



universität  
wien

# Diplomarbeit

Titel der Arbeit

**Der Einfluss von übereinstimmend angeschauten  
Objekten auf die Wiedererkennungsleistung von  
natürlichen Szenen nach Perspektivsprüngen.**

Verfasserin

**Bettina Schuckert**

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im Juli 2011

Studienkennzahl: 298  
Studienrichtung: Psychologie  
Betreuer: V.-Prof. Dr. Ulrich Ansorge



## DANKSAGUNG

---

An dieser Stelle möchte ich all jenen Personen meinen Dank aussprechen, ohne die das Schreiben dieser Diplomarbeit nicht möglich gewesen wäre.

Als erstes möchte ich mich bei Prof. Ulrich Ansorge für die hervorragende Betreuung in dieser Zeit bedanken. Vielen Dank für die Unterstützung beim Erstellen dieser Arbeit und für die immer rasche Hilfestellung bei all meinen Fragen und Anliegen.

Weiters gilt ein großer Dank meiner Kollegin Nina Glösel. Das gemeinsame Arbeiten zu Beginn hat mir den Einstieg sehr erleichtert und ich konnte mich bis zum Ende jederzeit an dich wenden.

Ein besonderes Dankeschön möchte ich meiner Familie, insbesondere meinen Eltern aussprechen. Durch ihre jahrelange finanzielle Unterstützung haben sie mir das Studium der Psychologie erst ermöglicht.



*“It is one of the most difficult things in the world  
to look at anything simply.”*

(Jiddu Krishnamurti, 1969)



## ABSTRACT

---

How could scenes be recognized if the visual focus has changed? Which influence does observed objects exert? In the present thesis this problem formulation was tested with the eyetracking-method. For this purpose, some different color photographs were represented in two research blocks. First of all, there was a learning phase to memorize the showed pictures, which were tested in the following recognition phase. Thereby the recognition was complicated as only a part of the photographs was identic. The remaining pictures were presented from a different perspective. The result interpretation took place with the help of the objects, which were visually focused in both blocks. In the two conditions of outcomes show an ambivalent influence of the repeated observed objects to the recognition performance. While identic scenes are better recognized, scenes with changed perspective have a negative effect. Especially people who fixated the same objects reached worse performance on recognition.

*Key words:* scene recognition, object recognition, fixations on objects, changed perspective





# INHALTSVERZEICHNIS

---

1	EINLEITUNG.....	1
2	THEORETISCHER HINTERGRUND.....	4
2.1	Visuelle Wahrnehmung.....	4
2.2	Objekterkennung.....	6
2.2.1	Strukturbasierter, objektzentrierter Ansatz.....	9
2.2.2	Ansichtsbasierter, beobachterzentrierter Ansatz.....	12
2.2.3	Verbindung der beiden Ansätze.....	14
2.3	Sonderfall der Gesichtserkennung.....	14
2.4	Szenenerkennung.....	16
2.5	Perspektivwechsel bei Objekten und Szenen.....	20
3	METHODE.....	25
3.1	Hypothesen.....	25
3.2	Stichprobe.....	26
3.3	Instrumente und Messgeräte.....	26
3.4	Reizmaterial.....	27
3.5	Design.....	28
3.6	Durchführung.....	29
3.7	Datenanalyse.....	31

4	ERGEBNISSE.....	35
5	DISKUSSION.....	39
6	ZUSAMMENFASSUNG.....	44
	LITERATURVERZEICHNIS.....	45
	ABBILDUNGSVERZEICHNIS.....	53
	CURRICULUM VITAE.....	54

# 1 EINLEITUNG

---

Täglich müssen wir uns in der Welt orientieren und uns an verschiedenen Orten zurechtfinden, seien dies neue oder bekannte Orte, Straßen oder diverse Räumlichkeiten. Dabei können wir innerhalb eines Augenblickes erkennen, ob die vor uns liegende Szene bekannt ist oder wir vor einer unbekanntem Umgebung stehen. Diese Leistung stellt einen wichtigen Aspekt unseres Lebens dar. Wie sollten wir ansonsten die Straße, in der wir leben, zur Schule gehen oder Freunde wohnen, von anderen Straßen unterscheiden können? Schließlich weisen sie alle ähnliche Merkmale auf: in der Mitte befindet sich der Straßenbelag, daneben ein Gehsteig und anschließend auf beiden Seiten Häuserzeilen. Größere Unterschiede gibt es bei der Bepflanzung vor den Häusern, der Anordnung der Häuser und natürlich am Erscheinungsbild der Häuser selbst.

Darüber hinaus müssen wir einen Ort nicht nur aus einer Position heraus wiedererkennen können, sondern auch aus verschiedenen Blickwinkeln. Dies stellt eine besondere Herausforderung dar, weil hierbei fast alle Objekte in einer anderen Form erscheinen, neue Details und Objekte hinzukommen und alte teilweise verdeckt werden oder ganz aus dem Blickfeld verschwinden. Es scheint dabei fast so, als ob wir eine andere Szene betrachten, die dennoch in Teilen dieselbe ist.

Die Leistung, Objekte aus verschiedenen Perspektiven wahrzunehmen und als dieselben erkennen zu können, wird im alltäglichen Leben als selbstverständlich angesehen. Dass dies jedoch nicht so einfach ist, zeigt eine Untersuchung von Gregory & Wallace (2001). Sie bekamen die Möglichkeit mit einem Mann zu arbeiten, der durch eine Operation nach über 50 Jahren seine Sehkraft wiedererlangt hat. Dieser Patient verlor sein Augenlicht bereits im Alter von wenigen Monaten. Dadurch lernte er seine Umwelt als Blinder kennen und verstehen. Als nunmehr Sehender stand er vor dem Problem, seine Umgebung neu zu erlernen. Er musste das, was er mit den Augen wahrnahm, mit dem, was er über seine Umwelt wusste, vereinen und sich merken. Einige Monate nach der Operation zeigte sich, welche Schwierigkeiten es bereiten kann Orte und Objekte wahrzunehmen und wiederzuerkennen speziell wenn sie aus einem anderen Blickwinkel betrachtet werden. Der Patient wurde dabei beobachtet, wie er

eine Lampe betrachtete, um sie herumging und von neuem anschaute. Hierbei erstaunte ihn am Meisten das veränderte Aussehen nach einem Perspektivwechsel.

*„Quite recently he had been struck by how objects changed their shape when he walked round them. He would look at a lamp post, walk round it, stand studying it from a different aspect, and wonder why it looked different and yet the same.“*

(Gregory & Wallace, 2001, S. 33)

Der Mann konnte die Objekte wiedererkennen, solange sie in derselben Form repräsentiert wurden, wie er sie als nunmehr Sehender gelernt hatte. Änderte sich allerdings die Perspektive bei der Betrachtung, so schien es, als ob ein neues Objekt vor ihm liegen würde. Dementsprechend musste er die neuen Informationen lernen und zu seinem bestehenden Wissen über den Gegenstand hinzufügen.

Dieser Fall zeigt deutlich, dass der Prozess der Wiedererkennung bei der Betrachtung von Objekten nach Änderung der Blickrichtung schwieriger ist als man selbst vermutet und aus eigener Erfahrung kennt. Daher ist es umso erstaunlicher, dass wir in der Lage sind ganze Szenen nach einem Perspektivwechsel ohne große Schwierigkeiten wiedererkennen zu können.

### **Aufbau der Arbeit**

Die vorliegende Diplomarbeit beschäftigt sich mit der Wiedererkennungsleistung von Objekten nach einem Perspektivsprung, besonders im Hinblick auf übereinstimmend angeschauten Objekten in gleichen oder ähnlichen Szenen. Das Ziel dieser Studie ist deshalb die Frage zu klären, ob die Betrachtung von Objekten Einfluss auf die Wiedererkennung von ganzen Szenen nimmt.

Zur besseren Veranschaulichung dieser Fragestellung, wird zunächst ein theoretischer Hintergrund dargeboten. Dieser besteht anfangs aus einem kurzen Einblick in den biologischen Aspekt der visuellen Wahrnehmung (siehe Abschnitt 2.1). Anschließend wird auf die zwei verschiedenen Ansätze zur Objektwahrnehmung (siehe Abschnitt 2.2) näher eingegangen. Diese unterscheiden sich vor allem in den entgegengesetzten Sichtweisen über den Blickpunkt. Abschnitt 2.2.1 behandelt die blickpunktunabhängigen Theorien, während im nächsten Abschnitt (2.2.2) auf die

blickpunktabhängigen Theorien eingegangen wird. Anschließend wird versucht die beiden Ansätze miteinander zu vereinen (siehe Abschnitt 2.2.3). Ebenfalls wird kurz die Wahrnehmung von Gesichtern (siehe Abschnitt 2.3) angeführt, da es sich bei diesem Prozess um einen Sonderfall der Wahrnehmung und Wiedererkennung handelt. Da die einzelnen Objekte auch im Kontext ihrer Umgebung betrachtet und untersucht werden, beschreibt der Abschnitt 2.4 die Wiedererkennung von ganzen Szenen. Zum Abschluss des Theorieteils werden noch genauer die Schwierigkeiten und Hintergründe, welche eine Änderung der Perspektive von Objekten und ganzen Szenen mit sich bringt, beleuchtet (siehe Abschnitt 2.5).

Darauf aufbauend wird die Methode der Untersuchung dargestellt (siehe Abschnitt 3). Es wird zunächst eine Beschreibung von der Durchführung des Experiments gegeben bis hin zur Schilderung der Bearbeitung von den Daten. Im Abschnitt 4 werden die Ergebnisse der Studie angeführt und erläutert. Ein Zusammenspiel des theoretischen Hintergrundes mit den Ergebnissen aus der Untersuchung folgt in der Diskussion (siehe Abschnitt 5), sowie ein Ausblick auf mögliche weitere Fragestellungen für zukünftige Studien.

## 2 THEORETISCHER HINTERGRUND

---

### 2.1 Visuelle Wahrnehmung

Um den Prozess des Erkennens von Objekten und ganzen Szenen besser verstehen zu können, gibt es zu Beginn eine kurze Einführung in das menschliche visuelle System und die damit verbundene visuelle Wahrnehmung der Umwelt (Anderson 2001; Birbaumer & Schmidt 2003; Goldstein 2002). Schließlich handelt es sich bei der Wahrnehmung nicht nur um die Aufnahme und weitere Verarbeitung der vom Auge kommenden visuellen Stimuli, sondern auch um deren Erkennung und Identifikation anhand von Informationen aus dem Gedächtnis. Folglich kommt die Leistung der visuellen Wahrnehmung durch ein Zusammenspiel der Augen mit dem Gehirn zustande.

Die Aufnahme der Umwelt erfolgt durch die ausgehende elektromagnetische Strahlung mit einer Wellenlänge zwischen 400 und 700 Nanometer. Diese durchqueren das Auge, bestehend aus der Hornhaut, dem Kammerwasser, der Pupille und dem Glaskörper, und erzeugen auf der hinten liegenden Netzhaut (Retina) ein Abbild der Umgebung. Das Bild wird auf dem Kopf stehend, seitenverkehrt und stark verkleinert projiziert. Die Retina besteht aus Photorezeptoren, den Zäpfchen und Stäbchen. Während die Stäbchen eine sehr hohe Lichtempfindlichkeit aufweisen, besteht die Aufgabe der Zapfen im Farbsehen. Diese zwei Typen von Rezeptoren sind auf der Netzhaut unterschiedlich verteilt und befinden sich aufgrund ihrer Eigenschaften an verschiedenen Orten. Die Fovea centralis, auch gelber Fleck genannt, beinhaltet fast nur Zapfen, daher ist sie die Stelle des schärfsten Sehens im Auge. Wird ein Objekt ins Visier genommen, so kommt es zu einer kurzen Fixationsdauer des Blickes auf den Gegenstand. Aufgrund dieser Verweildauer bewegen wir unsere Augen so zum Objekt, dass es in den Bereich des schärfsten Sehens fällt. Die Informationen über die Abbildung der Umwelt werden von den Photorezeptoren an die Bipolarzellen und anschließend an die Ganglienzellen weitergegeben. Diese bilden den Sehnerv, welcher die Verbindung zum Gehirn darstellt. Im Chiasma opticum kreuzen sich die zwei Sehnerven und tauschen die Hälfte ihrer Bahnen so aus, dass von jedem Auge die Informationen über die linke Gesichtsfeldhälfte in die rechte Hirnhälfte weitergeleitet wird und bei der rechten Gesichtsfeldhälfte genau umgekehrt.

Nach dem Austausch im Chiasma opticum gibt es mehrere Verzweigungen der Sehnerven. Ein solcher visueller Pfad führt zum Corpus geniculatum laterale, dessen Funktion vor allem in der Feinwahrnehmung und dem Erkennen von Objekten besteht. Ein zweiter Pfad führt von der Sehkreuzung zum Colliculus superior. Dieser Teil des Hirns gibt Auskunft über die Lokalisation von Objekten und ist für die Wahrnehmung von groben Mustern zuständig. Die nun vorverarbeitete Information wird an die nachgeschalteten Hirnregionen, dem visuellen Cortex, weitergeleitet. Durch diesen Teil des visuellen Systems wird die visuelle Wahrnehmung erst möglich. Hierbei ist vor allem die primäre Sehrinde (Areal 17) wichtig, da sie unmittelbar das Abbild der gegenüberliegenden Gesichtsfeldhälfte reproduziert. Weitere Sehzentren stellen Areal 18 und 19 dar.

In den primären visuellen Cortex wird eine vollständige Information über die Umwelt übermittelt. Die Aufgabe der Sehrinde besteht darin, die erhaltenen Daten an verschiedene extrastriäre visuelle kortikale Areale zu übertragen. Jedes Areal ist auf die Analyse eines anderen Merkmals spezialisiert. Welche Farben in dem gesehenen Ausschnitt präsentiert werden, welche Formen die einzelnen Objekte haben oder den Kontrast zwischen den Objekten. Wobei die Spezialisierung der Areale immer genauerer und voneinander differenzierter wird.

Von diesen Arealen werden Signale in die visuellen Assoziations- und Integrationsregionen, den Temporal- und Parietallappen, übertragen. Für die Objektidentifikation sind die Assoziationsfelder im unteren Temporallappen zuständig, wohingegen die Assoziationsfelder des Parietallappens Fragen nach der räumlichen Lokalisation der Gegenstände beantworten müssen.

Zusammengefasst stellen die wichtigsten Punkte beim Erkennen der Umwelt zunächst das Auge für die Aufnahme und Bündelung der Lichtstrahlen, der Sehnerv zum Übermitteln der Abbildungen, der primäre visuelle Cortex zum Ordnen der Informationen und die Assoziationsfelder zum Erkennen der Umwelt dar.

Ein Problem, welches sich bei der Wahrnehmung von Objekten ergibt, ist jenes ihrer Übertragung. Die Netzhaut ist aufgrund ihres Aufbaus nur in der Lage zweidimensionale Bilder zu übermitteln, die wir aber als dreidimensionale Gebilde wahrnehmen. Die Augen liegen etwas voneinander entfernt und dadurch nehmen sie zwei leicht voneinander unterschiedliche Abbildungen der Umwelt auf. Durch diese kleinen Differenzen, auch Querdisparation genannt, kann der Sehapparat mehr

Botschaften über seine Umgebung wie die Raum-Lage-Beziehung der Objekte oder deren Tiefeninformationen gewinnen und das räumliche Sehen unterstützen. Aus diesem Grund, und kombiniert mit unserem vorhandenen Wissen, können wir Objekte als dreidimensional erkennen. (Zur weiteren Vertiefung siehe Anderson 2001; Birbaumer & Schmidt, 2003; Goldstein, 2002)

## 2.2 Objekterkennung

Die Thematik der Objekterkennung scheint auf den ersten Blick einfach zu sein. Objekte werden wahrgenommen, als solche abgespeichert und können später wieder abgerufen werden. Gleich zu Beginn dieser Aufzählung merkt man, dass es hierbei große Schwierigkeit gibt, da unsere Augen nur zweidimensionale Bilder auf der Retina ablichten können, die wir trotzdem dreidimensional wahrnehmen (Tarr & Bülthoff, 1998). Daher stellt sich die Frage, wie wir unsere Umgebung wahrnehmen und wiedererkennen, wenn sie selbst als solche nicht abgespeichert werden kann.

*„The problem of recognition can be summarized as determining how perceptual mechanisms map infinitely varying two-dimensional images into representations of a finite set of three-dimensional objects. Both learning and memory are intrinsic to this process; [...] Thus, any account of the visual recognition of complex objects is an account of how observers learn about and remember the visual world.“*

(Barenholtz & Tarr, 2006, S. 157)

Zahlreiche Forscher haben sich mit dieser Problematik beschäftigt und haben hierzu einige Theorien aufgestellt. Diese lassen sich in zwei Gruppen gliedern: den objektzentrierten (in weiterer Folge auch strukturbasiert oder blickpunktunabhängig genannt) und den beobachterzentrierten (in weiterer Folge auch ansichtenbasiert oder blickpunktabhängig genannt) Ansatz. Sie unterscheiden sich vor allem dadurch, wie die verschiedenen Informationen von Objekten wie im Gehirn abgespeichert werden. Wobei sich die Frage stellt, welche dieser beiden Konstrukte das Wahre ist (Biederman & Bar, 1999; Biederman & Gerhardstein, 1993, 1995; Hayward & Tarr, 2000; Tarr & Bülthoff, 1995). Es gibt allerdings auch Annahmen die für ein Zusammenspiel beider Theorien sprechen (Tarr & Bülthoff, 1995; Lawson & Humphreys, 1998).



Diesen Ansätzen zur Erforschung der Objekterkennung liegen zunächst die Figur-Grund-Trennung und die Gestaltgesetze zugrunde. Eine bedeutende Leistung bei der visuellen Wahrnehmung stellt die Strukturierung und Segmentierung der gesehenen Objekte dar (Wertheimer, 1912, 1923). Gegenstände müssen unabhängig von ihrer Umgebung als solche erkannt werden können. Daher ist es wichtig sie von ihrem Hintergrund zu lösen und davon unabhängig wahrzunehmen. Peterson (1994) zeigte, dass dieser Prozess einige Schwierigkeiten birgt speziell bei dreidimensionalen Darstellungen. Eine weitere Grundlage der Objekterkennung erörtert die Gestaltpsychologie mit ihren Gestaltgesetzen der Wahrnehmungsorganisation. Bereits Anfang des 20. Jahrhunderts beschäftigten sich vor allem deutsche Psychologen, wie Max Wertheimer, Kurt Koffka und Wolfgang Köhler, mit der Frage nach der Organisation von Merkmalen zu Einheiten und Objekten. Man muss wissen, wo hört das eine Objekt auf und wo fängt das Nächste an, bevor man überhaupt beurteilen kann was man sieht. Ein Problem beim Erkennen von Gegenständen liegt darin, dass sie von ihrer Umgebung oder anderen Utensilien teilweise verdeckt werden oder ineinander überlaufen. Mithilfe der Gestaltgesetze gelingt es die vor einem liegenden Objekte zu gliedern und zu vervollständigen. Hierbei lautet die Grundregel aller Gestaltpsychologen „Das Ganze ist mehr als die Summe seiner Teile“. Dieser Leitsatz bedeutet, dass die Wahrnehmung einzelner Teile als eine Einheit mehr Bedeutung und dadurch eine höhere Qualität bekommt, wie wenn man sie alleine betrachten würde. Im Laufe der Zeit entstanden daraus viele Gesetzmäßigkeiten, daher fasste Metzger (1975) sie auf insgesamt sieben zusammen, welche Palmer (1992) sowie Palmer und Rock (1994) noch um zwei erweiterten. Diesen allen ist der Wunsch nach einem klaren und prägnanteren Wahrnehmungseindruck gemeinsam, daher wurde das „Gesetz der guten Gestalt“ verwirklicht. Es werden immer jene Einheiten zu einer Gestalt zusammengefasst, welche eine einfache und übersichtliche Form repräsentieren. Dadurch werden sie leichter erkannt und können besser abgespeichert werden.

Mit welchen Schwierigkeiten der Prozess der Objekterkennung verbunden ist, zeigt Abbildung 1. Objekte müssen unter differierenden Verhältnissen als dieselben erkannt werden. In der unten abgebildeten Szene erkennt man einen Raum mit einer Vielzahl an Stühlen. Diese Sitzmöglichkeiten variieren sehr stark voneinander, da sie unterschiedliche Größenverhältnisse haben, aus anderen Blickwinkeln abgebildet sind

und unter verschiedenen Lichtbedingungen stehen. Trotzdem kann man alle Stühle, bis auf einen abweichenden Stuhl in der Mitte, als dasselbe Modell erkennen, obwohl manche teilweise verdeckt werden oder anders zusammengesetzt wurden wie im Bild rechts außen. Eine weitere Leistung unserer Wahrnehmung ist die Unterscheidung der Stühle anhand ihrer Sitzmöglichkeit. Wir erkennen rechts an der Wand ein Bild sowie ein Schattenbild eines Stuhls und wissen, dass wir auf diesen nicht Platz nehmen können. Ebenso können wir mit einem Blick auf die Abbildung feststellen, dass der Stuhl auf dem Schreibtisch nicht zum Sitzen geeignet ist, obwohl er dieselbe Größe hat wie jener im Hinterzimmer auf dem man gleichwohl Platz nehmen kann.



Abbildung 1: Darstellung einer komplexen Szene. Diese beinhaltet viele Stühle, die unterschiedliche Größen aufweisen, aus verschiedenen Blickwinkeln abgebildet sind und unter anderen Lichtverhältnissen stehen (aus Wallis & Bühlhoff, 1999).

Diese kurze Einführung soll als Grundlage für die weitere Darstellung der Erforschung der Objekterkennung dienen, die im nachfolgenden näher behandelt wird. Für einen

Überblick über die Entwicklung der Objekterkennung in den letzten Jahren bis zum heutigen Forschungsstand sei auf Peissig und Tarr (2007) verwiesen.

### **2.2.1 Strukturbasierter, objektzentrierter Ansatz**

Der erste Ansatz zur Objekterkennung ist der so genannte blickpunktunabhängige bzw. objektzentrierte. Der Name verdeutlicht schon den Ausgangspunkt dieser Theorie: das Objekt steht im Mittelpunkt. Die Abspeicherung der Umgebung erfolgt unabhängig vom Blickpunkt aus dem sie betrachtet wird. Vertreter dieser strukturbasierten Theorien sind Marr sowie Biederman, welche anschließend näher erläutert werden. Grundlage dieser zwei Ansätze ist die Wiedererkennung anhand einfacher Komponenten, die aus jeder Perspektive gleich gut funktioniert.

#### *Generalized cylinders*

Marr (1982) entwickelte einen algorithmischen Ansatz zur Wahrnehmung und Erkennung von Objekten. Grundlage dieser computationalen Theorie ist, dass das Sehen ein Prozess der Verarbeitung von zweidimensionalen Bildern auf der Netzhaut ist, der schlussendlich zu einer dreidimensionalen Beschreibung der Umwelt führt. Dieser mathematische Verarbeitungsprozess besteht aus insgesamt vier Stufen, die mit zunehmendem Grad immer detailreicher werden.

Ausgangspunkt stellt die Abbildung des vor einem liegenden Objektes oder Szene auf der Retina dar. Anschließend gibt es zwei Zwischenstufen, welche anschließend in einer dreidimensionalen Repräsentation enden.

Die erste Zwischenstufe wird als „primäre Rohskizze“ bezeichnet, was soviel wie der erste Entwurf des Wahrgenommenen bedeutet. Darunter versteht Marr eine zweidimensionale Beschreibung des Objektes, die sich durch die Analyse von Lichtintensitätsunterschieden ergibt. Somit können Kanten und Konturen, sowie zusammenhängende Flächen und Umrisse des betrachteten Gegenstandes erkannt werden.

In der zweiten Phase „zweieinhalbdimensionale Skizze“ werden Elementarmerkmale gruppiert und verarbeitet. Mit Hilfe der Informationen aus der primären Rohskizze als auch durch die divergierenden Netzhautbilder, werden Orientierung und Tiefe des Objektes ermittelt.

Die „dreidimensionale Skizze“ vereint nun die ganzen Verarbeitungsschritte und setzt sie miteinander in Beziehung. Um das Objekt zu erkennen, wird dieses entstandene Modell mit bereits abgespeicherten Skizzen verglichen und es kommt zur Wahrnehmung eines dreidimensionalen Objektes. Der Gegenstand wird, unabhängig von der betrachteten Perspektive, immer als derselbe erkannt.

Die Repräsentation der Umgebung erfolgt anhand von „generalized cones“, generalisierten Zylindern (Marr & Nishihara, 1978). Bei der dreidimensionalen Darstellung beschreiben die Hauptachsen der Zylinder die Ausrichtung der Objekte. Diese Modellachse gibt auch Aufschluss über die Größe und Breite der Gegenstände, welche durch Änderung dieser diverse Figuren veranschaulichen. Damit feinere Unterschiede der Objekte zum Ausdruck kommen, kann ein Zylinder wiederum in mehrere kleinere Zylinder zerlegt werden. Ein Beispiel dafür wird in Abbildung 2 gezeigt, in der durch Teilung eines einzelnen Zylinders eine immer präzisere Darstellung eines Menschen entsteht.

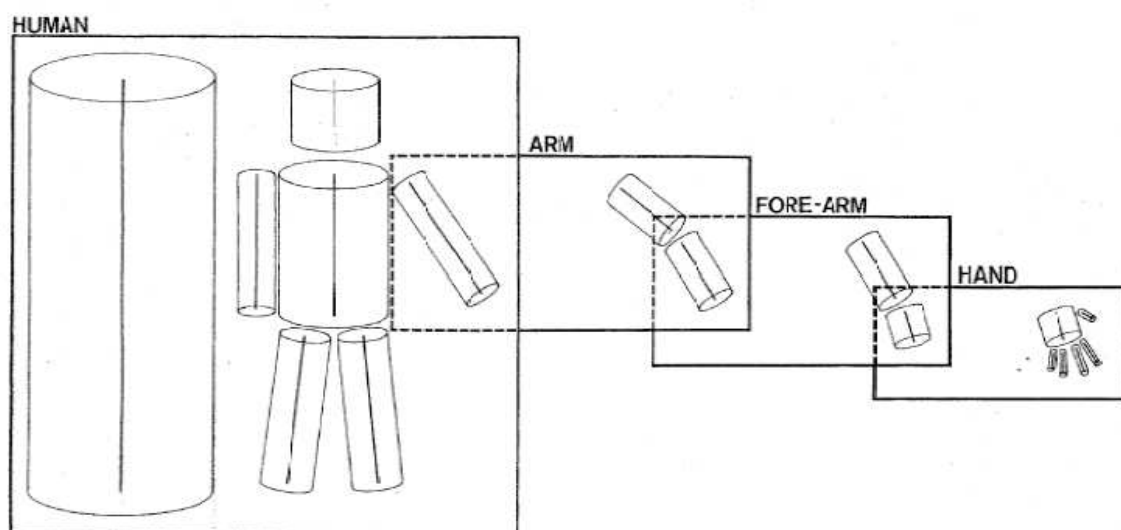


Abbildung 2: Darstellung eines hierarchischen 3-D Modells eines Menschen; von einem Zylinder, der die Hauptachse eines Körpers repräsentiert, bis hin zu kleineren Zylindern für die Handfläche und die einzelnen Finger einer Hand (aus Marr & Nishihara, 1978).

### *Recognition - by - Components*

Auf Grundlage von Marr's computationalen Ansatz entwickelte Biederman (1987;

Biederman & Gerhardstein 1993; für neuere Entwicklungen siehe Biederman 2001) seine Theorie der „Recognition-by-Components“. Diese besagt, dass mit Hilfe von elementaren Teilkörpern alle Objekte, laut Biederman an die 30 000, erkannt werden. Unter den Grundelementen werden Zylinder, Quader, Kegel, gekrümmte Bögen und noch weitere Körper verstanden, die so genannten „Geons“ (von geometrical Icon abgeleitet). Insgesamt gibt es 36 verschiedene Elemente, mit denen, durch die unterschiedliche Art und Weise ihrer Zusammensetzung und vielfältigen Kombinationen, alle Objekte der Welt beschrieben werden können (siehe Abbildung 3). Die Geons sollen möglichst blickpunktunabhängige Eigenschaften besitzen, damit die daraus zusammengesetzten Gegenstände ebenfalls aus jeder Perspektive erkenntlich sind.

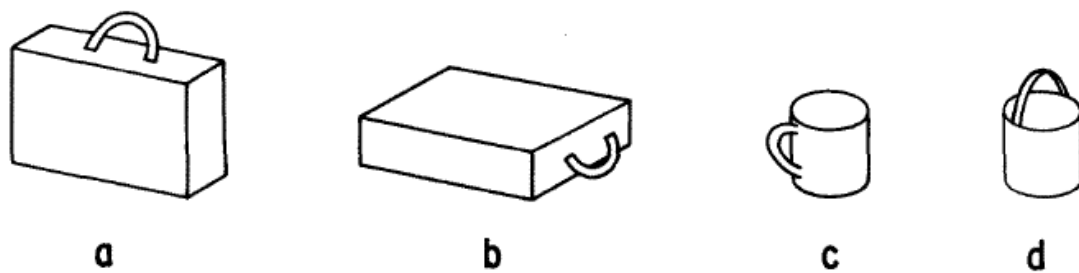


Abbildung 3: Darstellung von Möglichkeiten der Zusammensetzung von Geons. Die unterschiedliche räumliche Anordnung gleicher Geons kann zu verschiedenen Objekten führen. Allen Gegenständen ist der gekrümmte Bogen gleich, a und b haben zusätzlich einen Quader und c und d einen Zylinder. Durch die andersartige Zusammensetzung entstehen ein Aktenkoffer (a), eine Geldkassette (b), eine Tasse (c) und ein Kübel (d) (aus Biederman, 1987).

Bei der Wahrnehmung von Objekten stellten sich speziell Kanten, Ecken und Krümmungen, auch nicht-zufällige Eigenschaften genannt, als wichtig heraus. Dies untersuchte Biederman (1987) indem er Probanden Bilder vorgab, bei denen einmal Kanten und Ecken fehlten, ein anderes Mal die zufälligen Eigenschaften von Gegenständen. Das Erkennen gelang im ersteren Fall besser, da sie wieder rekonstruierbar waren. Ebenfalls gelingt die Identifikation der Objekte, wenn nur einige ihrer Geons sichtbar sind, solange es sich um spezifische Elemente der Gegenstände handelt.

Das Erkennen bei der Recognition-by-Components Theorie erfolgt durch einen Vergleich der zusammengesetzten Geons mit gespeicherten Repräsentationen. Bei dem Prozess hat die Größe der Objekte keinerlei Einfluss, da nur die Kombination der Elemente wichtig ist.

### **2.2.2 Ansichtsbasierter, beobachterzentrierter Ansatz**

Einen anderen Ansatz bei der Erforschung von Objektwahrnehmung veranschaulicht der beobachterzentrierte oder auch blickpunktabhängige genannt. Wie die Benennung bereits andeutet, wird hier vom Beobachter ausgegangen. Die wichtigste Frage die sich stellt ist die, wie zweidimensionale Abbildungen von Objekten auf der Netzhaut erscheinen. Es werden mehrere Ansichten von Objekten und Szenen abgespeichert, mit denen die Dreidimensionalität repräsentiert wird. Somit entfällt bei diesen Theorien die Speicherung von dreidimensionalen Strukturen.

Palmer, Rosch & Chase (1981) stellten Untersuchungen zur Erkennung von Objekten aus unterschiedlichen Perspektiven an. Sie fotografierten Gegenstände aus verschiedenen Blickwinkeln und ließen Probanden diese nach ihrer Vertrautheit und Typikalität beurteilen. Es zeigte sich, dass Fotografien aus einer bestimmten Perspektive bevorzugt wurden und zwar jene der sogenannten „kanonischen Ansicht“. Anschließend präsentierten sie Versuchspersonen Bilder mit unterschiedlichen Objekten, aufgenommen aus diversen Perspektiven, die sie so schnell wie möglich benennen sollten. Die Ergebnisse zeigten, dass jene Objekte schneller erkannt wurden, die aus der kanonischen Ansicht abgebildet waren. Die Leistung der Probanden verschlechterte sich mit zunehmender Abweichung von diesem Blickwinkel.

Die Erkenntnisse über kanonische Ansichten sowie weitere Studien (Edelman & Bülthoff, 1992; Tarr & Bülthoff, 1998) legen die Vermutung nahe, dass Objekte anhand mehrerer zweidimensionaler Ansichten abgespeichert werden und nicht als vollkommene dreidimensionale Gebilde. Edelman und Bülthoff (1992) erhöhten bei ihrer Untersuchung den Schwierigkeitsgrad und ersetzten bei ihrem Versuchsaufbau natürliche Objekte mit ungewohnten, komplexen Figuren. Ebenfalls fanden Lawson

und Humphreys (1996) in ihren durchgeführten Experimenten Hinweise auf ein Zusammenspiel von Objekterkennung mit Perspektivänderung.

Diese blickpunktabhängigen Repräsentationen erfordern vom menschlichen Hirn große Speicherkapazitäten, da von jedem Objekt alle möglichen Blickwinkel berücksichtigt und abgespeichert werden müssen. Ebenso gibt es Schwierigkeiten bei der Erkennung der Gegenstände aus einer ungewohnten Perspektive, da diese zweidimensionale Ansicht noch nicht im Gehirn verankert wurde. Daher wurden verschiedene Theorien vorgeschlagen, um von einer unbekanntem auf eine bekannte Ansicht generalisieren zu können.

Tarr und Pinker (1989) schlugen ein Modell der mentalen Rotation vor. Der Gegenstand wird solange um seine Achse gedreht, bis es mit einer gespeicherten Ansicht übereinstimmt. Je mehr die neue Ansicht von dem bereits gespeicherten Abbild abweicht, desto mehr Zeit und Aufwand wird zur korrekten Identifikation benötigt.

Eine zweite Möglichkeit zur Generalisierung stellt die Alignment-Theorie von Ullman (1989) dar. Bei der Erkennung durch Anpassung findet ein Abgleich der zweidimensionalen Ansicht mit den abgespeicherten dreidimensionalen Repräsentationen aufgrund von Merkmalen, die beide Bilder enthalten, statt. Durch Berechnungen der Kennzeichen kann die neue Ansicht demjenigen Modell zugeordnet werden, mit dem es die größtmögliche Deckung erzielt.

Ein mathematischer Zugang findet mit dem Modell der „Linearkombination von zweidimensionalen Ansichten“ (Ullman & Basri, 1991) statt. Die Autoren nehmen an, dass von einem Objekt mehrere detaillierte, zweidimensionale Bilder abgespeichert werden. Um die, im Blickfeld liegenden aber aus der einen Perspektive noch nie gesehenen, Objekte mit Repräsentationen aus dem Gehirn abgleichen zu können, müssen die bereits abgespeicherten Bilder des Gegenstandes mittels Linearkombinationen verrechnet werden. Nach Ullman und Basri genügen insgesamt fünf gespeicherte Ansichten damit das Objekt korrekt identifiziert werden kann.

Ein weiterer Ansatz der in den Bereich der zweidimensionalen Repräsentationen fällt, ist jener des „View Interpolation“ (Poggio & Edelman, 1990). Dieser untersucht den Abgleich von Objektrepräsentationen in einem Koordinatensystem. In diesem Raum gibt es, wie im Modell von Ullman und Basri (1991), ein paar abgespeicherte

Darstellungen des Gegenstandes. Die unterschiedlichen Perspektiven des Objektes werden als fixe Punkte im Koordinatensystem angesehen. Kommen Abbildungen aus neuen Betrachtungswinkeln hinzu, werden sie mithilfe von Funktionen abgeglichen, aufsummiert und anschließend erkannt.

### **2.2.3 Verbindung der beiden Ansätze**

Neuere Untersuchungen legen die Vermutung nahe, dass sowohl der blickpunktabhängige als auch der blickpunktunabhängige Ansatz der Wirklichkeit entsprechen (Lawson & Humphreys, 1998). Die zwei Modelle sind aber für verschiedene Wiedererkennungsaufgaben zuständig (Tarr & Bühlhoff, 1995).

Tarr, Williams, Hayward und Gauthier (1998) stellten Untersuchungen der Frage betreffend, welcher dieser beiden Ansätze nun stimmt, an. Unter verschiedenen Bedingungen mussten Versuchspersonen dreidimensionale Objekte wiedererkennen. Tarr et al. fanden heraus, dass die Wiedererkennung von Objekten bei einfachen Vorgaben (z.B. Unterscheidung von Kategorien wie Hund oder Katze) unabhängig von der Perspektive stattfindet. Bei schwierigeren Aufgaben (z.B. Unterscheidung innerhalb von Kategorien: Welche Katze?) jedoch die blickpunktabhängigen Theorien zum Einsatz kommen (Tarr & Bühlhoff, 1995).

## **2.3 Sonderfall der Gesichtserkennung**

In unserem Alltag müssen nicht nur Objekte wiedererkannt und zugeordnet werden, sondern auch Gesichter bekannten und unbekanntem Menschen. Die Schwierigkeiten der Gesichtserkennung lassen sich in sechs Problemfelder (Gross, Shi & Khon, 2001) einteilen, die aufgrund ihrer wesenhaften und stark veränderlichen Art nur zum Teil mit jenen der Objektwahrnehmung übereinstimmen. Beiden Gruppen ist die Wiedererkennung unter verschiedenen Perspektiven und Lichtverhältnissen gleich. Genauso wie Utensilien werden Antlitze erkannt, wenn sie teilweise - durch Sonnenbrillen oder ähnliches - verdeckt werden. Im Gegensatz zur Wahrnehmung von Objekten, muss die Gesichtswahrnehmung Änderungen in der Mimik und im Aussehen bewältigen. Beim Gesicht handelt es sich um kein starres Objekt, sondern um viele bewegliche Partien mit denen abwechslungsreiche Mimik erzeugt wird und



Gesichtsausdruck einhergeht. Durch Altern verändert sich im Laufe der Zeit das Antlitz, jedoch kann es ebenso durch Frisuren, Make-up, Brillen usw. von Tag zu Tag variieren. Eine weitere Rolle bei der Gesichtswahrnehmung spielt die Differenzierung zwischen den Geschlechtern, aber auch zwischen ethnischen Gruppen, da sie sich in den einzelnen Merkmalen als auch in den Relationen derselben unterscheiden.

Eine Untersuchung von Pegna, Khateb, Michel und Landis (2004) zeigte, dass Objekt- und Gesichtserkennung unterschiedliche Wahrnehmungsgeschwindigkeiten als auch biologische Prozesse zugrunde liegen. Während zur Erkennung von Objekten ungefähr 200 ms benötigt werden, können Gesichter bereits nach 100 ms korrekt identifiziert werden. Im Gehirn gibt es jeweils für beide Vorgänge spezialisierte Hirnareale. Während bei Gesichtern die rechten temporalen Areale aktiviert werden, gibt es für Gegenstände eine erhöhte Aktivität auf beiden Seiten, wobei die rechte dominiert.

Ein weiterer Beleg, der für die eine Abgrenzung der Gesichtserkennung spricht, stellt eine Störung namens „Prosopagnosie“ dar. Durch Schädigung oder Verletzung bestimmter Kortexareale haben Menschen Probleme Gesichter zu erkennen und zuzuordnen, jedoch die Fähigkeit der Objektwahrnehmung bleibt bestehen (Barton, Press, Keenan & O’Connor, 2002; Duchaine & Nakayama, 2005). Dies deutet auf eine getrennte Abspeicherung im Gehirn hin (Pitcher, Charles, Devlin, Walsh & Duchaine, 2009). Personen, die unter dieser Krankheit leiden, können die Gesichter von bekannten Menschen nicht erkennen und darüber hinaus nicht einmal mehr ihr eigenes. Die Merkmale können voneinander getrennt benannt und ebenfalls erkannt werden, es können jedoch keine Relationen hergestellt werden. Das Gesicht als Ganzes hat keine Aussagekraft mehr. Burton, Young, Bruce, Johnston und Ellis (1991) untersuchten Patienten mit Wahrnehmungsstörungen in Bezug auf Gesichter. Für vergleichende Untersuchungen von Prosopagnosie mit offenen und verdeckten Bedingungen sei auf Schweinberger & Burton (2003) verwiesen.

Das bekannteste Modell zur Gesichtserkennung stammt von Bruce & Young (1986). Es handelt sich hierbei um ein modulares Konstrukt, welches Annahmen bezüglich der Gesichtsverarbeitung anstellt. Es zeigt einen Informationspfad an, der vom Input

durchlaufen werden muss. Gleichzeitig gibt er auch an, welche Zwischenschritte unabhängig passieren und welche abhängig voneinander sind. Bruce und Young stellten acht Komponenten auf, die einen wichtigen Beitrag zur Gesichtserkennung leisten. Diese Komponenten beinhalten nicht nur allgemeine und spezifische Details zum Äußeren, sondern darüber hinaus allgemeine Informationen über die Person wie deren Name, Interessen und noch mehr.

Die vorherrschende Annahme, wie Gesichtserkennung funktioniert, ist jene der holistischen Verarbeitung. Gesichter werden nicht aufgrund ihrer spezifischen Merkmale wie Augen, Nase und Mund erkannt, sondern anhand ihrer konfiguralen Beschreibungen (Bruce & Humphreys, 1994). Farah, Wilson, Drain & Tanaka (1998) zeigten in ihren Experimenten, dass bei der Analyse von Gesichtern eine holistische Verarbeitung stattfindet. Probanden mussten in mehreren Experimenten Gesichter, invertierte Gesichter, Wörter oder Häuser wiedererkennen, nachdem eine Maskierung oder eine Teilmaskierung der Aufgabe erfolgte. Die Ergebnisse berichteten über keine Unterschiede zwischen den Bedingungen in den Kategorien Gesichter, Häuser und Wörter, da sich die Wiedererkennung an Details orientiert. Bei invertierten Gesichtern erfolgte die Wahrnehmung mehr an den Relationen zwischen den einzelnen Merkmalen. Dieses Resultat wurde von Leder und Bruce (2000) bestätigt. In ihrer Untersuchung mussten sich Probanden Gesichter merken, die sie anschließend aufrecht oder auf dem Kopf gestellt wiedererkennen sollten. Sie konnten wie Farah et al. zeigen, dass vor allem die Abstände zwischen den Gesichtsmerkmalen die Wahrnehmung beeinflussen.

## **2.4 Szenenerkennung**

In unserem alltäglichen Leben erblicken wir Objekte nicht nur als solche alleine, da sie zumeist in ihre Umgebung eingebunden sind. Wir sind daher in der Lage ganze Szenendarstellungen mit ihren darin enthaltenen Objekten wahrzunehmen und wiederzuerkennen. Somit erkennen wir Objekte meist nicht einzeln, sondern als Teil der uns umgebenden Welt. Dementsprechend besteht eine Szene aus mehreren Objekten, die in einer sinnvollen Art und Weise angeordnet sind.

Palmer (1975) untersuchte, welchen Einfluss der Kontext auf die Erkennung von Objekten ausübt. Den Teilnehmern wurden diverse kontextabhängige Szenen (z.B. Küche) präsentiert und daran anschließend sahen sie ein Objekt aus einer der drei Versuchsbedingungen: adäquat (z.B. Brotlaib), inadäquat (z.B. Briefkasten), ohne Kontext. In letzterer Bedingung wurde anstatt der kontextabhängigen Abbildung eine leere Szene angezeigt. Die Aufgabe der Probanden bestand darin, die Objekte so schnell wie möglich zu benennen. Die Ergebnisse zeigten deutlich einen Vorteil gegenüber der adäquaten Versuchsbedingung, es wurden mehr Objekte erkannt als in den beiden anderen. Einen weiteren Beleg für die Kontextabhängigkeit von Objekten lieferten Biederman, Glass und Stacy (1973). In ihrer Untersuchung arbeiteten sie mit natürlichen Szenen und Szenen, die in Stücke geteilt und neu zusammengesetzt wurden. Nach einer kurzen Präsentationsdauer der Darstellungen, zeigte ein Pfeil auf eine Stelle des Bildes, wo zuvor ein Gegenstand gewesen ist. Die Teilnehmer sollten das zuvor abgebildete Objekt richtig erkennen, was in der Bedingung der natürlichen Szenen besser gelang.

*„A scene composed of contextually related objects is more than just the sum of the constituent objects. Objects presented in a familiar context are faster to localize and recognize.“*

(Oliva & Torralba, 2007, S. 525)

Ist in einem Bild ein Kontext vorhanden, so wirken die darin enthaltenen Gegenstände nicht mehr nur für sich alleine, sondern durch ihr Zusammenspiel bekommen sie eine tiefere Bedeutung (für einen Überblick siehe Oliva & Torralba, 2007). Dies ist ein Vorteil beim Anblick von gewohnten Szenen: man erkennt Objekte schneller und auch dann, wenn sie großteils verdeckt und nicht gleich offensichtlich sind. Wie die Wahrnehmung von Objekten vom Kontext beeinflusst wird, zeigt Abbildung 4 anhand von Schriftzeichen. In Abhängigkeit von den umgebenden Buchstaben wird der gleiche Buchstabe in der Mitte jedes Wortes anders interpretiert. Im ersten Wort wird er zunächst als „H“ gelesen, beim zweiten Wort jedoch als „A“ aufgefasst.

# TAE CAT

Abbildung 4: Darstellung der Einflussnahme des Kontexts. In der Mitte jedes Wortes befindet sich derselbe Buchstabe, jedoch wird er im Kontext der anderen Buchstaben verschieden interpretiert (aus Selfridge, 1955).

Einige Forscher gehen davon aus, dass Szenen anhand ihres "gist", der Quintessenz einer Szene, gemerkt und wiedererkannt werden. Bereits Mary Potter (1976) zeigte mit ihrem Experiment von einer schnellen Abfolge von Bildern, dass eine Repräsentation von 100 ms ausreicht um das gesehene Bild zu erfassen. Weiters genügt eine Darstellungsdauer von 300 ms um Szenen im Gedächtnis abzuspeichern. Aber schon bei einer Darbietungszeit von 19 bis 67 ms kann eine Szene zu 75 % richtig erkannt werden und bei einer Dauer von ungefähr 100 ms nahezu alle Bilder (Greene & Oliva, 2009). Dies ist durch die Erkennung des Gist möglich, wobei meist ihre Kategorien wie Küche, Straße oder Parks gemerkt werden. Ebenso können mit Hilfe der Quintessenz Annahmen darüber gestellt werden, wohin ein Objekt in einer Szene gehört und wo es demnach zu finden sein wird (Henderson & Hollingworth, 1999).

Dobel, Gunnior, Bölte und Zwitserlood (2007) stellten Untersuchungen zur schnellen Verarbeitung des Gist in Szenen an. Sie gaben Probanden diverse maskierte Bilder vor, auf denen eine aktive Handlung stattfand wie ein Indianer, der mit Pfeil und Bogen auf ein Tier zielte. Diese Darstellungen hatten unterschiedlich lange Präsentationsdauer, zwischen 100 und 300 ms. Die Aufgabe der Teilnehmer bestand in der Wiedergabe des Inhalts, der abgebildeten Person und der Objekte. Darüber hinaus mussten sie entscheiden, ob die Szene eine sinnvolle (Indianer zielt auf Tier) oder eine sinnlose Handlung (Indianer zielt in die entgegengesetzte Richtung) beinhaltete. Die Ergebnisse von Dobel et al. zeigen aufgrund der hohen Anzahl an richtigen Antworten in Bezug auf die Identifizierung der Personen, dass diese vor allem im Zusammenhang mit der Stimmigkeit in Szenen geschieht. Infolgedessen vermuten sie eine hohe Empfänglichkeit für sinnvolle Szenen und ihrer Quintessenz.

Wird aus einer Szene nur der Gist verarbeitet, so werden meist die spezifischen Merkmale eines Bildes vernachlässigt und es kann zu Verwechslungen von Szenen

führen, die denselben Gist besitzen (Sampanes, Tseng & Bridgeman, 2008). Diese Verkennung wird auch „change blindness“ genannt. Darunter versteht man eine Fehlfunktion bei der Erfassung von auffälligen Veränderungen innerhalb einer Szenendarstellung, nachdem es eine kurze visuelle Störung gegeben hat. Simons, Franconeri und Reimer (2000) konnten anhand von zwei Experimenten darlegen, dass der Effekt des change blindness auch ohne eine visuelle Unterbrechung auftreten kann.

Bei der Betrachtung von Bildern offenbart sich der kognitive Faktor als wichtige Komponente. Loftus und Mackworth (1978) konnten mit ihrem Experiment zeigen, dass die kognitive Verarbeitung eine große Rolle spielt. Sie gaben Untersuchungsteilnehmern verschiedene Bilder vor, die sich in solche gliedern, die in dieser Form eher wahrscheinlich in der Wirklichkeit anzutreffen sind, und solche, bei denen dies eher unwahrscheinlich ist. Unter ersterem wurden Szenen verstanden, in denen nur Objekte abgebildet waren, die mit hoher Wahrscheinlichkeit in der Realität in dieser Kombination auftreten wie ein Traktor in einer Farmszene. Im Gegensatz dazu wurden Szenen vorgegeben, die vermutlich in dieser Zusammensetzung von Objekten nicht in der Wirklichkeit erscheinen wie ein Oktopus in einer Farmszene. Sie fanden heraus, dass Objekte, die offensichtlich nicht ins Geschehen passen, eher früher, öfter und länger fixiert werden als natürlich vorkommende, da sie offensichtlich nicht zum restlichen Bild passen. Die kognitive Komponente nimmt somit bei der Aufnahme von Informationen, als auch bei der späteren Wiedererkennung von Szenen einen wichtigen Platz ein.

*„Human gaze control is intelligent in that it draws not only on currently available visual input but also on cognitive knowledge structures, including short-term and episodic memory for the current scene; stored long-term visual, spatial, and semantic information about other similar scenes; and the goals and plans of the viewer.“*

(Henderson, 2007, S. 220)

Die Erfassung einer Szenendarstellung erfolgt durch schnelle Bewegungen mit den Augen, Sakkaden genannt (Hayhoe & Ballard, 2005). Bis zu drei Mal innerhalb einer Sekunde kann von einem Blickpunkt zum nächsten gewechselt werden. Informationen aus der vor einem befindlichen Szene werden vor allem von Fixationen zwischen den

Sprüngen aufgenommen (für einen kurzen Überblick siehe Henderson, 2007). Diese zwei Prozesse, Bewegung und Fixation, werden auf zwei unterschiedliche visuelle Vorgänge zurückgeführt, die jedoch nur gemeinsam zur Wiedererkennung von Szenen führen können (Pannasch, Helmert, Roth, Herbold & Walter, 2008).

Pannasch et al. (2008) wollten anhand von vier Experimenten erforschen, ob es einen Zusammenhang zwischen Fixationsdauer und Sakkadensprung bei der Betrachtung von natürlichen Stimuli gibt. Hierzu verglichen sie die Augenbewegungen, die in der Anfangsphase der Betrachtung gemacht wurden, mit denen, die in deren Endphase vorgekommen sind. Die Experimente unterschieden sich in den vorgegeben Szenen, welche differenzierenden Einfluss, wie emotionale Stimuli oder ähnliche Stimuli, ausüben sollten. Die Ergebnisse zeigten, dass im Laufe der Betrachtungszeit einer Szene die Dauer der Fixationen zunimmt und die Sakkadensprünge reduziert werden. Daraus schlossen sie, dass es eine wechselnde Balance der beiden Prozesse bei der Verarbeitung des visuellen Inputs gibt.

Trotz Fixation auf ein bestimmtes Objekt, wird auch die Umgebung des Gegenstandes erfasst. Aber wieviel nehmen wir davon wirklich wahr? Irwin und Zelinsky (2002) stellten sich diese Frage und überprüften sie mit einfachen Szenen, die nur wenige Objekte enthielten. Die Dauer der Betrachtungszeit war für jeden Teilnehmer vorgegeben, indem ihm nur eine gewisse Anzahl an Fixationen auf dem Bild erlaubt war. Waren die gestatteten Blickpunkte aufgebraucht, wurde das Bild ausgeblendet. Anschließend wurde überprüft, ob Gegenstände und ihre Position richtig wiedererkannt werden. Es interessierten die Autoren speziell jene Objekte, auf denen keine Fixation lag, aber aufgrund der Sakkadensprünge trotzdem wahrgenommen wurden. Die Resultate zeigten deutlich eine Beschränkung der Anzahl an Objekten, die nebenbei bemerkt werden, auf nur fünf Items.

## **2.5 Perspektivwechsel bei Objekten und Szenen**

*„Because most objects or situations are unlikely to recur in the exact form and context in which they are first experienced, it is important for human and nonhuman animals alike to have mental processes that can generalize from past to present experiences.“*

(Friedman, Waller, Thrash, Greenauer & Hodgson, 2011, S.229)

Eine zentrale Problemstellung bei der Objekt- und Szenenwahrnehmung stellt jene dar, welche nach einer Änderung des Blickpunktes auftritt. Objekte und Szenen erscheinen bei neuerlicher Betrachtung oftmals abweichend von dem vorherigen Bild, da man sie aus einem anderen Winkel ansieht. Dadurch tritt die Form der Gegenstände in veränderter Weise auf und überdies kann sich der Kontext der Darstellung ändern. Trotzdem haben wir nur geringe Schwierigkeiten sie als dieselben zu identifizieren. Aus diesem Grund ergibt sich die Frage, wie es gelingt, Objekte wiederzuerkennen, wenn sich die Perspektive auf sie verändert hat (Tarr, Bühlhoff, Zabinski & Blanz, 1997; Wang & Simons, 1999).

Kommt es zu einer Änderung der Perspektive auf einen Gegenstand, so kann dies auf zweierlei Arten geschehen. Zum einen kann das Objekt selbst rotieren und es entsteht eine Änderung der Orientierung des Gegenstandes (im weiteren Verlauf „Rotationsprobleme“ genannt). Zum anderen kann der Betrachter sich um ein Objekt oder in einer Szene bewegen und es kommt zu einer Änderung des Blickpunktes auf diese (im weiteren Verlauf „Perspektivproblem“ genannt) (Huttenlocher & Presson, 1973; Simons & Wang, 1998).

Huttenlocher und Presson (1973) untersuchten die zwei unterschiedlichen Aufgabenstellungen an Kindern und entdeckten, dass Beiden ein anderes Muster an Fehlern zugrunde liegt. Darüber hinaus gibt es größere Schwierigkeiten beim Zurechtkommen von Perspektivproblemen, als bei Rotationsproblemen. Des Weiteren fanden Simons und Wang (1998) durch ihre Experimente heraus, dass Rotationen von Objekten blickpunktabhängig sind, während Wiedererkennung aufgrund eines Blickpunktwechsels durch Bewegung der Person blickpunktunabhängig ist. Letzteres tritt besonders in dem Fall auf, wenn sich Personen in einer bekannten Umgebung neu orientieren müssen und es hierfür zweckdienliche Hinweise aus der alten Perspektive gibt. Hingegen kamen Motes, Finlay und Kozhevnikov (2006) mit ihren zwei Untersuchungen zu anderen Ergebnissen. Die Aufgabe der Probanden bestand darin, sich Szenen mit jeweils 11 oder 5 Objekten einzuprägen. Anschließend wurden ihnen aus verschiedenen Winkelperspektiven Szenen präsentiert, indem diese entweder rotiert wurden oder die Person um sie herum ging. Aufgrund der protokollierten Reaktionszeiten und genauen Beurteilung der gesehenen Szenendarstellungen kamen Motes et al. zu dem Schluss, dass es, unabhängig vom gesehenen Winkelabstand zur

gelernten Darstellung, keinen Vorteil gegenüber der Bedingung Perspektivsprung infolge der Bewegung der Personen gibt. Darüber hinaus folgerten sie, dass sich die mentale Repräsentation der gelernten Szene nicht gut an die neue Ansicht anpassen lässt. Daher gibt es Schwierigkeiten bei der Vorstellung wie die bekannte Szenenanordnung aus einer neuen Sicht erscheint.

Mit Hilfe von mentaler Rotation erfolgt ein Vergleich des vor einem liegenden Objektes mit bereits abgespeicherten Repräsentationen von anderen Objekten, bis eine Übereinstimmung von Objekten zustande kommt. Bereits Shepard und Metzler (1971) nahmen sich dem Problem der mentalen Repräsentation und Rotation an. Um letzteres überprüfen zu können, arbeiteten sie mit zweidimensionalen Darstellungen von dreidimensionalen Figuren. Durch Vorgabe von denselben, nur durch Rotation veränderten, oder unterschiedlichen Figuren, konnten Schlüsse über die mentale Rotation gezogen werden. Es zeigte sich, dass die Reaktionszeiten abhängig vom Rotationswinkel erfolgten. Je mehr zwei Figuren voneinander gedreht dargestellt waren, desto länger dauerte die Reaktionszeit. Dementsprechend gibt es einen linearen Zusammenhang zwischen dem Rotationswinkel der Objekte und der Zeitspanne, bis sie korrekt ihrem gelernten Blickwinkel zugeordnet werden können (Diwadkar & McNamara, 1997; Hayward, Zhou, Gauthier & Harris, 2006). Die mentale Transformation versucht immer den kürzesten Weg der Rotation zur gelernten Ansicht oder aber zu ihrer kanonischen Ansicht zu finden (Tarr & Pinker, 1989). Boer (1991) konnte die Ergebnisse der mentalen Rotation bestätigen und nahm, auf diesen aufbauend, an, dass die Erkenntnisse besonders bei Perspektivproblemen ihre Verwendung finden, wenn man sich seine Umgebung mental aus einem anderen Blickwinkel vorstellt.

Im Gegensatz zu der bisherigen Annahme, dass Objekte schneller wiedererkannt werden, wenn der Rotationswinkel gering ist, folgerten Niimi und Yokosawa (2008) aus ihren Versuchen, dass die Wiedererkennung abhängig von der dargestellten Richtung der Gegenstände ist. Die Vorderseite als auch die Rückseite von Objekten werden schneller erkannt als deren seitliche Repräsentation.

Die vorherrschende Vermutung im Bereich mentaler Rotation und rotierten Objekten besagt, dass Beiden dieselben Prozesse zugrunde liegen, da ihre Reaktionszeiten beim



Wiedererkennen nicht sehr voneinander differieren. Daraus folgerten viele Forscher, dass diese zwei Konstrukte miteinander in Verbindung stehen und Objekte mithilfe mentaler Rotation wiedererkannt werden (Hayward, Zhou, Gauthier & Harris, 2006). Jedoch zeigte sich auch, dass eine neue Perspektive einer Szene nicht nur durch Transformation zu bereits abgespeicherten Ansichten erklärt werden kann (Christou & Bühlhoff, 1997). Hayward et al. (2006) stellten sich die Frage, inwieweit mentale Rotation und Objekterkennung tatsächlich zusammenhängen und kamen zu dem Ergebnis, dass mentale Rotation linear erfolgt, aber Objektwiedererkennung nur bei kleinen Drehungen gelingt. Aufgrund ihrer Resultate und Erkenntnissen aus Gehirnstudien, vermuteten sie, dass Objektwiedererkennung eher auf einem Abgleich von Bildmerkmalen basiert als auf mentalen 3d-Transformationen.

Um seine Umgebung aus verschiedenen Blickwinkeln wiedererkennen zu können, plädieren einige Forscher für die Theorie der „view combination“ (Friedman & Waller, 2008; Friedman et al., 2011). Darunter versteht man, dass, durch Verbindung von bereits abgespeicherten Repräsentationen, Szenen aus neuen Perspektiven besser wiedererkannt werden, je mehr davon bereits abgespeichert sind. Friedman und Waller (2008) ließen Untersuchungsteilnehmer Szenen aus zwei verschiedenen Blickwinkeln lernen. Anschließend mussten sie diese als auch neue Perspektiven von den gelernten Szenen wiedererkennen. Die Ergebnisse zeigen, dass in Abhängigkeit der Winkeldifferenz der Bilder voneinander, neue Szenen gleich gut oder schlechter erkannt werden. Daher sind die Autoren der Ansicht, dass die Theorie der view combination als besseres Erklärungsmodell für neue Blickwinkel dient als jenes, der Transformation zum nächsten abgespeicherten Bild. Darüber hinaus stellt view combination einen allgemeinen und grundlegenden Mechanismus zur Wiedererkennung dar (Friedman et al., 2011).

Wie bereits in den vorherigen Kapiteln dargelegt gibt es im Bereich der Objektwahrnehmung zwei differente Modelle, dem strukturbasierten und dem ansichtsbasierten (für genauere Erläuterungen siehe Kapitel 2.2.1 und Kapitel 2.2.2). Diese unterscheiden sich besonders in der ungleichen Auffassung der Repräsentation von Objekten und, davon abhängig, ob sie aus jeder Perspektive gleich gut erkannt werden oder nicht. Foster und Gilson (2002) zeigten, dass beide Ansätze nahezu gleich

abhängig vom Blickpunkt sind und sie sich somit in Anbetracht eines Perspektivwechsels nicht voneinander unterscheiden. Daher plädieren sie für ein Zusammenwirken dieser zwei, voneinander unabhängigen, Prozesse. Diese Erkenntnis entspricht den Annahmen, die bereits einige andere Forscher aufgestellt haben (vergleiche hierfür Kapitel 2.2.3).

### 3 METHODE

---

In dem nachfolgenden Experiment wird der Einfluss von übereinstimmend angeschauten Objekten auf die Wiedererkennungslleistung nach Perspektivsprüngen und Spiegelbildern untersucht. Wobei in dieser Arbeit speziell auf die Wiedererkennung nach Änderung des Blickwinkels eingegangen wird. Zur genaueren Betrachtung der Ergebnisse für Bilder mit Spiegelungen sei auf die Diplomarbeit von Nina Glösel verwiesen.

Für die Untersuchung werden unterschiedliche Farbfotografien von natürlichen Szenen verwendet, welche sich die Teilnehmer in der Lernphase einprägen sollen. Anschließend sollen sie diese unter verschiedenen Bedingungen wiedererkennen: identisches Bild, Szenen mit Perspektivsprung, Szenen mit gespiegelten Ansichten und neue Bilder. Die Aufgabe der Untersuchungsteilnehmer in der zweiten Phase des Experiments besteht darin per Tastendruck anzugeben, ob sie das vorgegebene Bild bereits kennen oder nicht. Dabei gelten identische Szenen und solche mit Perspektivsprung als bereits gesehen, die anderen zwei Bedingungen werden als noch nicht gesehen klassifiziert.

In die Ergebnisse fließen sowohl die Antworten „gesehen“ und „nicht gesehen“ der Probanden mit ein, als auch die übereinstimmend betrachteten Objekte der beiden Phasen.

#### 3.1 Hypothesen

Daraus ergibt sich nun die Fragestellung, ob Szenen besser erkannt werden, wenn Objekte sowohl in der Lernphase als auch in der Wiedererkennungsphase betrachtet werden. Aus dieser Untersuchung lassen sich folgende drei Hypothesen ableiten:

H0<sup>1</sup>: Wiederholt angeschaute Objekte in der Bedingung „identische Bilder“ verbessern die Wiedererkennungslleistung nicht.

H1<sup>1</sup>: Wiederholt angeschaute Objekte in der Bedingung „identische Bilder“ verbessern die Wiedererkennungslleistung.

H0<sup>2</sup>: Wiederholt angeschaute Objekte in der Bedingung „Perspektivsprung“ verbessern die Wiedererkennungslleistung nicht.

H1<sup>2</sup>: Wiederholt angeschaute Objekte in der Bedingung „Perspektivsprung“ verbessern die Wiedererkennungslleistung.

H0<sup>3</sup>: Bei identischen Bildern gibt es keine bessere Wiedererkennungslleistung als bei Bildern mit Perspektivsprung.

H1<sup>3</sup>: Bei identischen Bildern gibt es eine bessere Wiedererkennungslleistung als bei Bildern mit Perspektivsprung.

### **3.2 Stichprobe**

An dieser Studie nahmen insgesamt 12 Psychologiestudenten der Universität Wien teil, wobei die Geschlechter ungleich verteilt waren. Ein Teilnehmer war ein Mann, die restlichen 11 waren Frauen. Das Durchschnittsalter der Probanden betrug 23 Jahre. Die Studenten konnten sich die Teilnahme an dieser Untersuchung für ihr Studium anrechnen lassen. Um an diesem Experiment teilnehmen zu können, mussten alle Studenten zwei Bedingungen erfüllen. Zunächst war es erforderlich, dass sie eine normale oder zumindest auf normal korrigierte Sehfähigkeit hatten. Weiters mussten sie eine schriftliche Einverständniserklärung unterzeichnen um bei dem Experiment mitwirken zu können.

### **3.3 Instrumente und Messgeräte**

Das Experiment wurde an einem Computer in der Fakultät für Psychologie im Testraum „TR-K7 - Eyetracking Labor“ durchgeführt. Die Programmierung des Experiments erfolgte mit Experiment Builder (SR Research, 2009), die Datenanalyse mit Data Viewer (SR Research, 2009). Für die Auswertung der Daten wurde das Statistikprogramm SPSS herangezogen. Die Stimuli wurden an einem 19'' (48.3 cm) CTR Farbmonitor (Sony Multiscan G400) mit einer Abtastfrequenz von 100 Hz und einer Bildschirmauflösung von 1024×768 (72 dpi) präsentiert.

Am Computer wurde ein sogenanntes Gamepad, ein Eingabegerät für beispielsweise Computerspiele, angeschlossen, mit welchem das Antwortverhalten der Versuchspersonen aufgenommen wurde. Durch das Drücken von nur zwei Tasten mit

dem Zeigefinger der rechten oder linken Hand konnten die Antworttypen „ja“ oder „nein“ aufgezeichnet werden. Weiters wurde das Gamepad dazu benutzt um nach den Instruktionen mit einer der beiden Phasen des Experiments zu beginnen, indem eine beliebige Taste zu betätigen war.

Zusätzlich musste für das Erforschen der Blickrichtung ein Eyetracker eingesetzt werden. Dieses EyeLink 1000 Desktop Mount eye-tracking System (SR Research, 2009) zeichnete mit einer Abtastfrequenz von 1000 Hz beidseitige Augenbewegungen auf. Hierfür war es notwendig, dass der Kopf bzw. die Augen die ganze Untersuchung hindurch eine gleich bleibende Entfernung zum Bildschirm einhielten. Daher wurden Stirn und Kinn der jeweiligen Probanden 57 cm vom Bildschirm entfernt auf einer Stütze fixiert.

### **3.4 Reizmaterial**

Die vorgegebenen Stimuli bestanden aus digitalen Farbfotografien von Innen- und Außenräumen, welche bei Tageslicht aufgenommen wurden. Die verwendeten Bilder lassen sich sechs Kategorien zuordnen: Straßen/Wege, Treppen, Innenräume, Häuserfassaden, Naturansichten und Parks (siehe Abbildung 5). Hierbei sollte das Verhältnis der verwendeten realen Szenen ausgewogen sein. Ebenfalls wurde darauf geachtet, dass es eine gleichmäßige Verteilung der Innenraumszenen gab, welche sich in Küchenszenen, Badezimmerszenen und Wohnzimmerszenen aufteilten. Um diese Ausgewogenheit beizubehalten, wurden jedem Teilnehmer in jeder Bedingung drei verschiedene Szenen aus jeder Kategorie vorgegeben. Jede Szene wurde aus zwei verschiedenen Blickwinkeln aufgenommen und eine Szene davon wurde auf dem Computer gespiegelt. Aufgrund dessen durften auf den Aufnahmen keine Buchstaben oder Wörter sichtbar sein, da sie dadurch Hinweise auf die Spiegelbedingung gegeben hätten. Auch sollten weder Menschen noch Tiere im Vordergrund abgebildet sein, da diese einerseits das Erkennen einer Szene vereinfacht hätten, andererseits bei der Wahrnehmung von Gesichtern andere Prozesse beteiligt sind (vergleiche hierfür Abschnitt 2.3). Wenn es aber nicht vermeidbar war, dann sollten die Personen höchstens als nicht identifizierbar im Hintergrund abgebildet sein. Die vorgegebenen Szenen sollten von durchschnittlicher Schwierigkeit für das Wiedererkennen sein. Aus diesem Grund sollten die Fotografien keine besonders auffälligen Objekte oder

Personen aufweisen, jedoch einige markante Punkte als Orientierungshilfen beinhalten, die sie von ähnlichen Szenen unterscheidbar machen.



Abbildung 5: Szenenbeispiele der sechs verwendeten Kategorien

Insgesamt gab es einen Bilderpool von 108 unterschiedlichen Fotografien, die sich in 36 Szenenfotos, 36 dazugehörige Fotografien mit Perspektivsprung und 36 Spiegelbilder aufteilen. Letztgenannte Bedingung setzt sich aus 18 Spiegelungen von den Szenenfotos und aus 18 Spiegelungen von Perspektivsprüngen zusammen. Das Reizmaterial wurde von Psychologiestudenten des Forschungspraktikums II unter der Leitung von Prof. Dr. Ulrich Ansorge im Wintersemester 2009/2010 selbst fotografiert und für die Untersuchung ausgewählt.

### 3.5 Design

Die Testung teilte sich in 2 Blöcke auf, einer Lernphase und einer daran anschließenden Wiedererkennungsphase, auch Transferphase genannt. Für jede Phase musste ein Bildset angefertigt werden. Das Lernset bestand aus 3×18 und das Wiedererkennungsset aus 4×18 farbigen Digitalfotos. Für die Hälfte der Probanden wurde aus dem gleichen Bilderpool ein zweites Reizset angefertigt. Hierfür mussten

Bilder vertauscht und in anderen Bedingungen eingesetzt werden. Die identischen Szenenbilder wurden mit den neuen Szenenbildern ausgetauscht, ebenso die Fotografien mit Perspektivsprung mit jenen der Spiegelbedingung. Somit gab es bei der Durchführung insgesamt vier Versuchsbedingungen: identisches Bild, Perspektivsprung, Spiegelung und neues Bild. In dieser Untersuchung gibt es zwei abhängige Variablen, welche miteinander korreliert werden. Zum einen die übereinstimmend angeschauten Objekte in beiden Versuchsbedingungen, zum anderen die Wiedererkennungslleistung dieser.

### **3.6 Durchführung**

Die Untersuchungen wurden in einem Testraum der Fakultät für Psychologie durchgeführt. Hierfür saßen die Teilnehmer alleine in einem abgedunkelten Raum mit schwacher Beleuchtung vor einem Bildschirm und mussten sich Fotografien einprägen und anschließend wiedererkennen. Aus dem Bilderpool wurden 2 unterschiedliche Sets an Fotografien zusammengestellt, welche den Teilnehmern zufällig zugeordnet wurden. Die Sitzung dauerte pro Teilnehmer etwa 30-40 Minuten und gliederte sich in zwei Blöcke.

Zu Beginn konnten sich die Probanden am Bildschirm die Instruktionen für das nachfolgende Experiment durchlesen. Weiters wurde ihnen vom Versuchsleiter erklärt, dass sich die Testung aus einer Lernphase und einer Wiedererkennungsphase zusammensetzt. Ihre Aufgabe bestehe darin die präsentierten Bilder aus der Lernphase im anschließenden Durchgang wiederzuerkennen: identische Bilder oder Bilder mit Perspektivsprüngen sind als gesehen zu beurteilen. Erscheinen die Bilder als Spiegelung oder sind es neue Bilder, so seien sie als noch nicht gesehen zu bewerten.

Der Lernblock wurde von der Testperson selbst eingeleitet, indem sie eine beliebige Taste des Gamepads drückte. Anschließend wurde jedes Bild für 3,000 ms präsentiert. Insgesamt wurden in der ersten Phase 63 Trials, bestehend aus 54 Durchgängen mit Fotografien und 9 Catch-Trials, vorgegeben. Catch-Trials waren graue Scheiben, die für 1,500 ms auf dem Bildschirm eingeblendet wurden. Sie sollten sicherstellen, dass die Aufmerksamkeit der Teilnehmer während des Versuchs aufrecht bleibt und nicht abdriftet. Daher sollten die Probanden darauf mit einem bestimmten Tastendruck reagieren. Die Anweisung, welche Taste zu drücken sei, variierte über die

Testpersonen und wurde für die Probanden zufällig ausgewählt. Nach jedem Durchgang erschien in der Mitte des leeren Bildschirms ein Fixierungskreuz, auf welches der Blick gerichtet werden musste. Dies war auch der Grund, warum die Verwendung des Eye-tracking Gerätes notwendig war.

Nach Beendigung der Lernphase konnten sich die Teilnehmer kurz erholen, bevor es mit einer erneuten Instruktion am Bildschirm mit der zweiten Phase der Testung weiterging. Nun bestand ihre Aufgabe darin zu beurteilen, ob sie die präsentierten Szenen bereits in der ersten Phase wahrgenommen haben oder nicht. Die jeweilige Antwort auf diese Frage musste innerhalb der 3,000 ms erfolgen, in der die Bilder präsentiert werden. Ein Teil der Probanden musste für die Antwort „ja“ die rechte Taste des Gamepads drücken, der andere Teil der Personen musste für dieselbe Antwort die linke Taste drücken. Daher variierte die Anweisung für die Antwort „nein“ ebenfalls zwischen den beiden Tasten. Zusätzlich wurden sie darauf hingewiesen, dass wieder Catch-Trials in manchen Bildern erscheinen werden und sie mit dem Drücken einer bestimmten Taste reagieren sollten. Die Anweisung ob sie mit der linken oder rechten Taste darauf antworten sollten, variierte zwischen den Teilnehmern.

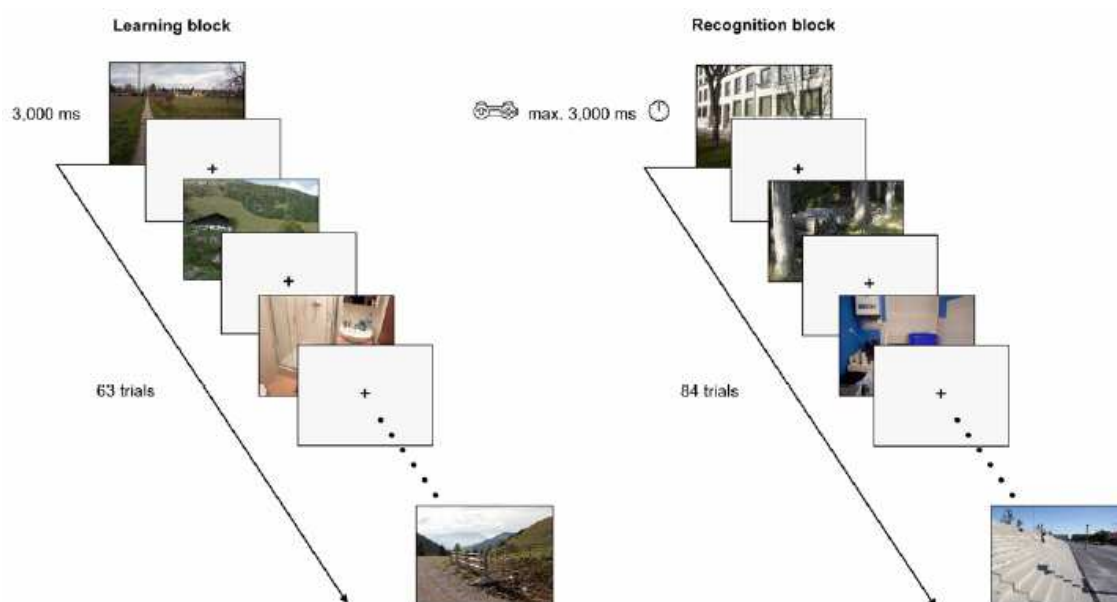


Abbildung 6: Schematische Darstellung der Reizabfolge der zwei Blöcke



Der Transferblock begann, wie bereits die erste Phase, mit dem Drücken einer beliebigen Taste des Probanden. Die Teilnehmer mussten über insgesamt 84 Trials urteilen, welche sich aus 72 Bildern und 12 Catch-Trials zusammensetzten. Die grauen Scheiben wurden zufällig zwischen den einzelnen Szenen eingefügt. Die Bilder setzten sich aus 18 wiederholten Szenen, 18 Fotografien von neuen Szenen, 18 Spiegelbildern von bekannten Szenen und 18 Fotografien von bekannten Szenen aus einer anderen Perspektive zusammen. Die Probanden bekamen während der Durchführung des Experiments kein Feedback über die Korrektheit ihrer Antworten.

### **3.7 Datenanalyse**

Zur Beantwortung der Forschungsfrage nach wiederholt angeschauten Objekten in den verschiedenen Bedingungen wurde ein Eyetracker eingesetzt. Dieser nahm die Blickbewegungen und einzelne Fixationen der Teilnehmer bei der Betrachtung des Bildmaterials auf. Für die Datenanalyse fanden die Fixationen ihre Verwendung, die Daten der Blickbewegungen waren jedoch nicht von Bedeutung. Die Auswertung der Blickpunkte erfolgte mit Data Viewer (SR Research, 2009).

Insgesamt wurden von allen 12 Probanden in der Lern- und Wiedererkennungsphase Daten aus 1296 vorgegeben Bildern gewonnen. Catch-Trials wurden bei der Analyse ausgeschlossen, da sie für die Aufrechterhaltung der Aufmerksamkeit während der Testung vorgegeben wurden und daher keine Aussagekraft über die Leistung bei der Wiedererkennung haben. Ebenfalls musste aus diversen Gründen auf einige Fixationen für die weitere Bearbeitung verzichtet werden. Aufgrund des Versuchsaufbaus mit dem Betrachten des Fixationskreuzes vor jeder Szene, wurden alle Fixationen unter 100 ms ausgeschlossen, da sie meist mittig anzutreffen waren. Weiters wurden Blickpunkte, welche außerhalb des Bildes lagen, in die Berechnungen nicht miteinbezogen. Einige Fixationen, deren Anblick nur von kurzer Dauer war und von nur einer Testperson vorgenommen wurde, waren von der weiteren Bearbeitung ebenso ausgenommen.

Mit Hilfe des Data Viewer konnten die einzelnen Bilder mit ihren dazugehörigen Fixationspunkten angezeigt werden. Die Bearbeitung jedes Szenenfotos erfolgte einzeln. Auf jedem Bild wurden die Fixationen einer Testperson präsentiert, die innerhalb der vorgegebenen Zeit von 3,000 ms getätigt wurden.

### *Markierung der Regions of Interest*

Die Bearbeitung des Bildmaterials wurde von 2 Bewertern vorgenommen, wobei die Daten hier jeweils zur Hälfte übernommen wurden. Ihre Aufgabe bestand darin, Regionen auf denen eine Fixation lag zu markieren und zu benennen. Diese Regionen sind die sogenannten „Interest Areas“ oder auch „Regions of Interest“ (ROI) genannt. Interest Areas stellen somit jene Bereiche dar, welche von den Probanden betrachtet wurde. Jeder Blickpunkt wurde herangezogen und analysiert, auf welchem Objekt er liegt. Daraufhin erfolgte eine Markierung dieses Objektes und es bekam eine entsprechende Benennung von den Bewertern. Die Markierung erfolgte mit Hilfe von Umrandungen der ROIs. Für die Umrandung standen drei mögliche Formen – Vierecke, Ellipsen oder von den Bewertern Freihand gezeichnet – zur Auswahl. Es sollte jene Form ausgesucht werden, welche das Objekt am besten wiedergab. Auch sollten alle Fixationen, die auf dem Objekt lagen, darin enthalten sein.

### *Label / Benennung der Regions of Interest*

Als nächster Schritt erfolgte die Benennung der Interest Areas. Es war wichtig, dass in jedem Bild die Bezeichnung einer bestimmten Region nur einmal pro Bild vorkam, außer es handelte sich um dasselbe Objekt. Die äquivalenten Szenen aus dem Lernblock und Transferblock mussten ebenfalls dieselben Benennungen der einzelnen Regionen erhalten um diese später miteinander vergleichen zu können.

Bei der Benennung der einzelnen Regionen musste auf einige Schwierigkeiten geachtet werden. Zunächst musste man auf den Unterschied zwischen Einzahl und Mehrzahl der Objekte aufpassen. Wurde nur eines betrachtet oder mehrere und hat diese Differenz Einfluss auf die Auswertung? Kam ein Objekt in einer Szene öfters vor, so musste die Bezeichnung gleich bleiben auch wenn es einmal alleine und an anderer Stelle in einer Gruppe erschien. Weiters musste man beim Benennen auf die Unterscheidung zwischen „dasselbe“ und „das gleiche“ Objekt achten. Dieselben Objekte erhielten auch denselben Namen. Waren die Merkmale zwischen den Objekten jedoch gravierend und die Objekte daher nur ähnlich, so musste der Name der betrachteten Region ein anderer sein. Um diese Problematik zu verdeutlichen sei kurz ein Beispiel angeführt. Stellt man sich einen Wohnraum vor, der sowohl einen Essplatz beinhaltet als auch einen Platz zum Arbeiten mit einem Computer bietet, so befinden sich in diesem Raum mehrere Stühle und zumindest zwei Tische. Es handelt sich

hierbei um gleiche Objekte – Tische sowie Stühle, jedoch nicht immer um dieselben. Einer der Tische wird zum Essen benutzt, auf dem anderen steht der Computer und weist daher auch andere spezifische Merkmale auf. Die Stühle beim Esstisch sind alle dieselben, aber der Stuhl beim Computertisch unterscheidet sich von ihnen. Ein weiteres Problem bei der Benennung stellten Fixationen dar, die genau zwischen mehreren Objekten angesiedelt waren. Es war nicht offensichtlich auf welchem der Objekte tatsächlich der Blick der Probanden gerichtet war. Deswegen war es nötig spezielle Bezeichnungen einzuführen, die diese Sonderfälle behandelten. Lag die Fixation genau auf der Linie zwischen einem Baum vor einem Haus, so wurde dieser Punkt als „BaumVorHaus“ gelabelt.

### *Bearbeitung der Regions of Interest*

Wie bereits weiter oben erwähnt, wurden die Bilddaten zwischen den beiden Bewertern aufgeteilt. Daher war es wichtig eine optimale Abstimmung bei der Bearbeitung des Materials zwischen den zwei Personen zu erreichen. Das wurde dadurch bewirkt, indem zu Beginn einige Fotografien gemeinsam durchgenommen wurden, und erst daran anschließend einzeln gearbeitet wurde. Bei der gemeinsamen Bearbeitung wurden mögliche Unklarheiten in Bezug auf die Markierung und Benennung beseitigt. Wie sollen die einzelnen Markierungen am besten gemacht werden? Wie groß ist die fixierte Region? Welche Fixationen gehören zu dem einen Objekt, welche schon zu einem anderen? Welche Form umschreibt das Objekt am schönsten? Wie benennt man die Interest Areas?

Da mit dieser Arbeit die Frage nach dem Einfluss von übereinstimmend angeschauten Objekten auf die Wiedererkennungslleistung behandelt wird, mussten die markierten Regionen der Lernphase und Wiedererkennungsphase gleich sein. Aus diesem Grund war es wichtig, nicht nur Regionen zu markieren auf denen eine Fixation lag, sondern auch bestimmte Regionen ohne einen Blickpunkt zu kennzeichnen. Daher wurden im Data Viewer aus jeder Bedingung dieselben Bilder als Gruppe geordnet und danach einzeln bearbeitet. Anschließend wurden die Markierungen von jedem Bild in der Gruppe auf die anderen übertragen, sodass schlussendlich auf jeder Fotografie mit demselben Motiv dieselben Markierungen zu sehen waren. Nach der Bearbeitung der Szenen aus der Lernphase wurden die entsprechenden Bilder der Transferphase herangezogen. Es mussten wieder alle Areas of Interest gefunden und gekennzeichnet

werden. Ebenso mussten die Markierungen über alle Bilder mit demselben Motiv gleich sein, auch wenn nicht immer eine Fixation darauf lag. Die Fragestellung erforderte darüber hinaus eine Übereinstimmung der ROIs der beiden Blöcke. Aus diesem Grund war es notwendig die Interest Areas der beiden Phasen miteinander abzugleichen und, wenn erforderlich, neue Markierungen hinzuzufügen.

### ***Bearbeitung der Bilder mit Perspektivsprung***

Das Wiedererkennen von Szenen wird oftmals durch den Wechsel des Blickpunktes erschwert. Dies stellt eine große Herausforderung dar, weil dieselbe Szene oftmals einen anderen, neuen Eindruck hinterlässt. Objekte ändern ihre Form, obwohl es noch immer dieselben sind. Durch den Perspektivwechsel werden Teile von Objekten verdeckt oder verschwinden sogar ganz aus dem Blickfeld. Dafür treten neue Objekte in die Szenendarstellung. Daher wirken manche Bilder nach einer Änderung des Blickpunktes als noch nie gesehen, obwohl sich die Umgebung nur teilweise ändert. Aufgrund dieser Merkmale eines Perspektivsprungs traten bei der Bearbeitung der Daten Schwierigkeiten auf. Objekte die vollkommen aus der Szene verschwunden waren, konnten in der Wiedererkennungsphase nicht markiert werden. Jedoch musste für neue Objekte, auf denen eine Fixation lag, ein Label gefunden werden. Handelte es sich um ein anderes, jedoch selbes Objekt, bekam es denselben Namen wie in der Lernphase. Waren Objekte hinreichend ähnlich zu Objekten aus der ersten Bedingung, so erhielten sie ihre alte Bezeichnung. Traten neue Objekte ins Bild, welche komplett neue Merkmale aufwiesen, erhielten sie ein neues Label.

### ***Bearbeitung der neuen Bilder***

Probleme anderer Art traten bei der Bearbeitung der neuen Bilder in der Wiedererkennungsphase auf, da es für diese Daten keine bestimmten Vergleichsszenen aus der Lernphase gab. Daher wurden die Bilder aller Lernbedingungen herangezogen und jene ausgewählt, welche am ähnlichsten waren. Für mehr Informationen zu der Bearbeitung der Spiegelbilder und den neuen Bildern, sowie deren Ergebnissen sei auf die Diplomarbeit „Der Einfluss von übereinstimmend angeschauten Objekten auf die Wiedererkennungsleistung von natürlichen Szenen nach Spiegelung“ von Nina Glösel hingewiesen.

## 4 ERGEBNISSE

---

Bereits während der Datenanalyse kam es zum Ausschluss von Daten. Wie im vorherigen Abschnitt genauer besprochen, wurden alle Durchgänge mit Catch-Trials ausgesondert, sowie von einigen Fotografien Fixationen nicht miteinbezogen. Ansonsten musste bei den Berechnungen auf keine weiteren Daten verzichtet werden. Bei der Ergebnisdarstellung wird auf nur zwei Versuchsbedingungen näher eingegangen, den wiederholten Bildern und Bildern mit Perspektivsprung.

Die Daten wurden aus dem Data Viewer in das Statistikprogramm SPSS importiert um anschließend die Berechnungen durchzuführen. Hierbei wurde mit Varianzanalysen (ANOVA) und mit bivariaten Korrelationen gearbeitet. Letztere wurden vor allem verwendet um einen möglichen Zusammenhang von übereinstimmend angeschauten Objekten der verschiedenen Bedingungen zu erhalten. Das Signifikanzniveau (Irrtumswahrscheinlichkeit) betrug bei allen Auswertungen  $\alpha = .05$ .

### Urteilsgenauigkeit

Für die Berechnung der Urteilsgenauigkeit wurden neben den Catch-Trials auch die Lerndurchgänge nicht miteinbezogen, da die Untersuchungsteilnehmer erst in der zweiten Phase des Experiments Urteile über das Wiedererkennen von Szenen abgeben mussten.

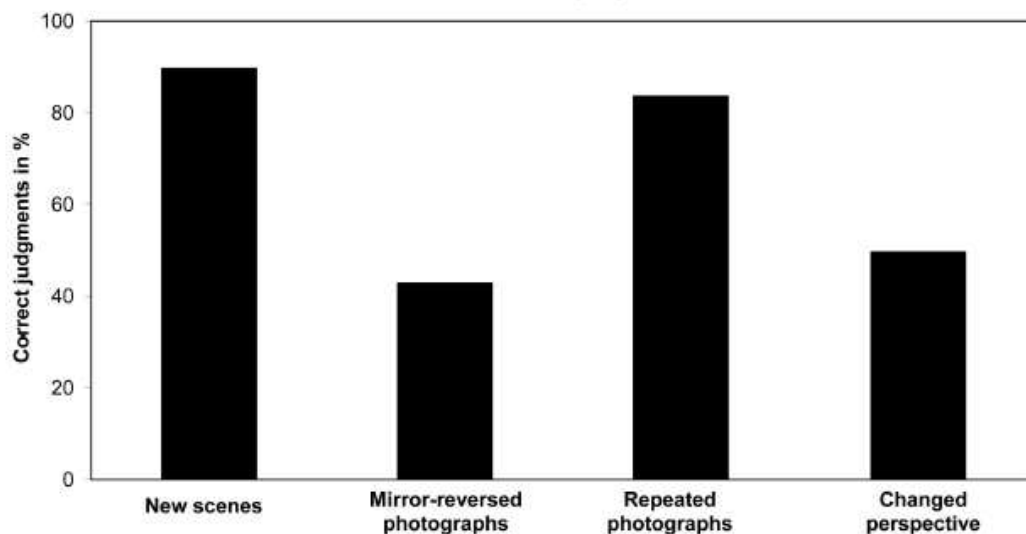


Abbildung 7: Rate der richtigen Urteile in den vier Versuchsbedingungen

Die Effekte der unterschiedlichen Versuchsbedingungen auf das Wiedererkennen von Szenen wurden mit einer univariaten ANOVA untersucht. Abbildung 7 zeigt, dass der Prozentsatz an korrekten Antworten in den Bedingungen wiederholte (83,6%) und neue (89,7%) Szenen höher war als bei gespiegelten Fotografien (42,8%) oder Bildern mit Perspektivsprung (49,5%). Die Varianzanalyse zeigte, dass der Bildtyp einen signifikanten Effekt auf die Richtigkeit der Urteile hat,  $F(3, 44) = 16.96$ ,  $p < .001$  mit einer Effektgröße  $\eta^2 = .536$ .

In dem Experiment hatten die Teilnehmer 3,000 ms Zeit um ein Urteil über die dargestellte Szene abzugeben. Diese Dauer wurde manches Mal überschritten. Daher wurde eine zweite ANOVA mit dem Prozentsatz der fehlenden Urteile durchgeführt. Der Prozentsatz fehlender Urteile schwankt je nach Bedingung zwischen 4% und 8,2%. Bei keiner Versuchsbedingung konnte ein signifikanter Effekt nachgewiesen werden,  $F(3, 44) = .66$ , n.s.,  $\eta^2 = .043$ .

### Einfluss der Fixationen auf die Urteilsrichtigkeit

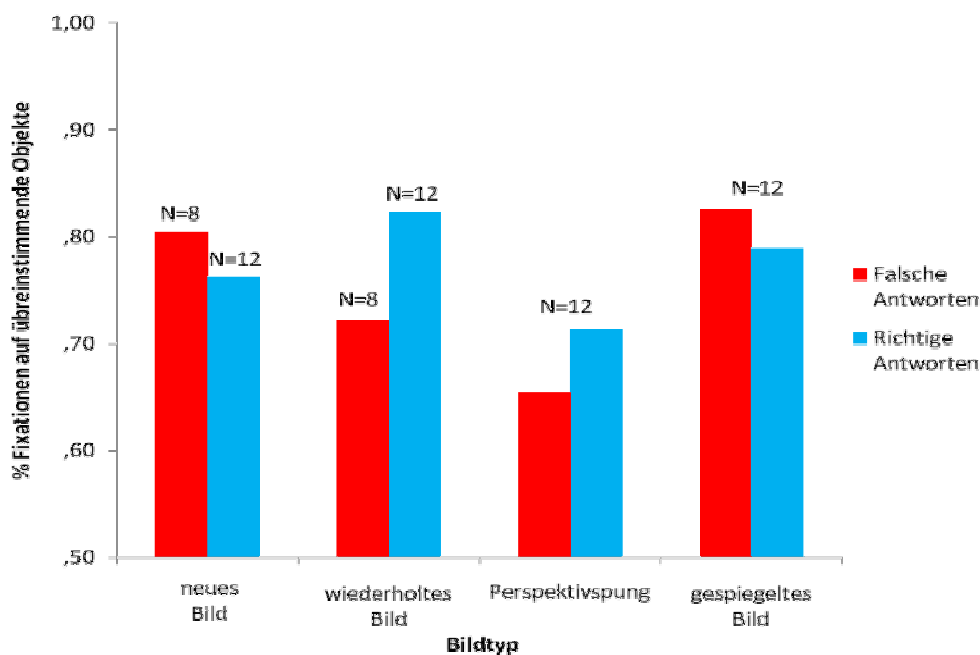


Abbildung 8: Prozentsatz Fixationen (relativ zu allen Fixationen) auf übereinstimmenden Objekten in Lern- und Transferbildern als Funktion der Richtigkeit der Antwort. (N = 8, da 4 Teilnehmer alle neuen und wiederholten Bilder richtig erkannt haben)

Um zu untersuchen, ob die Blickpunkte, die auf den übereinstimmenden Objekten der beiden Phasen gerichtet waren, die Richtigkeit der Antwort beeinflusst, wurde eine Pearson-Korrelation mit den Mittelwerten über alle teilnehmenden Personen ausgeführt. Vier der zwölf Probanden wurden bei den Berechnungen der falschen Antworten in den Bedingungen neues Bild und wiederholtes Bild nicht miteinbezogen, da sie in diesen nur korrekte Antworten abgegeben haben.

In den Versuchsbedingungen „wiederholtes Bild“ und „Perspektivsprung“ sind richtig beurteilte Bilder auch mit höheren Prozentraten wiederholt betrachteter Objekte verbunden. Die Wiedererkennungslleistung verbessert sich in diesen zwei Bedingungen mit der Betrachtung derselben Objekte. Bei neuen und gespiegelten Szenen hingegen korrelieren falsche Urteile mit einem höheren Prozentsatz an Fixationen.

#### **Berechnung zu den Bedingungen „Wiederholtes Bild“ und „Perspektivsprung“**

Die Ergebnisse zu diesen beiden Versuchsbedingungen wurden mit einer Pearson-Korrelation der individuellen Leistungen ermittelt.

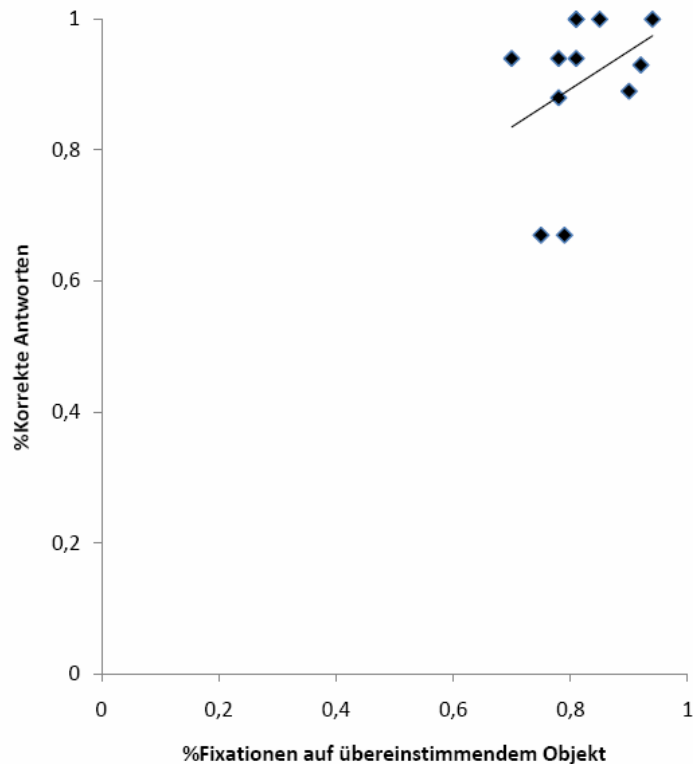


Abbildung 9: Prozentsatz korrekter Antworten in Abhängigkeit des Prozentsatzes von Fixationen auf übereinstimmendem Objekt in der Bedingung „Identische Bilder“

Die bivariate Korrelation ergibt einen positiven Zusammenhang zwischen dem Prozentsatz korrekter Antworten und dem Prozentsatz wiederholt angeschauter Objekte von  $r(12) = .34$ , n.s. Dieses Ergebnis ist nicht signifikant.

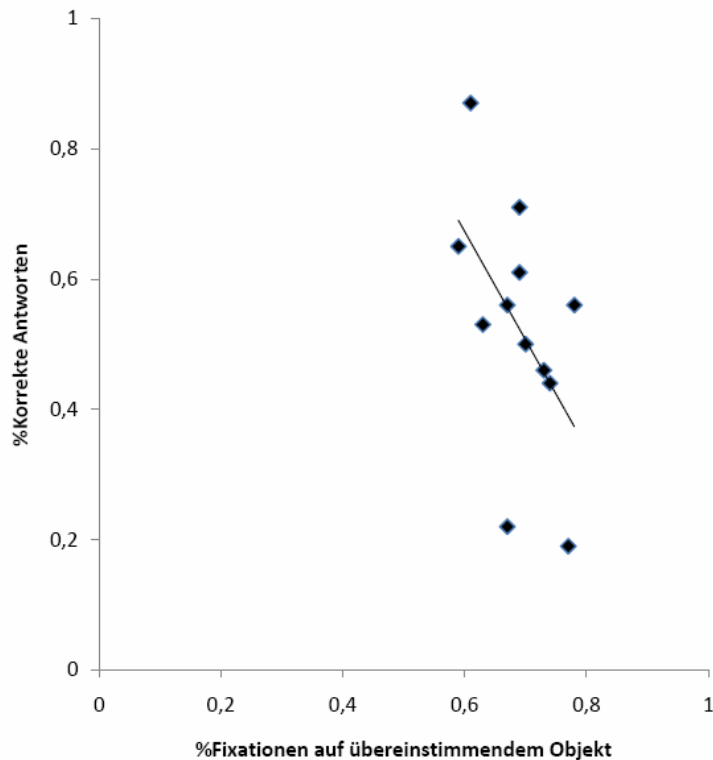


Abbildung 10: Prozentsatz korrekter Antworten in Abhängigkeit des Prozentsatzes von Fixationen auf übereinstimmendem Objekt in der Bedingung „Perspektivsprung“

Die Berechnung zeigt einen negativen Zusammenhang der Procente korrekter Antworten mit den Prozenten wiederholt angeschauter Objekte,  $r(12) = -.50$ ,  $p < .05$  (einseitig). Diese Korrelation ergibt einen signifikanten Effekt. Personen, die oftmals in beiden Bedingungen dieselben Objekte betrachtet haben, erzielten beim Wiedererkennen besonders schlechte Leistungen.



## 5 DISKUSSION

---

Das Ziel der vorliegenden Studie war die Untersuchung der Wiedererkennungsleistung von Szenendarstellungen anhand von Objekten speziell wenn eine Änderung des Blickpunktes vorlag. Dieser Forschungsfrage liegt die Annahme zugrunde, dass Orte besser wiedererkannt werden, wenn die darin enthaltenen Objekte sowohl bei der ersten Betrachtung der Szene als auch bei der Aufgabenstellung der neuerlichen Erkennung angeschaut werden. Aus diesem Grund liegt das Hauptaugenmerk der Überprüfung der Wiedererkennung auf den wiederholt betrachteten Objekten, welche mit Hilfe der Eye-tracking Methode umgesetzt wurde. Um den Effekt der Wiedererkennung zu erzeugen wurde im Experiment mit zwei Phasen gearbeitet, einer Lernphase und einer anschließenden Transferphase. Zunächst sollten sich Probanden Bilder einprägen, die sie anschließend unter verschiedenen Bedingungen – identisches Bild, Spiegelbilder, Bilder mit Perspektivsprung und neue Bilder – korrekt als „gesehen“ oder „nicht gesehen“ identifizieren sollten. Wobei zur Auswertung die Daten der fixierten Objekte herangezogen wurden. Als Wiedererkennungsleistung wird im folgenden Teil die Korrektheit der Antworten der Teilnehmer interpretiert. Nachfolgend soll auf die Ergebnisse des Experiments näher eingegangen und darüber hinaus diskutiert und miteinander in Verbindung gebracht werden.

Zunächst wurde überprüft, ob die abgegebenen Urteile „gesehen“ und „nicht gesehen“ dem Bildmaterial richtig zugeordnet werden konnten. Es interessierte vor allem die Frage nach dem Einfluss der vier Bedingungen auf das korrekte Antwortverhalten der Probanden. Die Ergebnisse zeigen einen signifikanten Effekt bei der Richtigkeit der gegebenen Urteile. Dies bedeutet, dass die Korrektheit der gegebenen Antworten von der gezeigten Bedingung abhängt. Es fiel den Probanden leicht identische und neue Szenen richtig zuzuordnen, jedoch hatten sie große Probleme bei der Unterscheidung der Bilder mit Perspektivsprung und Spiegelung. In diesen Bedingungen erfolgte die richtige Identifizierung nur bei ungefähr der Hälfte aller Bilder. Das Ergebnis offenbart somit, dass geänderte Szenenansichten der genauen Erkennung Schwierigkeiten bereiten.

Ein wichtiger Faktor dieser Untersuchung stellten die wiederholt betrachteten Objekte in beiden durchgeführten Phasen dar. Aus diesem Grund war eine Überprüfung des Einflusses der auf den Objekten liegenden Fixationen auf die Richtigkeit der Urteile notwendig. Die Ergebnisse der Mittelwerte über alle Probanden zeigten hierbei einen deutlichen Effekt zwischen den verschiedenen Bedingungen. Während bei wiederholten Bildern und Szenen mit Perspektivsprung die richtigen Urteile mit einem höheren Prozentsatz von Fixationen auf den übereinstimmend betrachteten Objekte korrelieren, ist dies in den zwei anderen Bedingungen der gegenteilige Fall. Sie zeigen einen Zusammenhang der Blickpunkte auf den Objekten mit falschen Urteilen. Interpretiert man die richtigen Antworten als Wiedererkennungslleistung und den Prozentsatz an Fixationen auf den übereinstimmend angeschauten Objekten als Objekte, lässt sich anhand dieser Ergebnisse folgendes darlegen: die Wiedererkennungslleistung von Objekten verbessert sich in den zuerst genannten Bedingungen, dagegen verschlechtert sie sich unter neuen und gespiegelten Bildern.

Eine weitere Annahme in der Untersuchung bezog sich auf den kennzeichnenden Teil dieser Arbeit, der Änderung der Perspektive in Szenen. Es wurde die Vermutung zugrunde gelegt, dass identische Bilder eine bessere Leistung im Erkennen von bereits bekannten Darstellungen erzielen als Bilder, die einen Blickpunktwechsel beinhalten. Um diese These zu überprüfen, wurden in der zweiten Phase des Experiments diverse Bedingungen eingeführt. Im Folgenden werden die individuellen Leistungen der identischen Bilder und Bilder mit Perspektivsprung näher erläutert. Die Ergebnisse bei wiederholten Szenen zeigen eine leichte Tendenz in Richtung bessere Leistung bei wiederholt angeschauten Objekten ( $r(12) = .34$ ). Dieser positive Zusammenhang war jedoch nicht signifikant. Daraus lässt sich schlussfolgern, dass bei den individuellen Leistungen die Betrachtung derselben Objekte keinen Einfluss auf die Wiedererkennung der Szene unter Berücksichtigung der Irrtumswahrscheinlichkeit hat. Differente Resultate bekommt man bei der Analyse des Bildmaterials mit Perspektivsprung. Die erwarteten Ergebnisse einer positiven Verbindung von wiederholt angeschauten Objekten mit richtigen Urteilen konnte nicht bestätigt werden ( $r(12) = -.50$ ,  $p < .05$ , einseitig). Die gefundenen Resultate sprechen eher für den gegenteiligen Effekt. Je öfter eine Person in beiden Phasen der Untersuchung ein

Objekt fixiert hatte, desto fehlerhafter war das abgegebene Urteil und somit ihre Wiedererkennungsleistung in dieser Bedingung.

Beachtet man nur die ersteren Ergebnisse, die Mittelwerte über alle Probanden, dieser Untersuchung, so finden unsere Hypothesen darin ihre Bestätigung. Neuerliche Betrachtung derselben Objekte verbessert die Wiedererkennungsleistung in beiden Bedingungen. Darüber hinaus werden bessere Leistungen bei den identischen Bildern erzielt als bei Bildern mit Perspektivwechsel.

Werden die Ergebnisse jedoch in Hinsicht auf die individuellen Leistungen genauer überprüft, so finden sie darin keine Bestätigung. Mehr noch, sie zeigen entweder keinen (Bedingung „identisches Bild“) oder einen negativen Effekt (Bedingung „Perspektivsprung“). Die Wiedererkennung funktionierte bei Personen, die übereinstimmende Objekte in beiden Phasen betrachteten, äußerst schlecht. Diese Leistung könnte zustande gekommen sein, wenn Szenen besonders schlecht eingepägt wurden. Wäre dies der Fall gewesen, so hätte es eine große Unsicherheit bezüglich der Beurteilung der Szenen als gesehen oder nicht gesehen gegeben. Personen hätten aus diesem Grund dieselben Objekte häufiger angeschaut, aber nur mit niedriger oder mittlerer Korrektheit der Antworten. Wird angenommen, dass diese Strategie – dieselben Objekte öfters zu betrachten um Szenen wiedererkennen zu können – notwendig ist, damit Szenen mit Perspektivsprung als bereits gesehen identifiziert werden können, so würden bereits einige wenige korrekte Antworten, die nicht auf Zufall beruhen, genügen um sie zu bestätigen. Dies würde auch erklären, warum die Mittelwerte insgesamt bessere Leistungen aufgrund von übereinstimmend betrachteten Objekten vorhersagen. Daraus lässt sich schließen, dass sich die Wiedererkennung nach einer Änderung des Blickwinkels in dieser Untersuchung als schwierig herausstellte, zudem die Ergebnisse der Mittelwerte nur knapp über der Zufallsgrenze lagen.

Aus diesen Ergebnissen lässt sich schließen, dass übereinstimmend angeschaute Objekte die Wiedererkennungsleistung allgemein positiv beeinflussen, irrelevant ob sie identisch sind oder aus einer anderen Perspektive erscheinen. Allerdings gibt es dahingehend Unterschiede, inwieweit die Beeinflussung in den zwei Bedingungen erfolgt. Identische Szenen weisen gute Resultate auf, hingegen gibt es in der Bedingung „Perspektivsprung“ ambivalente Resultate. Diese sprechen daher für die

Annahme der blickpunktabhängigen Theorie (Edelman & Bühlhoff, 1992; Lawson & Humphreys, 1996; Tarr & Bühlhoff, 1998). Diese besagt, dass Szenendarstellungen aus neuen Ansichten schlechter wiedererkannt werden, da die veränderte Repräsentation noch nicht abgespeichert ist.

Bereits aus der Planung und Auswertung dieser Untersuchung ergeben sich mehrere Schwierigkeiten, die im Folgenden näher erläutert werden.

Das erste Problem stellte die Anzahl der Antwortmöglichkeiten bei der Beurteilung der Bilder dar, da es für vier Bedingungen nur zwei Antworten, gesehen oder nicht gesehen, gegeben hat. Dies lässt aber keine Chance für die Unterscheidung innerhalb dieser Antwortkategorien zu. Man weiß nicht, wenn Probanden eine Szene mit Perspektivsprung als bereits gesehen beurteilt haben, ob sie diese tatsächlich als eine Szene mit Blickpunktwechsel identifiziert haben oder, ob sie diese Szene als identische Szene angesehen haben und deswegen ihre Antwort „gesehen“ lautete.

Eine weitere Schwierigkeit bei der Planung ergibt sich aus den Bildern, die einen Perspektivwechsel aufweisen. Die Literatur zeigte bereits, dass die Winkeldifferenz von Objekten bei der Wiedererkennung einen wichtigen Faktor darstellt (Friedman & Waller, 2008; Tarr & Pinker, 1989). Dies wurde bei der Auswahl der visuellen Stimuli nicht berücksichtigt. Objekte werden besser wiedererkannt, wenn der Grad zwischen gelernter und präsentierter Ansicht möglichst gering ist. Jedoch gibt es auch bestimmte Blickrichtungen auf Gegenstände, die deren Erkennung erleichtern wie die kanonische Ansicht (Palmer, Rosch & Chase, 1981; Tarr & Pinker, 1989). Objekte werden aus einem bestimmten Blickwinkel besser erkannt als aus anderen. Das verwendete Stimulusmaterial wurde aus zwei verschiedenen Perspektiven fotografiert, die allerdings in keinem bestimmten Winkel voneinander erfolgten. Weiters wurden manche vorgegebenen Bilder von einem eher ungewohnten Blickpunkt fotografiert, was die Wiedererkennung dieser Szenen eventuell erschwert hat.

Nach Änderung der Perspektive gehen vielerlei Informationen aus einem Bild verloren. Objekte verschwinden ganz oder teilweise aus Szenen, erscheinen aus einem anderen Blickwinkel und verändern somit das ganze Erscheinungsbild. Bei den verwendeten Fotografien wurde nicht darauf geachtet, wieviele Objekte nach einem Perspektivsprung in der neuen Darstellung noch zu erkennen waren, und welchen Einfluss die veränderte Anzahl an Objekten ausübt.

Die Auswertung der Daten erfolgte durch insgesamt zwei Beurteiler. Zwar wurde versucht die Beiden möglichst gut aufeinander abzustimmen, damit die Bilder annähernd identisch bearbeitet werden konnten, trotzdem treten Abweichungen in der Bearbeitung des Materials auf wie bei der Begrenzung von Markierungen und Benennungen. Fasst man die Objekte in einem Oberbegriff (z.B. Kosmetikartikel) zusammen oder markiert sie einzeln (z.B. Zahnbürste, Duschgel, Lotion)? Da die Ergebnisse sehr knapp ausgefallen sind, könnte man sich die einzelnen Ausarbeitungen genauer ansehen, ob die kleinen Unterschiede bei der Bearbeitung Einfluss auf die Ergebnisse ausüben.

Bei der Untersuchung wurde vor allem auf die Fixationen auf den einzelnen Fotografien eingegangen. Diese wurden als wichtiger Faktor für die Wiedererkennung angesehen. Lag eine Fixation auf einem Objekt, so wurde dieses als gesehen deklariert. Gegenstände, auf denen sich keinerlei Blickpunkte befanden, wurden als nicht gesehen aufgefasst und hatten daher in dieser Studie keinerlei Einfluss auf die Wiedererkennungsleistung. Irwin und Zelinsky (2002) stellten aber fest, dass durch Sakkadensprünge viel an Informationen aufgenommen wird. Sie tragen somit auch zur Wiedererkennung bei. Pannasch et al. (2008) sind darüber hinaus der Ansicht, dass es ein Zusammenspiel von Fixationen und Sakkadensprüngen bei dem Prozess der Erkennung gibt.

Zusammenfassend stellt diese Studie einen weiteren Beleg für die blickpunktabhängige Theorie bei Objekt- und Szenenwiedererkennung dar. Nach einem Wechsel der Perspektive auf dieselbe Szene werden schlechtere Leistungen erzielt als bei identischen. Als wichtige Funktion wurden in dieser Untersuchung die Fixationen auf den Objekten angesehen. Während wiederholte Szenen leichter wiedererkannt werden, wenn dieselben Objekte in beiden Phasen fixiert werden, stellt sich dies bei Szenen nach einem Perspektivsprung als schwierig dar. Trotzdem war die Wiedererkennung anhand von fixierten Objekten etwas besser als es nach Zufall zu erwarten gewesen wäre. Diese Ergebnisse erfordern weitere Untersuchungen speziell im Hinblick auf die Winkeldifferenzen zwischen zwei Szenen derselben Umgebung. Ebenso ist der Einfluss von Sakkadensprüngen noch nicht geklärt oder die Anzahl an Objekten, die für eine korrekte Wiedererkennung erforderlich sind.

## 6 ZUSAMMENFASSUNG

---

Woran können Szenen wiedererkannt werden, wenn ein Blickpunktwechsel stattgefunden hat? Welchen Einfluss üben hierbei die betrachteten Objekte aus? Diese Fragestellung wurde in der vorliegenden Diplomarbeit mithilfe des Eyetracking - Verfahrens überprüft. Zu diesem Zweck wurden diverse Farbfotografien in zwei Untersuchungsblöcken vorgegeben. Zuerst erfolgte eine Lernphase zum Einprägen der angezeigten Bilder, die anschließend in einer Wiedererkennungsphase abgeprüft wurden. Allerdings wurde das Erkennen erschwert, indem nur ein Teil der Fotografien identisch erschien. Die restlichen Bilder wurden aus einer anderen Perspektive präsentiert. Die Auswertung der Ergebnisse erfolgte anhand der Objekte, die in beiden Blöcken von den Augen fixiert wurden. Die Resultate der zwei Bedingungen zeigen eine ambivalente Beeinflussung der wiederholt betrachteten Objekte auf die Wiedererkennungsleistung. Während identische Szenen besser erkannt werden, weisen Szenen mit einem Perspektivwechsel einen negativen Effekt auf. Speziell Personen, die häufiger dieselben Objekte fixiert haben, erzielten bei der Wiedererkennung schlechtere Leistungen.

*Schlüsselwörter:* Szenenwiedererkennung, Objektwiedererkennung, Fixationen auf Objekten, Perspektivsprung

## LITERATURVERZEICHNIS

---

- Anderson, J. R. (2001). *Kognitive Psychologie* (3. Auflage). Heidelberg, Berlin: Spektrum Akademischer Verlag.
- Barenholtz, E. & Tarr, M. J. (2006). Reconsidering the role of structure in vision. *Psychology of Learning and Motivation*, 47, 157-180.
- Barton, J. J., Press, D. Z., Keenan, J. P. & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, 58, 71-78.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147.
- Biederman, I. (2001). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision*, 13, 241-253.
- Biederman, I. & Bar, M. (1999). One-shift viewpoint invariance in matching novel objects. *Vision Research*, 39, 2885-2899.
- Biederman, I. & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 1162-1182.
- Biederman, I. & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: reply to Tarr and Bülthoff (1995). *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1506-1514.
- Biederman, I., Glass, A. L. & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology*, 97, 22-27.

- Birbaumer, N. & Schmidt, R. F. (2003). *Biologische Psychologie* (5. Auflage). Berlin, Heidelberg, New York: Springer-Verlag.
- Boer, L. C. (1991). Mental rotation in perspective problems. *Acta Psychologica*, 76, 1-9.
- Bruce, V. & Humphreys, G. W. (1994). Recognizing objects and faces. *Visual Cognition*, 1, 141-180.
- Bruce, V. & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305-327.
- Burton, A. M., Young, A. W., Bruce, V., Johnston, R. A. & Ellis, A. W. (1991). Understanding covert recognition. *Cognition*, 39, 129-166.
- Christou, C. G. & Bühlhoff, H. H. (1997). *View-direction specificity in scene recognition after active and passive learning* (Technical Report No. 53). Tübingen, Germany: Max-Planck-Institut für biologische Kybernetik.
- Diwadkar, V. A. & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, 8, 302-307.
- Dobel, C., Gummior, H., Bölte, J. & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica*, 125, 129-143.
- Duchaine, B. & Nakayama, K. (2005). Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience*, 17, 249-261.
- Edelman, S. & Bühlhoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, 32, 2385-2400.
- Farah, M. J., Wilson, K. D., Drain, M. & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, 105, 482-498.



- Foster, D. H. & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 269, 1939-1947.
- Friedman, A. & Waller, D. (2008). View combination in scene recognition. *Memory & Cognition*, 36, 467-478.
- Friedman, A., Waller, D., Thrash, T., Greenauer, N. & Hodgson, E. (2011). View combination: A generalization mechanism for visual recognition. *Cognition*, 119, 229-241.
- Goldstein, E. B. (2002). *Wahrnehmungspsychologie*. Aus dem Amerikanischen übersetzt von Gabriele Herbst und Manfred Ritter (2. deutsche Auflage herausgegeben von Manfred Ritter). Heidelberg, Berlin: Spektrum Akademischer Verlag.
- Greene, M. R. & Oliva, A. (2009). The briefest of glances. The time course of natural scene understanding. *Psychological Science*, 20, 464-472.
- Gregory, R. L. & Wallace, J. G. (1963/2001). Recovery from early blindness: A case study. *Experimental Psychology Society Monograph*, 2, 1-44. Online verfügbar unter: [http://www.richardgregory.org/papers/recovery\\_blind/contents.htm](http://www.richardgregory.org/papers/recovery_blind/contents.htm) (Stand 11.07.2011)
- Gross, R., Shi, J. & Cohn, J. (2001). Quo vadis face recognition? - The current state of the art in face recognition. *Forschungsbericht*. Pennsylvania, Pittsburgh: Carnegie Mellon University, Robotics Institute.
- Hayhoe, M. & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9, 188-194.
- Hayward, W. G. & Tarr, M. J. (2000). Differing views on views: comments on Biederman and Bar (1999). *Vision Research*, 40, 3895-3899.

- Hayward, W. G., Zhou, G., Gauthier, I. & Harris, I. M. (2006). Dissociating viewpoint costs in mental rotation and object recognition. *Psychonomic Bulletin & Review*, 13, 820-825.
- Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, 16, 219-222.
- Henderson, J. M. & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, 50, 243-271.
- Huttenlocher, J. & Presson, C. C. (1973). Mental rotation and the perspective problem. *Cognitive Psychology*, 4, 277-299.
- Irwin, D. E. & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics*, 64, 882-895.
- Lawson, R. & Humphreys, G. W. (1996). View-specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 395-416.
- Lawson, R. & Humphreys, G. W. (1998). View-specific effects of depth rotation and foreshortening on the initial recognition and priming of familiar objects. *Perception & Psychophysics*, 60, 1052-1066.
- Leder, H. & Bruce, V. (2000). When inverted faces are recognised: the role of configural information in face recognition. *Quarterly Journal of Experimental Psychology: Section A: Human Experimental Psychology*, 53, 513-536.
- Loftus, G. R. & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 565-572.

- Marr, D. (1982). *Vision. A computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman.
- Marr, D. & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 200, 269-294.
- Metzger, W. (1975). *Gesetze des Sehens* (3. Auflage). Frankfurt am Main: Waldemar Kramer.
- Motes, M. A., Finlay, C. A. & Kozhevnikov, M. (2006). Scene recognition following locomotion around a scene. *Perception*, 35, 1507-1520.
- Niimi, R. & Yokosawa, K. (2008). Determining the orientation of depth-rotated familiar objects. *Psychonomic Bulletin & Review*, 15, 208-214.
- Oliva, A. & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11, 520-527.
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3, 519-526.
- Palmer, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology*, 24, 436-447.
- Palmer, S. & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, 1, 29-55.
- Palmer, S. E., Rosch, E. & Chase, P. (1981). *Canonical perspective and the perception of objects*. In Long, J. & Baddeley, A. (Eds.), *Attention and Performance*, IX (pp. 135-151). Hillsdale, NJ: Erlbaum.

- Pannasch, S., Helmert, J. R., Roth, K., Herbold, A.-K. & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 2(2):4, 1-19.
- Pegna, A. J., Khateb, A., Michel, C. M. & Landis, T. (2004). Visual recognition of faces, objects and words using degraded stimuli: Where and when it occurs. *Human Brain Mapping*, 22, 300-311.
- Peissig, J. J. & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology*, 58, 75-96.
- Peterson, M. A. (1994). Object recognition processes can and do operate before figure-ground organization. *Current Directions in Psychological Science*, 3, 105-111.
- Pitcher, D., Charles, L., Devlin, J., Walsh, V. & Duchaine, B. (2009). Triple dissociation between faces, bodies, and objects in extrastriate cortex. *Current Biology*, 19, 319-324.
- Poggio, T. & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343, 263-266.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, 2, 509-522.
- Sampanes, A. C., Tseng, P. & Bridgeman, B. (2008). The role of gist in scene recognition. *Vision Research*, 48, 2275-2283.
- Schweinberger, S. R. & Burton, A. M. (2003). Covert recognition and the neural system for face processing. *Cortex*, 39, 9-30.
- Selfridge, O. G. (1955). Pattern recognition and modern computers. *Proceedings of the Western Joint Computer Conference*. New York: Institute of Electrical and Electronics Engineers.

- Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Simons, D. J., Franconeri, S. L. & Reimer, R. L. (2000). Change blindness in the absence of a visual disruption. *Perception*, 29, 1143-1154.
- Simons, D. J. & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science*, 9, 315-320.
- SR Research (2009). Eyelink 1000 [Apparatus und software]. Mississauga, Ontario, Canada: SR Research.
- Tarr, M. J. & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1494-1505.
- Tarr, M. J. & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, 67, 1-20.
- Tarr, M. J., Bülthoff, H. H., Zabinski, M. & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, 8, 282-289.
- Tarr, M. J. & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21, 233-282.
- Tarr, M. J., Williams, P., Hayward, W. G. & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, 1, 275-277.
- Ullman, S. (1989). Aligning pictorial description: An approach to object recognition. *Cognition*, 32, 193-254.

- Ullman, S. & Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 992-1005.
- Wallis, G. & Bülthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences*, 3, 22-31.
- Wang, R. F. & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition*, 70, 191-210.
- Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, 61, 161-265.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung*, 4, 301-350.

## ABBILDUNGSVERZEICHNIS

---

Abbildung 1: Darstellung einer komplexen Szene (aus Wallis & Bülthoff, 1999) .....	8
Abbildung 2: Darstellung eines hierarchischen 3-D Modells eines Menschen (aus Marr & Nishihara, 1978) .....	10
Abbildung 3: Darstellung von Möglichkeiten der Zusammensetzung von Geons (aus Biederman, 1978) .....	11
Abbildung 4: Darstellung der Einflussnahme des Kontexts (aus Selfridge, 1955) .....	18
Abbildung 5: Szenenbeispiele der sechs verwendeten Kategorien (Quelle: Psychologische Fakultät Wien; Forschungspraktikum II, WS 2009/2010; Ulrich Ansorge) .....	28
Abbildung 6: Schematische Darstellung der Reizabfolge der zwei Blöcke (Quelle: Psychologische Fakultät Wien; Forschungspraktikum II, WS 2009/2010; Ulrich Ansorge) .....	30
Abbildung 7: Rate der richtigen Urteile in den vier Versuchsbedingungen .....	35
Abbildung 8: Prozentsatz Fixationen (relativ zu allen Fixationen) auf übereinstimmenden Objekten in Lern- und Transferbildern als Funktion der Richtigkeit der Antwort .....	36
Abbildung 9: Prozentsatz korrekter Antworten in Abhängigkeit des Prozentsatzes von Fixationen auf übereinstimmendem Objekt in der Bedingung „Identische Bilder“ .....	37
Abbildung 10: Prozentsatz korrekter Antworten in Abhängigkeit des Prozentsatzes von Fixationen auf übereinstimmendem Objekt in der Bedingung „Perspektivsprung“ .....	38

## CURRICULUM VITAE

---

### Angaben zur Person

<b>Name</b>	Bettina Schuckert
<b>Staatsangehörigkeit</b>	Österreich
<b>Geburtsdatum</b>	31. Oktober 1986
<b>Geburtsort</b>	Mistelbach

### Ausbildung

<b>1992 - 1996</b>	Volksschule Wildendürnbach
<b>1996 - 2004</b>	Bundesgymnasium Laa/Thaya
<b>seit 2004</b>	Psychologiestudium an der Universität Wien

### Zusatzqualifikationen

<b>Sprachen</b>	Deutsch, Muttersprache Englisch, in Wort und Schrift Russisch, Grundkenntnisse
<b>Kenntnisse</b>	MS Office SPSS
<b>Praktikum</b>	Kuratorium für Verkehrssicherheit August und September 2008