. 13**).

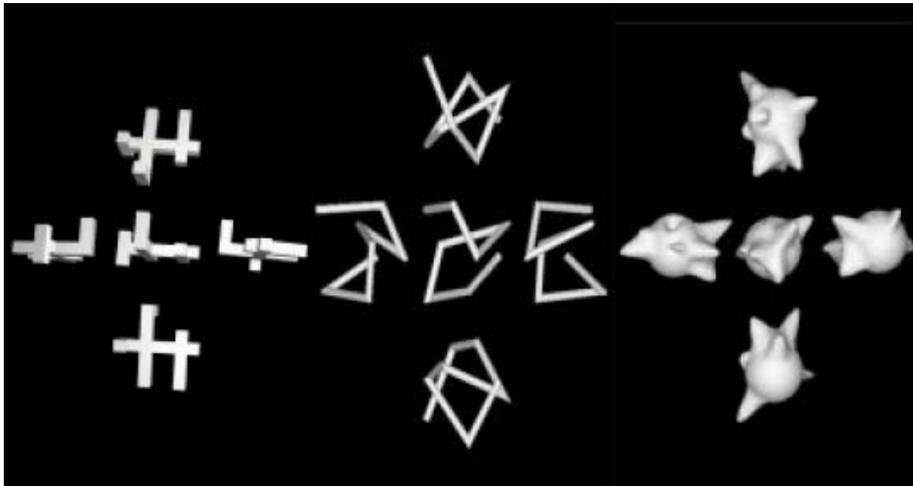


Abb. 13 Testobjekte. Links und Mitte: Drahtmodelle und “Paperclips”, rechts „Amöben“ (aus Bülthoff et al., 1995).

Variierte Priming-Bedingungen. Eine zweite Möglichkeit, das Vorwissen der Probanden über den Objektsatz zu kontrollieren, besteht darin, den Erwerb von Objektvorwissen kontrolliert in den Ablauf des Experiments mit einzubeziehen (Osman et al., 2000; Rentschler, Juttner, Osman, Muller & Caelli, 2004).

Dieser experimentelle Ansatz variiert gezielt den Erwerb der Kontextinformation in verschiedenen experimentellen Bedingungen (*Priming conditions*), hält aber um die Vergleichsmöglichkeit zu garantieren den restlichen Teil des Paradigmas konstant. Das durch die Primingbedingungen unterschiedlich vorhandene Objektvorwissen kann dann als Inhalt des Langzeitgedächtnisses die Erkennungsleistung nur durch den direkten Einfluss auf die interne Objektrepräsentation modifizieren. So lässt sich ein direkter Rückschluss auf die Eigenschaften der internen Repräsentation ableiten.

Tarr & Pinker (1989) ließen ihre Probanden die Objekte ausführlich lernen, bevor sie sie dem Experiment unterzogen. Sie hielten die Probanden dazu an, den Linienverlauf der verschiedenen Stimuli zu verfolgen (*trace*), den zugehörigen Namen aufzuschreiben und danach die Stimuli selbst zu zeichnen. Indem sie dies ohne Zeitdruck durchführen und so lange wiederholen sollten, bis sie zweimal in Folge alle drei Teststimuli richtig gezeichnet hatten, gewährleisteten die Autoren, dass die Probanden bereits ein aktives Verständnis der Objektstrukturen hatten, noch *bevor* sie das Experiment begannen (Tarr & Pinker, 1989, S. 244). Es kann sich im anschließenden experimentellen Test daher nur um ein Abrufen von bereits explizit erworbenem Objektwissen handeln: Die Objektrepräsentation ist bereits gebildet.

Um die Natur der Objektrepräsentation zu charakterisieren, ist jedoch interessant zu untersuchen, wie die Repräsentationen überhaupt gebildet werden, und welche Art von Objektwissen durch das Verfolgen der Linien und das Zeichnen (*tracking* und *tracing*) erworben wurde.

4.4.2 Relative Konfiguration der Testobjekte

Die Unterscheidung von Objekten, wie Flasche, Banane und Stuhl bietet kein Problem (Biederman, 1987). Solche Objekte tragen stark unterschiedliche Merkmale und werden auf den ersten Blick erkannt (entry-level-Kategorisierung, vgl. Kap. 2.2.2).

Zur experimentellen Untersuchung der subordinaten Ebene der Kategorisierung wurden deshalb sehr ähnliche Objekte verwendet. Die Objekte eines Objektsatzes sollten untereinander alle ähnliche Teile aufweisen, die nur in verschiedenen räumlichen Relationen zueinander angeordnet sind. So ist die Möglichkeit minimiert, die Objekte nach einmaligen spezifischen (*diagnostischen*) Merkmalen zu unterscheiden. Ein solches Verhältnis wird als typisch für die subordinate Kategorisierungsebene angesehen (Bülthoff et al., 1995).

In Verbindung mit einer Delayed-Matching-to-Sample-Technik haben sich in verschiedenen Labors gleiche Paradigmen durchgesetzt. Nach Bülthoff et al.(1995) gehören dazu die folgenden Punkte:

- Beurteilung des Lernfortschrittes: Messungen der anfänglichen Erkennungsleistung eines „novel“ Objektes wie auch der Erkennungsleistung nach ausführlichem Training.
- Anfängliches Training nur unter einer beschränkten Anzahl von Blickrichtungen, um den Grad der Abhängigkeit von der gelernten Blickrichtung in der Erkennungsleistung zu untersuchen.
- Das Verwenden von unbekanntem „Test-Ansichten“, um die zugrunde liegende Organisation der Ansichten, die durch das Lernen aufgebaut wurde, einzuschätzen.

Ein spezielles Problem der relativen Objektsatzkonfiguration bzw. der subordinaten Kategorisierung ist dabei das Auftreten von *chiralen Objektpaaren*. In den Experimenten zur mentalen Rotation ist das Auftreten von solchen Paaren von vorneherein bekannt und wird demzufolge vom Probanden erwartet (Shepard & Metzler, 1971; Tarr & Pinker, 1989; Gauthier et al., 2002).

4.4.3 Die Objektumgebung

Als dritter Punkt gehört die „Landschaft“, in der das Objekt auftritt, die Objektumgebung, zum Kontext. Ursprünglich wurden in den meisten Arbeiten isolierte Objekte vor einfarbigem Hintergrund gezeigt. In diesem Fall spielen der Kontrast und die Farbigekeit eine Rolle. Mittlerweile ist jedoch auch die Rolle der „natürlichen“ Objektumgebung untersucht und ihr positiver Einfluss auf die Erkennungsleistung aufgezeigt worden (Christou et al., 2003; Bar, 2004).

Bei einem kontextuellen Zusammenhang sowohl räumlicher als auch nicht-räumlicher Art fanden Bar & Aminoff (2003) eine Aktivierung derselben Gehirnareale, die auch für die Verarbeitung von räumlicher Information und für das episodische

sche Gedächtnis verantwortlich gemacht werden (parahippokampalen Kortex und des retrosplenialen Kortex, Areae 26, 29 und 30). Die Autoren vermuten deshalb, dass dort eine Brücke zwischen diesen verschiedenen Leistungen liegt.

5 Ziele der vorliegenden Untersuchung

In den vorangehenden vier Kapiteln wurde dargestellt, dass die neuronale Grundlage des Erkennungsvorgangs nicht als ein System der einfachen Feedforward-Analyse des eingehenden Bildes aufgefasst werden kann. Sie muss vielmehr als ein mehrstufiger Vorgang des Schlussfolgerns betrachtet werden, der vom jeweiligen Kontext abhängige Adaptionsvorgänge beinhaltet und im Zusammenhang mit anderen Sinnesinformationen, insbesondere der Tasterfahrung, steht. Eine Objekterkennung ist nur möglich, wenn dem wahrgenommenen Objekt eine Bedeutung beigemessen, wenn es in einen kategoriellen Zusammenhang eingegliedert werden kann. Dazu bedarf es des Abgleichs des Eingangsbildes mit der gespeicherten Repräsentation (Matching-Prozess). Der Fokus der meisten psychophysischen Arbeiten zur Objekterkennung war auf die Natur der Repräsentation gerichtet, ohne zu berücksichtigen, dass sowohl der Gehalt an Bildmerkmalen wie auch das vorhandene Kontextwissen die gemessene Erkennungsleistung wesentlich beeinflussen.

In früheren Untersuchungen unserer Arbeitsgruppe zur Objekterkennung ist deshalb ein Klassifikationsparadigma verwendet worden. Darin muss die Objektrepräsentation unmittelbar im Experiment aufgebaut werden. Zusätzlich wurde Kontextwissen in Form von wohldefiniertem Objektvorwissen (Priming) experimentell integriert. Es wurde gefunden, dass die Fähigkeit, wie Menschen die Strukturen von 3D-Objekten aus 2D-Abbildungen wahrnehmen, entscheidend vom impliziten Objektwissen abhängt (Osman et al., 2000) und sich erst spät im Jugendalter voll entwickelt (Rentschler et al., 2004).

Außerdem hatten diese Arbeiten ein bemerkenswertes Resultat erbracht: Probanden, die die Testobjekte mit verbundenen Augen tastend vorexploriert hatten (haptisches Priming) konnten neue Testansichten besser klassifizieren als Probanden, die die Objekte bereits gesehen, aber nicht angefasst hatten (visuelles Priming). Dieses Ergebnis war insbesondere unter der gängigen Annahme einer intramodalen Objektrepräsentation erstaunlich (vgl. Kap. 2.1.4). Setzte man aber eine supramodale oder visuohaptische Objektrepräsentation voraus (Binkofski et al., 1999; Zangaladze et al., 1999; James et al., 2002), ließen sich diese Ergebnisse besser einordnen. Dann könnten beide Modalitäten gleichermaßen auf die Objektrepräsentation zugreifen. Dennoch war hier nicht zu erklären, warum die haptische Vorerfahrung effektiver sein sollte als die visuelle.

5.1 Experiment I: Multimodale Repräsentation der 3D-Objektstruktur

Das Ziel der vorliegenden zwei Experimente ist eine Präzisierung der in früheren Arbeiten (Osman et al., 2000; Rentschler et al., 2004) gefundenen Überlegenheit von haptischem gegenüber visuellem Objektvorwissen bei der visuellen Klassifikation von Objektansichten. Zu diesem Zweck werden einige Elemente des dort verwendeten Paradigmas modifiziert: In den vorangehenden Arbeiten geschah das visuelle Priming (dort „visual“) durch Betrachten der Objekte am Bildschirm mit der Möglichkeit, diese mit der Maus zu ergreifen und nach Belieben hin und her zu drehen. Eine solche Möglichkeit der Einflussnahme schloss aber eine hilfreiche motorische Rückkopplung über einen Perception-Action-Zyklus (Fuster, 2003) nicht aus. Um die Trennung zwischen visuell und haptisch noch konsequenter durchzuführen, soll in den vorliegenden Experimenten das visuelle Priming ohne jede Möglichkeit der motorischen Einflussnahme auf die Position der Objekte auf dem Bildschirm vorgenommen werden (*vision-Priming*). So sind in beiden vorliegenden Experimenten die verschiedenen Priming-Bedingungen:

- (1) 5 Min. passives Betrachten der sich auf dem Bildschirm nacheinander in alle drei Raumesrichtungen drehenden Objekte (*vision-Priming*),
- (2) 5 Min. aktives Manipulieren realer Objektmodelle bei verbundenen Augen (*motor-Priming*)²,
- (3) eine Kontrollbedingung ohne jegliches Vorwissen (*control*).

Außerdem wird auf eine homogene Einteilung der Probanden in die experimentellen Gruppen geachtet. Dazu wird die Technik der „matched triples“ verwendet (vgl. Kap. 6.1.1). Der weitere Ablauf wird aus den oben genannten Arbeiten übernommen: Die Probanden werden nach dem Priming mittels überwachten Lernens darauf trainiert, einen Satz von Ansichten der Testobjekte zu klassifizieren. Nachdem mit dem Erreichen eines Kriteriums für alle Probanden gleiche Ausgangsbedingungen garantiert sind, treten sie in den Generalisationstest ein und beurteilen neue Testansichten. Anschließend werden die Probanden nach ihrer Introspektion befragt und müssen in einem letzten Test die Testobjekte als reale Modelle rekonstruieren, so wie sie sie sich vorstellen. Im Vergleich dieser Ergebnisse zeigt sich dann die unterschiedliche Wirkung des Objektvorwissens (Primings).

² Diese Form des Priming wurde hier *motor* genannt, um die Betonung auf die *freie Manipulation* des Objektes zu legen (vgl. Kap. 2.1.4 und auch (Lederman & Klatzky, 1998)). Dies, um eine klare Abgrenzung zu anderen Studien vorzunehmen, in denen haptisches Objektvorwissen nur durch statisches Handauflegen erworben wird (Newell, Ernst, Tjan & Bulthoff, 2001). Die hier verwendete Form des „motor-Priming“ entspricht aber in der Art seiner Durchführung dem „haptischen Priming“ der früheren Arbeiten (Osman et al., 2000; Rentschler et al., 2004).

5.2 Experiment II: Bildmerkmale und Lernbarkeit der 3D-Objektstruktur

Die Ergebnisse von Experiment I bestätigten erneut den Effekt der motorischen Verstärkung auf die Klassifikation von chiralen Objekten (Gschwind, Brettel, Osman & Rentschler, 2004). Ein fraglicher Punkt in der Interpretation dieser Ergebnisse ist jedoch: Betrifft dieser Effekt generell die rechts-links-Klassifikation chiraler Strukturen, oder ist er auch von weiteren Eigenschaften der Bildmerkmale abhängig?

Aufgrund der Ergebnisse von Experiment I scheint es deshalb angebracht, ein weiteres Experiment zur Präzisierung durchzuführen. Eine Anknüpfungsmöglichkeit bietet sich in der Beobachtung, dass Abbildungen der verwendeten Testobjekte nicht den Kriterien für „realistische Objekte“ zu genügen scheinen. So sind die Objekte in der OpenInventor-Simulation „nur“ mit Lambertschen und nicht mit strukturierten Oberflächen versehen, die farbliche Abgrenzbarkeit zum Hintergrund ist nicht stark, und insbesondere scheinen die Objekte während der Tiefenrotation auf dem Bildschirm keine feste Kohärenz zu besitzen und die einzelnen Teile (Kugeln) umeinander herum zu schweben.

Deshalb wird ein zu Experiment I äquivalentes Experiment II durchgeführt, in dem die bisher verwendeten Kugel-Objekte durch Objekte, die besonders „realistisch“ ausgestattet sind, ersetzt werden. Alle anderen Parameter der experimentellen Objektpräsentation, wie innere Objektstruktur, Beleuchtungsmodell, Wahl der Ansichten und Präsentationsablauf bleiben dabei wohlgeemerkt unverändert.

Die Begriffe „realistisch“ und „unrealistisch“ bezeichnen die subjektive Erfahrung eines Sachverhalts, den Liu, Knill & Kersten (1995) „reich“ (realistisch) bzw. „arm“ (unrealistisch) an ordnenden Merkmalen und klaren Verbindungen zwischen den Objektteilen nannten. Dies wirft die Frage auf, ob etwa die über das Priming induzierten Effekte des Objektwissens auch von der Verfügbarkeit der Bildmerkmale, die klassifikationsrelevant sind, abhängen (Informationsdichte des Stimulus, stimulus input information).

Indem die beiden Experimente gegeneinander kontrastiert werden, wird in Anlehnung an Liu und Mitarbeiter (Liu et al., 1999) ein Vorgehen gewählt, das die Effekte des impliziten Objektwissens auf die Kategorisierung zwischen partiellen Bildinformationen (Kugelobjekte) und vollständigeren Bildinformationen (Würfelobjekte) vergleicht.

6 Methoden

Die zwei vorliegenden Experimente I und II basieren auf identischen Methoden. Der einzige Unterschied zwischen den Experimenten ist die Variation der Stimuli (vgl. Kap. 6.2.3.): In Experiment I wurden „Kugel-Objekte“ in Experiment II „Würfel-Objekte“ verwendet. Die Methoden werden aus diesem Grund hier nur einmal dargestellt.

6.1 Probanden

Die Probanden waren insgesamt 60 Studenten im Alter zwischen 18 und 32 Jahren. Sie wurden per Listenaushang in der Universität und über Bekanntschaft rekrutiert und hatten zuvor noch nie an einem solchen Experiment teilgenommen. Den Hauptanteil (90%) der Probanden bildeten 54 Studierende der Humanmedizin, eine kleinere Untergruppe (10%) waren 6 Studierende der Architektur. Es handelte sich also um ein eng vorselektiertes Kollektiv. Sie wurden in der Reihenfolge ihrer Teilnahme (pseudozufällig) in die verschiedenen experimentellen Gruppen eingeteilt. Dabei wurde darauf geachtet, dass die Gruppen nach Geschlecht und Studienfachbereich gleich strukturiert waren (*matched triples*; gleichartige unverbundene Stichproben). In jeder Gruppe waren 6 Frauen und 4 Männer. All Probanden waren Rechtshänder und normalsichtig oder sehkorrigiert. Für die Teilnahme am Experiment wurden die Probanden mit 10,00 Euro/h aus der Finanzierung des DFG-Graduiertenkollegs 267 entlohnt.

6.1.1 *Homogene experimentelle Gruppen*

Eine Fallzahl von 10 Probanden in einer experimentellen Gruppe ist aus statistischer Sicht normalerweise zu klein, um für die Ergebnisse eine Normalverteilung anzunehmen. Es wurden jedoch zwei Dinge getan, um eine gewisse Homogenität der Gruppen zu erreichen: Bei der Probandenauswahl wurde aus einem engen vorselektierten Kollektiv geschöpft und bei der Einteilung in die experimentellen Gruppen wurde die Technik der *matched triples* angewendet: Auf den entsprechenden Plätzen in den Gruppen wurden die Probanden nach Geschlecht und Studienfach äquivalent eingeteilt. So kam beispielsweise auf Platz 1 in jeder Gruppe jeweils eine weibliche Studentin der Architektur, auf Platz 2 jeweils ein männlicher Student der Medizin usw.

Fig. 1 zeigt das Histogramm der Altersverteilung:

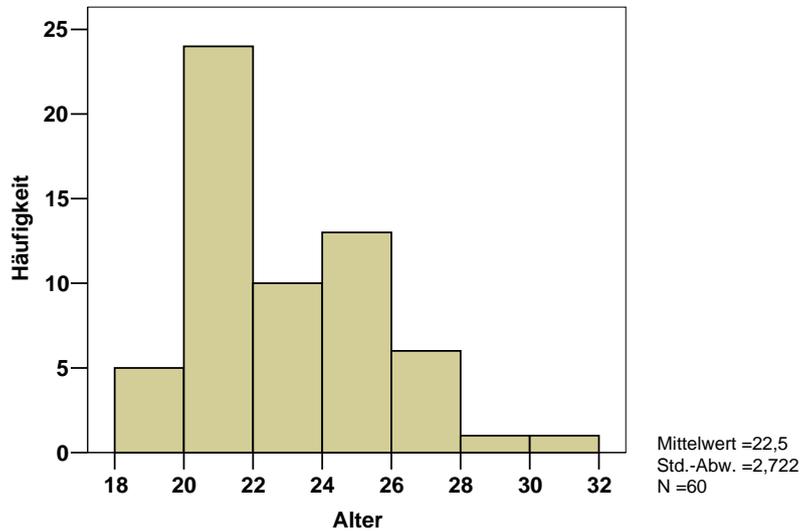


Fig. 1: Histogramm der Altersverteilung

Fig. 2 zeigt die Altersverteilung zwischen den einzelnen experimentellen Gruppen: Für Experiment I: 3 = Kontrollgruppe, 4 = Priming *vision*, 5 = Priming *motor*. Für Experiment II: 6 = Kontrollgruppe, 7 = Priming *vision*, 8 = Priming *motor*, 9 = Abbrecher

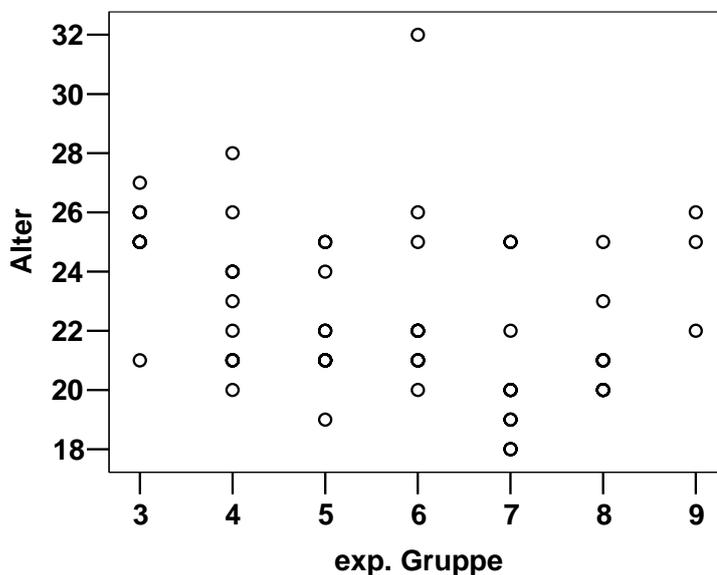


Fig. 2: Altersverteilung zwischen den einzelnen experimentellen Gruppen: Für Experiment I: 3 = Kontrollgruppe, 4 = vision Priming, 5 = motor Priming. Experiment II: 6 = Kontrollgruppe, 7 = vision Priming, 8 = motor Priming, 9 = Abbrecher

Die gleichmäßige Altersverteilung zwischen den Gruppen wurde mit einem nicht-signifikant ausgefallenen Mediantest bestätigt: $\chi^2(7) = 11,162$, $p = 0.132$.

6.2 Stimuli

Stimuli der Familie „Paperclip“ (vgl. Kap. 4.4.1) sind verschiedentlich als sehr realitätsfremd kritisiert worden (Biederman & Bar, 1999; Biederman, 2000). So glaubt z.B. Pizlo, dass „Paperclip“-Stimuli nicht die Merkmale zur Verfügung stellen, die das visuelle System braucht, um das Problem der 3D-Erkennung aus einer 2D-Ansicht zu lösen, und dass deshalb Resultate aus Experimenten mit „Paperclips“ nur wenig darüber aussagen können, wie das visuelle System das inverse Problem der Formwahrnehmung löst (Pizlo, 2001).

Paperclip-Objekte sind auch außerhalb der experimentellen Bedingungen in ruhiger Betrachtung kaum voneinander zu unterscheiden. Biederman und Mitarbeiter monierten die gänzliche Abwesenheit von NAPs (nonaccidental properties), die eine Geon-Extraktion unmöglich mache.

Die hier verwendeten Stimuli bestehen deshalb aus problemlos zu extrahierenden Komponenten (Kugeln, Würfel, Zylinder). Sie verhalten sich aber wie Paperclips, was ihre Teile und Relationen angeht: gleiche Teile in unterschiedlicher Anordnung und ein Verzicht auf Farb- und Texturmerkmale.

6.2.1 Die 3D-Testobjekte und die Konstruktion der Objektansichten

Es wurde ein Satz von 3 Testobjekten gewählt, die untereinander nur anhand ihrer strukturellen 3D-Organisation unterscheidbar (klassifizierbar) sind (Osman et al., 2000). Alle drei Objekte bestanden aus derselben Anzahl von gleichartigen Teilen (*parts*): Sie unterscheiden sich nur durch die Anordnung der Teile im Raum, aber nicht anhand der Merkmale dieser Teile („*part-specific*“ *features* oder *attributes*) (Abb. 14).

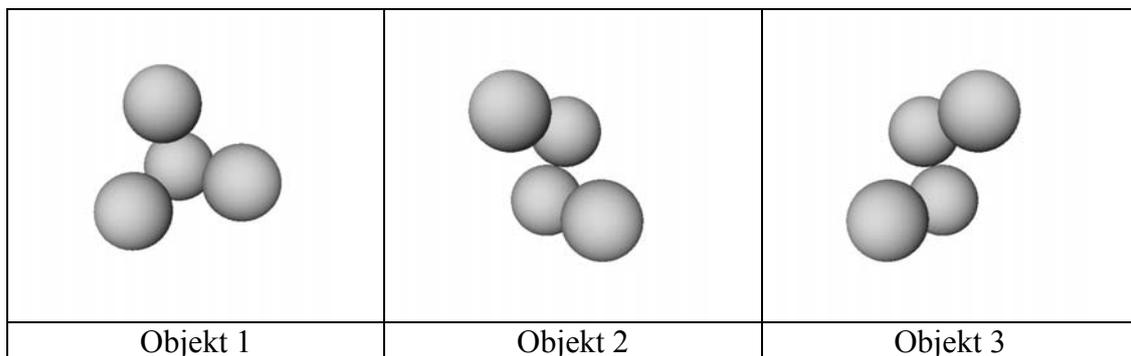


Abb. 14: Die Stimuli, drei nur strukturell unterscheidbare Objekte: das achirale Objekt 1 und das chirale Paar, Objekt 2 und 3 (Osman et al., 2000).

Für die Durchführung des überwachten Kategorienlernens und der Generalisierung der erlernten Klassifikation wurden jeweils 2D-Ansichten (Bilder) dieser 3D-Objekte aus verschiedenen Blickwinkeln computergraphisch erzeugt.

Lernansichten für das überwachte Lernen: Die Einheitskugel um das Objekt (viewing sphere) wurde in Schritte von 60° unterteilt: Drei gegeneinander jeweils um 60° geneigte Ebenen wurden ausgewählt, und auf diesen jeweils wieder drei um 60° gegeneinander geneigte Projektionsachsen bestimmt. Dadurch ergaben sich in der horizontalen Ebene vier und in den beiden schrägen Ebenen jeweils zwei zusätzliche, also insgesamt acht Projektionsachsen, die gegeneinander den Winkel von 60° haben (*Abb. 15*).

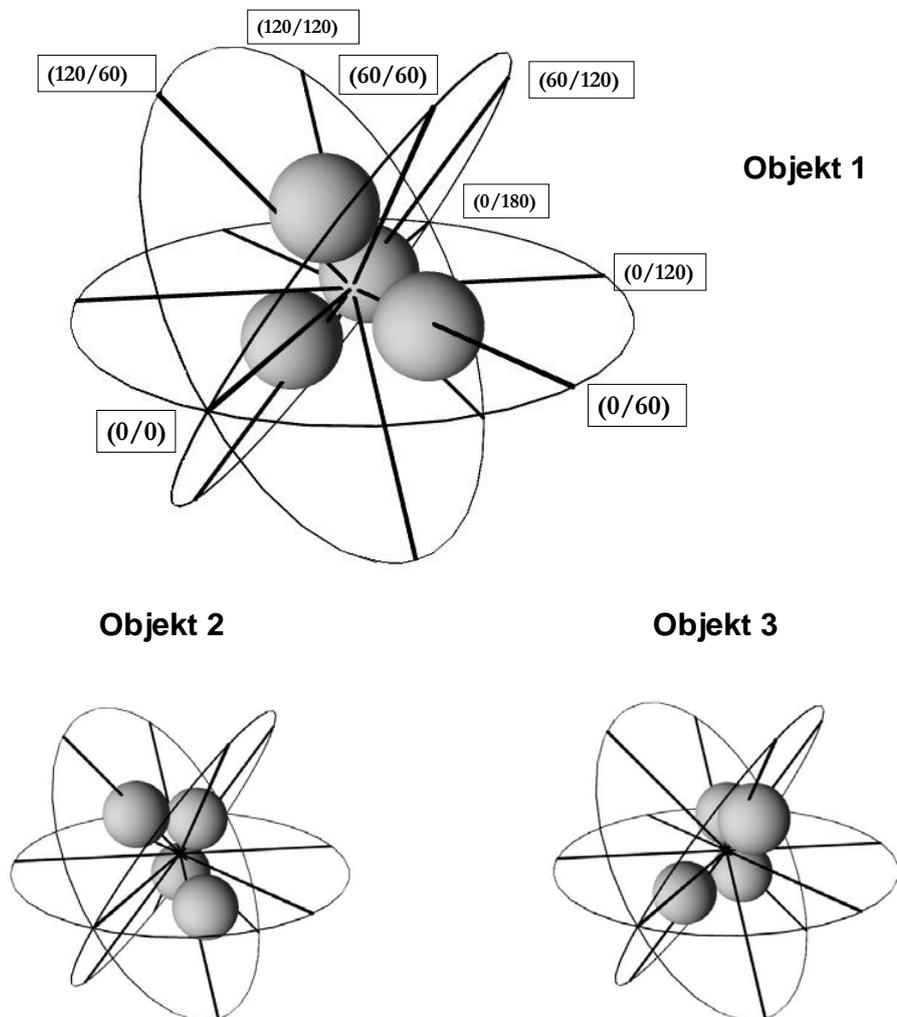


Abb. 15: Perspektivische Darstellung der „viewing sphere“ um die Objekte und der verschiedenen Blickrichtungen aus denen die Ansichten für den Lernsatz gewonnen wurden. Die angezeigten Geraden entsprechen in der experimentellen Präsentation immer der y-Achse des Bildschirms. Oben Objekt 1, unten links Objekt 2, unten rechts Objekt 3.

Die Objekte wurden mit ihrem Schwerpunkt im Zentrum der Einheitskugel platziert. Um das Entstehen spezieller Ansichten durch die Übereinstimmung der Objektachsen mit den Bildschirmachsen zu vermeiden, wurden sie so ausgerichtet, dass ihre Rotationssymmetrieachsen auf die Projektionsachse (0/0) zu liegen kamen (vgl. **Abb. 15**), und keine zwei Kugeln gleichzeitig auf eine Bildschirmachse zu liegen kamen (Rotation gegen den Uhrzeigersinn um die Achse (0/0) um zufällige $11,5^\circ$).

Aufgrund der dreizähligen Rotationssymmetrie von Objekt 1 (120°) waren vier der acht Ansichten, die sich auf diese Weise ergaben, paarweise gleich. Deshalb wurden die doppelten Ansichten jeweils eliminiert. Für alle 3 Objekte zusammen ergab sich so ein Lernsatz von insgesamt 22 Ansichten, der 6 Ansichten für Objekt 1 und jeweils 8 Ansichten für die Objekte 2 und 3 enthielt (vgl. **Abb. 18**). Die Bezeichnung der Ansichten folgte den aufsteigenden Gradzahlen der entsprechenden Azimuth- und Polarwinkel:

<u>1</u> : 0 / 0	<u>5</u> : 60 / 60
<u>2</u> : 0 / 60	<u>6</u> : 60 / 120
<u>3</u> : 0 / 120	<u>7</u> : 120 / 60*) ³
<u>4</u> : 0 / 180	<u>8</u> : 120 / 120*)

Testansichten für die Generalisierung: Zur Untersuchung der Generalisierungsleistung wurden zusätzliche Ansichten der Objekte erzeugt, indem die (halbe) Einheitskugel, die für den Lernsatz in 60° -Schritte unterteilt war, nunmehr in 30° -Schritte weiter unterteilt wurde. Dadurch erhöhte sich die Zahl der Ansichten für das Objekt 1 auf 21 und für die Objekte 2 und 3 auf jeweils 31.

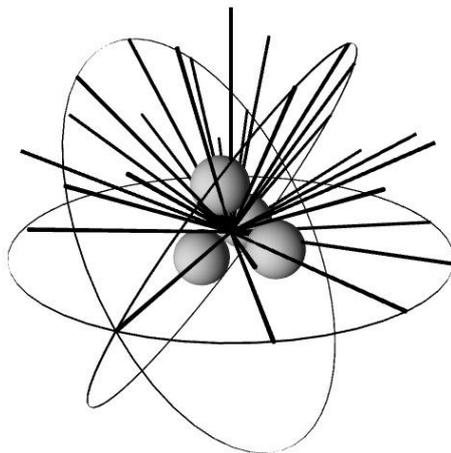


Abb. 16: Darstellung der Blickrichtungen für die 83 Ansichten der Generalisierung.

³ Die mit *) markierten Ansichten Nr. 7 und 8 sind bei Objekt 1 aus Gründen der inneren Objektsymmetrie doppelt. Sie wurden eliminiert.

Jeweils die Ansicht mit der Nummer 4 wurde im Testsatz der Generalisierung weggelassen, da diese mit dem Winkel (0 / 180) die Rückansicht der Grundansicht (0 / 0) darstellte. Von den übrig bleibenden 83 Ansichten waren 5 Ansichten von Objekt 1 schon vom Lernen her bekannt und jeweils 7 Ansichten der Objekte 2 und 3. Für die Untersuchung der Generalisierungsleistung standen also $16+24+24=64$ neue (novel) und zur Kontrolle $5+7+7=19$ bereits vom Lernen her bekannte (learned) Objektansichten zur Verfügung:

Nummerierung der Ansichten für die Generalisierung:

- 64 novel views:
 - O1: [2 4 6 7 8 9 10 11 12 14 16 17 18 19 20 21]
 - O2: [2 4 6 7 8 9 10 11 12 14 16 17 18 19 20 21 22 24 26 27 28 29 30 31]
 - O3: [2 4 6 7 8 9 10 11 12 14 16 17 18 19 20 21 22 24 26 27 28 29 30 31];
- 19 learned views:
 - O1: [1 3 5 13 15]
 - O2: [1 3 5 13 15 23 25]
 - O3: [1 3 5 13 15 23 25]

6.2.2 Objektansichten für rein-strukturbasierte Klassifikation.

Um eine ausschließlich strukturbasierte Klassifikation der Ansichten zu gewährleisten, müssen zwei Bedingungen erfüllt sein (vgl. Kap. 4.4.2): (1) Die Objekte sollten sich nur aufgrund ihrer räumlichen Struktur unterscheiden und (2) die 2D-Ansichten sollten keine hervorstechenden diagnostischen Merkmale aufweisen.

Durch die Konstruktion der hier verwendeten 3D-Testobjekte aus jeweils vier identischen Kugelteilen kam den Merkmalen der räumlichen Anordnung eine besondere Rolle für die Klassifikation dieser Objekte zu. Die 3 Objekte unterschieden sich dabei strukturell in zwei Komplexitätsebenen:

1. **Ebene: achiral versus chiral** – Objekt O1 versus Objekte O2 und O3.
 Objekt O1 besitzt eine zentrale Kugel, die die drei anderen Kugeln berührt. Die Objekte O2 und O3 sind beide lineare, gewundene Kugelketten. Objekt O1 trägt so ein besonderes strukturelles Merkmal, das es bei den anderen nicht gibt. Die Unterscheidung dieses Objektes von den beiden anderen kann durch Detektion oder Ausschluss dieses *diagnostischen Merkmals* erfolgen.
2. **Ebene: chirale Objekte gegeneinander** – Objekt O2 versus Objekt O3.
 Die beiden Objekte sind sog. händige oder chirale Objekte: Sie sind für sich genommen beide asymmetrisch in Bezug auf einander, jedoch gehen sie durch Spiegelung an einer Ebene im 3D-Raum auseinander hervor. Das bedeutet, dass Objekte O2 und O3 nicht nur alle Merkmale ihrer Teile (*part-specific features*), sondern auch alle (binären) Relationen je zweier Teile gemeinsam haben. Abbildungen von händigen Objekten haben per Definition identische spezifische und

binäre Merkmale. Die strukturelle Beschreibung von solchen Objekten erfordert deshalb eine Orientierung an einem übergeordneten Raum (Prelog, 1975; Gross & Bornstein, 1978).

6.2.2.1 Randomisierte Kamerarotation und Minimum-Distance-Classifier

Um dieser Möglichkeit einer Klassifikation der verschiedenen Ansichten aufgrund von Unterschieden in der Verteilung der Bildenergie (Monopol- und Dipolstatistiken der Grauwert-Verteilungen) entgegenzuwirken, wurde bei der Erzeugung der 2D-Ansichten für jede Ansicht eine zusätzliche *Kamerarotation* um die Blickrichtungsachse eingeführt, die für jede Ansicht zufällig einen der Werte 0° , 60° , 120° oder 180° einnahm. Ein Einfluss der statistischen Momente der Verteilung der Bildenergie erster und zweiter Ordnung (Caelli & Julesz, 1978) auf die Klassifikation der 2D-Ansichten konnte dadurch ausgeschlossen werden. Dies zeigt sich, wenn man ihre Klassifikation mithilfe eines sog. *Minimum-Distance-Classifiers* (MDC) simuliert (Osman et al., 2000; Osman, in Vorbereitung) (**Abb. 17**). Dieser verwendet als Abstandsmaß den Kehrwert des Maximums der Kreuzkorrelationsfunktion je zweier Objektansichten. Die Klassifikationswahrscheinlichkeit einer Objektansicht in eine der drei Objektklassen ergibt sich dabei als mittlerer Abstand dieser Ansicht zu allen Ansichten der gewählten Objektklasse, dividiert durch den mittleren Abstand zwischen allen Ansichten des ganzen Ansichtssatzes (Ahmed & Rao, 1975).

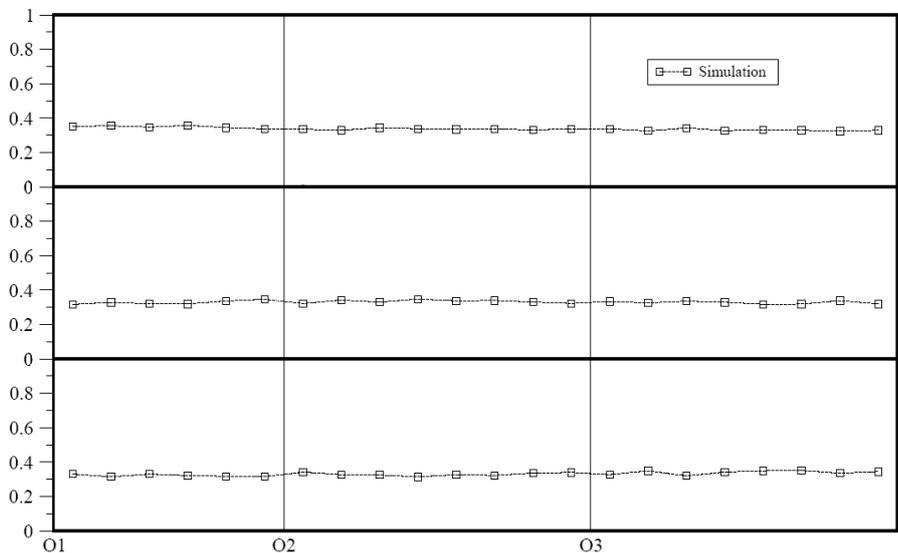


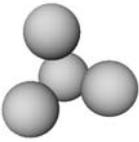
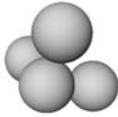
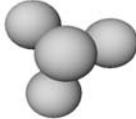
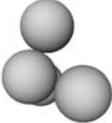
Abb. 17: Simulation der Klassifikation der Ansichten des Lernsatzes mit einem Minimal-Distance-Classifier. Für alle Ansichten des Lernsatzes zeigt sich eine gleichwertige Klassifikationswahrscheinlichkeit in alle drei Objektklassen (aus Osman, in Vorbereitung).

Wie man **Abb. 17** entnehmen kann, ergeben die Klassifikationsvoraussagen des MDC für jede Ansicht des gesamten Lernsatzes (vgl. **Abb. 18**) eine fast vollkommene Gleichverteilung der Klassifikationswahrscheinlichkeiten zu den drei Objektklas-

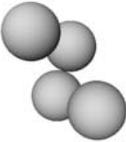
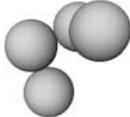
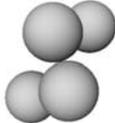
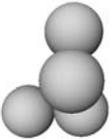
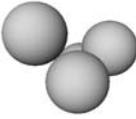
sen. Mit anderen Worten, die Objektansichten sind mithilfe von Bildmerkmalen (erster und) zweiter Ordnung nicht zu klassifizieren.

Das von uns verwendete Paradigma für die 3D-Objekterkennung kann daher auch in Analogie zum Begriff der *Texturunterscheidung bei identischen Bildstatistiken zweiter Ordnung* (iso-second-order texture discrimination nach Caelli & Julesz, 1978) als *strukturbasierte* Objektklassifikation bezeichnet werden.

Objekt 1

			
1_0: (0 / 0 / 120*)	1_1: (0 / 60 / 60)	1_2: (0 / 120 / 120)	1_3: (0 / 180 / 60)§
			
	1_4: (60 / 60 / 60)	1_5: (60 / 120 / 120))	

Objekt 2

			
2_0: (0 / -0 / 180*)	2_1: (0 / -60 / 60)	2_2: (0 / -120 / 60)	2_3: (0 / -180 / 180*)§
			
2_4: (60 / -60 / 0)	2_5: (60 / -120 / 180)	2_6: (120 / -60 / 180)	2_7: (120 / -120 / 60)

Objekt 3

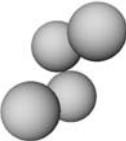
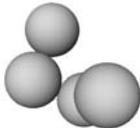
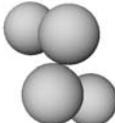
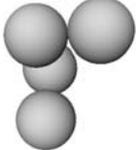
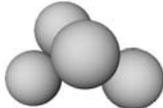
			
3_0: (0 / 0 / 0)	3_1: (0 / 60 / 120)	3_2: (0 / 120 / 120)	3_3: (0 / 180 / 0)§
			
3_4: (60 / 60 / 0)	3_5: (60 / 120 / 120)	3_6: (120 / 60 / 0)	3_7: (120 / 120 / 120)

Abb. 18: Die 22 Ansichten der drei Objekte des Lernsatzes. Mit § gekennzeichnete Ansichten wurden beim Generalisierungssatz aus Redundanzgründen weggelassen (vgl. Kap. 6.2.1).

6.2.3 Variation der Stimuli für Experiment II

Für das Experiment II wurden die Stimuli von Experiment I folgendermaßen modifiziert (vgl. *Abb. 19*):

- **Würfel:** Die Kugeln als identische Objektteile wurden durch Würfel (Kantenlänge = 1, Zwischenraum = 1) ersetzt. Damit erhielten die Objekte mehr sichtbare Linien, Kanten, T- und Y-Junctions, Parallelen – also so genannte „nonaccidental properties“ (Biederman & Bar, 1999).
So vereinfachte sich die Abschätzung der Position der Teile und damit der globalen Objektgeometrie beim Betrachten der Ansichten. Außerdem ist der Würfel mit seinen 90°-Winkeln eine sehr vertraute Struktur in unserem Kulturkreis.
- **Verbindungsstäbchen:** Die verschiedenen Würfel wurden mit zylindrischen Stäbchen verbunden (Länge = 1, Durchmesser = 0,2). Dadurch bestand über den Kontakt und die Relation zwischen den Würfeln absolute Klarheit.
- **Mehr Teile:** Durch die Kombination von Würfeln und Stäbchen wurde die Anzahl der Elemente für ein Objekt von 4 auf 7 erhöht. Dadurch ist in jeder Ansicht eines Objektes mehr Einzelinformation verfügbar.
- **Textur:** Die Oberfläche der Würfel wurde mit einer holzähnlichen braunen Textur überzogen („wood.jpg“ aus dem After Step Classic Paket (asclassic) des Windows Manager Vers. 1.1b, Debian Linux). Die Stäbchen blieben im Kontrast dazu in der silberartigen lambertschen Farbe der Grundeinstellung.
- **Farbe:** Die Textur war dunkelbraun und hob sich somit deutlich vom grauen Hintergrund ab. Auch dies ist eine Hilfe für die Figur-Hintergrund-Trennung.

Wie *Abb. 19* zeigt, bleibt die globale Geometrie der Objekte trotz der Modifikationen erhalten, und die 2D-Projektionen ergeben immer noch Ansichten für eine „nur-Struktur-Klassifizierung“.

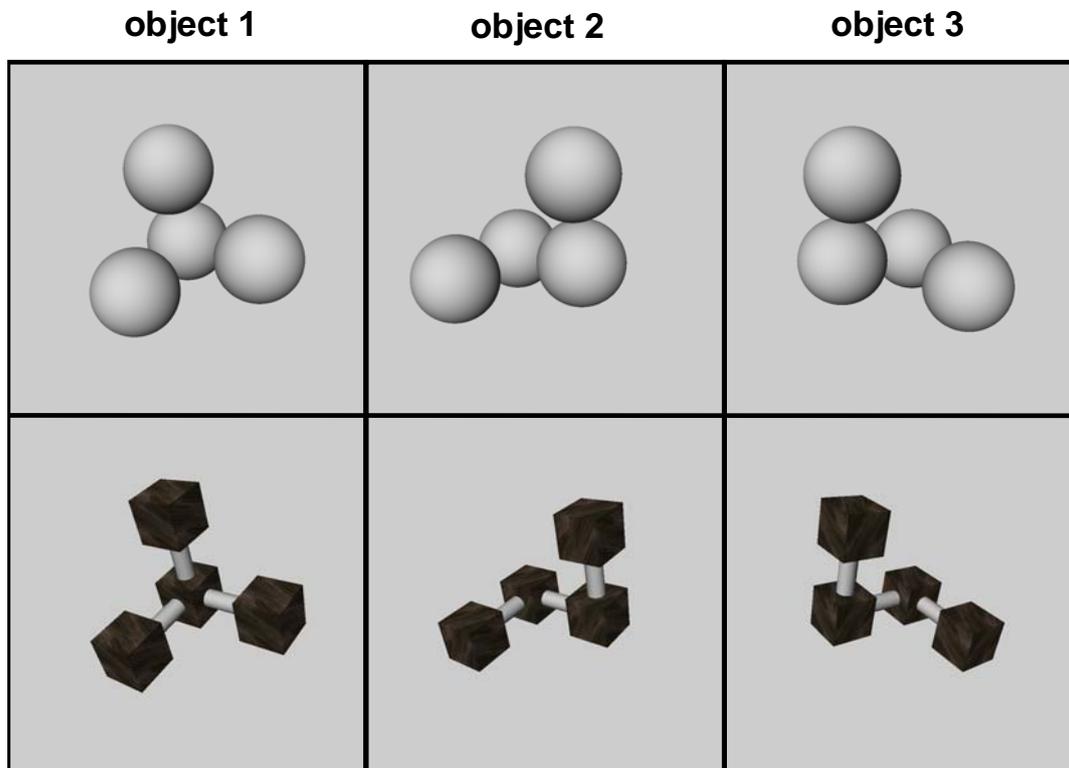


Abb. 19: Die modifizierten Würfel-Objekte für Experiment II (darüber die entsprechenden Kugelobjekte von Experiment I).

6.2.4 Experimentelle Präsentation der Ansichten

Die Ansichten der Objekte wurden computergrafisch mit dem Open Inventor Grafiktoolkit (Silicon Graphics Inc.) auf einem Intel Pentium III Prozessor mit einer NVIDIA Quadro4 200 NVS-Grafikkarte unter *Debian Linux* erzeugt und auf einem 27 x 34 cm Bildschirm präsentiert⁴. Der Augen-Bildschirmabstand betrug etwa 1 m, die Objekte hatten eine mittlere Bildschirmgröße von etwa 13 cm, was einem mittleren Sehwinkel von $\alpha = 7,4^\circ$ entsprach.

Das benutzte Beleuchtungsmodell verwendete das default-Beleuchtungsmodell des OpenInventors mit einem Headlight (Licht in Blickrichtung) der Stärke 1 und einer Umgebungshelligkeit (ambient intensity) der Stärke 0,2. Die Oberflächen der Objekte waren weiß und hatten lambertsche Eigenschaften (keine gerichteten Reflexionseffekte). Der Hintergrund hatte einen gleichmäßigen mittleren Grauwert der Stärke 60%. Es gab keine Schlagschatten, weder auf dem Hintergrund, noch auf den Kugeln selbst.

⁴ Autor dieser Implementation war PD Dr. Hans Brettel, École Nationale Supérieure de Télécommunication, Paris

Die perspektivische Darstellung verwendete eine perspektivische Kamera mit dem Höhenwinkel von $0,79 \pi/4$ (entspricht etwa $35,1^\circ$). Zu den Einstellungen vgl. auch **Tab. 1.:**

X11-server		
	Gamma-Correction	0,8
X-Resources		
	BackgroundColor (line 101&103)	gray60
OpenInventor		
DirectionalLight	Direction	(0,0 0,0 -1,0) (=Headlight)
	Intensity	1
Environment	ambientIntensity	0,2
PerspectiveCamera	heightAngle	0,785398 ($\pi/4$)
RotationSpeed*		1 Umdr = 10sec

Tab. 1: Einstellungen zur Objekt-Präsentation (*Rotation Speed: Umdrehungsgeschwindigkeit beim visuellen Priming)

6.3 Objektvorwissen: Experimentelle Bedingungen

Nachdem allen Probanden eine kurze Einführung in die Problemstellung gegeben und die Aufgabe erklärt worden war (s. Anhang 11.1), wurden sie gemäß der experimentellen Gruppe, denen sie jeweils angehörten, unter unterschiedlichen Bedingungen mit den Objekten vertraut gemacht.

1. Kontrollgruppe (control). Diese Gruppe hatte kein Priming. Sie wurde zum Zwecke des Vergleichs der Priming-Effekte überhaupt geführt.

2. Visuelle Gruppe (vision). Die Probanden lernten die Objekte durch 5 min (passives) Anschauen auf dem Bildschirm kennen (*Abb. 20*):

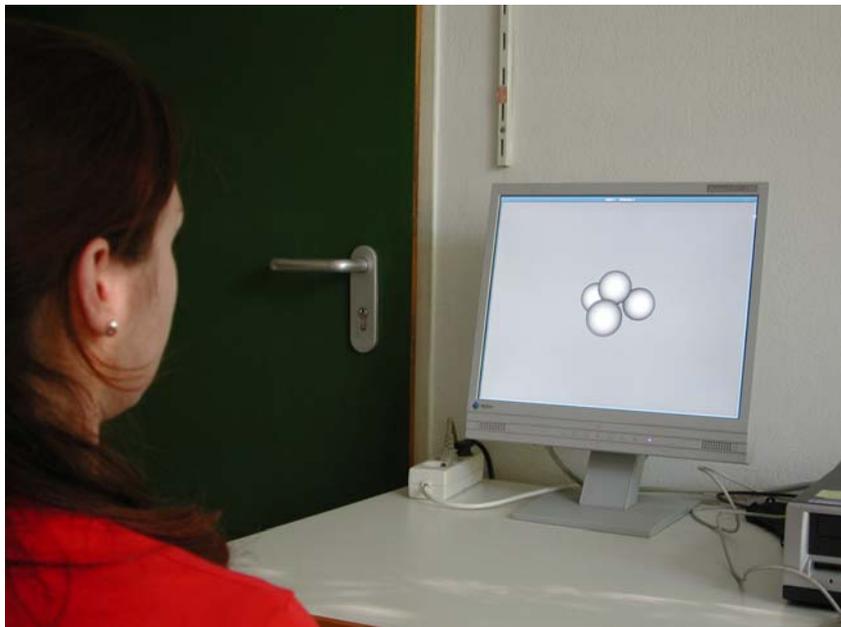


Abb. 20: Visuelles Priming: Die Probandin schaute während 5 Min. nacheinander auf die drei Objekte, die jeweils mehrmals hintereinander um alle drei Raumachsen rotierten.

Die Objekte rotierten nacheinander um alle drei Raumachsen (x, y, z) im Uhrzeigersinn mit einer Drehgeschwindigkeit von 6 Umdrehungen/Minute. Die Probanden sahen jeweils 3 Umdrehungen einer Rotationsrichtung (in 30s), dann folgte die nächste Rotationsrichtung. Die Reihenfolge der Rotationsrichtungen war immer die gleiche:

1. Zyklus: O1 xyz⁵ - O2 zyx - O3 xyz (jeweils 30 Sek. eine Drehrichtung)
2. Zyklus: O1 zyx - O2 xyz - O3 zyx. (jeweils 10 Sek. eine Drehrichtung)

Der Versuchsleiter wechselte die Objekte und Rotationsrichtungen nach diesem Präsentationsschema über die Tasten der PC-Tastatur.

⁵ x, y, z entsprechen den Rotationsachsen im Koordinatenkreuz.

3. Motorische Gruppe (motor⁶). Die Probanden erkundeten die Objekte durch manuelles Abtasten mit verbundenen Augen (*Abb. 21*). Die realen Objektmodelle waren aus Styropor-Kugeln (d = 6 cm) zusammengeklebt. Der Versuchsleiter gab den Probanden die Objekte nacheinander gemäß dem Präsentationsschema einzeln in die Hand. Die Probanden konnten die Objekte nach Belieben drehen und wenden.



Abb. 21: Motorisches Priming: Die Probandin betastete mit verbundenen Augen nacheinander die drei Objekte während insgesamt 5 Min.

Das Präsentationsschema. Für die beiden Gruppen mit Priming (*vision* und *motor*) erfolgte die Priming-Präsentation der drei Objekte (O1, O2, O3) genau identisch nach folgendem Zeitschema:

O1 (90s) => O2 (90s) => O3 (90s) => O1 (10s) => O2 (10s) => O3 (10s)

⁶ Diese Form des Priming wurde hier *motor* genannt, um die Betonung auf die *freie Manipulation* des Objektes zu legen (vgl. Kap. 2.1.4 und auch Lederman & Klatzky, 1998). Dies, um eine klare Abgrenzung zu anderen Studien vorzunehmen, in denen haptisches Objektvorwissen nur durch statisches Handauflegen erworben wird (Newell et al., 2001). Die hier verwendete Form des „motor-Priming“ entspricht aber in der Art seiner Durchführung dem „haptischen Priming“ der früheren Arbeiten (Osman et al., 2000; Rentschler et al., 2004).

6.4 Psychophysische Methode: Das Klassifikations-Paradigma

Es wurde ein Klassifikations-Paradigma gewählt (vgl. Kap. 4.1), in dem die Probanden in zufälliger Folge unverbundene Ansichten von verschiedenen Objekten diesen zuordnen mussten. Die Zuordnung zu den drei Objektklassen musste nach Kriterien erfolgen, die den Probanden zu Anfang unbekannt waren und erst während des überwachten Lernens erarbeitet werden konnten. Es wurde ein großer Bereich von möglichen Blickrichtungen abgedeckt. In Verbindung mit den Kriterien in Kap. 6.2.2. (Objektansichten für rein-strukturbasierte Klassifikation), die die Ansichten der Objekte erfüllten, sollte es damit möglich sein, die Entstehung der Strukturvorstellung während dieses Lernvorganges zu untersuchen.

6.4.1 Überwachtes Kategorielernen (*supervised category learning*)

Das unmittelbar auf das Priming folgende überwachte Kategorielernen bestand aus den drei Elementen:

1. Lernphase
2. Testphase
3. Feedback

Diese drei Elemente bildeten zusammen eine Lerneinheit (LE, engl. learning unit).

1. Lernphase. Die Lernphase wurde vom Probanden selbst über Tastendruck gestartet. Während ihres Ablaufes wurden alle 22 Ansichten des Lernsatzes in zufälliger Reihenfolge einmal präsentiert. Der zeitliche Ablauf für jede einzelne Ansicht war der folgende (vgl. **Abb. 22**, erste Zeile): Nach einer Maske von 800 ms (leerer Bildschirm in gleichmäßig mittlerem Grauwert) folgte die Ansicht des Objektes mit einer Präsentationszeit von 250 ms. Nach einer zweiten Maske von 500 ms folgte für 1000 ms die Lösung in Form der Objektzahl 1, 2 oder 3. Danach begann der Ablauf mit der 800 ms-Maske für die nächste Ansicht wieder von vorne.

2. Testphase. Zwischen der Lernphase und der Testphase bestand eine offene Pause und die Testphase wurde ebenfalls vom Probanden gestartet. Die Testphase begann genau wie die Lernphase mit Maske und randomisierter Ansicht. Danach wurde die Zeit gemessen, bis der Proband über die PC-Tastatur seine Antwort eingab (Response Time). Den Probanden wurde gesagt, sich die nötige Zeit zu nehmen, aber nicht absichtlich zu verlangsamen (vgl. **Abb. 22**, zweite Zeile).

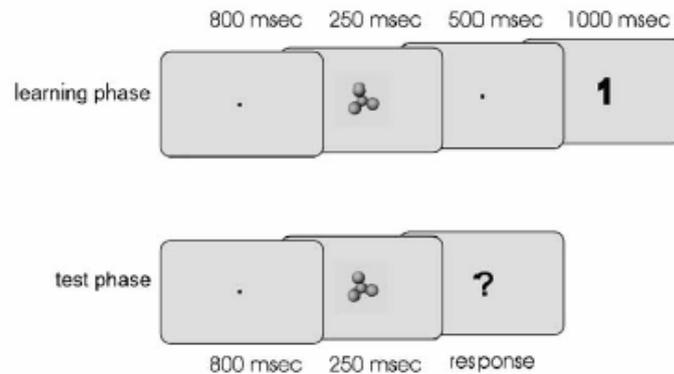


Abb. 22: Abfolge der Präsentationszeiten in der Lernphase (1. Zeile) und der Testphase (2. Zeile). Eine Lernphase und eine Testphase zusammen ergeben eine Lerneinheit (nach Osman et al., 2000). Insgesamt dauerte das Durchlaufen einer Lerneinheit etwa 6100 s.

3. Feedback. Nach einmaliger Darbietung aller 22 Ansichten des Lernsatzes in der Testphase wurde die Klassifikationsmatrix der aktuellen Lerneinheit mit 3x3 Feldern in einem Fenster gezeigt. So konnte der Proband sehen, wie viele Ansichten der einzelnen Objekte er verwechselt hatte (**Abb. 23**).

	Antwort = 1	Antwort = 2	Antwort = 3	
Objekt 1	4	1	1	66,66%
Objekt 2	1	4	3	50%
Objekt 3	0	3	5	62,5%
Total:				59,72%

Abb. 23: Beispiel für eine Klassifikationsmatrix einer Lerneinheit. Die Zahlen beziehen sich auf die Anzahl beurteilter Ansichten (ihre Summe ist 22). In der Zeile die Objekte, in der Spalte die gegebenen Antworten.

Die durchschnittliche Dauer einer kompletten Lerneinheit (LE, learning unit) von 22 Ansichten war etwa 6 Minuten. Die Lerneinheiten wurden so lange wiederholt, bis der Proband eine mittlere Klassifikationsleistung (Total) von mindestens 90% erreicht hatte.

Zwischen den einzelnen Lerndurchgängen stand es den Probanden frei, eine kurze Pause zu machen. Die einzelnen Sitzungen wurden auf 1,5 Std. begrenzt. Falls ein Proband in dieser Zeit das Kriterium von 90% noch nicht erreicht hatte, wurde eine weitere Sitzung innerhalb der nächsten Tage durchgeführt. Falls die 90% gerade noch erreicht worden waren, aber keine Zeit mehr für den anschließenden Generalisierungstest vorhanden war, wurden zu Beginn der neuen Sitzung zuerst noch einmal einige Lerndurchgänge bis zum erneuten Erreichen des 90%-Kriteriums durchgeführt. Nach der Datenauswertung zeigte sich, dass die Zahl dieser zusätzlich einge-

schobenen Lerndurchgänge nie mehr als 4 betrug, und dass sie nie unter 75% korrekt begannen. Sie wurden bei der weiteren Datenauswertung nicht berücksichtigt.

6.4.2 Generalisierungstest

Der dritte Teil des psychophysischen Paradigmas gehörte dem Test auf Generalisierbarkeit des erlernten Objektwissens. Dieser Test diente der Unterscheidung, ob während des überwachten Lernens reines Memorieren der Ansichten des Lernsatzes zum Erfolg geführt hat oder ob ein generalisierbares Objektwissen erworben wurde. Die Generalisierungsfähigkeit gilt als ein Beweis für eine erfolgreiche Objekterkennung und somit für die Existenz einer präzisen mentalen Objektrepräsentation (vgl. Kap. 2.2.1).

Der Generalisierungstest bestand, genau wie die Testphase des überwachten Lernens, aus Maske, Testansicht und offenem Zeitfenster: Nach der grauen Maske von 800 ms folgte für 250 ms eine Ansicht aus dem Testsatz, die der Proband über die Tastatur einem der drei Objekte zuteilen musste. Auch hier wurde die Zeit, bis der Proband die Taste gedrückt hatte, als Antwortzeit (Response Time) gemessen.

Die 83 Ansichten des Testsatzes setzen sich aus den 64 Ansichten aus den neuen Blickrichtungen (novel views) und den 19 zufällig darunter gemischten Ansichten des Lernsatzes (learned views) zusammen und wurden in zufälliger Reihenfolge präsentiert.

6.4.3 Introspektionsberichte der Probanden

Nach Abschluss der Generalisierungsphase wurden die Probanden befragt, was ihnen schwer und was ihnen leicht gefallen war. Diese Berichte wurden stichwortartig notiert und gesammelt. Sie sind in Anhang 11.2 aufgeführt.

6.4.4 Praktische Rekonstruktion der Objekte im Modell

Zum Schluss der Sitzung wurden die Probanden aufgefordert, die Objekte aus der Erinnerung zu konstruieren. Dazu stand ein kleines Magnetspiel mit Kugeln und Kontaktmagnetstäbchen zur Verfügung („SuperMag 24“ der Firma *PlastWood*).

Die Konstruktionsleistung wurde nach folgendem Score bewertet:

- 4 = sofortiges und richtiges Zusammenbauen eines Objektes,
- 3 = richtiges Zusammenbauen eines Objektes, jedoch mit längerem Überlegen und evtl. Nachfragen,

- 2 = für richtige Anzahl Kugeln und sternartige Konstruktion (bei O1) bzw. schlangenartige Konstruktion (bei O2 und O3), aber falsche Winkel (Objekte flach),
- 1 = Punkte für alle Versuche mit 5 und mehr verwendeten Kugeln oder bei kompletter Ratlosigkeit.

Das Nichterkennen der Spiegelsymmetrie/Händigkeit der Objekte O2 und O3 wurde mit der 2 beim falsch gebauten Objekt (meist Objekt O3) zusammen mit einem gesonderten Hinweis bewertet.

6.5 Datenauswertung

Datenanalyse und Datenauswertung erfolgten mit den Programmen MATLAB (Organisation der Rohdaten und Grafiken), MS Excel und Powerpoint (Grafiken) und SPSS 12.0 (Statistik).

6.5.1 Variablen

Bei der Datenauswertung wurden die zwei Abschnitte *Lernen* und *Generalisierung* gesondert betrachtet.

Lernen. Die Ergebnisse wurden an drei Punkten im Ablauf des Experiments analysiert: Die Ergebnisse der ersten Testphase der 1. Lerninheit (1. LE) des überwachten Kategorienlernens dienen zur Einschätzung der unmittelbaren Effekte des Priming. Außerdem wurde die Anzahl der notwendigen Lerneinheiten bis zum Erreichen des Kriteriums von 90% erfasst.

Generalisierung. Bei den Ergebnissen der Generalisierung wurden die Ansichten des Testsatzes (aus neuen Blickwinkeln, novel views) und die randomisiert untergemischten Ansichten des bekannten Lernsatzes (learned views) getrennt analysiert.

So ergaben sich über das gesamte Paradigma drei Messungen für jeden Probanden und jedes Objekt:

- 1. Lerneinheit (1. LE)
- Anzahl notwendige Lerneinheiten bis zum Kriterium
- Novel views
- Learned views

6.5.1.1 Erkennungsleistung.

In der Testphase (bei Lernen und Generalisierung) wurden jeweils die Klassifikationsantworten des Probanden zeilenweise in eine dreispaltige Matrix eingetragen, mit den Antworten 1,2 oder 3 als Einsen in den entsprechenden Spalten und den nummerierten Objektansichten als Zeilen (**Abb. 24**).

Die vierte Spalte zeigt die für die Klassifikationsentscheidung gemessene Antwortzeit in [s] fest. Sie wird kurz als RT (Response Time) abgekürzt.

	<i>Ansicht</i>	Antwort=1	Antwort=2	Antwort=3	RT[s]
<i>Objekt1</i>	1	1	0	0	0.34
	2	0	1	0	1.23
	3	0	0	1	0.45
	4	1	0	0	2.40
	5	1	0	0	0.23
	6	1	0	0	0.34
<i>Objekt2</i>	1	0	1	0	2.23
	2	0	1	0	1.45
	3	0	1	0	2.30
	4	0	1	0	0.73
	5	0	1	0	0.34
	6	0	1	0	1.83
	7	0	1	0	0.35
	8	0	1	0	2.47
<i>Objekt3</i>	1	0	0	1	0.27
	2	0	0	1	3.34
	3	0	0	1	1.23
	4	0	0	1	0.45
	5	0	0	1	2.49
	6	0	0	1	0.23
	7	0	0	1	3.41
	8	0	0	1	0.23

Abb. 24: Beispiel für eine Klassifikationsmatrix des gesamten Lernsatzes. (Der Generalisierungssatz hatte 83 Ansichten und folglich 83 Zeilen bei ebenfalls 4 Spalten).⁷

Diese Klassifikationsmatrizen wurden über die Lerndurchgänge und innerhalb einer experimentellen Bedingung kumuliert, so dass eine Matrix relativer Häufigkeiten für die Erkennungsleistung der drei Ansichtsklassen (Objekte) für jede Lerneinheit oder für jede Generalisierungstesteinheit gewonnen wurde (**Abb. 25**).

⁷ Lesehilfe: Zeile 1: 1. Ansicht von Objekt 1 vom Probanden in 0.34 s richtig klassifiziert.

Zeile 2: 2. Ansicht von Objekt 1 vom Probanden fälschlicherweise mit Objekt 2 verwechselt, bei einer Antwortzeit von 1.23 s.

Zeile 3: 3. Ansicht von Objekt 1, vom Probanden fälschlich als Objekt 3 klassifiziert usw.

	rel. Häufigk. 1	R.H. 2	R.H. 3
Objekt 1	0.73	0.12	0.15
Objekt 2	0	0.5	0.5
Objekt 3	0.17	0.46	0.37

Abb. 25: Beispiel für eine Matrix relativer Häufigkeiten für die Erkennungsleistung der drei Ansichtsklassen (Objekte).

6.5.1.2 d' Berechnung

Die Klassifikationshäufigkeiten wurden zunächst als Prozentkorrekt-Werte berechnet (nur die richtigen Klassifikationen werden berücksichtigt, in **Abb. 25** fettgedruckt). Aus diesen Hitraten (Hit; fett) und den Falschalarmraten (FA, Summe der nichtfetten Einträge in jeder Spalte) wurden anschließend nach der Signal-Entdeckungstheorie als Maß für die Klassifikationsleistung d'-Werte berechnet (Green & Sweets, 1966). Dies geschah mit Hilfe von dafür geschriebenen MATLAB-Skripten mittels folgender Formel:

$$d' = \text{norminv}(\text{Hit}) - \text{norminv}(\text{FA})$$

wobei der MATLAB-Befehl „norminv“ die invertierte Normalverteilungsfunktion berechnet (Probit-Funktion).

Da die Probit-Funktion bei einer relativen Häufigkeit von 0 oder 1 jeweils einen negativen oder positiven unendlichen Wert erreichen würde (asymptotischer Verlauf der Normalverteilung), wurde sie zusätzlich beschränkt: Die Werte 0 und 1 wurden durch 0.025 und 0.975 ersetzt. Dies ergab bei einer Hit-Rate von 1,0 (nun also 0.975) und einer FA-Rate von 0,0 (nun 0.025) ein maximales d' von 4,3101. Diese Ersatzwerte wurden nach MacMillan & Creelman (1991) folgendermaßen berechnet: $1/(2N)$ für die relative Häufigkeit von 0 und $1-1/(2N)$ für die relative Häufigkeit von 1.

N ist dabei die „Körnigkeit“ der relativen Häufigkeiten und entspricht hier: (6 resp. 8 Ansichten/Objekt) x (3 Wiederh. beim Generalisieren) = 18 resp. 24. Um es nicht unnötig zu verkomplizieren wurde ein mittlerer Wert von N = 20 verwendet. Die d'-Werte wurden innerhalb einer Priming-Gruppe für jedes Objekt gemittelt.

6.5.1.3 Antwortzeiten (Response Time)

Die Antwortzeiten waren nicht beschränkt. Die Probanden hatten als einzige Anweisung zu antworten, sobald sie die Antwort wüssten. Die Antwortzeiten wurden für alle Antworten (falsche und richtige) ausgewertet. Damit war das Problem umgangen, dass für eine falsch klassifizierte Ansicht kein einziger „richtiger“ RT-Wert existieren würde, falls ein Proband in einer einzigen Lerneinheit das Kriterium erreicht hatte.

Für jede Klassifikation einer Ansicht wurde eine Antwortzeit in Sekunden (Response Time, RT) registriert. Sie wurde für jedes Objekt über alle Ansichten und die Probanden innerhalb einer experimentellen Bedingung gemittelt.

6.5.2 *Dynamik des Lernens*

Der Verlauf der Lernkurve im überwachten Lernen kann Auskunft darüber geben, auf welche Weise die Objektrepräsentation beim Probanden entsteht. Im Bestreben, auch etwas über diese Dynamik aussagen zu können, wurden neben der Einstiegsleistung (1. Balken von links) und der Anzahl Lerneinheiten noch zwei weitere Maße entwickelt.

Relapse. Würde man davon ausgehen, dass der Lernvorgang monoton ansteigend verläuft, so sollten die Resultate jeder nachfolgenden Lerneinheit besser ausfallen als die der vorangehenden. Ein Rückschritt in der Leistung käme nicht vor. Die Lernschritte könnten unterschiedlich groß sein, wären aber nie negativ. Beim Betrachten der in Anhang 11.5 dargestellten Beispiele von Lernkurven fällt jedoch auf, dass häufig Rückschritte gemacht wurden und der Prozentkorrektwert der nachfolgenden Lerneinheit unter demjenigen der vorangegangenen Lerneinheit liegt. Um ein Maß für die Häufigkeit von Rückschritten zu erhalten, wurden für jeden Probanden die Beträge der negativen Lernfortschritte ($LE_2 - LE_1 < 0$) addiert und die Summe über die Anzahl Lerneinheiten (LE) normiert. Dies ergab das Maß *Relapse* mit der Einheit [Prozentkorrekt pro LE]. NB: Im Unterschied zu einer bekannten Regressionsgeraden mit quadratischen Abständen wird hiermit nicht die Streubreite insgesamt gemessen, sondern nur die *negativen* Fortschritte. Eine zwar stark konkave, aber stetig ansteigende Kurve hätte mit der Regression einen hohen Wert, das Maß *Relapse* jedoch wäre dennoch gleich null (vgl. zur Plausibilität des Maßes *Relapse* den Anhang 11.4).

Progress. Die andere Größe, die interessiert, ist der Fortschritt des Lernens. Er wurde als Differenz zwischen dem zu erreichenden Kriterium von 90% und dem Leistungsniveau der ersten Lerneinheit, normiert über die Anzahl der Lerneinheiten (LE) berechnet (P, Progress). Das ergibt dann den mittleren Fortschritt pro Lerneinheit. Er ist ein Maß für die Steigung der Lernkurve (Prozentkorrekt pro LE).

6.5.3 *Statistische Methoden*

Aufgrund (1) der homogenen Gruppen und (2) nach erfolgtem Nachweis normalverteilter Daten (Kolmogorov-Smirnov-Test) wurde trotz der kleinen Fallzahl parametrische Statistik angewendet.

Nur im Experiment II ergab sich im Abschnitt der Generalisierung aufgrund der sehr guten Leistungen ein Deckeneffekt („ceiling effect“), so dass die Daten nicht

normalverteilt waren. Dort wurde auf nichtparametrische Statistik zurückgegriffen. Im Zweifelsfall wurden beide Tests parallel gerechnet.

Die jeweiligen Ergebnisse für die chiralen Objekte O2 und O3 wurden mittels gepaarter t-Tests auf Unterschiedlichkeit geprüft. Da diese t-Tests ausnahmslos nicht-signifikant ausfielen, konnten die Daten zum Zweck der Datenreduktion gemittelt werden. Sie wurden anstelle der Variablen O2 und O3 als neue Variable O23 in die Berechnungen miteinbezogen.

Die Ergebnisse der 1. Lerneinheit (1. LE) wurden mithilfe von einfaktoriellen Varianzanalysen (oneway-ANOVAs) jeweils für die Objekte O1 und O23 auf Unterschieds-Effekte zwischen den Priming-Gruppen untersucht. Zur weiteren Ausdifferenzierung der Signifikanzen wurden post-hoc-Tests nach Bonferroni berechnet (Field, 2000). Dies wurde für die d' -Werte und für die Antwortzeiten (RT) jeweils getrennt durchgeführt.

Zur Analyse der Ergebnisse der Generalisierung wurde eine 2x2x3 Varianzanalyse gemischten Typs (Mixed Factorial ANOVA) durchgeführt mit den Innersubjekt Faktoren „Objekt“ (O1, O23) und „Messung“ (novel views, learned views) und dem Zwischensubjektfaktor „Priming-Condition“ (control, vision, motor). Zur weiteren Ausdifferenzierung der Signifikanzen wurden post-hoc-Tests nach Bonferroni berechnet.

Alle Tests wurden für die d' -Werte und für die Antwortzeiten (RT) getrennt durchgeführt.

Wenn nicht anders vermerkt, zeigen die Fehlerbalken in den Grafiken einen Standardfehler an (± 1 S.E.).

7 Ergebnisse

7.1 Experiment I

Die Ergebnisse wurden für die experimentellen Abschnitte vor dem Kriterium von 90% (überwachtes Lernen) und nach dem Kriterium (Generalisierung) getrennt analysiert. Ebenso wurden die Ergebnisse für die mittlere Klassifikationsleistung (d') und die mittlere Antwortzeit (RT) getrennt analysiert.

Die Klassifikationsleistungen für die chiralen Objekte O2 und O3 unterschieden sich nicht (gepaarte t-Tests nicht signifikant). Sie wurden der Einfachheit halber gemittelt und in den statistischen Berechnungen jeweils als eine neue Variable O23 berücksichtigt.

Wenn nicht anders vermerkt, zeigen die Fehlerbalken in den Grafiken einen Standardfehler an (± 1 S.E.).

7.1.1 Lernen

Für die Beurteilung des Lernverhaltens standen drei Wertegruppen zur Verfügung: Die Klassifikationsleistung (d') und Antwortzeit (RT) der allerersten Lerneinheit (1. LE) sowie die mittlere Anzahl benötigter Lerneinheiten bis zum Erreichen des Kriteriums.

Klassifikationsleistung. *Fig. 3* zeigt die Klassifikationsleistung (Signal detection d' , vgl. Kap. 6.5.1.2) für die zwei Priming-Bedingungen (*vision* und *motor*) und die Kontrollgruppe (*control*) während der ersten Testphase des überwachten Lernens (1. Lerndurchgang).

Für die Kontrollgruppe liegt die Klassifikationsleistung erwartungsgemäß im Zufallsbereich (links). Die Priming-Gruppe *motor* lässt eine deutlich verbesserte Klassifikation besonders der Objekte O2 und O3 erkennen (rechts).

Die Klassifikationsleistung der ersten Lerneinheit wurde für O1 und O23 mit jeweils einer einfaktoriellen Varianzanalyse (oneway-ANOVA) für den Faktor „Priming-Bedingung“ analysiert.

Für das achirale Objekt O1 ergab sich kein signifikanter Effekt.

Für die chiralen Objekte O23 ergab sich jedoch ein signifikanter Unterschied zwischen den Priming-Gruppen ($F(2,24) = 9.04, p < 0.01$). Im post-hoc-Test nach Bonferroni zeigte sich ein signifikanter Unterschied der Priming-Gruppe *motor* gegenüber der Priming-Gruppe *vision* und der Kontrollgruppe (jeweils $p < 0.01$).

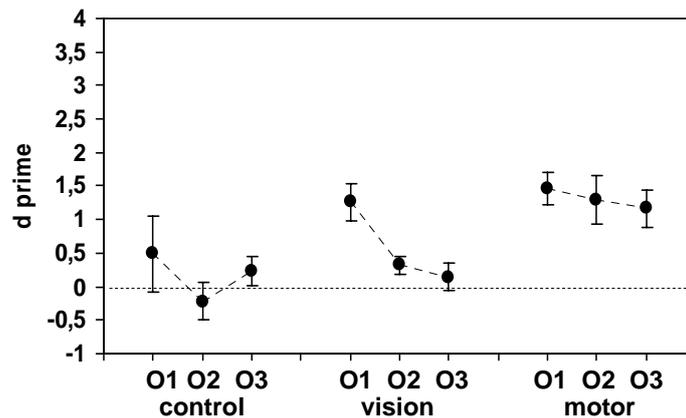


Fig. 3: Klassifikationsleistung für die Kontrollgruppe (*control*) und die zwei Priming-Bedingungen (*vision* und *motor*) während der ersten Testphase des überwachten Lernens (1. Lerndurchgang). Die Priming-Gruppe *motor* klassifizierte die chiralen Objekte O2 und O3 signifikant besser als die beiden anderen Gruppen (Fehlerbalken ± 1 S.E.).

Antwortzeiten. *Fig. 4* zeigt die dazugehörigen Antwortzeiten (RT): Die Kontrollgruppe hat die kürzesten Antwortzeiten, die Priming-Gruppe *motor* deutlich längere.

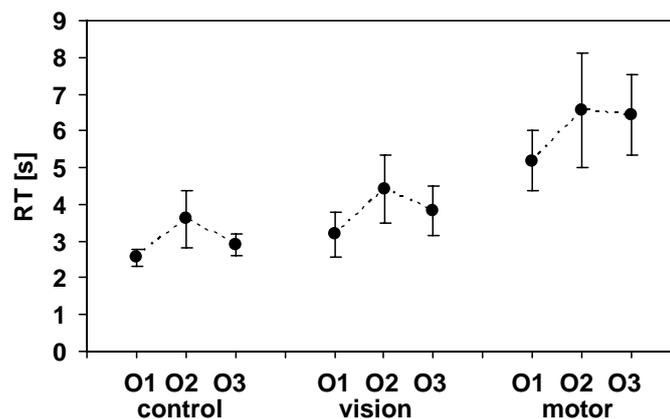


Fig. 4: Antwortzeiten (*Response Time*) für die Kontrollgruppe (*control*) und die zwei Priming-Bedingungen (*vision* und *motor*) während der ersten Testphase des überwachten Lernens. Die Priming-Gruppe *motor* benötigte signifikant mehr Zeit für die Klassifikation von Objekt 1 und nahezu signifikant mehr für die Objekte 2 und 3.

Wie die Klassifikationsleistungen (d') wurden auch die Antwortzeiten (RT) der ersten Lerneinheit für die Objekte O1 und O23 mittels jeweils einer einfaktoriellen Varianzanalyse (oneway-ANOVA) für den Faktor „Priming-Bedingung“ analysiert.

Sie zeigte einen Effekt der Priming-Bedingung für das achirale Objekt O1 ($F(2,24) = 4.182, P < 0.05$) und der post-hoc-Test nach Bonferroni zeigte einen signifikanten Unterschied zwischen den Gruppen *motor* und *control* ($p < 0,05$).

Für die chiralen Objekte O23 ergab sich nur ein zur Signifikanz tendierender Effekt ($F(2,24) = 3.13, p = 0.062$).

Anzahl Lerneinheiten. *Fig. 5* zeigt die mittlere Anzahl der benötigten Lern-durchgänge bis zum Kriterium von 90% korrekter Klassifikation aller Ansichten des Lernsatzes. In der Kontrollgruppe haben 3 Probanden das Kriterium von 90% nicht erreicht. Sie haben nach 34, 43 bzw. 44 Lerneinheiten das überwachte Lernen abgebrochen (zusätzliche Datenpunkte x). Der dargestellte Mittelwert wurde aus den restlichen sieben Probanden, die das Kriterium erreicht haben, berechnet (links).

Hier zeigt sich ein tendenzieller Vorteil der Priming-Gruppe *motor* gegenüber der Gruppe *vision* und der Kontrollgruppe. Der Test nach Kruskal-Wallis ist jedoch nur beinahe signifikant: $\chi^2(2) = 4.629, p = 0.099$.

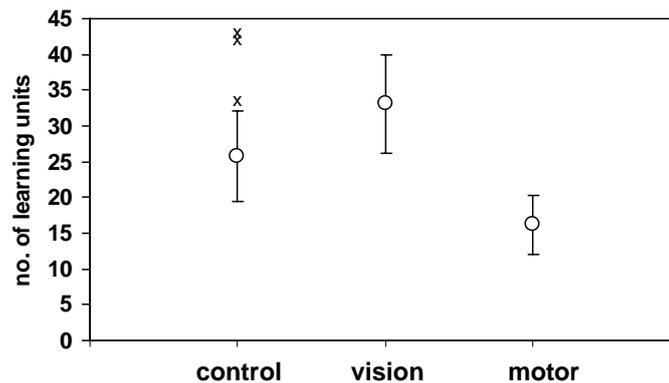


Fig. 5: Die mittlere Anzahl der benötigten Lerneinheiten des überwachten Kategorienlernens bis zum Kriterium von 90% korrekter Klassifikation für alle Ansichten. In der Kontrollgruppe haben 3 Probanden das Kriterium von 90% nicht erreicht. Sie haben nach 34, 43 bzw. 44 Lerneinheiten das Experiment abgebrochen (zusätzliche Datenpunkte x).

7.1.2 Generalisierung

Für die Beurteilung der Generalisierung standen zwei Wertegruppen zur Verfügung: Die Klassifikationsleistung (d') und Antwortzeit (RT).

Klassifikationsleistung. *Fig. 6* zeigt die Generalisierungsleistung (Signal detection d') für die 64 neuen Ansichten (novel views, schwarze Quadrate) und die Reklassifikationsleistung für die 19 bereits bekannten Ansichten des Lernsatzes (learned views, leere Rhomben) für die Kontrollgruppe *control* und die zwei Priming-Bedingungen *vision* und *motor*.

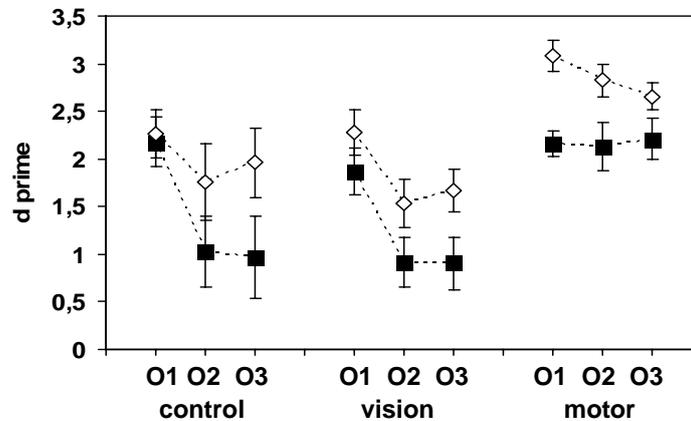


Fig. 6: Generalisierungsleistung für die 64 neuen Ansichten (novel views, schwarze Quadrate) und Reklassifikationsleistung für die 19 bereits bekannten Ansichten des Lernsatzes (learned views, leere Rhomben) für die Kontrollgruppe (control) und die zwei Priming-Bedingungen vision und motor. Die Priming Gruppe motor generalisierte alle drei Objekte gleich gut und die Objekte O2 und O3 signifikant besser als die anderen zwei Gruppen, zwischen denen kein Unterschied besteht.

Für die Beurteilung der Generalisierungsleistung und des Grades an View-dependence wurde eine 2x2x3-Faktoren Varianzanalyse des gemischten Typs (mixed type factorial ANOVA) durchgeführt mit den Innersubjektfaktoren Objekt (O1, O23) und Ansichten (novel, learned) und dem Zwischensubjektfaktor Priming-Bedingungen (control, vision, motor).

Für alle drei Faktoren zeigte sich je ein signifikanter Effekt:

Für den Unterschied zwischen dem achiralen Objekt O1 und den chiralen Objekten O23 ($F(1,24) = 21.37, p < 0.001$),

für den Unterschied zwischen den novel und den learned Ansichten ($F(1,24) = 31.91, p < 0.001$) und

für den Unterschied zwischen den Priming-Bedingungen ($F(2,24) = 6.90, p < 0.01$). Post-hoc-Tests nach Bonferroni ergaben einen signifikanten Unterschied zwischen den Gruppen control und motor ($p < 0.05$) und zwischen den Gruppen vision und motor ($p < 0.01$).

Außerdem zeigte sich eine signifikante Interaktion zwischen Objekttyp, Ansichtsart und Priming-Condition ($F(2,24) = 3.92, p < 0.05$). Das heißt, dass die Verhältnisse zwischen den einzelnen Faktoren (novel, learned und achiral, chiral) sich innerhalb der verschiedenen Priming-Gruppen unterschieden und innerhalb dieser jeweils gesondert betrachtet werden müssen. In Verbindung mit den Ergebnissen der post-hoc-Tests bedeutet dies, dass es im Unterschied zu den Gruppen control und vision in der Gruppe motor keinen Unterschied zwischen den Objekten gab, und dass die gelernten Ansichten (leere Rhomben) für alle drei Objekte signifikant besser als die neuen (schwarze Quadrate) klassifiziert wurden. Im Gegensatz dazu wurde in den beiden Gruppen control und vision das Objekt O1 signifikant besser als die Objekte

O2 und O3 klassifiziert und ein Unterschied zwischen den neuen und den bereits gelernten Ansichten bestand nur für die Objekte O23, nicht aber für Objekt O1.

Antwortzeiten. *Fig. 7* zeigt die dazugehörigen Antwortzeiten (Response Time, RT) für novel views (schwarze Quadrate) und learned views (leere Rhomben).

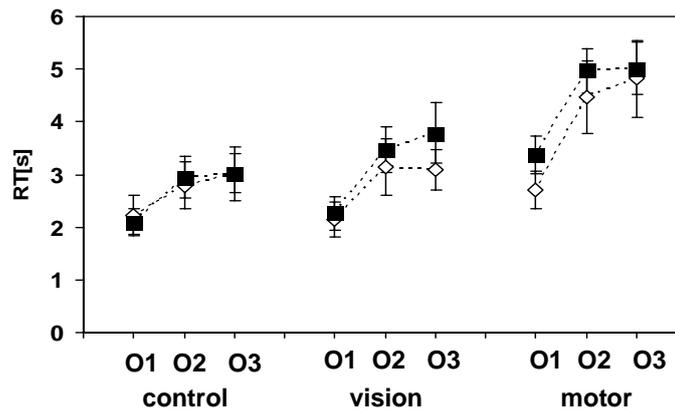


Fig. 7: Antwortzeiten (Response Time) für die 64 neuen Ansichten (novel views, schwarze Quadrate) und Reklassifikationsleistung für die 19 bereits bekannten Ansichten des Lernsatzes (learned views, leere Rhomben) für die Kontrollgruppe (control) und die zwei Priming-Bedingungen vision und motor. Die chiralen Objekte O2 und O3 wurden signifikant langsamer generalisiert als das achirale Objekt O1. Insgesamt benötigte die Priming-Gruppe motor signifikant mehr Zeit.

Die Kontrollgruppe zeigt auch hier die kürzesten Antwortzeiten, die Priming-Gruppe motor die längsten.

Für die Beurteilung der Antwortzeiten während der Generalisierung wurde ebenfalls eine 2x2x3-Faktoren Varianzanalyse des gemischten Typs (mixed type factorial ANOVA) durchgeführt mit den Innersubjektfaktoren Objekt (O1, O23) und Ansichten (novel, learned) und dem Zwischensubjektfaktor Priming-Bedingung (control, vision, motor).

Es zeigte sich nur ein signifikanter Effekt für den Unterschied zwischen dem achiralen Objekt O1 und den chiralen Objekten O23 ($F(1,24) = 20.89, p < 0.001$) und für einen Unterschied zwischen den Priming-Conditions ($F(2,24) = 4.73, p < 0.05$). Post-hoc-Tests nach Bonferroni ergaben einen signifikanten Unterschied zwischen den Gruppen motor und control ($p < 0.05$) und einen zur Signifikanz tendierenden Unterschied zwischen den Gruppen motor und vision ($p = 0.06$).

Es zeigten sich nur ein tendenzieller Effekt für den Unterschied zwischen den novel und den learned Ansichten ($F(1,24) = 3.55, p = 0.072$) und keine Interaktionen.

7.1.3 Objektkonstruktion

Fig. 8 zeigt die Ergebnisse der Rekonstruktionsaufgabe. Das Objekt O1 wird von allen drei Gruppen in über 80% schnell und sicher gebaut. Die Priming-Gruppe *motor* ist aber als einzige in der Lage, die chiralen Objekte in über 70% der Fälle schnell und sicher zu bauen. Die Priming-Gruppe *vision* erreicht wie die Kontrollgruppe nur etwa 30% und hat in weiteren 30% überhaupt keine Ahnung, wie diese chiralen Objekte aussehen.

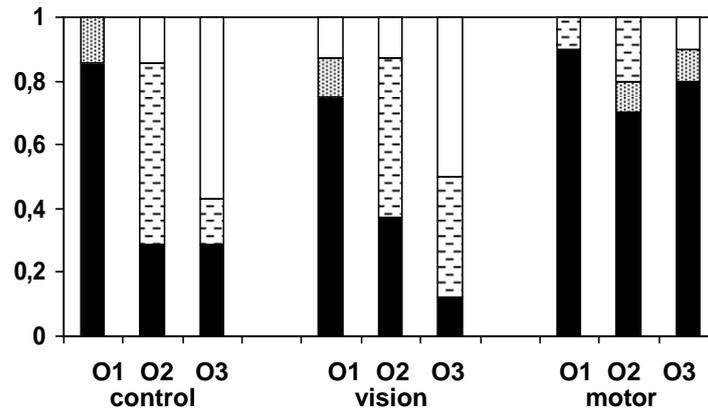


Fig. 8: Prozentuales Histogramm der Score-Ergebnisse der Rekonstruktionsaufgabe. Entschiedene und rasche Konstruktion = 4 (schwarz), zögerliche, aber richtige Konstruktion = 3 (gepunktet), richtige Anzahl Kugeln, aber falsche Winkel = 2 (gestrichelt), falsche Anzahl Kugeln = 1 (weiß). Nur in der Priming-Gruppe *motor* konnten alle drei Objekte von 80% der Probanden richtig konstruiert werden.

Die Rekonstruktion wurde vom Experimentator durch einen Rating-Score beurteilt: Entschiedene und rasche Konstruktion = 4 (schwarz), zögerliche, aber richtige Konstruktion = 3 (gepunktet), richtige Anzahl Kugeln, aber falsche Winkel = 2 (gestrichelt), falsche Anzahl Kugeln = 1 (weiß).

Die Probanden begannen die Konstruktion mit Objekt O1, dann folgte Objekt O2 und schließlich Objekt O3. Aus dieser durch die Nummerierung sich ergebenden Reihenfolge erklärt sich das durchweg schlechter ausgefallene Objekt O3: Die Probanden haben mit Mühe Objekt O2 zusammengebaut und waren dann gänzlich ratlos, wie sich Objekt O3 davon unterscheiden sollte. Wenn Objekt O2 oder O3 nicht in der richtigen chiralen Version sondern als das jeweilige Spiegelbild konstruiert wurde, wurde es mit dem Score von 2 beurteilt (richtige Anzahl Kugeln, aber falsche Winkel, gestrichelt).

7.1.4 Zusammenfassung der Ergebnisse

Die wichtigsten Ergebnisse sollen noch einmal zusammengefasst werden.

Lernen. Nur das motorische Priming ermöglichte eine äquivalente Klassifikationsleistung für alle drei Objekte. Diese Äquivalenz über alle Objekte in der Gruppe *motor* war schon zu Beginn des überwachten Lernens vorhanden und ging mit deutlich erhöhten Antwortzeiten einher.

Generalisierung. Der Generalisierungstest zeigte ein ähnliches Bild. Die Gruppe *motor* generalisierte als einzige alle drei Objekte äquivalent auf hohem Niveau – auch hier ging diese gute Leistung auf Kosten von Reaktionszeit. Die Gruppe *vision* dagegen zeigte wie die Kontrollgruppe zwar eine gute Generalisierung für Objekt O1, nicht aber für die chiralen Objekte O23. Die bereits bekannten Ansichten (learned views) wurden von allen drei Gruppen für beide Objekttypen besser klassifiziert als die neuen (novel views), wobei die Gruppe *motor* als einzige sehr hohe d' Werte für das Objekt O1 erreichte.

Objektkonstruktion. Auch die Ergebnisse des Rekonstruktionstests fügten sich in dieses Bild: Die Gruppe *motor* war als einzige in der Lage, alle drei Objekte gleichermaßen sicher und rasch zu bauen, während die anderen beiden Gruppen jeweils nur Objekt O1 in vergleichbarer Qualität rekonstruierten, die Objekte O2 und O3 jedoch nur zu einem geringen Anteil.

7.2 Experiment II

Wie für Experiment I wurden auch für Experiment II die Ergebnisse für die experimentellen Abschnitte vor dem Kriterium von 90% (überwachtes Lernen) und nach dem Kriterium (Generalisierung) getrennt betrachtet. Ebenso wurden auch die mittlere Klassifikationsleistung (d') und die mittlere Antwortzeit (RT) getrennt analysiert.

Auch bei Experiment II unterschied sich die Klassifikationsleistung für die chiralen Objekte O2 und O3 an keinem Punkt. Sie wurden deswegen wieder gemittelt und für die Statistik als eine neue Variable „O23“ berücksichtigt.

Wenn nicht anders vermerkt, zeigen die Fehlerbalken in den Grafiken einen Standardfehler an (± 1 S.E.).

7.2.1 Lernen

Für die Beurteilung des Lernverhaltens standen zwei Wertegruppen zur Verfügung: Die Klassifikationsleistung (d') und Antwortzeiten (RT) der allerersten Lerneinheit (1. LE).

Klassifikationsleistung. Die Werte für Experiment II zeigten sich für einige Variablen in den Gruppen *control* und *motor* als nicht normalverteilt (signifikantes Resultat im Kolmogorov-Smirnov-Test für die Gruppe *motor*), weswegen eine nicht-parametrische Statistik verwendet wurde.

Die Klassifikationsleistung der ersten Lerneinheit (vgl. **Fig. 9**) wurde für die Objekte O1 und O23 mit jeweils einer Rangreihenvarianzanalyse nach Kruskal-Wallis für die Priming-Bedingungen analysiert.

Für das achirale Objekt O1 ergab sich ein signifikanter Unterschied zwischen den Priming-Gruppen ($\chi^2(2) = 7.949$, $p < 0.05$). Die post-hoc durchgeführte Reihe von Mann-Whitney-Tests zeigte ein signifikantes Abfallen der Kontrollgruppe gegenüber den zwei anderen Priming-Gruppen auf (*control* vs. *vision*: $U = 19.5$, $p < 0.05$; *control* vs. *motor*: $U = 18.0$, $p < 0.05$).

Für die chiralen Objekte O23 ergab sich ein tendenziell signifikanter Effekt ($\chi^2(2) = 5.665$, $p = 0.059$).

Eine parallel gerechnete oneway-ANOVA ergab die gleichen Resultate. Für Objekt O1 zeigte auch sie einen signifikanten Unterschied: ($F(2,27) = 9.12$, $p < 0.01$). Im Post-Hoc-Test nach Bonferroni zeigte sich ein signifikanter Unterschied sowohl zwischen der Gruppe *control* und *vision* ($p < 0.01$) als auch zwischen der Gruppe *control* und *motor* ($p < 0.01$).

Für Objekt O23 ergab sich nur ein tendenziell signifikanter Unterschied ($F(2,27) = 3.14$, $p = 0.060$).

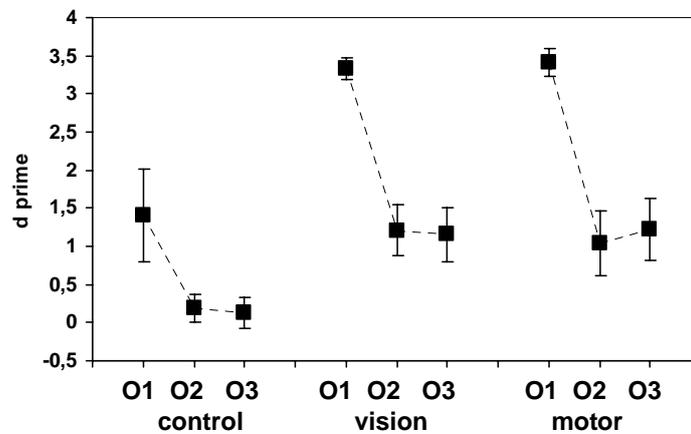


Fig. 9: Klassifikationsleistung für die Kontrollgruppe (control) und die zwei Priming-Bedingungen (vision und motor) während der ersten Testphase des überwachten Lernens. Die Kontrollgruppe klassifizierte alle drei Objekte signifikant schlechter als die beiden anderen Gruppen.

Antwortzeiten. Die Antwortzeiten der ersten Lerneinheit (vgl. **Fig. 10**) waren alle normalverteilt (Kolmogorov-Smirnov-Test nicht signifikant). Sie wurden für die Objekte O1 und O23 jeweils mit einer einfaktoriellen Varianzanalyse (oneway-ANOVA) für die Priming-Bedingung analysiert.

Für das achirale Objekt O1 ergab sich kein signifikanter Effekt ($F(2,27) = 1.12$, $p = 0.341$).

Für die chiralen Objekte O23 zeigte sich ein signifikanter Effekt ($F(2,27) = 5.63$, $p < 0.01$). Gemäß einem post-hoc-Test nach Bonferroni bestand dieser Effekt in einem signifikanten Unterschied der Kontrollgruppe zu den zwei Priming-Gruppen (jeweils $p < 0.01$).

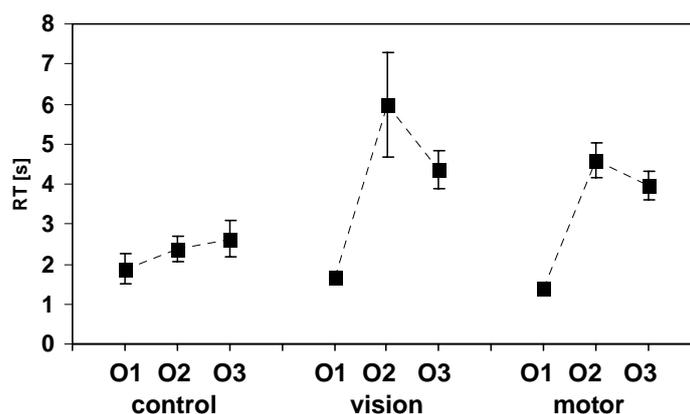


Fig. 10: Antwortzeiten (Response Time) für die Kontrollgruppe (control) und die zwei Priming-Bedingungen (vision und motor) während der ersten Testphase des überwachten Lernens. Die Kontrollgruppe benötigte für die chiralen Objekte O2 und O3 signifikant weniger Zeit als die beiden anderen Gruppen.

7.2.2 Generalisierung

Klassifikationsleistung. Die Klassifikationsleistungen in der Generalisierungsphase von Experiment II (**Fig. 11**) waren alle ausnehmend gut, so dass ein starker Deckeneffekt (ceiling-effect) eine Normalverteilung der Daten verhinderte.

Eine wegen der fehlenden Normalverteilung durchgeführte nicht-parametrische Rangreihenvarianzanalyse nach Kruskal-Wallis zeigte für beide Objekttypen keinen Effekt der Priming-Gruppen, weder für die neuen Ansichten (novel, für O1: $\chi^2(2) = 3.337$, $p = 0.189$; bzw. für O23: $\chi^2(2) = 4.049$, $p = 0.132$) noch für die bekannten Ansichten (learned, für O1: $\chi^2(2) = 2.247$, $p = 0.325$; bzw. für O23: $\chi^2(2) = 0.705$, $p = 0.703$).

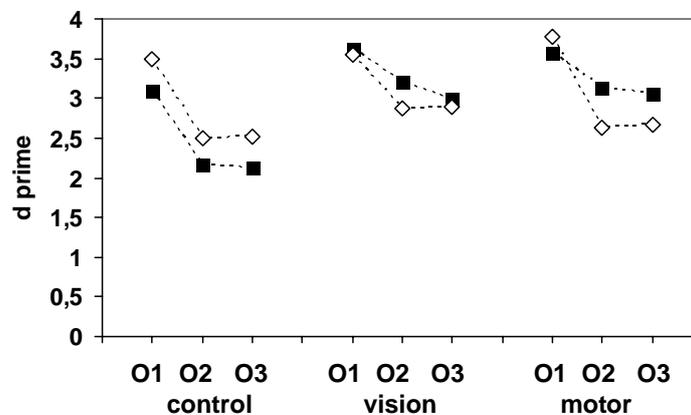


Fig. 11: Generalisierungsleistung für die 64 neuen Ansichten (novel views, schwarze Quadrate) und Reklassifikationsleistung für die 19 bereits bekannten Ansichten des Lernsatzes (learned views, leere Rhomben) für die Kontrollgruppe (control) und die zwei Priming-Bedingungen vision und motor. Die Daten waren aufgrund des ceiling-effects nicht normalverteilt. Die Mittelwerte sind nur zur Orientierung dargestellt und die Fehlerbalken sind weggelassen. Es bestand kein Unterschied zwischen den Gruppen.

Antwortzeiten. Für die Beurteilung der Antwortzeiten während der Generalisierung (**Fig. 12**) wurde eine 2x2x3-Faktoren Varianzanalyse des gemischten Typs (mixed type factorial ANOVA) durchgeführt mit den Innersubjektfaktoren Objekt (O1, O23) Ansichten (Novel, Learned) und dem Zwischensubjektfaktor Priming-Condition (control, vision, motor).

Es zeigte sich nur ein signifikanter Effekt für den Unterschied zwischen dem achiralen Objekt O1 und den chiralen Objekten O23 ($F(1,27) = 76.43$, $p < 0.001$).

Es zeigte sich aber kein Unterschied für die Priming-Conditions ($F(1,27) = 1.83$, $p = 0.181$) und kein Effekt für den Unterschied zwischen den novel und den learned Ansichten ($F(1,27) = 0.091$, $p = 0.765$) und ebenso keine Interaktionen.

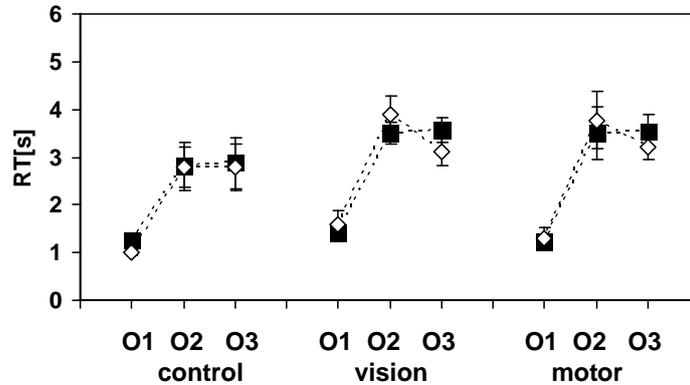


Fig. 12: Antwortzeiten (Response Time) für die 64 neuen Ansichten (novel views, schwarze Quadrate) und Reklassifikationsleistung für die 19 bereits bekannten Ansichten des Lernsatzes (learned views, leere Rhomben) für die Kontrollgruppe (control) und die zwei Priming-Bedingungen vision und motor. Die chiralen Objekte O2 und O3 wurden von allen Gruppen signifikant langsamer generalisiert als das achirale Objekt O1.

7.2.3 Objektkonstruktion

Fig. 13 zeigt die Ergebnisse der Rekonstruktionsaufgabe. Von beiden Priming-Gruppen konnten alle drei Objekte in hohem Maße schnell und sicher gebaut werden. Die Kontrollgruppe fällt dagegen ab, für alle drei Objekte.

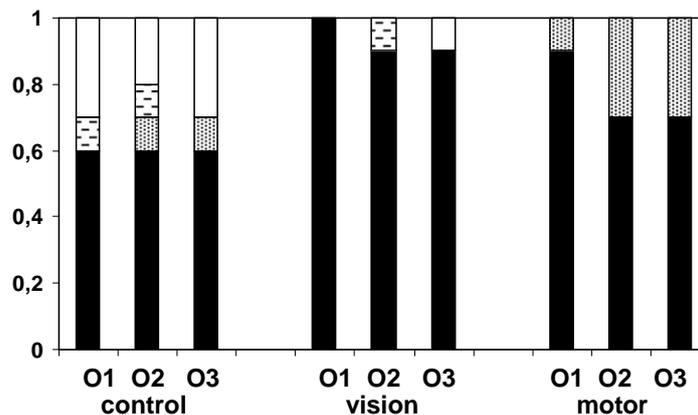


Fig. 13: Prozentuales Histogramm der Score-Ergebnisse der Rekonstruktionsaufgabe. Entschiedene und rasche Konstruktion = 4 (schwarz), zögerliche, aber richtige Konstruktion = 3 (gepunktet), richtige Anzahl Kugeln, aber falsche Winkel = 2 (gestrichelt), falsche Anzahl Kugeln = 1 (weiss). In der Kontrollgruppe vermögen nur 60% der Probanden alle Objekte zielstrebig richtig zu bauen.

Die Rekonstruktionsleistung wurde vom Experimentator mit Hilfe eines Rating Scores beurteilt: entschiedene und rasche Konstruktion = 4 (schwarz), zögerliche, aber richtige Konstruktion = 3 (gepunktet), richtige Anzahl Kugeln, aber falsche Winkel = 2 (gestrichelt), falsche Anzahl Kugeln = 1 (weiß).

7.2.4 Zusammenfassung der Ergebnisse

Die Ergebnisse von Experiment II sollen hier noch einmal kurz zusammengefasst werden:

Lernen. Beide Priming-Bedingungen ermöglichten ein sehr gutes Klassifizieren von O1 schon gleich von Anfang an (über 90%, $d' > 3$). Die Kontrollgruppe fiel dagegen signifikant ab.

Die Antwortzeiten (RT) zeigen keinen Effekt für O1 aber eine starke Verzögerung für die chiralen Objekte O2 und O3 in den beiden Priming-Gruppen *vision* und *motor*, während die Kontrollgruppe sich signifikant davon unterschied und alle drei Objekte gleich schnell behandelte.

Generalisierung. In der Generalisierung ist die Kontrollgruppe zu den beiden anderen Gruppen aufgerückt: Es gibt nun keine Unterschiede mehr zwischen den Gruppen, auch keine Unterschiede zwischen den Objekten, und keinen Unterschied zwischen neuen (novel) und bekannten (learned) Ansichten. Da alle d' -Werte um 3 und darüber liegen, muss von einem „ceiling effect“ ausgegangen werden, die Aufgabe von Experiment II war zu leicht.

Dieses Bild wird durch die Antwortzeiten (RT) bestätigt, dort gibt es auch keine Unterschiede zwischen den Gruppen, aber die Objekte bilden sich nach gewohnter Weise ab: Die Generalisierung der achiralen Objekte O2 und O3 ist nun in allen Gruppen deutlich verzögert. Das rasche Antworten der Kontrollgruppe ist verschwunden.

Objektkonstruktion. In der Objektrekonstruktionsaufgabe fällt wiederum die Kontrollgruppe gegen die zwei anderen Gruppen ab.

7.3 Vergleich von Experiment I und Experiment II: Kugeln versus Würfel

Nachdem beide Experimente einzeln ausgewertet worden waren, wurde auch noch ein paarweiser Vergleich der korrespondierenden Gruppen durchgeführt. Verglichen wurde die Kontrollgruppe I (Kugeln) mit der Kontrollgruppe II (Würfel), die Gruppe *vision I* (Kugeln) mit der Gruppe *vision II* (Würfel) und die Gruppe *motor I* (Kugeln) mit der Gruppe *motor II* (Würfel). Aus dem unterschiedlichen Verhalten zwischen den korrespondierenden Gruppen in beiden Experimenten lassen sich die essentiellen Schlüsse ziehen. Um diesen Unterschied beurteilen zu können wurden an ausgewählten Stellen wegen der nicht immer gegebenen Normalverteilung der Daten nicht-parametrische Rangreihenvarianzanalysen nach Kruskal-Wallis durchgeführt. Parallel dazu wurden auch oneway-ANOVAs gerechnet, die bis auf eine Ausnahme jedoch immer die gleichen Signifikanz-Ergebnisse lieferten.

7.3.1 Lernen

Fig. 14 zeigt die Gegenüberstellung der Klassifikationsleistungen (d') und der entsprechenden Antwortzeiten (RT) der 1. Lerneinheit von Experiment I (Kugeln) und Experiment II (Würfel).

Klassifikationsleistung. Die Klassifikationsleistung für Objekt O1 und für Objekt O23 wurde jeweils zwischen den beiden korrespondierenden Gruppen in Experiment I und Experiment II verglichen:

Die Kontrollgruppen verhielten sich in beiden Experimenten in Bezug auf achirale und chirale Objekte gleich. Weder Kruskal-Wallis-Test (für O1: $\chi^2(1) = 0.343$, $p = 0.558$; bzw. für O23: $\chi^2(1) = 0.288$, $p = 0.591$) noch oneway-ANOVA (für O1: $F(1,15) = 1.11$, $p = 0.31$ bzw. für O23: $F(1,15) = 0.39$, $p = 0.543$) zeigen einen Unterschied.

Das vision Priming bewirkte bei den Würfeln für beide Objekttypen eine starke Verbesserung der Leistung. Kruskal-Wallis-Test für O1 ($\chi^2(1) = 13.845$, $p < 0.001$) bzw. für die Objekte O23 ($\chi^2(1) = 6.223$, $p < 0.05$) und oneway-ANOVA für O1 ($F(1,18) = 43.53$, $p < 0.001$) bzw. für O23 ($F(1,18) = 6.61$, $p < 0.05$) stimmen überein.

Das motorische Priming bewirkte bei den Würfelobjekten im Vergleich zu den Kugelobjekten noch eine Steigerung für das achirale Objekt 1, nicht aber für die Objekte O23: Kruskal-Wallis für O1 ($\chi^2(1) = 12.122$, $p < 0.001$) und für O23: ($\chi^2(1) = 0.242$, $p = 0.623$). Oneway-ANOVA für O1: $F(1,18) = 44.33$, $p < 0.001$ und für O23: $F(1,18) = 0.037$, $p = 0.849$.

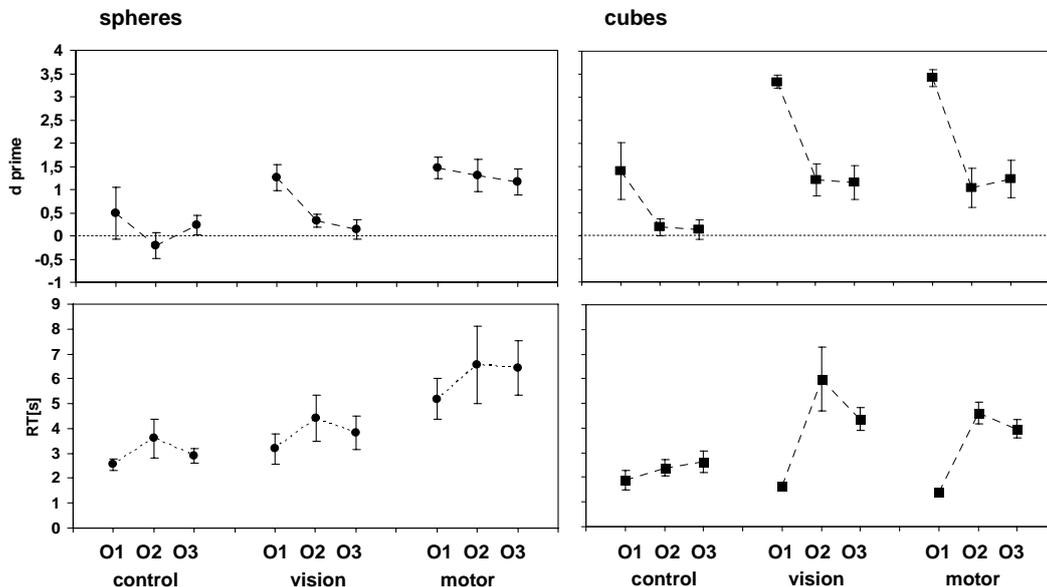


Fig. 14: Gegenüberstellung der Klassifikationsleistungen (d') und der entsprechenden Antwortzeiten (RT) der 1. Lerneinheit von Experiment I (links, spheres) und Experiment II (rechts, cubes). Klassifikation: Die Kontrollgruppen unterscheiden sich nicht; die Gruppe vision ist in allen Objekten signifikant besser; die Gruppe motor ist nur bei Objekt O1 signifikant besser. RT: Die Kontrollen unterscheiden sich nicht; die Gruppe vision ist bei Objekt O1 tendenziell schneller, die Gruppe motor ist bei Objekt O1 signifikant schneller.

Antwortzeiten. Der Vergleich der Antwortzeiten für Objekt O1 und Objekt O23 zwischen den korrespondierenden Gruppen der beiden Experimente zeigte folgendes Ergebnis:

Die beiden Kontrollgruppen unterschieden sich auch bezüglich der Antwortzeiten für alle Objekte nicht. Kruskal-Wallis-Test für O1, $\chi^2(1) = 2.752$, $p = 0.097$ bzw. für O23: $\chi^2(1) = 0.952$, $p = 0.329$. Oneway-ANOVA für O1: $F(1,15) = 1.76$, $p = 0.204$ bzw. für O23: $F(1,15) = 1.627$, $p = 0.222$.

Das vision Priming bewirkt bei den Würfeln tendenziell für das achirale Objekt O1 eine leichte Verkürzung der Antwortzeit, keinesfalls aber für O23: Der Kruskal-Wallis-Test zeigte für O1: $\chi^2(1) = 3.291$, $p = 0.070$ bzw. für O23: $\chi^2(1) = 0.966$, $p < 0.326$. und die oneway-ANOVA zeigt für O1: $F(1,18) = 5.97$, $p < 0.05$ bzw. für O23: $F(1,18) = 0.836$, $p = 0.373$.

Kruskal-Wallis-Test und ANOVA stimmen hier ausnahmsweise nicht überein. Der Kruskal-Wallis-Test gilt als konservativer als die ANOVA, und insofern ist dieses Ergebnis als Tendenz zur Signifikanz zu verstehen.

Das motorische Priming bewirkt bei den Würfelobjekten im Vergleich zu den Kugelobjekten noch eine deutliche Beschleunigung für das achirale Objekt O1, nicht aber für die chiralen Objekte O23: Der Kruskal-Wallis-Test ist signifikant für O1 $\chi^2(1) = 9.143$, $p < 0.01$, nicht signifikant für O23 $\chi^2(1) = 2.766$, $p = 0.096$. Desglei-

chen die oneway-ANOVA signifikant für O1: $F(1,18) = 20.62$, $p < 0.001$, für O23: $F(1,18) = 3.174$, $p = 0.092$.

Anzahl benötigte Lerndurchgänge. Fig. 15 zeigt die mittlere Anzahl der benötigten Lerndurchgänge für Experiment I (Kugelobjekte, Kreise) und Experiment II (Würfelobjekte, Quadrate). Die Anzahl der benötigten Lerndurchgänge ist ein grobes Maß für die Effizienz des Lernvorganges. Der statistische Mittelwertsvergleich wurde mit Mann-Whitney-Tests durchgeführt.

Die zwei Kontrollgruppen zeigten keinen Unterschied ($U = 33.5$, $p = 0.887$).

Die beiden Priming-Bedingungen profitierten jedoch in hohem Maß von den Würfelobjekten und zeigten jeweils eine deutliche Reduktion der notwendigen Anzahl Lerneinheiten: *vision*: ($U = 8.5$, $p < 0.01$) und *motor* ($U = 15.0$, $p < 0.01$).

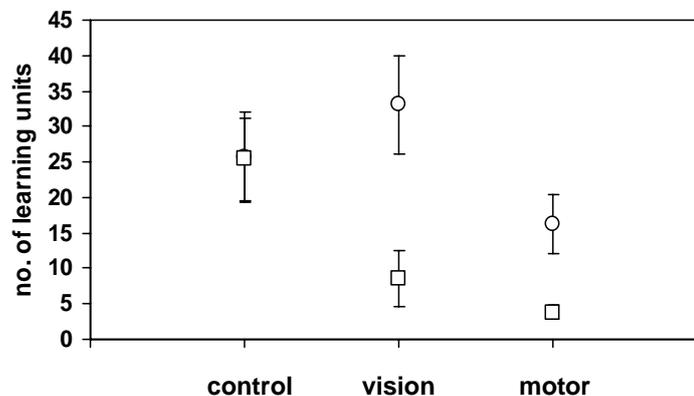


Fig. 15: Die mittlere Anzahl der benötigten Lerneinheiten des überwachten Kategorienlernens bis zum Kriterium von 90% korrekter Klassifikation für alle Ansichten. Die Gruppen *vision* und *motor* benötigen in Experiment II (Würfelsymbole) signifikant weniger Lerneinheiten.

7.3.2 Generalisierung

Fig. 14 zeigt die Gegenüberstellung der Klassifikationsleistungen (d') und der entsprechenden Antwortzeiten (RT) für die Generalisierung von Experiment I (Kugeln) und Experiment II (Würfel).

Klassifikationsleistung. Die Generalisierungsleistung für die neuen Ansichten (novel views) wurde für Objekte O1 und für Objekt O23 jeweils zwischen den Gruppen verglichen:

Die Würfel-Kontrollgruppe zeigt im Vergleich zur Kugel-Kontrollgruppe eine deutliche Verbesserung der Generalisierungsleistung in Bezug auf beide Objekttypen (achiral, chiral). Kruskal-Wallis-Test für O1: $\chi^2(1) = 5.038$, $p < 0.05$ bzw. für O23: $\chi^2(1) = 4.610$, $p < 0.05$.

Das vision Priming bewirkt bei den Würfeln für beide Objekttypen eine starke Verbesserung der Generalisierung. Kruskal-Wallis-Test für O1: $\chi^2(1) = 14.296$, $p < 0.001$ bzw. für O23: $\chi^2(1) = 12.091$, $p < 0.01$.

Das motorische Priming bewirkt bei den Würfelobjekten im Vergleich zu den Kugelobjekten noch eine starke Verbesserung für beide Objekttypen: Kruskal-Wallis für O1: $\chi^2(1) = 14.372$, $p < 0.001$ bzw. für O23: $\chi^2(1) = 8.251$, $p < 0.01$.

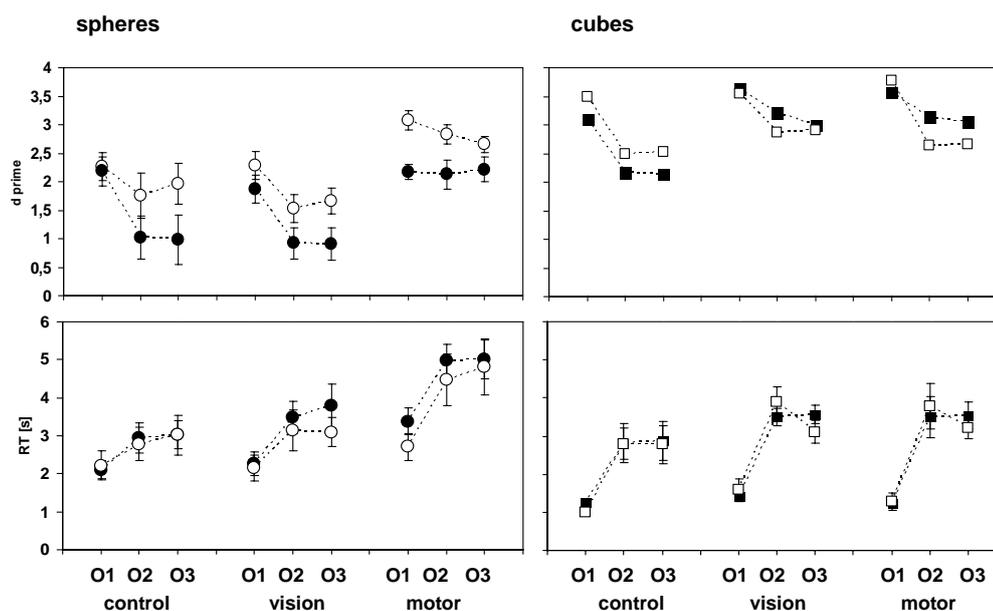


Fig. 16: Gegenüberstellung der (d') und der entsprechenden Antwortzeiten (RT) der Generalisierung (novel views, schwarze Symbole) und der Reklassifizierung der bekannten Ansichten (learned views, leere Symbole) von Experiment I (links, spheres) und Experiment II (rechts, cubes).

Antwortzeiten. Die Antwortzeiten für die neuen Ansichten (novel views) wurden für Objekte O1 und für Objekt O23 jeweils zwischen den Gruppen verglichen.

Die Kontrollgruppe II (Würfelobjekte) erreichte im Vergleich zur Kontrollgruppe I (Kugelobjekte) eine Verringerung der Antwortzeit für das achirale Objekt O1, nicht aber für die chiralen Objekte: Kruskal-Wallis-Test für O1: $\chi^2(1) = 6.943$, $p < 0.01$ bzw. für O23: $\chi^2(1) = 0.610$, $p = 0.435$. Oneway-ANOVA für O1: $F(1,15) = 13.355$, $p < 0.01$ bzw. für O23: $F(1,15) = 0.043$, $p = 0.838$.

Das vision Priming bei den Würfeln bewirkte im Vergleich zu den Kugeln ebenfalls nur für das achirale Objekt O1 eine leichte Verkürzung der Antwortzeit, keinesfalls aber für die Objekte O23: Der Kruskal-Wallis-Test zeigte für O1: $\chi^2(1) = 4.806$, $p < 0.05$, bzw. für O23: $\chi^2(1) = 0.000$, $p = 1.0$. Die oneway-ANOVA zeigte für O1: $F(1,18) = 5.364$, $p < 0.05$ bzw. für O23: $F(1,18) = 0.034$, $p = 0.856$.

Das motorische Priming bewirkt bei den Würfelobjekten im Vergleich zu den Kugelobjekten noch eine deutliche Beschleunigung für beide Objekttypen: Kruskal-Wallis für O1: $\chi^2(1) = 14.286$, $p < 0.001$, bzw. für O23 $\chi^2(1) = 5.143$, $p < 0.05$. Desgleichen die oneway-ANOVA für O1: $F(1,18) = 35.218$, $p < 0.001$, bzw. für O23: $F(1,18) = 5.387$, $p < 0.05$.

7.3.3 Dynamik des überwachten Lernens

Fig. 17 zeigt den mittleren Lernfortschritt *Progress* über dem Rückfallsmaß *Relapse* (vgl. Kap. 6.5.2) aufgetragen für jeden einzelnen Probanden aus Experiment I (volle Symbole) und Experiment II (leere Symbole).

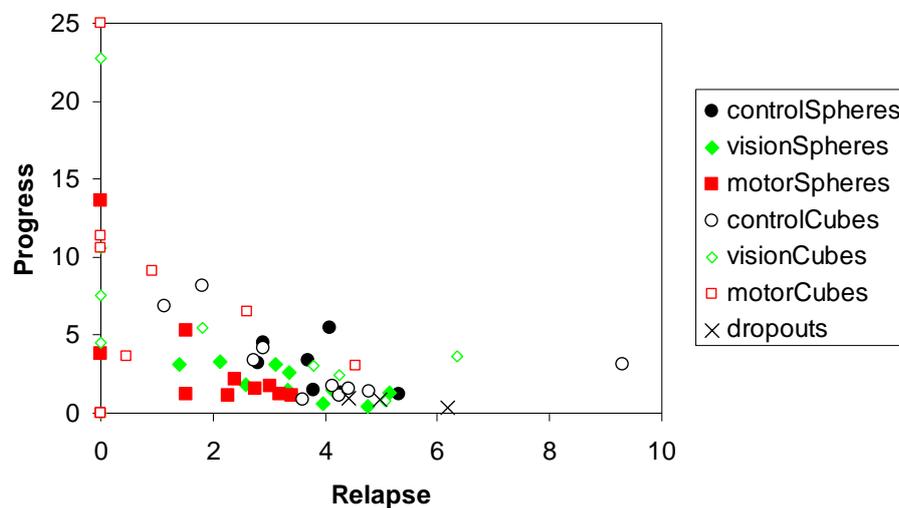


Fig. 17: Verhältnis von mittlerer Steigung der Lernkurve (*Progress* [Prozentkorrekt pro LE]) und mittlerer Rückschritt pro Lerneinheit (*Relapse* [negative Prozentkorrekt pro LE]) für jeden einzelnen Probanden von Experiment I und Experiment II.

Je größer der Wert von *Progress* ist, desto steiler ist der Anstieg der Lernkurve. Je größer der Wert von *Relapse* ist, desto größer war die Tendenz des Probanden, Lerneinheiten mit guten Ergebnissen auch wieder solche mit schlechten folgen zu lassen. Liegt der Punkt (*Relapse/Progress*) im Ursprung, so erreichte der Proband schon in der ersten Lerneinheit das Kriterium. Die Darstellung der Gruppenmittelwerte zeigt **Fig. 18** (man beachte die abweichende Skalierung!).

Es fällt auf, dass die Werte ein antikorrelatives Bild ergeben, und zwar über alle sechs experimentellen Gruppen hinweg: Mit abnehmendem *Relapse* nimmt der *Progress* zu. Es zeigen sich die bereits bekannten Muster: Die Kontrollgruppen bei Kugeln und Würfeln und die Priming-Gruppe *vision* der Kugeln zeigen ein gleiches Verhalten: eine starke Rückfallstendenz in der Leistungskurve (*Relapse*) bei einem flachen globalen Anstieg (*Progress*). Die Priming-Gruppen *motor* für Kugeln sowie die Gruppen *vision* und *motor* für Würfel zeigen ein anderes Verhalten: eine geringere Rückfallstendenz im Lernfortschritt (*Relapse*) und einen steilen Anstieg (*Pro-*

gress). Der Wechsel von Kugelobjekten zu Würfelobjekten bewirkte bei den Priming-Gruppen *vision* und *motor* eine Verringerung der Rückfälle und einen Anstieg der Lernkurven (Verschiebung nach oben links), nicht aber bei der Kontrollgruppe.

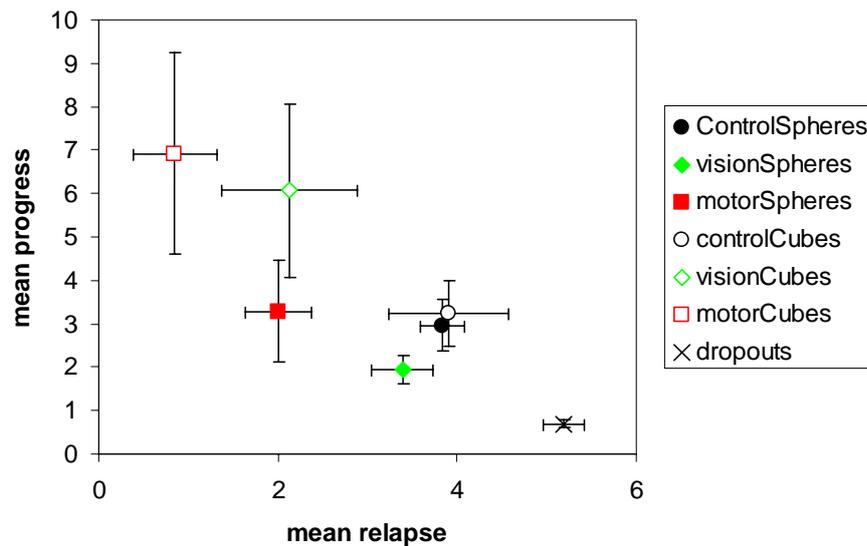


Fig. 18: Gruppenmittelwerte des Verhältnisses von mittlerer Steigung der Lernkurve (Progress) und mittlerem Rückschritt pro Lerneinheit (Relapse), wie sie für jeden einzelnen Probanden in Fig. 17 dargestellt ist. Der Wechsel von Kugelobjekten zu Würfelobjekten erbrachte bei den Priming Gruppen eine Verschiebung nach oben links (Abnahme der Rückschritte und Zunahme der Steigung), nicht aber bei der Kontrollgruppe (Fehlerbalken: ± 1 S.E.).

Diese Ergebnisse lassen sich gut zu früheren Ansätzen in Bezug setzen, visuelles Lernen als Testen von Arbeitshypothesen zu interpretieren (Unzicker, 2001; Jüttner et al., 2004). Jeder Rückfall zieht demzufolge einen Wechsel der Arbeitshypothese über die Objektstruktur nach sich. Die monotonen Teile der Lernkurve hingegen entsprechen einem fortgesetzten "Elaborieren" der Arbeitshypothese.

Vor diesem Hintergrund lassen sich die Ergebnisse dahingehend interpretieren, dass im Fall des vorausgegangenen visuellen Primings und durch den Übergang von Kugelobjekten zu den Würfelobjekten jeweils der "Hypothesenraum" der Probanden bezüglich der Beschaffenheit der Objekte so weit eingeschränkt wird, dass sie schneller und mit weniger Rückfällen (Relapse) lernen.

7.3.4 Zusammenfassung des Vergleiches zwischen Experiment I und Experiment II

Der Vergleich der korrespondierenden experimentellen Gruppen zwischen Experiment I und Experiment II erbrachte folgende Ergebnisse:

Kontrollgruppen. Die Kontrollgruppen verhielten sich in der 1. Lerneinheit bei allen Objekten gleich. Auch bei der mittleren Anzahl notwendiger Lerneinheiten bis zum Kriterium trat kein Unterschied auf. Allerdings war bei den Würfeln die Generalisierungsfähigkeit im Vergleich zu den Kugeln wesentlich besser, was mit einer signifikanten Verkürzung der Antwortzeiten für O1, nicht aber für die chiralen Objekte O2 und O3, einherging.

Vision Priming. Die Priming-Gruppe *vision* zeigte in der 1. Lerneinheit bei den Würfeln im Vergleich zu den Kugeln eine wesentlich bessere Klassifikationsleistung für beide Objekttypen. Auch die mittlere Anzahl notwendiger Lerneinheiten bis zum Kriterium war signifikant reduziert. Die Antwortzeiten blieben aber mit denen bei den Kugeln vergleichbar. In der Generalisierung zeigte sich das gleiche Bild: eine starke Verbesserung für beide Objekttypen bei gleich bleibenden Antwortzeiten.

Motor Priming. Die Priming-Gruppe *motor* bei den Würfeln war im Vergleich zu den Kugeln für das achirale Objekt O1 wesentlich besser. Dies ging auch mit kürzeren Antwortzeiten für O1, nicht aber für die chiralen Objekte O2 und O3 einher. Die mittlere Anzahl der notwendigen LE war signifikant geringer. In der Generalisierung zeigte sich eine bessere Leistung für beide Objekttypen, die ebenfalls mit deutlichen kürzeren Antwortzeiten einherging.

Zusammengefasst hatte der Wechsel von Kugelobjekten auf Würfelobjekte folgende Hauptwirkungen:

- Die Kontrollgruppe gelangte mit Hilfe des überwachten Lernens zu einer guten Generalisierungsfähigkeit.
- Die Gruppe *vision* zeigte nun von Anfang an eine ebensolche Klassifikationsleistung wie die Gruppe *motor*.
- In der Gruppe *motor* verschwanden die starken globalen Verzögerungen der Antwortzeiten für alle Objekte.

8 Diskussion

8.1 Experiment I

Zusammengefasst ergab Experiment I: Nur das motorische Priming ermöglichte während des überwachten Kategorienlernens den Erwerb eines gleichermaßen guten Objektverständnisses für alle drei Kugelobjekte. Dieses in der Generalisierung und der Rekonstruktion zielführende Objektwissen zeigte sich bereits zu Anfang der Lernprozedur und ging für alle Objekte mit deutlich erhöhten Antwortzeiten einher. Außerdem ermöglichte es sehr hohe d' -Werte für die gelernten Ansichten aller Objekte (learned views). Das visuelle Priming dagegen vermochte diese Qualität von Objektwissen nur für das achirale Objekt zu vermitteln, nicht aber für die zwei chiralen Objekte. Diese Ergebnisse bestätigten damit den in früheren Arbeiten der Arbeitsgruppe (Osman et al., 2000; Rentschler et al., 2004) gefundenen Vorteil von haptischem Priming (Objekt*manipulation*, vgl. Kap. 2.1.4) gegenüber visuellem Priming bei der Klassifikation dieses Kugelobjektsatzes.

Die Kontrollgruppe und die Priming-Gruppe *vision* zeigten eine hohe Klassifikationsleistung für das achirale Objekt O1 bei schlechten Leistungen für das Paar der chiralen Objekte. Diese Leistungsverteilung ist konsistent mit der Annahme, dass die Objekterkennung von Repräsentationen vermittelt wird, die gegenüber Spiegelungen invariant sind (reflexion invariant), wie es verschiedentlich vorgeschlagen wurde (Poggio & Edelman, 1990; Riesenhuber & Poggio, 1999). Auch erwiesen sich die Objektrepräsentationen in allen Gruppen, in Übereinstimmung mit früheren Vorschlägen (Tarr & Pinker, 1989; Poggio & Edelman, 1990; Tarr, 1995; Riesenhuber & Poggio, 1999; Gauthier et al., 2002), als ansichtenabhängig (view-dependent): Sie führten in der Generalisierung zu durchweg höheren Werten für die gelernten Ansichten (learned views) als für die nicht-gelernten Ansichten (novel views; vgl. Kap. 3.3.3).

Die erhöhten Antwortzeiten für die chiralen Objekte O2 und O3 sprechen zunächst für die Annahme, dass die Händigkeit von Objekten durch mentales Drehen der Objekte entschieden werde („rotation-for-handedness“-Hypothese).

Die „Rotation-for-handedness“-Hypothese (Tarr & Pinker, 1989; Tarr, 1995; Gauthier et al., 2002) liefert jedoch aus dem folgenden Grund keine Erklärung für die hier dargestellten Ergebnisse. Die Hypothese der mentalen Rotation gründet sich auf der Beobachtung einer linearen Abhängigkeit zwischen Drehwinkel und Antwortzeiten beim Vergleich händiger Objekte anhand von Bildern. Wie in Kap. 6.2 beschrieben, unterschieden sich die von uns benutzten Objektansichten aber nicht nur, wie in

psychologischen Studien seit Shepard & Metzler (1971) üblich, in *einem* Drehwinkel um eine vorgegebene Achse im Raum. Sie wurden vielmehr durch die Variation von *drei* Winkeln erzeugt. Die drei Winkel und die zugehörigen Drehachsen waren den Probanden während des Experimentes jedoch nicht bekannt, und die Prozedur des überwachten Kategorienlernens bot auch keinerlei Information darüber. Aus drei Freiheitsgraden kann kein einfacher skalarer Wert wie die proportional erhöhte Antwortzeit mit Eindeutigkeit abgeleitet werden. Die Probanden können deshalb in dem hier dargestellten Experiment in keiner systematischen Weise mentale Rotation zur Klassifikation der Testansichten benutzt haben. Vor diesem Hintergrund rufen die erhöhten Antwortzeiten für die chiralen Objekte O2 und O3 einen Hinweis von Shepard & Metzler (1971) in Erinnerung. Diese Autoren stellten fest, dass die gemessenen Reaktionszeiten nicht nur durch die mentale Drehung bestimmt sind, sondern notwendigerweise auch die gesamte Zeit beinhalten, die der Proband benötigt, um zu entscheiden, wie das Bild zu verarbeiten sei (S. 703). Darunter fällt folglich jede Art der Bildanalyse und Bildtransformation. Das bedeutet, dass die Reaktionszeiten vom Grad der Vertrautheit des Beobachters mit dem Objekt abhängen. Dass mentale Rotation keine alleinige Erklärung für das Verhalten der Probanden sein kann, geht auch aus folgender Tatsache hervor: Zu Beginn des Kategorienlernens haben die Probanden der Kontrollgruppe überhaupt kein Objektwissen. Die Anwendung mentaler Rotation setzt jedoch die Existenz einer – mindestens vorläufigen – Repräsentation voraus.

Die in der Priming-Gruppe *motor* für achirale und chirale Objekte insgesamt erhöhten Antwortzeiten widersprechen dieser Interpretation nicht. Aber sie führen zur Frage, was unter dem motorischen Priming so viel mehr Zeit kostete und warum das motorische Priming die Klassifikation und Konstruktion besonders der chiralen Objekte so stark verbesserte. Das visuelle Priming hatte diese Wirkung nicht und unterschied sich damit nicht von der Kontrollbedingung.

Aus der Beobachtung heraus, dass die Probanden mit motorischem Priming häufig ihre Hände bewegten, um ihre visuelle Klassifikationstätigkeit während der Aufgabe zu unterstützen, könnte man argumentieren, dass die Handposition ihnen einen referenziellen Rahmen lieferte, um die Händigkeit des Objektes zu bestimmen. Doch auch dies würde die Frage nicht beantworten, warum eine solche Strategie nicht ebenfalls nach dem visuellen Priming benutzt werden konnte.

An dieser Stelle helfen die Ergebnisse aus Experiment II weiter.

8.2 Experiment II

Das visuelle Priming konnte im Fall der Ansichten der Kugelobjekte von Experiment I keine bessere rechts-links Klassifikation ermöglichen als die Kontrollbedingung ohne Priming. Das war anders bei den Würfelobjekten von Experiment II, wo das visuelle Priming ebenso wie das motorische Priming zu einem guten und für alle

drei Objekte äquivalenten Verständnis führte. Außerdem verschwanden in Experiment II die starken Verzögerungen der Antwortzeiten (RT) in der Priming-Gruppe *motor* und auch die Kontrollgruppe gelangte durch das überwachte Lernen zu einer guten Generalisierungsfähigkeit. All diese Effekte müssen auf den Einfluss des Unterschiedes zwischen den Kugel- und den Würfel-Stimuli zurückgeführt werden: das Vorhandensein klassifikationsrelevanter Merkmale in den 2D-Abbildungen der Würfelobjekte.

Die Verteilung der RT in Experiment II gibt weitere Hinweise: In der Generalisierung war die Klassifikationsleistung der Kontrollgruppe ebenso gut wie die der beiden Priming-Gruppen, dabei trat die gleiche RT-Verzögerung für die Objekte O2 und O3 auf wie dort. In der 1. Lerneinheit gab es bei der Kontrollgruppe jedoch keine RT-Verzögerung, die Klassifikationsleistung aber war schlecht. Es lässt sich daraus schließen, dass sich diese kurzen RT in der 1. Lerneinheit der Kontrollgruppe von den kurzen RT für Objekt O1 in den Priming-Gruppen insofern unterscheiden, als sie ein Zeichen der Orientierungslosigkeit sind. Bei der kurzen experimentellen Präsentationszeit konnte ohne Objektvorwissen nur wenig Information aus den einzelnen Ansichten extrahiert werden, so dass die Probanden hofften, mit der durch den Antwort-Tastendruck aufgerufenen nächsten Objektansicht mehr relevante Bildinformation zu bekommen.

Tatsächlich legt das Verschwinden des Unterschiedes zwischen dem visuellen und dem motorischen Priming bei den Würfelobjekten nahe, dass das Finden des Referenzrahmens für die Unterscheidung der chiralen Objekte – wie es aufgrund der unterstützenden Handbewegungen bei den Probanden der Priming-Gruppen *motor* hätte angenommen werden können – gar nicht der kritische Punkt der Kategorisierungsaufgabe war. Stattdessen scheint es – bestätigt durch Rückmeldung der Probanden des *visuellen* Primings in Experiment I – dass die Abwesenheit von Textur und deutlicher Verbindungen zwischen den Objektteilen bei den Kugelobjekten die Probanden daran hinderte eine klare 3D-Struktur aus den 2D-Ansichten abzuleiten, denn Unbestimmtheiten der Verbindungen zwischen den Objektteilen verursachten eine störende relative Scheinbewegung innerhalb des tiefenrotierenden Objektes. Diese Interpretation ist konsistent mit der Feststellung von Liu und Mitarbeitern, dass Objekte, die nur aus Kugeln zusammengesetzt sind, wenig Zusammenhalt zu haben scheinen und keine ordnende Hilfe beim Aufbau von Korrespondenzen zwischen verschiedenen Ansichten geben (Liu et al., 1995).

Obwohl die Objekte während des motorischen Primings für den Probanden nicht sichtbar waren, ließ die motorische Manipulation in beiden Experimenten Objektrepräsentationen entstehen, die, auch im Fall der nicht-eindeutigen Bildreize (2D-Darstellungen der Kugel-Stimuli), eine erfolgreiche visuelle Klassifikation von Objektansichten ermöglichte, sogar von chiralen Objekten. Diese Repräsentationen müssen ausschließlich über den haptisch/motorischen Weg aufgebaut werden und können keine *visuellen* Merkmale enthalten, sondern nur gemeinsam verwertbare, supramodale Merkmale: Anzahl der Kugeln, gleiche Größe der Kugeln, gleiche Abstände, die Anordnung in rechten Winkeln, die Orientierung der Chiralität im

Raum, bezogen auf die asymmetrische Ausprägung der rechten und linken betastenden Hand.

Das motorische Priming verbesserte nicht nur die Klassifikation von neuen Ansichten in der Generalisierung. Es erleichterte in beiden Experimenten auch den Erwerb von Objektwissen aus den 2D-Ansichten des Lernsatzes. Dies zeigte sich in einer für alle Objekte höheren Klassifikationsleistung in der 1. Lerneinheit und einer größeren Effektivität des Lernens. Es gab weniger Einbrüche in der Lernkurve (*Relapse*, vgl. Kap. 7.3.3), so dass deutlich weniger Lerneinheiten bis zum Erreichen des Kriteriums notwendig waren. Die Kenntnis der haptisch-motorisch erfassten relationalen Merkmale muss also essentiell für das Lösen der Kategorisierungsaufgabe gewesen sein. Diese Ergebnisse sind konsistent mit den Ergebnissen von Zangaladze et al. (1999), die mittels transkranieller Magnetstimulation (TMS) und evozierten Potenzialen (EP) nachwiesen, dass der visuelle Kortex an der haptischen Objekterkennung beteiligt ist.

In einer fMRI-Studie, die als Erweiterung der hier berichteten Experimente anschließend durchgeführt wurde (vgl. Kap. 8.4 und Anhang 11.6), fanden wir nach haptischem Priming unter anderem eine erhöhte Aktivierung des visuellen Kortex als Zeichen eines erfolgreicherer visuellen Verarbeitens bei der Klassifikation von Objektansichten (Duhoux, Gschwind, Vuilleumier, Rentschler & Schwartz, 2005a,b). Die haptisch erworbenen und visuell nutzbaren relationalen Objektmerkmale könnten so eine Eingrenzung der Möglichkeiten bewirkt haben, aus der 2D-Ansicht des Objektes ein hypothetisches 3D-Modell abzuleiten. Dadurch waren die Probanden in der Lage, das überwachte Kategorienlernen zielstrebig und mit weniger Rückfällen zu absolvieren. Eine solche Interpretation ist konsistent mit früheren Ansätzen, den Vorgang visuellen Lernens als Testen von Arbeitshypothesen (hier in Form der hypothetischen 3D-Modelle) zu interpretieren (Unzicker, 2001; Jüttner et al., 2004).

8.3 Schlussfolgerung

Die wichtigsten Erkenntnisse aus den hier dargestellten Experimenten sind die folgenden: (1) Das Erkennen der 3D-Struktur aus einer 2D-Abbildung eines Objektes ist wesentlich abhängig von der Verfügbarkeit klassifikationsrelevanter Bildmerkmale (Experiment II). (2) Durch das haptisch-motorische Priming erworbenes, supramodal verfügbares Kontextwissen kann Mehrdeutigkeiten der bildlichen Darstellung eingrenzen und die visuelle Klassifikation verbessern (Experiment I).

Im Rahmen des hier angewandten Klassifikationsparadigmas, in dem unbekannte Objekte nach unbekanntem Klassenmerkmalen kategorisiert werden mussten, zeigte sich der Aufbau eines mentalen, zunächst hypothetischen 3D-Objektmodells aus dem 2D-Bildreiz als die entscheidende Voraussetzung für die Klassifikation händiger Strukturen. Wenn das Objekt aus bekannten Komponenten mit klarer Struktur besteht (Beispiel Würfel), ist die Erzeugung eines hypothetischen Modells ein-

fach, weil allgemeines Vorwissen über geometrische Körper zur Klärung genutzt werden kann. Wenn aus einer 2D-Ansicht die 3D-Struktur nicht eindeutig abzuleiten ist, können die Probanden das Objektwissen aus dem Kontext der Klassifikationsaufgabe heranziehen. Sind dies Merkmale wie die Größe und Anzahl der Objektteile, Raumwinkel zwischen Paaren von Objektteilen etc., die durch das haptisch-motorische Priming bekannt sind, so ist das besonders hilfreich. Steht weder Vorwissen noch externe Information zur Verfügung, so muss in einer zweiten Stufe das hypothetische 3D-Modell möglichen Transformationen unterzogen werden, bis deren Abbilder mit dem 2D-Eingangsbild abgeglichen werden können (transformation and matching-to-fit-cycle). Nach dem erfolgreichen Abgleich besteht eine Strukturrepräsentation. Solange keine neuen, damit inkompatiblen Objektansichten auftreten, die eine Änderung der aktuellen Strukturhypothese erfordern, bleibt diese gültig.

Im Falle des überwachten Lernens muss diese 3D-Repräsentation an weiteren Objektansichten getestet werden. Dabei tritt bei den unverbundenen Ansichten des hier verwendeten Paradigmas das Problem auf, dass schwer erkennbar ist, welcher Objektteil einer Ansicht welchem Objektteil der anderen Ansicht entspricht (Korrespondenzproblem). Im Laufe des überwachten Lernens kommt es so öfter zu Hypothesenänderungen, die nach Rückschlägen in der Lernkurve vorgenommen werden. Wenn die zweite Phase mit Transformation und Abgleich über längere Zeit nicht erfolgreich verläuft, kann vermutet werden, dass es sich um ein Paar chiraler Objekte handeln könnte.

Ob die Probanden die 3D-Struktur der Objekte in ihre internen Objektrepräsentationen integrieren konnten, zeigt sich in Form einer erfolgreichen Generalisierung. Auch bei den hohen Generalisierungsleistungen traten deutliche RT-Verzögerungen für die händigen Objekte auf. Diese Tatsache und die Berichte der Probanden im Feedback lassen auf die Anwendung mentaler Rotation schließen.

Nach den hier berichteten Ergebnissen könnte sich der Objekterkennungsprozess in verschiedenen Stufen vollziehen. Die erste Stufe besteht in der Erzeugung eines hypothetischen 3D-Objekt-Modells aus dem 2D-Bildreiz (Lösung des „inversen Problems“ (Pizlo, 2001)). Dies ist ein Prozess des Bildverstehens. Er ist wesentlich abhängig von der Verfügbarkeit geeigneter Bildmerkmale. In Abwesenheit von mehrdeutigen Lösungen reicht er für eine Klassifikation des Objektes aus. Ist zum Beispiel die 3D-Struktur sehr einfach oder wird ihre Kenntnis zur Klassifikation des Objektes nicht benötigt, weil ein diagnostisches Objektmerkmal vorhanden ist, so besteht keine Unklarheit. Hier ist der Platz der Theorie der „Recognition-by-components“ (Biederman, 1987), die fordert, dass explizite 3D-Modelle direkt aus den Konturen der 2D-Abbildung abgeleitet werden können. Reicht eine einfache Ableitung eines 3D-Modells jedoch nicht aus, um eine Klassifikation zu ermöglichen, weil das zu erkennende Objekt strukturell mehrdeutig bleibt oder kein diagnostisches Merkmal besitzt, schließt sich eine zweite Phase an.

Die zweite Phase besteht aus dem internen Transformieren des hypothetischen 3D-Modells und dem „Abgleichen“ der daraus hervorgehenden Projektionen mit dem Eingangsbild (Prinzip des „*forward modelling*“, „*matching-to-fit cycle*“).

Ihre Durchführung gelingt dem Probanden dann besser, wenn das hypothetische Modell detailliert ist und mental eine klare Form angenommen hat. Selbstverständlich kann der Proband hier eine Kombination aus Translation, Reflexion und Rotation verwenden – obwohl der Experimentator beim Erstellen der Testansichten vielleicht nur eine einzige Rotation verwendet hat. Diese Transformationen bis zum Abgleich werden vielleicht sogar parallel ausgeführt. Das so bestätigte hypothetische 3D-Modell bleibt als Objektrepräsentation gültig, bis eine Zusatzinformation seine Modifikation erfordert.

Ein besonderes Problem stellt die Klassifikation chiraler Objekte dar. Abbildungen von chiralen Objekten haben per Definition identische unäre und relationale Merkmale (so lange die letzteren nicht räumlich gerichtet sind). Die strukturelle Beschreibung solcher Objekte erfordert deshalb eine Orientierung an einem externen räumlichen Bezugssystem (vgl. Kap. 6.2.2). Die Existenz chiraler Objektpaare muss deutlich zutage treten (zum Beispiel durch gleichzeitiges Auftreten beider Versionen) oder vom Probanden explizit erwartet werden (zum Beispiel aufgrund der Information vor dem Experiment (Tarr & Pinker, 1989) oder provoziert durch fortgesetzte Erfolglosigkeit in der Transformationsphase). Erst dann können in einer dritten Phase rechte und linke Versionen voneinander unterschieden werden. In den Experimenten zur „mentalen Rotation“ von Tarr (1995) oder Gauthier et al. (2002) wissen die Probanden schon, dass sie rechte und linke Versionen von Objekten sehen werden, die sie davor kennen gelernt hatten (vgl. Kap. 3.3.3). Die Erzeugung eines hypothetischen 3D-Modells stellt dann keine Schwierigkeit dar, weil die mentalen Objektrepräsentationen bereits vorhanden sind. Die Betonung in der Erkennungsaufgabe wird folglich auf der dritten Phase liegen: Die Probanden versuchen, die hypothetischen Objektmodelle mental zu transformieren, bis ihre Projektion mit dem wahrgenommenen Eingangsbild übereinstimmt. Dies läuft im Falle händiger Objekte auf eine Drehung um eine geeignete Achse im Raum hinaus.

Wenn die Erzeugung des 3D-Modells sich jedoch schwierig gestaltet, wie in Experiment I mit den Kugelobjekten (hohe Komplexität), gibt es nur eine Möglichkeit, wie der Chiralitätsdefinierende Raum bestimmt werden konnte: Er wird in der Kodierung durch das motorische Gedächtnis (Fuster, 2003) als Folge der Fingerbewegungen entlang der Objekte aufgebaut (motorisches Priming). Das visuelle Priming verschaffte diese Möglichkeit nicht. Diese Interpretation ist konsistent mit den Ergebnissen von Binkofski et al. (1999), die zeigten, dass beim Betasten eines komplexen Objekts mit der Hand, die Aktivierung eines fronto-parietalen Kreislaufs auftrat (vgl. Kap. 2.1.4). Dies legt die Annahme nahe, dass eine fronto-parietale Interaktion im Gehirn die zeitliche Integration der kontinuierlichen Fingerkonfigurationen in die Repräsentation der geometrischen Makrostruktur und der Chiralität des komplexen Objektes leistet. Die unter den Kugelobjekten beobachteten allgemeinen Verzögerungen in den Antwortzeiten bei der Priming-Gruppe *motor* würden dann bedeuten, dass solche Repräsentationen nicht „fest-verdrahtete“ 3D-Modelle zur Erkennung per se liefern, sondern dass diese in Echtzeit an die ansichtenabhängige visuellen Repräsentationen gebunden werden. Unter den Würfelobjekten wurden solche global verzögerten Antwortzeiten nicht beobachtet. Dies ließe sich durch die

niedrigere Komplexität der Objekte erklären, d. h. das Vorhandensein von Bildmerkmalen zur Lösung des Korrespondenzproblems (alignment cues). Die so verminderte Mehrdeutigkeit vereinfachte die Generierung des hypothetischen Objektmodells und machte einen Einbezug des motorischen Gedächtnisses überflüssig.

Trotz der sehr hohen Generalisationsleistung bei den Würfelobjekten in Experiment II wurde immer noch eine deutliche Verzögerung der Antwortzeiten für die chiralen Objekte gemessen. Dies bedeutet, dass trotz einer strukturbasierten Objektrepräsentation keine *spontane* Klassifikation der händigen Objekte möglich war. Bei jeder zu klassifizierenden 2D-Ansicht war deshalb eine Bezugnahme auf ein *nicht*-objektzentriertes Bezugssystem notwendig. Dies bedeutet wiederum, dass Chiralität ein Strukturmerkmal darstellt, das von anderen strukturellen Merkmalen deutlich zu unterscheiden ist. Ob sich diese Eigenschaft ausschließlich auf die Chiralität beschränkt oder ob es noch andere vergleichbare Objektmerkmale geben könnte, lässt sich hier nicht endgültig entscheiden.

Zusammengefasst deuten die Ergebnisse auf einen mehrstufigen Vorgang der Objekterkennung hin. Die verschiedenen unvereinbar scheinenden Theorien der Objekterkennung wie „Recognition-by-Components“ (Biederman, 1987), „view-dependent“-Theorien (Bülhoff & Edelman, 1992; Tarr, 1995) und die „Rotation-forhandedness-Hypothese“ (Tarr & Pinker, 1989; Gauthier et al., 2002) könnten darin integriert werden. Sie beschreiben so jeweils nur einen Teilbereich des gesamten Erkennungsvorgangs. Die Auftrennung in verschiedene unabhängige Erkennungsprozesse wie „perzeptive Erstzuordnung“, „Objekterkennung“ oder „mentale Rotation“, wie sie diese Theorien jeweils beschreiben, macht vor dem Hintergrund der hier berichteten Ergebnisse wenig Sinn, da zu Beginn eines Erkennungsvorgangs normalerweise nicht bekannt ist, welcher Prozess verwendet werden soll. Die Integration dieser Teilbereiche in einen mehrstufigen Prozess hingegen ist konsistent mit den Konzepten der maschinellen Bilderkennung (Ballard & Brown, 1982). Auch dort wird ein solcher mehrphasiger Ablauf des Erkennungsvorgangs mit Erfolg angewendet. Außerdem ist sie konsistent mit neueren Ergebnissen bildgebender Untersuchungen (Vuilleumier et al., 2002), die ebenfalls ein Nebeneinander von ansichtenabhängigen und -unabhängigen Objektrepräsentationen nachweisen konnten.

Dieser mehrstufige Vorgang der Objekterkennung wird bei jeder Erkennungsaufgabe jedoch nur soweit durchgeführt, als es die zweifelsfreie Einteilung des Objektes in die entsprechende Objektklasse erfordert. Wann dieses Ziel erreicht ist, hängt vom Kontext der jeweiligen Erkennungsaufgabe ab.

8.4 Kritik und Ausblick

Obwohl mit der Auswahl der Probanden nach der Regel von „matched triples“ eine homogene Verteilung der Probanden auf die experimentellen Gruppen erreicht wurde, gab es doch große Leistungsunterschiede innerhalb der Gruppen. Diese Unter-

schiede führten in den Kontrollgruppen sogar zu einer tendenziell zweigipfligen Leistungsverteilung. Hier wäre einerseits an eine Erhöhung der Probandenanzahl zu denken. Andererseits besteht der Verdacht, dass hier Vorkenntnisse der Probanden doch eine wesentliche Rolle spielten: eine allgemeine Vertrautheit mit geometrischen Strukturen. Zur weiteren Untersuchung wäre es deshalb sinnvoll, vor dem Experiment einen geeigneten Test über geometrisches und räumliches Denken durchzuführen. Die Ergebnisse daraus könnten dann mit den Ergebnissen aus dem Experiment in Beziehung gesetzt werden.

Viele Probanden hatten während der Durchführung des Experimentes eine Art „Aha-Erlebnis“, wobei sie die spiegelsymmetrische Struktur der Objekte O2 und O3 plötzlich erkannten. Danach wurde das Kriterium des überwachten Lernens meistens schnell erreicht. Die Lerndauer bis zum Auftreten dieses „Aha-Erlebnisses“ variierte jedoch stark. Um diesen Punkt weiter zu untersuchen, wurde ein Nebenexperiment zur Kontrollgruppe durchgeführt. Die Probanden erhielten gleich zu Anfang die Information, dass sich die Aufgabe lösen lässt, wenn das Prinzip der Chiralität berücksichtigt wird (vgl. Anhang 11.3.1). Die Probanden berichteten aber, dass sie lange Zeit nichts mit dieser Information anfangen konnten, weil die präsentierten Bilder keinen Anhaltspunkt für ihre Anwendung boten. Sie sahen nur einen unstrukturierten Kugelhaufen (vgl. Anhang 11.3.2). Diese Beobachtung bestätigt die Schlussfolgerung, dass der erste Schritt im Vorgang der Objekterkennung das Aufbauen einer hypothetischen 3D-Objektstruktur sein muss.

Die Auswertung der Daten auf der Ebene der einzelnen Ansichten verspricht weitere Einsicht in die Lerndynamik. Interessieren würden dabei besonders zwei Fragestellungen: Einerseits die „Anatomie der Ansichten“ – welche waren die schwierigsten Ansichten und welche die leichtesten, und wodurch könnten diese Eigenschaften jeweils charakterisiert werden? Andererseits stellt sich die Frage: Wie verhält sich der Proband bezüglich der einzelnen Ansicht über mehrere Lerneinheiten hinweg? Gibt es Ansichten, die er immer falsch beurteilt und gibt es solche, die wechselnd beurteilt werden? Kann man eventuell das oben beschriebene „Aha-Erlebnis“ mit einer bestimmten Abfolge von Ansichten in Bezug setzen?

fMRI-Studie. Von großem Interesse ist das Korrelieren des hier dargestellten verhaltensexperimentellen Ansatzes mit neurophysiologischen Vorgängen oder einem bildgebenden Verfahren. Dazu wurde bereits in Zusammenarbeit mit der neurologischen Universitätsklinik Genf⁸ ein fruchtbarer Anfang unter der Verwendung funktioneller Kernspintomographie gemacht (Duhoux, Gschwind, Vuilleumier, Rentschler & Schwartz, 2005a,b).

In dieser fMRI-Studie wurde das in der vorliegenden Arbeit entwickelte experimentelle Paradigma mit motorischem Priming, überwachtem Lernen und Generalisieren verwendet, mit zusätzlichen Anpassungen an die Bedingungen des Scanners. So war es möglich, die intermodale Wirkung von motorischem Priming auf die

⁸ Zusammen mit S. Duhoux, Dr. S. Schwartz und Prof. Dr. P. Vuilleumier

visuelle Verarbeitung darzustellen, wobei zwischen der Verarbeitung von (1) neuen Ansichten eines neuen Objektes, (2) neuen Ansichten eines bekannten Objektes und (3) bekannten Ansichten eines bekannten Objektes unterschieden wurde.

Es zeigte sich mit einer bilateralen Aktivierung des Hippokampus und des parahippokampalen Kortex die besondere Rolle des Gedächtnisses beim Abrufen gelernter Ansichten bekannter Objekte. Eine bilateral verstärkte Aktivierung der superioren parietalen Regionen deutet darüber hinaus darauf hin, dass beim Generalisieren bereits gelernter (chiraler) Objekte ein Prozess der mentalen Transformation (mental rotation) aktiviert wird.

Es wurde aber auch die Frage untersucht, welche Verarbeitung erfolgreiches Lernen ermöglichte. Dazu wurden die Daten in Funktion der Leistung der Probanden in der Lern- und Generalisierungsphase analysiert und die Probanden in zwei Gruppen eingeteilt. Die zehn besten Probanden wurden der Gruppe von „guten Lernern“ zugeordnet, die sich ein besseres 3D Objektwissen erarbeitet hatten. Die verbleibenden zehn Probanden bildeten die Gruppe von „schlechten“ Lernern.

Der Vergleich der zerebralen Aktivierung der „guten Lerner“ mit der zerebralen Aktivierung der „schlechten Lerner“ zeigte, dass während der Lernphase der rechte Hippokampus und inferiore frontale Regionen eine erhöhte Aktivität aufwiesen, was ein Indiz für eine bessere Gedächtnisleistung bei den guten Probanden ist. Ebenso zeigte ein Vergleich der bekannten (learned) Ansichten mit den neuen (novel) eine erhöhte ansichtenspezifische Aktivität bei den guten Lernern in den frontalen Regionen, im rechten superioren temporalen Sulcus und ebenso im visuellen Cortex. Dies bedeutet, dass auch eine verstärkte frontale Kontrolle und ein intensiveres visuelles Verarbeiten bei den guten Probanden zu der besseren Leistung geführt haben.

So konnte das in dieser Arbeit dargestellte verhaltensexperimentelle Paradigma eine fruchtbare Grundlage für eine fMRI-Studie zur supramodalen Objektrepräsentation bilden. Die Ergebnisse werden an einem anderen Ort ausführlich dargestellt (vgl. auch Anhang 11.6 Poster).

9 Zusammenfassung

Die dreidimensionale Struktur eines Objektes ist aus einer einzigen Blickrichtung niemals vollständig sichtbar. Um sie zu verstehen, muss die dreidimensionale Struktur aus zweidimensionalen Projektionen heraus interpretiert werden, denn die Netzhaut liefert nur 2D-Abbildungen an die neuronale Weiterverarbeitung. Dieses Problem besteht insbesondere beim Verständnis von unbekanntem unbewegten Bildern, wenn kein Vorwissen, keine Stereoskopie und keine Bewegungsparallaxe zusätzliche räumliche Informationen liefern können.

Mit den hier berichteten zwei Verhaltensexperimenten wurden die Bedingungen für eine Repräsentation der Struktur von 3D-Objekten untersucht. Dazu wurde der Kontext des Wahrnehmungsvorgangs in die Mitte des Interesses gerückt. Ein Klassifikationsparadigma mit anfangs unbekanntem Klassenkriterien, einer Lern- und einer Generalisierungsphase wurde verwendet, das Einblicke in die Entstehung der Objektrepräsentationen erlaubte. Als Klassifikationsaufgabe wurde eine „nur-Struktur“-Kategorisierung verwendet. Die Technik des kontrollierten Erwerbs von Objektvorwissen (Priming) zeigte die Abhängigkeit der Erkennungsleistung von der Art des Objektvorwissens. Dieses blieb nicht auf das visuelle System begrenzt, sondern es wurde motorisch-haptisches Vorwissen miteinbezogen.

Die Ergebnisse zeigten, dass es den Probanden möglich war, aus unzusammenhängenden 2D-Abbildungen eine strukturbasierte Objektrepräsentation aufzubauen. Es zeigte sich eine deutliche Kontextabhängigkeit: Die Aufgabe war ohne Objektvorwissen (*control*) und für Objekte mit hoher Komplexität (Kugel-Stimuli) sehr viel schwieriger zu bewältigen. Als wesentlich stellten sich die Verfügbarkeit von klassifikationsrelevanten Bildmerkmalen (*alignment cues*) und das Erfassen der relationalen Objektmerkmale heraus. Standen solche Merkmale nur spärlich zur Verfügung, und war die Objektstruktur deshalb nicht eindeutig (hohe Komplexität), konnte der Mangel an visueller Information mit Hilfe von transmodal verfügbarem Objektwissen kompensiert werden (*motor-Priming*). Dieses Wissen scheint jeweils nach Bedarf in Echtzeit auf ansichtenabhängige (*view-dependent*) Repräsentationen übertragen zu werden.

Die Ergebnisse legen die Existenz mehrerer Phasen für den Vorgang der Objekterkennung nahe: Die erste Phase leistet das Bildverstehen und besteht aus der Erzeugung eines hypothetischen 3D-Strukturmodells des zu erkennenden Objektes. Erst wenn ein solches hypothetisches Objektmodell aufgebaut ist, kann es in der zweiten Phase verschiedenen Transformationen und Rotationen unterzogen werden, um eine Übereinstimmung mit dem Bildeingangssignal zu erreichen (*matching-to-fit-cycle*). Wenn diese zweite Phase nicht erfolgreich ist – weil beispielsweise das

Objekt chiral ist – wird das Objektmodell in einer dritten Phase zur Bestimmung seiner Chiralität (rechte oder linke Version) zusätzlich auf einen äußeren Referenzrahmen bezogen.

In diesem mehrstufigen Prozess könnten die verschiedenen, bisher unvereinbar scheinenden Theorien der Objekterkennung integriert werden. So beschreiben die „Recognition-by-Components“ (Biederman, 1987), die „view-dependent“-Theorien (Bülthoff & Edelman, 1992; Tarr, 1995) und die „Rotation-for-handedness-Hypothese“ (Tarr & Pinker, 1989; Gauthier et al., 2002) vor dem Hintergrund der Ergebnisse jeweils nur einen Teilbereich des Erkennungsvorgangs. Dieser mehrstufige Vorgang wird bei jeder Erkennungsaufgabe jedoch nur soweit durchgeführt, wie es notwendig ist, um eine zweifelsfreie Klassifikation des Objektes zu ermöglichen. Dies wiederum hängt vom Kontext der jeweiligen Erkennungsaufgabe ab.

Hier konnte damit gezeigt werden, dass experimentelle Paradigmen, die den Kontext des Erkennungsvorgangs – Vorwissen, Aufbau der Repräsentation und Beschaffenheit der Bildinformation – miteinschließen, es erlauben, die verschiedenen bestehenden Theorien der Objekterkennung in einen größeren Zusammenhang zu integrieren.

10 Literaturverzeichnis

- Ahmed, N. & Rao, K. (1975). *Orthogonal Transforms for Digital Signal Processing*. Berlin, New York, Springer.
- Albright, T. D. (1995). 'My most true mind thus makes mine eye untrue'. *Trends in Neurosciences* **18**(8): 331-3.
- Amedi, A., Malach, R., Hendler, T., Peled, S. & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience* **4**(3): 324-30.
- Amedi, A., Jacobson, G., Hendler, T., Malach, R. & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex* **12**(11): 1202-12.
- Ballard, D. & Brown, C. M. (1982). *Computer Vision*. Prentice Hall, Englewood Cliffs, NJ.
- Bar, M. & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron* **38**(2): 347-58.
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience* **5**(8): 617-29.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition* **11**(3): 211-27.
- Biederman, I. (1987). Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review* **94**: 115-47.
- Biederman, I. & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *J Exp Psychol Hum Percept Perform* **19**(6): 1162-82.
- Biederman, I. & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research* **39**: 2885-99.
- Biederman, I. (2000). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision* **13**: 241-53.
- Binkofski, F., Buccino, G., Posse, S., Seitz, R. J., Rizzolatti, G. & Freund, H. (1999). A fronto-parietal circuit for object manipulation in man: evidence from an fMRI-study. *European Journal of Neuroscience* **11**(9): 3276-86.
- Blanz, V., Tarr, M. J. & Bülthoff, H. H. (1999). What objects attributes determine canonical views. *Perception* **28**(5): 575-99.

- Brown, R. (1958). How should a thing be called? *Psychological Review* **65**: 14-21.
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews* **36**(2-3): 96-107.
- Bülthoff, H. H. & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Science* **89**: 60-64.
- Bülthoff, H. H., Edelman, S. & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex* **5**(3): 247-60.
- Caelli, T. & Julesz, B. (1978). On perceptual analyzers underlying visual texture discrimination: Part I. *Biological Cybernetics* **28**(3): 167-75.
- Caelli, T., Rentschler, I. & Scheidler, W. (1987). Visual pattern recognition in humans. I. Evidence for adaptive filtering. *Biological Cybernetics* **57**(4-5): 233-40.
- Caelli, T., Johnston, M. & Robison, T. (1993). 3D Object Recognition: Inspirations and Lessons from Biological Vision. In: Jain, A. K., Flynn, P. J., *Three-dimensional Object Recognition Systems*. New York, Elsevier.
- Christou, C. G., Tjan, B. S. & Bülthoff, H. H. (2003). Extrinsic cues aid shape recognition from novel viewpoints. *Journal of Vision* **3**: 183-98.
- Desimone, R. & Ungerleider, L. G. (1989). Neural mechanisms of visual perception in monkeys. In: Boller, F., Grafman, J., *Handbook of Neuropsychology*. Amsterdam, Elsevier. **Vol. 2**: 267-99.
- Dickinson, S. J. (1993). Part-based Modeling and Qualitative Recognition. In: Jain, A. K., Flynn, P. J., *Three-dimensional Object Recognition Systems*. New York, Elsevier.
- Duhoux, S., Gschwind, M., Vuilleumier, P., Rentschler, I. & Schwartz, S. (2005a). *New fMRI views on 3D object recognition (Abstract)*. Cognitive Neuroscience Society Annual Meeting, New York.
- Duhoux, S., Gschwind, M., Vuilleumier, P., Rentschler, I. & Schwartz, S. (2005b). Neural correlates of 3D object learning. *Perception* **34**(suppl.): 116.
- Edelman, G. (1987). *Neural Darwinism: the theory of neuronal group selection*. New York, Basic Books.
- Felleman, D. J. & Van Essen, D. C. (1991). Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex* **1**: 1-47.
- Field, A. (2000). *Discovering Statistics*. London, Sage Publications.
- Franz, V. H. (2001). Action does not resist visual illusions. *Trends in Cognitive Science* **5**(11): 457-9.
- Fujita, I., Tanaka, K., Ito, M. & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature* **360**(6402): 343-6.

- Fuster, J. (2003). *Cortex and Mind*. Oxford, Oxford University Press.
- Fuster, J. M. (2001). The prefrontal cortex--an update: time is of the essence. *Neuron* **30**(2): 319-33.
- Gauthier, I., Skudlarski, P., Gore, J. C. & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience* **3**(2): 191-7.
- Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P. & Gore, J. C. (2002). BOLD activity during mental rotation and viewpoint-dependent object recognition. *Neuron* **34**(1): 161-71.
- Goldstein, E. B. (2002). *Wahrnehmungspsychologie (1. dt. Auflage)*. Heidelberg, Berlin, Spektrum, Akademischer Verlag.
- Green, D. M. & Sweets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, Wiley.
- Grill-Spector, K., Kourtzi, Z. & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research* **41**(10-11): 1409-22.
- Gross, C. G. & Bornstein, M. H. (1978). Left and right in science and art. *Leonardo* **11**: 29-38.
- Grüsser, O.-J. & Landis, T. (1991). *Visual Agnosias and Other Disturbances of visual Perception and Cognition*. London, The Macmillan Press.
- Gschwind, M., Brettel, H., Osman, E. & Rentschler, I. (2004). Structured but viewdependent representation for visual 3-D object classification. *Perception* **33**(suppl.): 73.
- Hubel, D. (1995). *Eye, Brain and Vision*. New York, W.H. Freeman & Company.
- Jain, A. K. & Hoffman, D. (1988). Evidence-based recognition of objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**: 783-802.
- James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S. & Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia* **40**(10): 1706-14.
- Jüttner, M. (2003). *Visuelles Lernen: Erwerb und Anwendung bildkategorialen Wissens*. Lengerich, Pabst Science Publishers.
- Jüttner, M. (2003). Denken und bildliches Vorstellen. In: Rentschler, I., Madelung, E. & Fauser, P., *Bilder im Kopf*. Seelze-Felber, Kallmeyersche Verlagsbuchhandlung: 42-63.
- Jüttner, M., Langguth, B. & Rentschler, I. (2004). The impact of context on category learning and representation. *Visual Cognition* **11**: 921-45.
- Kandel, E. R., Schwartz, J. H., Jessell, Th., M. (1996). *Neurowissenschaften*. Berlin, Spektrum Akademischer Verlag.

- Lakoff, G. (1987). *Women, Fire and dangerous Things*. Chicago, The University of Chicago Press.
- Langguth, B. (1998). Bildkategoriales Lernen und Generalisieren bei gesunden und Patienten mit umschriebenen Hirnverletzungen (Dissertation). *Institut für Medizinische Psychologie*. München, Ludwig-Maximilians Universität.
- Lederman, S. J. & Klatzky, R. L. (1998). The Hand as a Perceptual System. In: Connolly, K. J., *The Psychobiology of the Hand*. London, Mac Keith Press: 16-35.
- Lee, T. S., Yang, C. F., Romero, R. D. & Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nature Neuroscience* **5**(6): 589-97.
- Li, W., Piech, V. & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience* **7**(6): 651-7.
- Liu, Z., Knill, D. C. & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research* **35**(4): 549-68.
- Liu, Z., Kersten, D. & Knill, D. C. (1999). Dissociating stimulus information from internal representation - a case study in object recognition. *Vision Research* **39**: 603-12.
- Logothetis, N. K., Pauls, J., Bülthoff, H. H. & Poggio, T. (1994). view-dependent object recognition by monkeys. *Current Biology* **4**(5): 401-14.
- Loomis, J. M. & Lederman, S. J. (1986). Tactual perception. In: Boff, K., Kaufman, L. & Thomas, J., *Handbook of Perception and Human Performance*. New York, John Wiley & Sons. **Vol. II**: Chapt. 31.
- MacMillan, N. A. & Creelman, C. D. (1991). *Detection theory: a user's guide*. New York, Cambridge University Press.
- Marr, D. & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Biological Sciences* **200**: 269-94.
- Marr, D. (1982). *Vision*. San Francisco, H. Freeman & Co.
- Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience* **1**(1): 59-65.
- Milner, A. D. & Goodale, M. A. (1995). *The visual brain in action*. Oxford, Oxford University Press.
- Mishkin, M., Ungerleider, L. G. & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences* **6**: 414-17.
- Moore, C. & Engel, S. A. (2001). Neural response to perception of volume in the lateral occipital complex. *Neuron* **29**(1): 277-86.
- Neisser, U. (1996). *Kognition und Wirklichkeit*. Stuttgart, Klett Cotta.

- Newell, F. N., Ernst, M. O., Tjan, B. S. & Bulthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science* **12**(1): 37-42.
- Osman, E., Pearce, A. R., Juttner, M. & Rentschler, I. (2000). Reconstructing mental object representations: a machine vision approach to human visual recognition. *Spatial Vision* **13**(2-3): 277-86.
- Osman, E. (in Vorbereitung). Visual 3D Object Recognition in Brains and Machines (Dissertation). *Institut für Medizinische Psychologie*. München, Ludwig-Maximilians Universität.
- Palmer, S., Rosch, E. & Chase, P. (1981). Canonical perspective and the perception of objects. In: Long, J., Baddeley, A., *Attention and Performance IX*. Hillsdale, NJ, Erlbaum: 135-51.
- Parsons, L. M. (2003). Superior parietal cortices and varieties of mental rotation. *Trends in Cognitive Sciences* **7**(12): 515-7.
- Peters, A., Payne, B. R. & Budd, J. (1994). A numerical analysis of the geniculocortical input to striate cortex in the monkey. *Cerebral Cortex* **4**(3): 215-29.
- Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research* **41**(24): 3145-61.
- Poggio, T. & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature* **343**(6255): 263-6.
- Prelog, V. (1975). Chirality in chemistry. *Nobel Lecture*. Stockholm.
- Puce, A., Allison, T., Gore, J. C. & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of Neurophysiology* **74**(3): 1192-9.
- Rentschler, I., Juttner, M. & Caelli, T. (1994). Probabilistic analysis of human supervised learning and classification. *Vision Research* **34**(5): 669-87.
- Rentschler, I. & Juttner, M. (2000). Dynamics and context dependence of visual category learning. *Spatial Vision* **13**(2-3): 231-40.
- Rentschler, I., Juttner, M., Osman, E., Muller, A. & Caelli, T. (2004). Development of configural 3D object recognition. *Behavioural Brain Research* **149**(1): 107-11.
- Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of recognition in cortex. *Nature Neuroscience* **2**: 1019-25.
- Rock, I., Di Vita, J. & Barbeito, R. (1981). The effect on form perception of change of orientation in the third dimension. *Journal of Experimental Psychology: Human Perception and Performance* **7**(4): 719-32.
- Rosch, E. & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* **7**: 573-605.

- Rosch, E., Mervis, C., Gray, W. & Johnson, D. (1976). Basic objects in natural categories. *Cognitive Psychology* **8**: 382-439.
- Scheerer, E. (1993). Mentale Repräsentation in interdisziplinärer Perspektive. *Zeitschrift für Psychologie* **201**: 136-166.
- Schiller, P. H., Logothetis, N. K. & Charles, E. R. (1990). Functions of the color-opponent and broad-band channels of the visual system. *Nature* **343**: 68-70.
- Shepard, R. N. & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science* **171**(972): 701-703.
- Shepard, R. N. & Cooper, L. A. (1982). *Mental Images and Their Transformations*. Cambridge, MA, MIT Press/Bradford Books.
- Sherman, S. M. & Koch, C. (1986). The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus. *Experimental Brain Research* **63**(1): 1-20.
- Sigala, N. (2004). Visual categorization and the inferior temporal cortex. *Behavioural Brain Research* **149**: 1-7.
- Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron* **24**: 49-65.
- Squire, L., Ed. (1992). *Encyclopedia of Learning and Memory*. New York, Macmillan Publishing Company.
- Sturm, W., Herrmann, M. & Wallesch, C. W., Eds. (2000). *Lehrbuch der Klinischen Neuropsychologie*. Lisse, Swets & Zeitlinger Pub.
- Sugase, Y., Yamane, S., Ueno, S. & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature* **400**(6747): 869-73.
- Tanaka, K., Saito, H., Fukada, Y. & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology* **66**(1): 170-189.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience* **19**: 109-139.
- Tanaka, K. (1997). Columnar organization in the inferotemporal cortex. In: Rockland et al., *Cerebral Cortex*. New York, Plenum Press. **12**.
- Tanaka, K. (1997). Mechanisms of visual object recognition: Monkey and human studies. *Current Opinion in Neurobiology* **7**(4): 523.
- Tarr, M. J. & Pinker, S. (1989). Mental Rotation and Orientation-Dependence in Shape Recognition. *Cognitive Psychology* **21**: 233-282.
- Tarr, M. J. (1995). Rotating objects to recognize them: a case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin Reviews* **2**: 55-82.

- Tarr, M. J. & Vuong, Q. C. (2002). Visual Object Recognition. In: Pashler, H. & Yantis, S., *Stevens' Handbook of Experimental Psychology: Sensation and Perception*. New York, John Wiley & Sons, Inc. **1**: 287-314.
- Unzicker, A. (2001). Psychophysik und formale Beschreibung visueller Klassifikationsleistungen (Dissertation). *Institut für Medizinische Psychologie*. München, Ludwig-Maximilians Universität.
- Vuilleumier, P., Henson, R. N., Driver, J. & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience* **5**(5): 491-9.
- Wertheimer, M. (1922). Untersuchungen zur Lehre von der Gestalt. I. Prinzipielle Bemerkungen. *Psychologische Forschung* **1**: 47-58.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung* **4**: 301-50.
- Wittgenstein, L. (2003). *Philosophische Untersuchungen*. Frankfurt, Suhrkamp.
- Young, M. P. (2000). The architecture of visual cortex and inferential processes in vision. *Spatial Vision* **13**(2-3): 137-46.
- Zangaladze, A., Epstein, C. M., Grafton, S. T. & Sathian, K. (1999). Involvement of visual cortex in tactile discrimination of orientation. *Nature* **401**(6753): 587-90.
- Zigmond, M. J., Bloom, F. E., Landis, S. C., Roberts, J. L. & Squire, L. R., Eds. (1999). *Fundamental Neuroscience*. London, Academic Press.

11 Anhang

11.1 Anweisung an den Probanden

Experiment „visuelle 3D Objekt-Erkennung“

Einleitung

Sie nehmen an einem Experiment teil, das von uns aufgebaut wurde, um die Arbeitsweise des Gehirns beim Betrachten von dreidimensionalen (3D-) Objekten zu untersuchen.

Es geht darin um die Frage, in welcher Form diese 3D-Objekte im Gehirn repräsentiert sind. Dafür gibt es in der Wissenschaft gegenwärtig hauptsächlich 2 verschiedene Theorien: Die eine geht davon aus, dass die Objekte als unveränderliche 3D-Struktur, unabhängig von den dargebotenen Ansichten, repräsentiert werden, die andere Theorie fordert, dass das Gehirn eine Sammlung 2-dimensionaler Ansichten der Objekte – von allen möglichen Blickrichtungen her – als Ausgangspunkt für seine Verarbeitung nimmt.

Ihre Aufgabe

Wir haben drei verschiedene Objekte konstruiert und zeigen Ihnen nun am Bildschirm viele verschiedene Abbildungen aus verschiedenen Blickrichtungen und in zufälliger Reihenfolge.

Ihre Aufgabe ist es, diese Abbildungen den Objekten so gut wie möglich zuzuordnen, d.h. zu erkennen, welches Objekt Sie gerade abgebildet sehen.

Das Experiment besteht aus 3 Teilen

Im **1. Teil** zeigen wir Ihnen die Objekte 1, 2 und 3, für insgesamt 5 Min. Sie können sie kennen lernen, während sie einfach in verschiedene Richtungen auf dem Bildschirm rotieren. So können Sie die Objekte einmal von allen Richtungen her sehen.

Der **2. Teil** besteht darin, dass Sie alle Ansichten den jeweiligen Objekten zuordnen. Er besteht aus einem Lerndurchgang, einem Testdurchgang und einer Ergebnistabelle.

Im *Lerndurchgang* zeigt der Bildschirm Ihnen je eine Abbildung eines der Objekte, danach folgt die Zahl, die anzeigt, welches Objekt es war. Nachdem Sie alle 22 verschiedenen Abbildungen einmal gesehen haben, folgt der Testdurchgang.

Im *Testdurchgang* müssen Sie nun selbst auf den Tasten 1, 2, 3 angeben, welches Objekt das gebotene Bild eben gezeigt hat. Die Ansichten bleiben dabei die gleichen wie im Lerndurchgang. Es gibt beim Testdurchgang keinen Zeitdruck, Sie bestimmen das Tempo, indem Sie die Antwort geben.

Am Ende des Testdurchgangs erhalten sie ihre *Ergebnisse* in Form einer kleinen Tabelle.

Lern- und Testdurchgang werden nun abwechselnd so lange wiederholt, bis Sie 90% der Bilder richtig erkannt haben.

Im **3. Teil** wird getestet, wie gut Sie auch neue Ansichten derselben Objekte zuordnen können (Generalisieren). Er besteht aus 3 weiteren Testdurchgängen. Sie sehen darin zusätzlich zu den bereits bekannten, eine große Menge neuer Abbildungen der Objekte 1, 2 und 3, aus Blickwinkeln, die Sie bisher noch nicht gesehen haben (ca. 84 insgesamt).

Sie werden insgesamt 3 Durchgänge zu je 84 Abbildungen zuordnen.

11.2 Introspektion der Probanden aus Experiment I

11.2.1 Probanden der Kontrollgruppe

Allgemeiner Eindruck. Die Probanden antworten rasch und oft auf gut Glück.

Feedback der einzelnen Probanden, soweit notiert (jeder Punkt ein Proband):

- Trotz Erklärung war anfangs die Aufgabe nicht klar. Meinte, sie müsse sich etwas merken. Stellte sich die Objekte als Figuren im Kopf vor. Darstellung war für sie irritierend, da keine Schatten.
- Zuerst Kugel in Zweiergruppen gruppiert, versucht, Verhältnis zueinander zu bestimmen. Dann Wechsel zu 3 + 1 Kugel, ist entmutigt, da kein Erfolg. Erkennt keinen Unterschied zwischen O2 und O3. Abbruch.
- In den ersten Durchgängen des überwachten Lernens versucht, die Objekte als Modelle räumlich vorzustellen. Später nur noch als 2D-Bild mit besonderen Merkmalen. Weiß nicht, wie die Objekte gebaut sind.
- Ging schon mit der Absicht ans Experiment: „21 Bilder: das kann ich memorieren!“ – Bildete für O2 und O3 viele Eselsbrücken (Würmer, Zahl 4, Knäule, etc.) und memorierte sie als Bilder
- Versuchte, die Baupläne der Kugeln zu erkennen. Hat dazu im Bildschirm Achsen durch die Kugeln gezogen. O1 als Dreistern und O2 und O3 als rechts- und linksdrehende Helix erkannt.
- Erst gar keine Ahnung gehabt. Dann O1 unterscheidbar, O2 und O3 als Kette. Versuchte Bilder zu drehen.
- Merkte ab etwa dem 5. Durchgang für jedes Objekt ein Referenzbild. Rotierte die Objekte in die entsprechende Referenzposition. Hat viel Erfahrung mit darstellender Geometrie.
- O1 am eindeutigsten, erkennt „Mercedes-Stern“, O2 und O3 versucht mit Eselsbrücken zu unterscheiden.

11.2.2 Probanden der Priming-Gruppe vision

Allgemeiner Eindruck. Gleich wie Kontrollgruppe.

Feedback der einzelnen Probanden, soweit notiert (jeder Punkt ein Proband):

- Anfangs war Objekt 1 das einfachste, dann Objekt 2. Objekt 3 überhaupt nicht erkennbar. – Stellte sich die Objekte als Modelle vor, dann als sitzende Männchen mit Arm. – Merkte sich einzelne Darstellungen: O1: „Kugeln eng“ O2: „Kugeln weit“ – visuelle Vorerfahrung war nützlich.
- Merkte sich O1 als Bild, O2 als Struktur „relational“, O3 als Ausschluss.
- 1. Anfangssystem Winkel? 2. Auswendig lernen. Rotierte das System im Kopf
- Bildete Metaphern und merkte sich die Objektnummer dazu auswendig – „keine Vorstellung von Objekten“ – keine räumliche Erfahrung – Priming: „Objekte waren nicht starr!“
- Objekte waren „unlogische Anhäufung“. Hat keine Vorstellung von Objekten, kein 3D Effekt bei den Ansichten. Objektpräsentation beim Priming: „Objekte sind nicht echt!“ – Vorgehen: Versuchte Bilder global zu beurteilen, konnte sich keine Modelle vorstellen
- Unterschied O1 vs. O2/O3: „Wurmförmig oder nicht wurmförmig“? – O2 vs. O3: Rotierte Modell innerlich auf Referenzposition.
- Teilweise Rotation, teilweise einzelne Bilder auswendig gelernt.
- Versuchte, ob das Bild mit der Vorstellung übereinstimmt (mental Rotation). – Besonders schwierig, wenn O1 von ungewohnter Seite (Dreistern nicht sichtbar!) – einzelne Ansichten auswendig gelernt. – Vorerfahrung gibt Idee, dass O1 wie Stern und O2/O3 untereinander ähnlich sind.
- Anfangs O1 und O2 gut unterscheidbar, O3 vollkommen unklar. – Stellte sich Objekte als chemische C-Ketten-Moleküle vor. – Vergleich direkte Erinnerung ans eben gesehene Bild mit der vorgestellten Struktur.
- Objekte in Vorerfahrung waren nicht fest, Kugeln rotierten umeinander! - Versuchte im ersten Lerndurchgang, von der Vorerfahrung zu übertragen, etablierte danach aber ein eigenes Eselsbrückensystem nach Merkmalen des Bildes (keine Konstruktionsvorstellung), merkte sich einzelne Bilder.
- O1 „wie Atomzeichen“ - O2/O3 anhand von Metaphern auswendig gelernt und versucht diese 2D-Referenzansichten im Raum zu rotieren. Gelingt nicht immer!
- Nach Vorerfahrung O1 sofort wiedererkannt: „geschlossenes Objekt, Pyramide“ - O2/O3 nur zwei Ansichten (wo je eine Kugel vollkommen verdeckt ist) explizit auswendig gelernt. Hat noch nach Erreichen des 90% Kriteriums überhaupt keine Ahnung wie O2 und O3 aussehen: „O2 ist ordentlich, O3 ist verschoben“

11.2.3 Probanden der Priming-Gruppe motor

Allgemeiner Eindruck. Die Probanden nehmen sich erstaunlich lange Zeit, überlegen bei Unsicherheit lieber noch einmal, bevor sie die Antwort eingeben. Viele nehmen ihre Hände zur Hilfe um sich die räumlichen Verhältnisse klar zu machen, oder sich an die Tasterfahrung zurückzuerinnern.

Feedback der einzelnen Probanden, soweit notiert (jeder Punkt ein Proband):

- O2/O3 als Kette gemerkt. Wenn Kette gesehen, Bild in Vorstellung gedreht und „Beine“ geortet, so dass sie mit der Referenzposition verglichen werden konnten. – Vorstellung der Objekte von der Tasterfahrung her. Bilder haben „eher gestört“. Hat immer auf Tasterinnerung zurückgegriffen.
- Konnte Tasterinnerung nicht mit den Ansichten in Übereinstimmung bringen. Versuchte während der ersten 20 Min die „tasterinnerten“ Objekte in den Ansichten zu suchen, danach hat er nur noch Bilder auswendig gelernt. Die Tasterinnerung ist durch die vielen Bilder der Ansichten schließlich überlagert worden und stand nicht mehr zur Verfügung.
- Erkannte beim Betasten das Bauprinzip der drei Objekte (Drei Kugeln als Basis und Variation der vierten Kugel). – Drehte die Ansicht mental in die Referenzposition. Bis zuletzt war Tasterinnerung dabei.
- Nach Tasten schon klare Vorstellung der Konstruktion. Beschreibt zunehmende Übung, die auf den Ansichten gesehenen Objekte mental zu rotieren.
- Konnte im Tastdurchgang O2/O3 noch nicht unterscheiden. – Legte durch O2 u. O3 Achsen und rotierte sie. – Ansichten, bei denen einzelne Kugeln verdeckt waren, auswendig gelernt.
- Hatte nach Tasterfahrung gute Vorstellung von den Objekten, konnte sie aber auf den präsentierten Ansichten nicht erkennen, da zu kurz präsentiert! – Dachte in Kugeln und erst zum Schluss in Verbindungsachsen!
- Vom Tasten her O1 klar, O2 und O3 ununterscheidbar. - Beschreibt bei sich selbst bekannte Schwierigkeit, Objekte im Kopf zu rotieren (ist noch von der Schule bekannt aus Chemie und Mathematik). So auch hier. Deswegen manche Ansichten auswendig gelernt.
- Versuchte im Bildschirm die Objekte so drehen, dass die Referenzposition erreicht wurde.
- Konnte nur Tasterfahrung von O1 auf die Ansichten übertragen. – Merkte sich einige Bilder erst in der Lernphase mittels Analogien.
- Konnte anfangs die Tasterfahrung gut auf die Ansichten anwenden. Stellte dann aber auf Analogien und auswendig gelernte Bilder um. – Rotierte die Objekte im Kopf.
- Wusste nach Tasterfahrung genau, wie die Objekte aussahen, konnte dann aber mit der Lernphase nichts anfangen: der Ablauf war viel zu schnell. Nahm sich in

der Testphase extrem viel Zeit, die Objekte zu rotieren. Kam gar nicht auf die Idee, einzelne Bilder auswendig zu lernen. – Hat auch diejenigen Ansichten mit verdeckten Kugeln durch räumliches Nachvollziehen gelöst.

11.3 Neben-Experiment „Kugel-Inform“

11.3.1 Information zur Chiralität

Die Probanden bekamen von Anfang an nur untenstehende Information und sonst kein weiteres Priming (entspricht der Kontrollgruppe).

Berücksichtigen Sie bei der Versuchsdurchführung folgende Ratschläge!

- „Bei dieser Aufgabe spielt das Prinzip der Spiegelsymmetrie eine wesentliche Rolle.“
- „Das Verhältnis, wie es zum Beispiel Ihre rechte Hand zu Ihrer linken Hand hat, führt zum Ziel.“
- „Das Prinzip der Händigkeit/Chiralität ist bestimmend.“

11.3.2 Introspektion eines Probanden (Kugel-Inform)

- Am Anfang war der Hinweis zur Chiralität unnütz, da nicht anwendbar. – Sieht nur ununterscheidbare Kugelhaufen.

Bald kristallisiert sich Objekt O1 heraus (Kugel mit drei Berührungspunkten = „diagnostisches Merkmal“). – O2 und O3 als Schlange erkannt, aber ihre Unterscheidung unmöglich.

Eine gewisse Zeit lang geht es so weiter, bis er deutlich fühlt, dass er stagniert und sich nichts mehr ändert.

Jetzt liest er den Zettel mit dem Hinweis zur Chiralität erneut und kann ihn plötzlich verstehen. Nun ist die Frage, wie die Spiegelsymmetrie in den Bildern verwirklicht ist.

Wahl einer Referenzposition, die anhand eines besonders deutlichen Bildes der in der learning phase hängen geblieben ist. Unterstützung durch sprachliche Formulierung (wenn rechts, dann xy)

Bildet eine Hypothese des Unterschiedes zwischen O2 und O3 (wenn-dann-Beziehung).

Versucht seine Hypothese mittels Rotieren des erinnerten Bildes in die Referenzposition zu verifizieren. Wenn die Hypothese aufgeht, ist das Objekt erkannt; wenn sie nicht aufgeht, wird die Hypothese verworfen.

Schon im früheren Lernstadium hatte er versucht zu rotieren, es war aber nicht möglich gewesen, da das Objekt zu unklar war (konnte keine klare Referenzposition fassen).

Fazit. Anfangswahrnehmung ist Sortieren: Die einzelnen Elemente werden schnell erkannt (Kugeln). Die Beziehung der Elemente zueinander muss sortiert werden. Der Hinweis auf Spiegelsymmetrie nützt dem Probanden erst etwas, wenn er Hypothesen über die Relationen zwischen den Objektteilen bilden kann. Diese beruhen auf sortierten (klar erinnerbaren) Bildern. Klar erinnern können – wie viele Kugeln, welche vorne, welche hinten, wie verbunden, 2D-Ebene – ist jedoch nicht das gleiche, wie das Objekt zu erkennen. Mental Rotieren kann der Proband das Objekt erst, wenn er in der Lage ist, eine Referenzposition zu bezeichnen. Mental Rotieren dient zur Bestätigung einer 3D-Hypothese. Eine Hypothese kann er sprachlich formulieren (wenn-dann-Beziehung).

11.4 Plausibilität des Maßes „Relapse“

Der statistikbewanderte Leser wird sich fragen, warum hier ein neues Maß eingeführt und nicht die Streuung der einzelnen Punkte mit der Varianz (bzw. Standardabweichung) berechnet wurde. Es soll kurz gezeigt werden, warum die Varianz in diesem Fall nicht die beste Charakterisierung der in Frage stehenden Eigenschaften der Lernkurve liefert, und dass das hier entwickelte Maß Relapse plausibel ist.

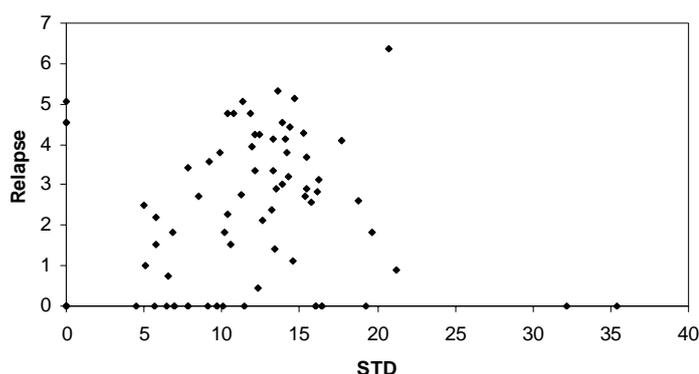
Unter Anhang 11.5 sind einige Lernkurven beispielhaft zusammengestellt. Untenstehende Wertetabelle bezieht sich auf die ersten acht Probanden (je eine Zeile). Sie zeigt die Werte für:

- Anzahl Lerneinheiten für einen Probanden (N)
- die normierte Differenz zwischen dem Kriterium und der Leistung in der 1. Lerneinheit (Progress)
- die Standardabweichung der Lernkurve
- die Summe der negativen Fortschritte
- die normierte Summe (Relapse)

Proband	N	Progress	Standard- Abweichung	S=Summe der neg. Fortschritte	S/N = Relapse
diafra	21	3,25	16,17	59,08	2,81
hanvog	16	3,41	15,47	59,09	3,69
ilili	10	5,46	17,7	40,91	4,1
jerhup	11	4,55	13,5	31,82	2,89
sidkad	47	1,26	13,6	250,01	5,32
stemue	7	6,49	18,82	18,18	2,6
stehue	5	9,09	21,17	4,55	0,91
nadnaga	43	0,74	11,31	218,17	5,07

Vergleicht man beispielsweise die Lernkurven der Probanden „sidkad und „stehue“ (fünfte und siebte) und in der Tabelle die entsprechenden Standardabweichungs- und Relapse-Werte, so zeigt sich für eine Kurve mit vielen Rückfällen („sidkad“) eine niedrige Standardabweichung (13.60) und ein sehr hoher Wert für Relapse (5.32). Für die Lernkurve mit wenigen Rückfällen („stehue“) zeigt sich eine hohe Standardabweichung (21.17) und ein niedriger Wert für Relapse (0.91).

Dass die Standardabweichung und das Maß Relapse durchaus unterschiedliche Dinge messen, wird auch aus der untenstehenden Grafik deutlich:

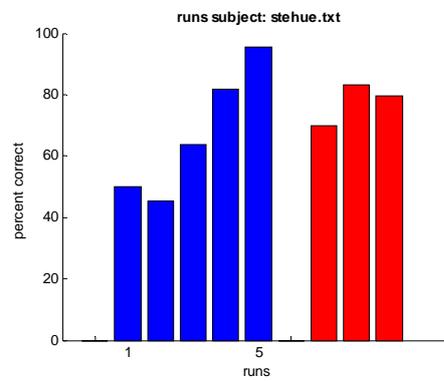
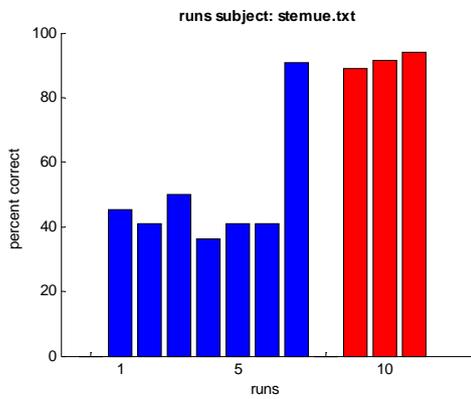
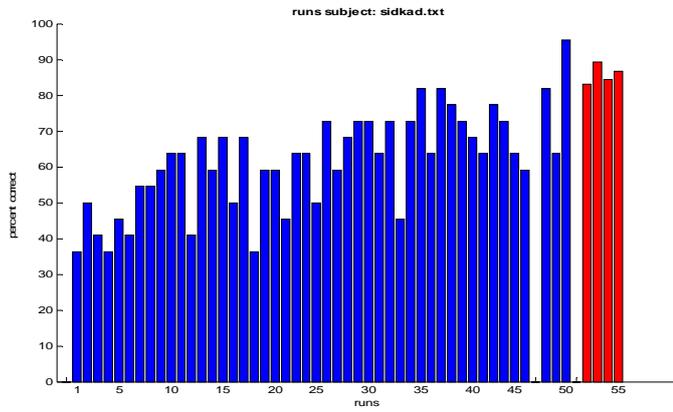
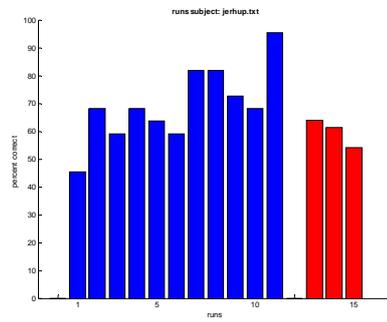
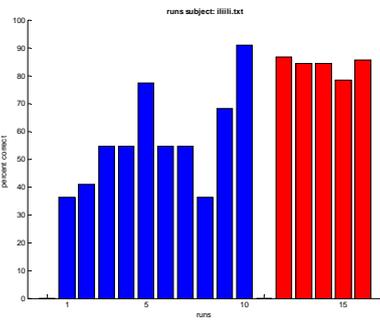
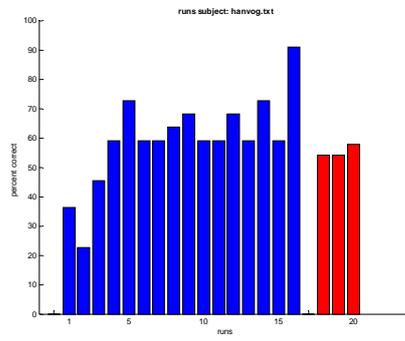
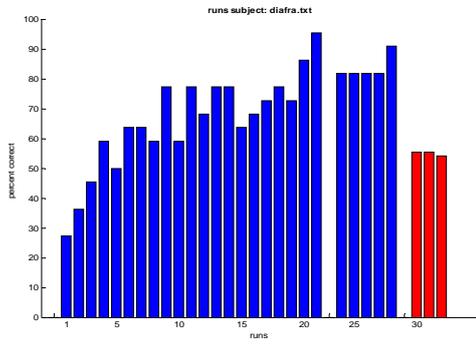


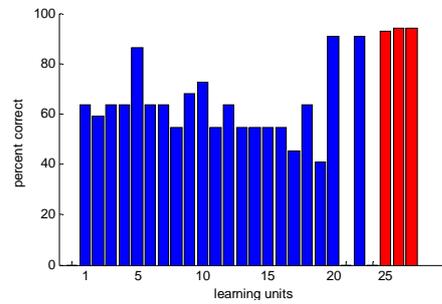
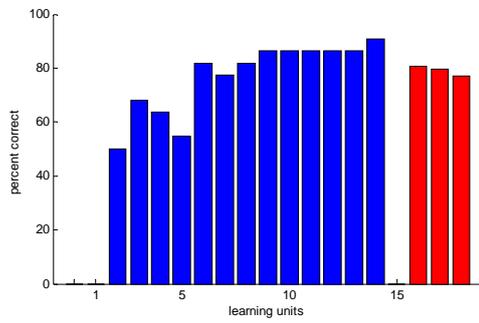
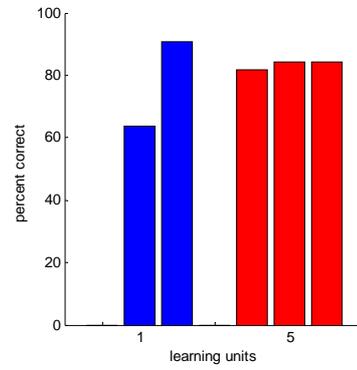
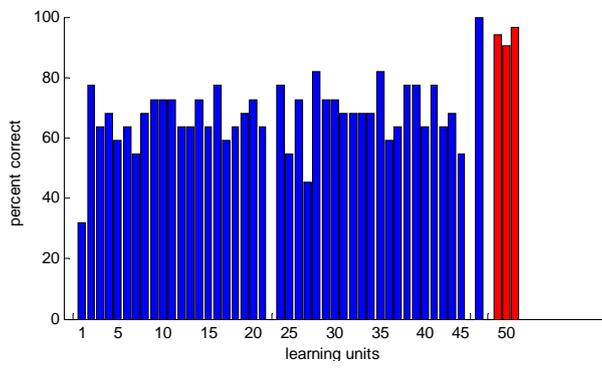
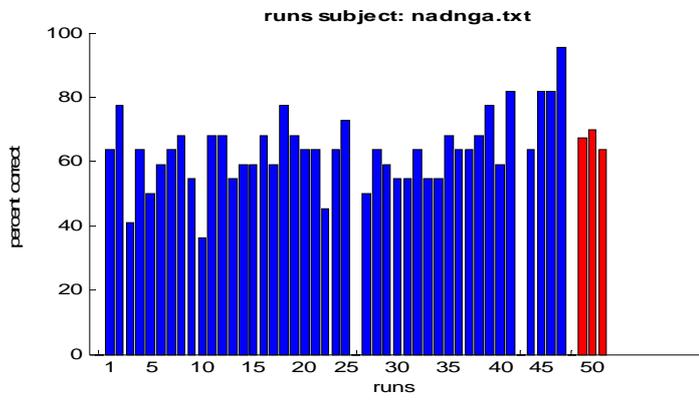
Hier scheint keine Korrelation vorhanden zu sein.

11.5 Beispiele zu den Lernkurven aus dem überwachten Kategorienlernen

Hier sind in willkürlicher Auswahl einige Lernkurven aus dem Ablauf des überwachten Kategorienlernens beispielhaft zusammengestellt (x-Achse: Anzahl Lerneinheiten N; y-Achse: Prozent korrekt).

Jeder Balken stellt den über die drei Objekte gemittelten Prozent-korrekt-Wert („Total“, vgl. Kap 6.4.1) für je eine Lerneinheit dar. Ein weggelassener Balken bedeutet Unterbruch und Fortsetzung des Experimentes in einer zweiten Sitzung. Die letzte abgesetzte Gruppe von Balken (rechts) zeigen die Durchgänge der Generalisierung. Davon wurden immer die ersten drei ausgewertet (einzelne Probanden hängten irrtümlich noch weitere Durchgänge an). Die Unterschiede in der Balkenbreite haben keine Bedeutung.





11.6 Poster

Im Folgenden sind die Poster angehängt, die zu dieser Arbeit entstanden sind:

1. European Conference on Vision & Perception (ECVP) August 2004, Budapest
2. Annual Meeting of the Cognitive Neuroscience Society, April 2005, New York⁹
3. European Conference on Vision & Perception (ECVP) August 2005, La Coruna
4. Europ. Conference on Vision & Perception (ECVP) August 2006, St. Peterburg

⁹ Zur fMRI-Studie vgl. Kap. 8.4 Kritik und Ausblick

Curriculum Vitae

Markus Andreas Gschwind

Geburtsdatum: 27. Juli 1974
Geburtsort: Basel/Schweiz
Nationalität: schweizerisch
Zivilstand: ledig

Studien

1999-2006 Studium der Humanmedizin an der Universität Witten/Herdecke, Witten
November 2006 III. Staatsexamen
März 2005 II. Staatsexamen
März 2002 I. Staatsexamen
März 2001 Physikum
1998-1999 Studium der Humanmedizin an der Universität Basel
1994-1998 Studium der Musik und Lehrdiplom für Klavier an der Musikhochschule Basel
1994 Matura, Gymnasium am Kohlenberg, Basel

Promotion

Seit 2003 Promotion über „Kontextabhängigkeit visueller Objekterkennung“ unter Prof. Dr. Ingo Rentschler am Institut für Medizinische Psychologie der LMU München
2003-2004 Stipendiat des DFG-Graduiertenkolleg 267 „Sensomotorische Interaktion“

Praktika

Juni - September 2006 PJ Chirurgie: Service de Chirurgie, CHU de Fort-de-France, Martinique
Februar - Mai 2006 PJ Neurologie: Service de Neurologie des Hôpitaux Universitaires de Genève, Schweiz
Oktober 2005 - Januar 2006 PJ Innere Medizin: Internal Medicine, MetroHealth Medical Center, Cleveland, Ohio, USA

2001	Praktikum der Allgemeinmedizin in London-Dagenham
2001 und 2000	2 Famulaturen im Service de Chirurgie Viscérale et Urologique, Centre Hospitalier de la Côte Basque, Bayonne, Frankreich

Aktivitäten

2005	Redaktion des Protokolls für das universitäre und interdisziplinäre „Bildwissenschaftliches Kolloquium“ von Iconic Turn (Akademie des 3. Jahrtausends der Hubert Burda Stiftung, München)
2003	Organisation des Video-gestützten Anamneseurses für das Curriculum der medizinischen Fakultät der Universität Witten/Herdecke (UWH),
2002-2003	Leitung der studentischen Veranstaltungsreihen von „HORIZONTE DER MEDIZIN“ (UWH) zu den Themen „Gehirn und Bewusstsein“ und „Der gute Arzt“
2000-2001	Wissenschaftliche Hilfskraft im Fachbereich Allgemeinmedizin der UWH

Publikationen

- M. Gschwind, H. Brettel, E. Osman, I. Rentschler (2004), Structured but viewdependent representation for visual 3D object classification; Perception Vol. 33, suppl. p.73
- S. Duhoux, M. Gschwind, P. Vuilleumier, I. Rentschler, S. Schwartz (2005a), New fMRI views on 3D object recognition (abstract), Cognitive Neuroscience Society, New York
- S. Duhoux, M. Gschwind, P. Vuilleumier, I. Rentschler, S. Schwartz (2005b), Neural correlates of 3D object learning, Perception Vol. 34, suppl. p.116
- H. Brettel, M. Gschwind, I. Rentschler, (2006), Contextual relativity of 3D object representations, Perception Vol. 35 suppl. p.27
- I. Rentschler, M. Gschwind, H. Brettel, E. Osman, T. Caelli, (in revision), Prior knowledge and learning enable structure-based 3D object classification from ambiguous images, Vision Research
- M. Gschwind, H. Brettel, I. Rentschler, (in preparation), Prior knowledge and learning in 3D object recognition. In: Osaka, N., Rentschler, I., & Biederman, I. Object Recognition, Attention, and Action, Springer Tokyo.