

Kontext ist nicht immer der Gewinner

Helmut Leder und Claus-Christian Carbon

Department of Psychology
Free University of Berlin
Habelschwerdter Allee 45
14195 Berlin
Germany
e-mail: lederh@zedat.fu-berlin.de

Abstract

Bei der Gesichtserkennung handelt es sich um einen einzigartigen kognitiven Mechanismus, der es uns erlaubt, einzelne Menschen mit hoher Trefferquote erfolgreich zu identifizieren. Es existieren konkurrierende Modelle, die die Verarbeitung und Repräsentation von Gesichtern zu erklären suchen: Das holistische Modell (Tanaka & Farah, 1993, QJEP) geht davon aus, dass Gesichter in Ihrem Ganzen verarbeitet werden und die konstituierenden Elemente nicht explizit repräsentiert werden. Alternative Modelle postulieren, dass lokale oder konfigurale Merkmale repräsentiert werden. In der vorliegenden Experimentenreihe wurden als Lernmaterialien einerseits isolierte Gesichtsmerkmale verwendet, andererseits die selben Gesichtsmerkmale, jedoch diesmal jeweils eingebettet in ein vollständiges Gesicht. Einer strengen holistischen Theorie zufolge müsste durch die Kontext-Lernbedingung ein deutlich besseres Enkodieren und damit ein besseres Wiedererinnern ermöglicht werden. In Experiment 1 war der Kontext in allen Gesichtern gleich: In diesem Fall wurden die vollständigen Gesichter besser als die isolierten Merkmale erinnert, wenn man jene gelernt hatte. Hatte man isolierte Merkmale gelernt, war man bei jenen – entgegen der holistischen Vorhersage – deutlich besser. Variierten die Kontexte wie bei realen Gesichtern dann zeigte sich ebenfalls kein Vorteil des Kontextes bei der Wiedererkennung isolierter Merkmale. Wurde der Kontext in allen Kontext-Bedingungen gleich gehalten in Form eines Standard-Kontextes, so konnte überhaupt kein positiver Kontexteffekt ausgemacht werden. Die Ergebnisse sprechen daher gegen eine strenge holistische Hypothese und verweisen auf die wichtige Rolle von Lern-Test Kompatibilitäten.

Einleitung

Die Gesichtserkennung ist Grundlage unseres täglichen Lebens. Sie liefert uns Information über Personen, denen wir im Alltagsleben begegnen, und ermöglicht dadurch Orientierung und effizientes Handeln. Verheerend ist die Vorstellung, wir könnten nicht unterscheiden, ob wir bestimmte Menschen kennen oder nicht, ob sie uns freundlich gesinnt sind oder ob eher Vorsicht angebracht ist. Die Gesichtserkennung bleibt nicht nur erste Stufe der Interaktion mit unseren Mitmenschen, sie vermittelt darüber hinaus soziale oder biologische Faktoren wie Geschlecht oder Alter und unter Umständen die Gefühlslage unseres Gegenübers. Auf Grund dieser Alltagsrelevanz handelt es sich bei der Gesichtserkennung um einen hochroutinisierten Prozess, der uns aufgrund seiner Verlässlichkeit fast selbstverständlich, ja nahezu trivial erscheinen mag. Nach Fodor (1983) wird der kognitive Apparat aus verschiedenen vertikalen Modulen gebildet, die als Eingabesystem für einen spezifischen Reiz zu verstehen sind.

Auch für die Gesichtsverarbeitung könnte ein solches Modul zur Verfügung stehen. Sobald der Stimulus Gesicht aus dem Hintergrund von umliegenden Umgebungsreizen gelöst wurde, durchfließt die Information funktionale Einheiten, welche unterschiedliche Informationsaspekte in Form von verschiedenen Kodierungen fassen (Bruce & Young, 1986). Bruce und Young (1986) fordern neben dem oben erläuterten Informationspfad, gesonderte Module zur Verarbeitung des Gesichtsausdrucks, der Mimik (d.h. die Bewegung des Gesichts bei der Kommunikation) und einer gerichteten visuellen Analyse, die Alter, Geschlecht und Kriterien wie Intelligenz, Attraktivität und Aufrichtigkeit mit einschließt.

Bislang umstritten ist immer noch die Frage, wie und welche Information auf der ersten Stufe kodiert wird, welche (vielleicht spezifischen) Merkmalsverarbeitung und Repräsentation bei der Gesichtserkennung zum Tragen kommt. Im Fokus der bisherigen Forschung stand die Frage, ob sich der Prozess auf holistische, oder Merkmals-basierte Verarbeitung stützt. Tanaka und Farah (1993) verstehen unter "holistisch", dass die Einzelkomponenten des Gesichts (Augen, Nase, Mund, Augenbrauen) keine explizite Beachtung beim Erkennen und Speichern von Gesichtern finden. Eine alternative Position beziehen Leder und Bruce (2000) oder Diamond und Carey (1986), die die besondere Bedeutung von konfiguralen oder relationalen Informationen postulieren, d.h. Positionen und Abstände der Gesichtsmerkmale werden als entscheidende Variablen beim Gesichtserkennen angesehen. Bruce und Leder (2000) stellten fest, dass isolierte konfigurale Merkmale erstaunlich gut wiedererkannt werden, wohingegen ein hinzugefügter Gesichtskontext die Wiedererkennung nicht verbessert.

Genau das sollte aber der Fall sein, folgt man der These von Tanaka und Farah (1993): Die Gesichtsverarbeitung ist ihres Erachtens am besten mit einer holistischen Repräsentation zu erklären. Die Effizienz der Gesichtserkennung beruhe auf Gestaltprinzipien, deren Effektivität sich nur innerhalb ganzer Gesichter, im Experiment also nur nach Präsentation kompletter Gesichter auswirken könne. Umgekehrt müsste ein Merkmal, das isoliert gelernt wurde, beispielsweise die Augenpartie, dann besser wiedererkannt werden, wenn es in einen Kontext, sprich einem Gesicht, eingebettet ist, und innerhalb dessen auch verarbeitet werden kann. Genau diese Fragestellung soll in der vorliegenden Untersuchung näher untersucht werden.

Zusätzlich wird untersucht, wie sich die Merkmalsverarbeitung in Abhängigkeit von der Orientierung des Gesichtes auswirkt. Diese Variation ist insofern interessant, da sie Rückschlüsse auf die Repräsentation von Gesichtern zulässt. Bis heute existiert ein Widerstreit in der Wissenschaftsdebatte, ob Objekterkennung auf einer betrachterunabhängigen Speicherart, d.h. in Form von vollständigen dreidimensionalen Repräsentationen, stattfindet, oder ob Objekte betrachterabhängig im Gedächtnis abgelegt werden. Solche vom Betrachter abhängigen Repräsentationsformen werden üblicherweise als zweidimensionale Einkodierungen aufgefasst. Weitreichende theoretische Arbeit leisteten hierzu Tarr und Pincker (1990), Shyns und Bülthoff (1993) und Hill, Shyns und Akamatsu (1997). Das wohl wichtigste betrachterunabhängige Modell der Objekterkennung wird dagegen von Marr (1982) vertreten. Interessanterweise konnte in Tarr und Pinker's (1990) Experimenten gezeigt werden, dass die Repräsentationsform von der Art der Aufgabe abhing: So zeigten sich typische betrachterabhängige Leistungen, wenn die Objekte nur durch die Relationen zwischen 2D-Elementen unterschieden werden konnten. Dagegen stellten sich zeigten sich betrachterunabhängige Leistungen ein, wenn die Unterscheidung durch eine eindimensionale Ordnung der Einzelelemente möglich war. In Gesichtserkennungs-Experimenten konnten reduzierte Erkennungs-Leistungen demonstriert werden, wenn zwischen Lern- und Testbedingung der Betrachtungswinkel verändert wurde (Bruce, 1982; Bruce, Valentine, & Baddeley, 1987). Schyns und Bülthoff (1993) untersuchten, welche Merkmale dafür verantwortlich sind, dass ein bestimmter Betrachtungswinkel besonders gute Erkennungsleistungen hervorruft. Mit Hilfe von 3D-Laserscans von Gesichtern als Reizmaterial konnte keine Präferenz des Betrachtungswinkels festgestellt werden, wenn alle in der Lernphase vorkommenden Winkel gleich häufig vorkamen. In einem zweiten Experiment fanden sie allerdings Evidenzen für einen Vorteil von Seitenansichten gegenüber Frontalansichten in puncto Einkodierung und Wiedererkennung von Gesichtern.

In der vorliegenden Studie wurde daher auch der Frage nachgegangen, ob die semantische Verarbeitung von Gesichtern von Dreiviertelansichten und Frontalansichten prinzipiell unterschiedlich ist. Immerhin unterliegen einige der als vermeintlich wichtigen Kandidaten für das Gesichtserkennen durch die Veränderung des Ansichtswinkels dramatischen Veränderungen: So verändert sich etwa der retinal wahrgenommene Augenabstand des zu identifizierenden Gesichts deutlich, wenn man das Gesicht seitlich schwenkt.

Experiment 1

Versuchspersonen: 16 freiwillige Psychologie-Studenten, die jeweils zufällig den Versuchsbedingungen zugeordnet wurden, und mit ihrer Teilnahme eine von zehn für das Grundstudium obligatorischen Versuchspersonenstunden gutgeschrieben bekamen.

Materialien: Digitalisierte Porträtaufnahmen von sechs, den Probanden unbekannt, männlichen Personen. Dieses Stimulusmaterial wurde auf zweierlei Methode variiert: Erstens über den Blickwinkel, d.h. die Gesichter wurden frontal ("frontal") bzw. um 20° gedreht ("seitlich") dargestellt (Faktor "Blickwinkel"), und zweitens über die Größe des angezeigten Gesichtsausschnittes, d.h. die Gesichter wurden vollständig ("Gesichter"; siehe Abbildung 1a) bzw. lediglich deren Augenpartien ("Augen"; siehe Abbildung 1b) dargestellt ("Ausschnittgröße"). Jeder dargebotenen Person wurde ein Name aus einer von zwei festen Listen zufällig zugeordnet.



Abbildung 1a



Abbildung 1b

Stimulusbeispiel (Frontalansicht) aus Experiment 1

Das Experiment teilte sich in drei Teile: Zuerst eine Lernphase, in der die Versuchspersonen mit dem Stimulusmaterial vertraut gemacht wurden. Die Lernphase teilte sich in drei Lernblöcke, in denen jeweils jedes Gesicht, der zugeordneten Lernbedingung visuell entsprechend, einmal dargeboten wurde. Anschließend wurde die Lernleistung durch eine Kontrollphase abgeprüft. Wenn eine Versuchsperson weniger als vier von insgesamt sechs Gesichtern richtig dem entsprechenden Namen zuordnen konnte, so wurde eine weitere Lernrunde, bestehend aus zwei Lernblöcken, durchgeführt. Schließlich folgte eine Testphase, bei der die Versuchspersonen den jeweils dargebotenen Gesichtern die entsprechenden Namen zuordnen sollten. Dies geschah über einen mit einer fest vorgegebenen Namensliste verbundenem Zahlenschlüssel von 1-6. Vor der Testphase wurden die Versuchspersonen davon in Kenntnis gesetzt, dass alle folgenden Personen bereits in der Lernphase präsentiert wurden, auch wenn möglicherweise die Darbietungsform nun unterschiedlich zu der beim Lernen werden würde.

Ergebnisse & Diskussion. Als abhängige Variable wurde jeweils die durchschnittliche Fehleranzahl pro Bedingung verwendet. Wie aus Abbildung 2 ersichtlich ist, werden innerhalb der Tests wesentlich weniger Fehler gemacht, wenn das Reizmaterial aus ganzen Gesichtern und nicht nur aus der Augenpartie besteht (1.8 zu 3.0 durchschnittliche Fehler). Durch einen entsprechenden t-Test lässt sich dieser Unterschied auch statistisch absichern.

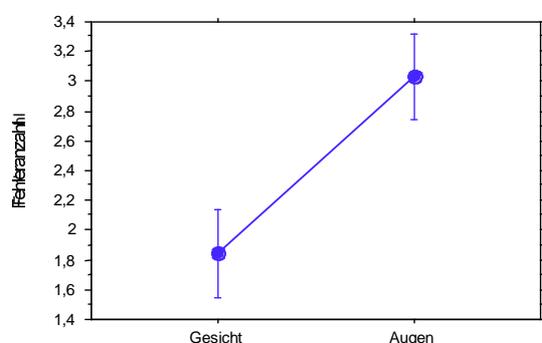


Abbildung 2. Ergebnisse aus Experiment 1: Teststimuli Gesichter vs. Augen

Dieser Befund könnte als Evidenz für eine holistische Theorie der Gesichtsverarbeitung gewertet werden, würde man nicht zusätzlich auch die Lernbedingung in die Analyse miteinbeziehen. Kontrolliert man die dargebotene Ausschnittsgröße (variiert zwischen Gruppen) entsprechend, so zeigt sich, dass sich lediglich dann ein deutlicher Vorteil von ganzen Gesichtern gegenüber isolierten Augenpartien ergibt, wenn auch in der Lernbedingung bereits ganze Gesichter gelernt werden sollten. Hat man dagegen nur isolierte Augenpartien als Lernmaterial erhalten, so sinken die Erkennungsleistungen von ganzen Gesichtern deutlich ab (siehe Abbildung 3).

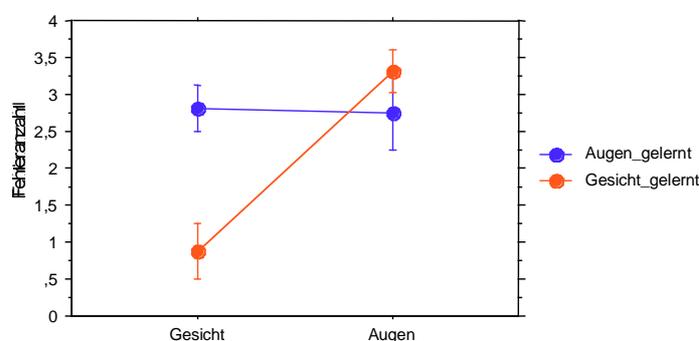


Abbildung 3. Ergebnisse aus Experiment 1: Teststimuli Gesichter vs. Augen in Bezug auf die jeweilige Lernbedingung

Eine ANOVA mit den 2x2 Lernbedingungen (Ausschnittsgröße, Blickwinkel) als *between subjects* Faktoren ergab keinen Haupteffekt "Lernbedingung". Die *within subjects* Faktoren (der Testbedingung) ergaben einen Haupteffekt "Ausschnittsgröße" in der Testphase: Wurden nur Augen präsentiert, so erreichten die Versuchspersonen eine durchschnittliche Fehlerzahl von 3.03, wurden hingegen ganze Gesichter präsentiert, so

sank die durchschnittliche Fehleranzahl auf 1.97. Dieser Unterschied konnte mit Hilfe eines t-Tests statistisch abgesichert werden. Neben der statistisch signifikanten Interaktion zwischen der Ausschnittsgröße in der Lern- und der Testbedingung konnte keine weitere Signifikanz erkannt werden. So blieb auch die Interaktion zwischen dem Blickwinkel in der Lern- und demjenigen in der Testphase knapp über dem festgesetzten -Niveau von 5%. Abbildung 4 veranschaulicht den Trend eines Lern-Test-Kompatibilitätseffekts in Bezug auf den Blickwinkel:

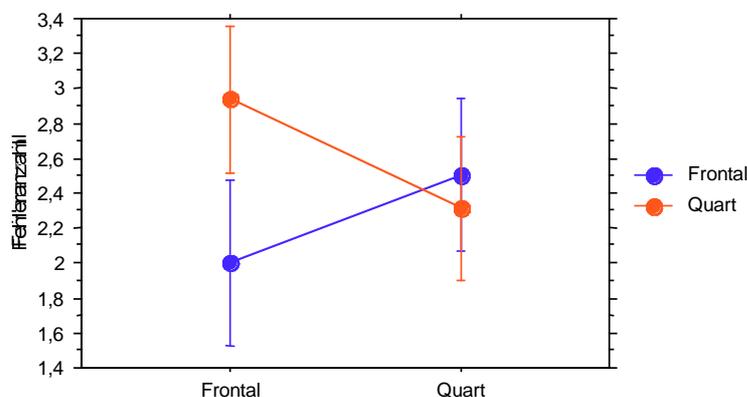


Abbildung 4. Ergebnisse aus Experiment 1: Effekte des "Blickwinkel"

Die Summe der Ergebnisse zeigt einen deutlichen Lern-Test-Kompatibilitätseffekt, d.h. dass die Versuchspersonen dann am wenigsten Fehler beim Identifizieren der Gesichter machten, wenn die Lern- und die Testbedingung übereinstimmten. Dieser Befund steht in Einklang mit Befunden aus anderen gedächtnispsychologischen Domänen. Betrachtet man nur die Lernbedingung "Augen", so zeigt sich jedoch ein wesentlich interessanterer Effekt. Denn entgegen einer Hypothese eines generell hilfreichen Kontextes, konnte im vorliegenden Experiment kein Vorteil eines Gesichtskontextes erkannt werden, wenn in der Lernphase das Stimulusmaterial lediglich über die Augenpartien dargeboten wurde. Mit Hilfe eines t-Tests der Testbedingung Ausschnittsgröße in der Gruppe mit der Lernbedingung "Augen" konnte kein signifikanter Unterschied zwischen den Präsentationsformen "Augen" und "Gesichter" festgestellt werden. Der Trend der Daten zeigt sogar einen entgegengesetzten Effekt, d.h. zusätzliche Information verschlechtert die Wiedererkennung in diesem Fall sogar tendenziell. Gesichter, die in der Lernphase komplett, in der Testphase jedoch nur noch als Augenpartie zu sehen waren, wurden erwartungsgemäß signifikant schlechter wiedererkannt. Mit dem, im Vergleich zur Lernphase, reduzierten Informationsgehalt des Stimulusmaterials geht auch eine Verschlechterung der Wiedererkennungseistung einher.

Experiment 2

Das zweite Experiment wurde zur Validierung des vorangegangenen Experiments verwandt. Hauptsächliches Anliegen war, dass der idiosynkratische Gesichtskontext (individuelle Frisur, Kinnform, Ohren) durch einen allgemeinen, also für alle Gesichter gleichen Standardkontext, ersetzt wird. Diese Manipulation sollte verhindern, dass einerseits Versuchspersonen mit Hilfe solcher zusätzlichen eigenständigen Merkmale eine verbesserte Wiedererkennungsrates per se erreichen konnten, andererseits aber auch gewährleisten, dass dennoch in der Lernphase die Hälfte aller Stimuli vollständige Gesichter waren, die gemäß holistischer Theorien prinzipiell bessere Enkodierung zeigen sollten, da nur sie als wirkliche Gesichter wahrgenommen werden und so von vermeintlich holistischen Enkodierungsstrategien profitieren sollten. Die Variation "Ausschnittsgröße", die im ersten Experiment lediglich die Augenpartie betraf, wurde nun erweitert auf die zusätzlichen Bereiche Mund und Nase. Außerdem erfolgte die Präsentation des Reizmaterials in der Testphase zu gleichen Teilen aufrecht als auch auf dem Kopf stehend. Dadurch sollten sog. Gesichts-Inversionseffekte (Yin, 1969) näher untersucht werden. Das restliche Versuchsvorgehen deckt sich mit dem des ersten Experiments.

Resultate & Diskussion. Der auffälligste Unterschied zwischen den Fehlerraten des ersten und des zweiten Experiments war die deutlich niedrigere durchschnittliche Fehleranzahl im zweiten Experiment. Betrachtet man nun die Fehlerraten innerhalb dieses Experiments, fällt jedoch auf, dass nun der standardisierte, und damit informationsirrelevante Gesichtskontext keinen Vorteil mehr gegenüber isolierten Merkmalen

einbrachte. Ganz im Gegenteil fiel die Leistung derjenigen Versuchspersonen statistisch signifikant ab, die in der Lernphase jenen Standardkontext zu sehen bekamen (siehe Abbildung 5).

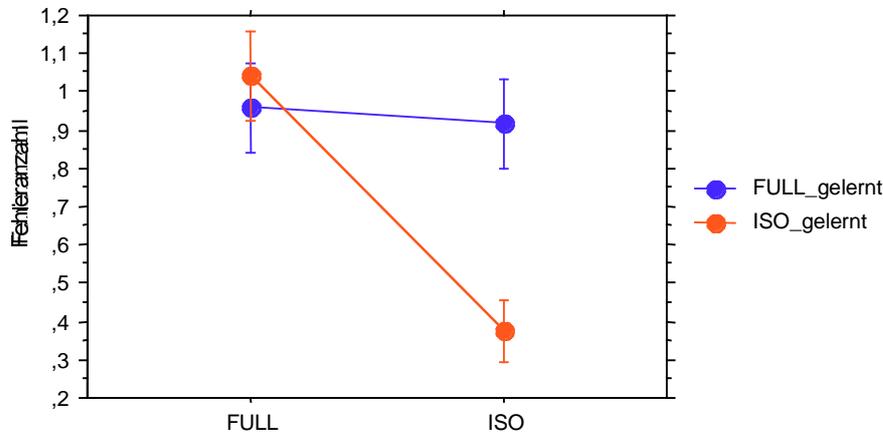


Abbildung 5. Experiment 2. Interaktion zwischen Ausschnittsgröße in Lern- und Testphase

Wie bereits aus unzähligen Forschungsergebnissen zuvor konnte ein deutliches und statistisch absicherbares Erkennungsdefizit bei invertierter Reizpräsentation festgestellt werden. Eine statistisch signifikante Interaktion zwischen Orientierung (aufrecht, invertiert) und Ausschnittsgröße in der Lernphase blieb zwar aus, jedoch ergab sich eine entsprechende Interaktion zwischen Orientierung und dargebotener Ausschnittsgröße in der Testphase. Dies bedeutet, dass trotz eines irrelevanten Kontextes in der FULL-Bedingung vermutlich beim Erkennen der Stimuli doch andere Prozesse verwendet wurden als in der isolierten Testbedingung, da der Inversionseffekt in unterschiedlicher Höhe auftrat (siehe Abbildung 6).

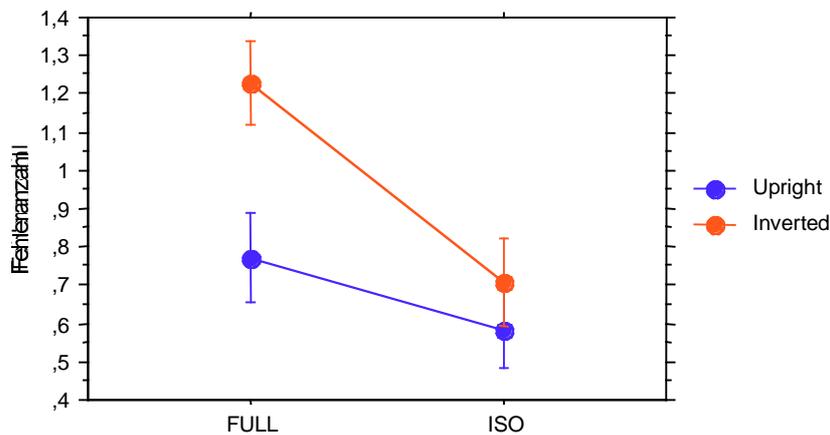


Abbildung 6. Ergebnisse aus Experiment 2: Interaktion zwischen den Testbedingungen Ausschnittsgröße und Orientierung auf die Fehleranzahl

Allgemeine Diskussion

In den vorliegenden Experimenten konnte ein deutlicher Einfluss von Lern-Test-Kompatibilitäten nachgewiesen werden: Wenn der Abfragemodus in der Testphase demjenigen der Lernphase entsprach, so wurden tendenziell bessere Wiedererkennungsleistungen erbracht. Es konnte gezeigt werden, dass das bloße Hinzufügen eines plausiblen Gesichtskontextes in der Testphase eher negative Auswirkungen auf die Wiedererkennung hatte, wenn den Versuchspersonen in der Lernphase lediglich die Augenpartien als Stimulusmaterial gezeigt worden waren. Dieser Befund spricht gegen die Theorie einer holistischen Gesichtsverarbeitung. Die Einbettung

isolierter Merkmale - in unserem ersten Experiment realisiert über die alleinige Darstellung der Augenpartie - in einen Gesichtskontext reicht alleine nicht zur Verbesserung der Wiedererkennung. Es sollte deshalb weiter untersucht werden, von welcher Art die verwendete isolierte Information am besten sein sollte, um die Wiedererkennung zu verbessern.

Sicher scheint auf jeden Fall zu sein, dass das Hinzufügen von Kontext an sich nicht immer bessere Wiedererkennungsleistungen evoziert, sondern dass dies nur dann hilfreich ist, wenn daraus zusätzliche Informationen gewonnen werden können. Vor allem die Resultate des zweiten Experiments demonstrieren eindrucksvoll, dass irrelevante Kontexte keinen Vorteil auf die Erkennungsleistungen besitzen, sondern diese sogar deutlich stören können. Der Kontext ist daher nicht immer der Gewinner! Andere Repräsentationsarten wie relationale Informationen scheinen eher wichtig für die Repräsentation und Wiedererkennung von Gesichtsmerkmalen zu sein. Dies deckt sich auch mit Befunden von Leder und Bruce (2000), bei denen deutliche Hinweise zu finden sind, dass isolierte relationale Informationen sehr gut wiedererkannt werden können. Statt einer holistischen Verarbeitung von Gesichtern scheint daher viel mehr eine die Enkodierung relationaler Merkmale wesentlich zur Wiedererkennung von Gesichtern zu sein.

Literatur

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, 73, 105-116.

Bruce, V. & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305-327.

Bruce, V., Valentine, T. & Baddeley, A. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, 1, 109-120.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, 115, 107-117.

Farah, M.J., Tanaka J.W. & Drain, H.M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, 21, 628-634.

Hill, H., Schyns, P.G. & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition*, 62, 210-221.

Leder, H. (1996). Line drawings of faces reduce configural processing. *Perception*, 25, 355-366.

Leder, H. & Bruce, V. (2000). When inverted Faces are recognised: the role of configural information in face recognition. *Quarterly Journal of Experimental Psychology*, 53a, 513-536.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

Schyns, P.G. & Bülthoff, H. (1993). Conditions for viewpoint dependent face recognition. *MIT A.I. Memo No. 1432*.

Tanaka, J.W. & Farah, M.J. (1993). Parts and Holes in Race Recognition. *Quarterly Journal of Experimental Psychology*, 46A, 225-245.

Tanaka, J.W. & Sengco, J.A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, 25, 583-592.

Tarr, M.J. & Pinker, S. (1990). When does human object recognition use a viewer centred reference frame? *Psychological Science*, 1 (42), 253-256.

Yin, R.K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81, 141-145.

Danksagung

Die Arbeiten wurden im Rahmen des Projektes "Merkmale und Repräsentationen bei der Gesichtserkennung" (LE 1286) von der DFG gefördert. Großer Dank gilt Florian Haubfleisch, Marc Seeman, Luka Klein und Adrian Schütte für die Mitkonzeption der Experimente und deren Durchführung.