

"Die sehen ja alle gleich aus !":

Einflussfaktoren der unterschiedlichen Wiedererkennensleistung
von Gesichtern der eigenen Ethnie und Gesichtern anderer Ethnien
(Cross-Race Bias)

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Philosophie des Fachbereichs Psychologie
der Justus-Liebig-Universität Gießen

vorgelegt von
Jürgen Gehrke

aus Kiel

2005

Dekan/in: Prof. Dr. Dr. Jürgen Hennig

1. Berichterstatter/in: Prof. Siegfried Ludwig Sporer, PhD

2. Berichterstatter/in: Prof. Dr. Gert Haubensak

Tag der Disputation: 24.5.2005

Ich sage nur ein Wort: Vielen Dank !

(Horst Hrubesch in seiner Abschiedsrede)

Es gibt eine große Anzahl an Personen, die mich bei dieser Dissertation unterstützt haben und denen ich dafür drei Wörter sagen möchte: Habt vielen Dank !

Zuerst ist da selbstverständlich mein Betreuer, früher nannte man eine solche Person etwas romantischer den 'Doktorvater', Prof. Siegfried Ludwig Sporer, PhD, der mich immer unterstützt hat und der mir Gelegenheit gab, meine Kenntnisse mit Apple Computern zu vertiefen. Weiterhin danke ich meinem Zweitgutachter, Prof. Dr. Gert Haubensak, neben dessen Büro ich zwei Jahre lang weilte und der sich freundlicherweise bereit erklärt hat, diese Arbeit zu begutachten.

Nicht zu vergessen, meine Kolleginnen: allen voran Dr. Tanja Stucke, die leider am 1. August 2004 im Alter von 32 Jahren verstorben ist, und bei uns eine große Lücke hinterlassen hat. Dipl.-Psych. Maike Breuer und Dipl.-Psych. Melanie Sauerland, die mir ebenfalls hilfreich zur Seite standen und JA, ich muss es an dieser Stelle noch einmal sagen: ich wünschte mir manchmal einen männlichen Kollegen mehr (Tanja übrigens auch).

Ein besonderer Dank gilt selbstverständlich meiner Dany, seit kurzem ebenfalls ein Dipl.-Psych., die mich durch diesen ganzen Prozess begleitet, Korrektur gelesen und sich meine stolzen Verkündigungen wie: "Ich bin jetzt bei 25000 Wörtern" angehört hat.

Ich danke weiterhin den Diplomanden, Christiane Henning, Andrea Schergung und Erdinç Dogan, für die Erhebung der Daten und viele anregende Gespräche.

Vielen Dank gebührt auch Herrn Gabriel, seines Zeichens zuständiger Beamter beim Bundesgrenzschutz für die Qualitätssicherung am Flughafen Frankfurt/Main und den, der anonymisierten Datenerhebung wegen, mir namentlich unbekanntem Beamten der Inspektion IV des Bundesgrenzschutzes am Flughafen Frankfurt für die freundliche und unkomplizierte Kooperation bei der Datenerhebung für das Experiment III. Außerdem danke ich der Arbeitsgruppe Sozialpsychologie der Philipps-Universität Marburg für die Bereitstellung des Labors für die Erhebung der Daten des Experiments II.

Ebenso danke ich meinen Eltern, die ich in der letzten Zeit leider etwas seltener gesehen habe, für ihre Unterstützung während der letzten Jahre. Zuletzt gilt mein Dank noch allen Teilnehmerinnen und Teilnehmern an den Experimenten. Letztendlich hängt von ihrem Tastendruck die Güte der Daten ab.....

Zusammenfassung

In der Forschung zur Wahrnehmung und zum Wiedererkennen menschlicher Gesichter ist der Cross-Race Bias Gegenstand vieler Studien und ein robustes Phänomen (Meissner & Brigham, 2001). Der Cross-Race Bias beschreibt eine schlechtere Wiedererkennungslleistung von Gesichtern, die nicht der eigenen Ethnie entstammen, sog. Fremd-Gruppen Gesichter, im Vergleich zu Gesichtern der eigenen ethnischen Gruppe. Trotz kleinerer gradueller Unterschiede im Ausmaß des Cross-Race Bias (Sporer, 1992), tritt dieser Effekt bei den meisten in der Literatur untersuchten Gruppen auf (vgl. Meissner & Brigham, 2001; Wright & Sladden, 2003). Zumeist sind es allerdings 'Weiße' und 'Schwarze', die zu den untersuchten Gruppen gehören. Ziel dieser Dissertation war, das In-Group/Out-Group Modell von Sporer (2001a) zu überprüfen, das aufbauend auf früheren theoretischen Überlegungen zur Erklärung des Cross-Race Bias, ein Modell darstellt, das die bisherigen Studien integriert und ein theoretisches Grundgerüst für die Erklärung dieses Phänomens liefert. Die zentrale Annahme des Modells ist, dass ausgehend von der Wahrnehmung einer Person als Mitglied der Eigen-Gruppe bzw. der Fremd-Gruppe, unterschiedliche Verarbeitungswege angestoßen werden, die zu einer schlechteren bzw. besseren Wiedererkennungslleistung führen. Eigen-Gruppen Gesichter sollten holistisch, Fremd-Gruppen merkmalsbasiert verarbeitet werden.

In Experiment I ($N = 42$) wurde mittels Objekten (Geldscheinen der DM, des Euro und des Britischen Pfunds) die Annahme der unterschiedlichen Verarbeitung und des Wiedererkennens von Gesichtern auf nicht-menschliche Stimuli ausgeweitet. Die Ergebnisse sprechen für eine Generalisierbarkeit des Modells auf Objekte, sofern eine Kategorisierung der Objekte, als der Eigen-Gruppe bzw. der Fremd-Gruppe zugehörig, möglich ist.

In Experiment II ($N = 128$) wurde die Wiedererkennungslleistung bei Gesichtern untersucht, denen äußere Merkmale (Haare, Ohren) entfernt wurden, die ein Wiedererkennen erleichtern. Teilnehmer waren je zur Hälfte türkische und deutsche Personen. Durch die Entfernung äußerer Gesichtsmerkmale sollte eine merkmalsbasierte Enkodierung der Gesichter erschwert werden und zu schlechteren Leistungen bei den Fremd-Gruppen Gesichtern führen. Zusätzlich wurde das Zeitintervall zwischen Lern- und Testphase variiert. Die schlechtere Leistung bei Gesichtern, denen äußere Merkmale entfernt wurden, wurde auch gefunden, allerdings für Eigen-Gruppen und Fremd-Gruppen Gesichter bei einem längeren Zeitintervall zwischen Lern- und Testphase gleichermaßen. Bei türkischen und deutschen Teilnehmern wurde bei maskierten Gesichtern ein Cross-

Race Bias bei sofortiger Testung gefunden, was die Annahme einer merkmalsbasierten Enkodierung von Fremd-Gruppen Gesichtern stützt.

In Experiment III ($N = 64$) wurde ebenfalls der Cross-Race Bias untersucht, indem zwei Berufsgruppen (BGS-Beamte und Studierende), die sich hinsichtlich ihrer Kontakthäufigkeit mit anderen Ethnien unterscheiden sollten, getestet wurden. Die beiden Gruppen unterschieden sich in der Gesamtleistung nicht voneinander, jedoch waren die BGS-Beamten bei afro-amerikanischen Gesichtern besser als die Studierenden, während die Studierenden bei deutschen Gesichtern bessere Leistungen zeigten als die BGS-Beamten. Die Ergebnisse stützen die Hypothese, dass durch eine höhere Kontakthäufigkeit der Cross-Race Bias verringert werden kann. Die Gesichter der Eigen-Gruppe wurden nach einem Wechsel der Gesichtsansicht (Frontalansicht vs. Halbprofil) besser identifiziert als die Gesichter der Fremd-Gruppen. Das Ergebnis wird als Beleg für die Annahme der merkmalsbasierten Enkodierung von Fremd-Gruppen Gesichtern gewertet, da diese Art der Enkodierung bei einem Ansichtswechsel des Stimulus zu einer Verschlechterung der Leistung führt.

1	Einleitung	1
1.1	Begriffs-Definitionen	2
1.1.1	Bezeichnungen für die in der Literatur untersuchten ethnischen Gruppen	3
1.1.2	Der Cross-Race Bias (CR-Bias)	3
1.1.3	Exkurs: Der Own-Age Bias und der Own-Sex Bias	7
1.1.4	Expertise in der Gesichterwahrnehmung	8
1.1.5	Stand der Forschung zum CR-Bias	16
1.2	Methoden in der Forschung zur Gesichtererkennung	30
1.2.1	Signal-Entdeckungstheorie	30
1.2.2	Testparadigmen	38
1.3	Integration der dargestellten Theorien und Phänomene	41
1.4	Fragestellung	44
1.4.1	Hypothesen für die Experimente zu einem möglichen Cross-Currency Effekt	44
1.4.2	Hypothesen zum Einfluss des Zeitintervalls und des Gesichtsausschnitts auf den Cross-Race Bias	46
1.4.3	Hypothesen zum Einfluss der Kontakthäufigkeit auf den Cross-Race Bias	49
1.5	Schlußbemerkungen zur Fragestellung	50
2	Die unterschiedliche Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Objekten (Cross-Currency Effekt)	52
2.1	Versuchsplan	53
2.1.1	Stichprobe	55
2.1.2	Material	55
2.1.3	Stimuli	56
2.2	Experiment I: Delayed-Matching-Aufgabe	60
2.2.1	Hypothesen	60
2.2.2	Stimulus-Darbietung	61
2.2.3	Versuchsdurchführung	62
2.2.4	Ergebnisse	63
2.2.5	Diskussion der Ergebnisse der Delayed-Matching-Aufgabe	68

2.3 Experiment II: AFC-Aufgabe	70
2.3.1 Hypothesen	70
2.3.2 Stimulus-Darbietung	71
2.3.3 Versuchsdurchführung	71
2.3.4 Ergebnisse.....	72
2.3.5 Diskussion der Ergebnisse der AFC-Aufgabe.....	75
2.4 Experiment III: Wiedererkennenstest.....	76
2.4.1 Hypothesen	77
2.4.2 Stimulus-Darbietung	77
2.4.3 Versuchsdurchführung	78
2.4.4 Ergebnisse.....	78
2.4.5 Diskussion der Ergebnisse des Wiedererkennenstests.....	84
2.5 Diskussion der Experimente zum Cross-Currency Effekt.....	86
3 Pilotstudie für die Auswahl der türkischen und deutschen Gesichter	90
3.1 Material.....	90
3.1.1 Stichprobe.....	91
3.1.2 Stimulus-Material	91
3.1.3 Skalen	93
3.1.4 Stimulus-Darbietung	95
3.1.5 Versuchsdurchführung	96
3.2 Ergebnisse.....	96
3.3 Diskussion.....	101
4 Der Einfluss von Behaltensintervall und Gesichtsausschnitt auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern... 104	
4.1 Hypothesen	108
4.2 Material.....	109
4.2.1 Stimuli	109
4.2.2 Aufteilung der Stimuli zu den Darbietungssets	111
4.2.3 Stimulus-Darbietung	111
4.2.4 Versuchsdurchführung	112
4.2.5 Versuchsablauf.....	113
4.2.6 Stichprobe.....	114

4.3	Auswertung.....	114
4.4	Ergebnisse.....	114
4.5	Diskussion.....	124
5	Pilotstudie für die Auswahl der latein-amerikanischen Gesichter	128
5.1	Material.....	129
5.1.1	Stichprobe.....	129
5.1.2	Stimulus-Material	129
5.1.3	Skalen.....	132
5.1.4	Stimulus-Darbietung	132
5.1.5	Versuchsdurchführung	133
5.2	Ergebnisse.....	134
5.3	Diskussion.....	136
6	Der Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung von Eigen- Gruppen und Fremd-Gruppen Gesichtern	137
6.1	Darstellung des Wiedererkennenstests und der Delayed-Matching Aufgabe	140
6.2	Hypothesen	140
6.2.1	Versuchsplan.....	142
6.2.2	Stichprobe.....	143
6.3	Material.....	144
6.3.1	Stimuli	144
6.3.2	Aufteilung der Stimuli auf die Darbietungssets und die Experimente.....	145
6.3.3	Stimulus-Darbietung	147
6.3.4	Versuchsdurchführung	147
6.3.5	Versuchsablauf.....	148
6.4	Ergebnisse.....	149
6.4.1	Kontaktfragebogen.....	149
6.4.2	Wiedererkennenstest	151
6.4.3	Diskussion des Wiedererkennenstests.....	155
6.4.4	Delayed-Matching-Aufgabe	157
6.4.5	Diskussion der Ergebnisse der Delayed-Matching Aufgabe.....	165
6.5	Diskussion der Experimente zum Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung bei Eigen-Gruppen und Fremd-Gruppen Gesichtern .	167

7	Zusammenfassende Diskussion und Ausblick.....	172
7.1	Experimente zur differentiell schlechteren Wiedererkennensleistung bei Fremd-Gruppen Objekten (Cross-Currency Effekt).	172
7.2	Pilotstudie für die Auswahl der türkischen und deutschen Gesichter.....	174
7.3	Der Einfluss von Behaltensintervall und Gesichtsausschnitt auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern	175
7.4	Pilotstudie für die Auswahl der latein-amerikanischen Gesichter.....	178
7.5	Der Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern.....	178
7.6	Integration der Befunde.....	180
7.7	Weitergehende Untersuchungen zu weiteren Modellannahmen.....	182
7.8	Praktische Implikationen	184
8	Literaturverzeichnis.....	186
9	Anhang.....	198

Die Ghanaer erkennen Sie an den gelben Stutzen!

(Marcel Reif beim Länderspiel Deutschland vs. Ghana)

1 Einleitung

Der Mensch sieht jeden Tag Hunderte von Gesichtern, sei es am Arbeitsplatz, im Verkehr, in Restaurants oder auch nur im Fernsehen, und jedes dieser Gesichter ist einzigartig, sieht man einmal von den Gesichtern eineiiger Zwillinge ab. Einige Gesichter sieht man mehrmals bis täglich während eines Lebens, andere nur einmal, die meisten allerdings nie. Trotz dieser Unmenge an Information, die das menschliche Gehirn verarbeiten muss, ist man in der Lage, sich an die Gesichter vieler Menschen zu erinnern und diese auch zu differenzieren. Jedoch, entgegen der weitläufigen Meinung, der Mensch sei in der Lage, Gegenstände oder Menschen fehlerfrei zu differenzieren und zweifelsfrei wiederzuerkennen, lehrt die einschlägige Literatur und auch der tägliche Umgang mit anderen Personen, dass dem nicht so ist. Als Beispiel sei eine eigene Erfahrung kurz beschrieben. Ich traf auf einer Geburtstagsfeier eine junge Frau, mit der ich mich auch kurz unterhalten habe. Ungefähr drei bis vier Monate später begegnete mir diese Frau wieder, allerdings hielt ich sie für einen Menschen, den ich bis dahin nicht gesehen hatte. Erst als sie die Geburtstagsfeier wieder erwähnte, erinnerte ich mich. Warum ich sie nicht wiedererkannte, lag in der Rückschau an ihrer Frisur, die in diesen drei bis vier Monaten von lang zu kurz gewechselt hatte. Nebenbei bemerkt war ich nicht der einzige, dem es an diesem Tag so ergangen ist.

Der Fähigkeit, Menschen voneinander zu differenzieren, kommt bei der interpersonalen Kommunikation eine wichtige Rolle zu. Schließlich muss die Information, die kommuniziert wird, auf die Person abgestimmt werden, mit der die Information ausgetauscht wird. Ist die Differenzierung nicht möglich, können Probleme u.a. im sozialen Zusammenleben auftreten, die die betroffenen Personen stark beeinträchtigen. Menschen, die beispielsweise an Prosopagnosie leiden, ist es nicht möglich, andere Menschen aufgrund ihres Gesichtes auseinander zu halten. Bei dieser Krankheit ist es den betroffenen Personen entweder von Geburt an oder aufgrund eines Schlaganfalles bzw. eines Unfalls, bei dem Teile des Gehirns insbesondere des visuellen Systems, beschädigt wurden, nicht möglich, Gesichter zu erkennen (Goldstein, 1989).

Ein weiteres Feld, in dem die Differenzierung und Identifizierung von Gesichtern und Personen eine bedeutende Rolle spielt, findet man im forensischen Kontext. Die

Zeugen¹ einer Straftat sollten den Täter nach Möglichkeit einwandfrei identifizieren, um nicht unter Umständen einen Unschuldigen zu verurteilen. Jedoch ist auch die Fehlerrate bei Gegenüberstellungen im polizeilichen Kontext teilweise sehr groß. Diese Fehlerrate variiert in Abhängigkeit von mehreren Faktoren wie z.B. der Zeit, die zwischen Tat und Gegenüberstellung liegt, der Art der Gegenüberstellung (sequentiell vs. simultan; siehe Sporer, 1994, Clifford & Bull, 1978) oder auch der Hautfarbe bzw. der ethnischen Herkunft des Verdächtigen und des Zeugen. Stammen Verdächtiger und Zeuge aus unterschiedlichen ethnischen Gruppen, steigt die Wahrscheinlichkeit eines Fehlers bei der Wiedererkennung des Verdächtigen verglichen mit Konstellationen, bei denen Täter und Zeuge aus gleichen ethnischen Gruppen kommen. D.h., dass entweder ein Unschuldiger mit höherer Wahrscheinlichkeit als Täter wiedererkannt wird oder ein Täter wird als nicht schuldig eingestuft. Dieser Umstand ist selbstverständlich nicht auf polizeiliche Gegenüberstellungen beschränkt, er tritt auch im täglichen Leben auf und ist darüber hinaus nicht auf eine Richtung beschränkt (z.B. Weiße erkennen Schwarze schlecht wieder), sondern auch Schwarze erkennen Weiße schlecht wieder. In der vorliegenden Arbeit soll dieses, in der Literatur als Cross-Race Bias (CR-Bias) bezeichnete Phänomen, näher beleuchtet werden.

1.1 Begriffs-Definitionen

In diesem Abschnitt sollen die in den späteren Abschnitten verwendeten Begriffe erläutert werden, die zum einen einer näheren Erklärung bedürfen und zum anderen neu eingeführt werden. Zunächst wird die Forschung zum Cross-Race Bias zusammengefasst. Anschließend wird auf Forschungen zur Expertise in der Gesichterwahrnehmung und der Trainierbarkeit der Gesichtererkennung näher eingegangen.

¹ Im Folgenden soll aus ökonomischen Gründen bei der Darstellung von Personengruppen die männliche Form verwendet werden. Sie schließt selbstverständlich die weibliche Form mit ein, zwar nicht grammatikalisch, jedoch inhaltlich.

1.1.1 Bezeichnungen für die in der Literatur untersuchten ethnischen Gruppen

In diesem Abschnitt sollen lediglich die Bezeichnungen für die in der Literatur untersuchten ethnischen Gruppen kurz dargestellt werden, weil diese, vor allem in den USA aber auch in Deutschland voreilig als 'politisch unkorrekt' angesehen werden können. Die meisten Studien, auf die in dieser Arbeit Bezug genommen wird, sowie die Experimente in den Abschnitten 4 und 6, benutzen Gesichter von afrikanischen Amerikanern, kaukasischen Amerikanern und Latein-Amerikanern. Es ist üblich, die Bezeichnungen 'Blacks' und 'Whites', bzw. 'Hispanics' zu verwenden (siehe z. B. MacLin & Malpass, 2001; Levin, 2000), allerdings ohne die ethnischen Gruppen diskriminieren zu wollen. Die Bezeichnungen 'Schwarze' und 'Weiße' werden im folgenden lediglich dazu genutzt, die zitierten Studien wiederzugeben. In den später beschriebenen Experimenten werden die Bezeichnungen *Afrikanische-Amerikaner* und *Kaukasische-Amerikaner* sowie *Latein-Amerikaner* eingeführt. Im Alltagsleben, zumindest in den USA, wird teilweise die Bezeichnung 'Amerikaner, afrikanischer (bzw. kaukasischer und latein-amerikanischer) Herkunft' benutzt.

Ein weiterer Grund für die Bezeichnungen *Afrikanische-Amerikaner*, *Kaukasische-Amerikaner* sowie *Latein-Amerikaner* ist der, dass der Begriff 'Weiße' mißverständlich sein kann, weil beispielsweise auch Latein-Amerikaner eine weiße Hautfarbe haben können. Die eingeführten Bezeichnungen für die ethnischen Gruppen stellen somit eine passendere Formulierung für den ethnischen Hintergrund der jeweiligen Gruppen dar als die eher simplen Bezeichnungen 'Schwarze' und 'Weiße', die sich im Prinzip nur auf die Hautfarbe als Kriterium der Zugehörigkeit zu einer Gruppe beziehen. Auch die türkischen und deutschen Gesichter, die in den später dargestellten Experimenten als Stimulusmaterial verwendet werden, können als *kaukasisch* bezeichnet werden (Wikipedia, n.d.).

1.1.2 Der Cross-Race Bias (CR-Bias)

Es wurden bereits in den vierziger Jahren des 20. Jahrhunderts Experimente zur Identifikation bzw. Klassifikation von Fremdgruppengesichtern durchgeführt, mit dem Ziel, mögliche Zusammenhänge von Vorurteilen gegenüber Personen anderer Gruppen mit der Einordnung dieser Personen zu anderen Gruppen zu erforschen (Allport & Kramer,

1946; Carter, 1948). Die Fähigkeit zum Wiedererkennen von Gesichtern allgemein und dabei das differentiell schlechtere Wiedererkennen von Fremdgruppen-Gesichtern wurde in den sechziger Jahren zu einem Gegenstand der psychologischen Forschung. Malpass und Kravitz (1969) wiesen das erste Mal bei Studierenden an Universitäten mit vornehmlich schwarzen bzw. weißen Studierenden eine Interaktion zwischen der Ethnie der Versuchspersonen und der Ethnie der Stimulusgesichter in einem Wiedererkennenstest nach.

Die Begriffe *Cross-Race Bias* oder auch *Own-Race Bias* bezeichnen Unterschiede in der Wiedererkennensleistung von Gesichtern der eigenen ethnischen Gruppe gegenüber Gesichtern anderer ethnischer Gruppen. Dabei werden Gesichter der eigenen ethnischen Gruppe besser wiedererkannt als Gesichter anderer Gruppen. Die beiden Begriffe bezeichnen das gleiche Phänomen, betrachten es jedoch aus unterschiedlichen Perspektiven (schlechtere bzw. bessere Wiedererkennung bei Fremd- bzw. Eigengruppengesichtern). Im weiteren Verlauf dieser Arbeit soll der Begriff *Cross-Race Bias* für das beschriebene Phänomen verwendet werden.

Der Cross-Race Bias konnte in einer Reihe von Studien repliziert werden (z.B. Bothwell, Brigham & Malpass, 1985; Meissner & Brigham, 2001), und wird als reliables Phänomen angesehen (siehe Meta-Analysen von Anthony, Copper & Mullen, 1992; Meissner & Brigham, 2001). Definiert ist ein Cross-Race Bias als Interaktion der Faktoren 'Ethnie der Versuchsperson' und 'Ethnie der Stimulusgesichter' in einem Wiedererkennensexperiment, in dem die Teilnehmer die Gesichter ihrer eigenen Gruppe besser wiedererkennen als die Gesichter der anderen Gruppe (siehe Abbildung 1).

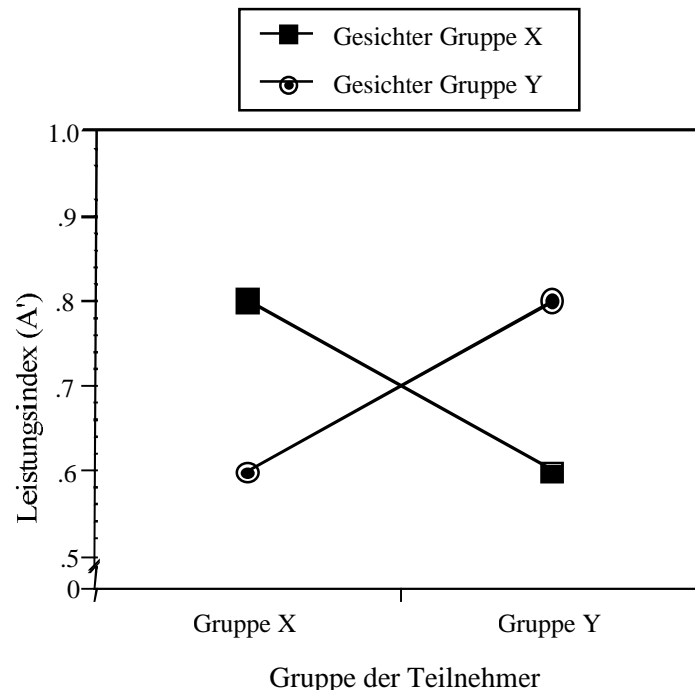


Abbildung 1. Hypothetische Interaktion eines Wiedererkennensexperiments mit zwei Teilnehmergruppen und den entsprechenden Stimulusgesichtern mit A' (siehe Abschnitt 1.2.1.1) als abhängiger Variablen

Neben einer schlechteren Wiedererkennensleistung bei Fremd-Gruppen Gesichtern berichten einige Autoren (Slone, Brigham & Meissner, 2000; Sporer, 1992) ebenfalls von einer Tendenz der Teilnehmer, bei Gesichtern von Fremd-Gruppen eher zu behaupten, sie hätten dieses Gesicht bereits vorher gesehen und zwar ungeachtet der Tatsache, ob das Gesicht bereits gesehen wurde, oder nicht. Solch eine Reaktionsneigung (engl. response bias B'') wird auch als *liberales Antwortkriterium* bezeichnet. Den beiden Maßen, d.h. der Wiedererkennensleistung A' und der Reaktionsneigung B'', ist ein eigener Abschnitt gewidmet, in dem die theoretische Konzeption und die Berechnung derselben erläutert werden.

Dass das Phänomen des Cross-Race Bias nicht alleine theoretisch von Interesse ist, zeigen u.a. Kriminalfälle, in denen die unterschiedlichen Ethnien von Zeuge und vermeintlichen Täter mit ein Grund für eine letztendlich falsche Identifizierung des Täters sein können.

In July 1984 Jennifer Thompson, a White woman, was raped by a Black assailant in her apartment. Ronald Cotton was arrested and identified

by Thompson in both a photo-spread and a live line-up. She appeared to be very confident in her identification as she claimed he had "a really distinctive nose". Based largely on this identification, in 1985 Cotton was convicted of raping Thompson. In 1987, while in prison, one of Cotton's fellow inmates confessed to the crime. This was not enough for Cotton's release given the weight placed on the eyewitness identification. It was not until 1995, when DNA evidence showed that Cotton was innocent, that his conviction was quashed. The DNA showed that the person who confessed was the culprit. After being in prison for ten and a half years Cotton was offered \$5000 compensation, which he refused (Connors, Lundregan, Miller, & McEwen, 1996; zit. nach Wright, Boyd & Tredoux, 2003, p. 365).

Solch ein Fall zeigt nicht, dass der Cross-Race Bias die Hauptursache für eine fälschliche Identifizierung ist, sondern vielmehr, dass bei unterschiedlichen Ethnien von Täter und Opfer (bzw. von Täter und Zeuge) die Möglichkeit einer falschen Identifizierung als Folge eines Cross-Race Bias beachtet werden sollte. Dass Jennifer Thompson so sicher war, den Täter identifiziert zu haben, da er eine 'auffällige Nase' ('a really distinctive nose') hätte, ist im Licht der Theorien zur Speicherung von Gesichtern (siehe Abschnitt 1.1.5) nicht ungewöhnlich, schließlich werden auffällige Gesichter besser wiedererkannt als typische Gesichter. Wenn aber mehrere Gesichter ein auffälliges Merkmal teilen, wie beispielsweise die Hautfarbe, ist dieses Merkmal nicht mehr für die Differenzierung geeignet. Darauf wird in den folgenden Kapiteln näher eingegangen. Die Tatsache, dass, wie in dem beschriebenen Fall, Zeuge und Täter aus verschiedenen Ethnien stammen, sollte bei Kenntnis des Cross-Race Bias eine Identifizierung eines mutmaßlichen Täters zumindest fragwürdig erscheinen lassen. In den USA wird die Möglichkeit einer falschen Identifizierung aufgrund unterschiedlicher Ethnien von Täter und Zeuge seit den achtziger Jahren von den Juristen in Betracht gezogen (Brigham & Wolfskeil, 1983)

In den meisten Studien, vor allem aus dem englischsprachigen Raum, wurden weiße und schwarze Versuchspersonengruppen miteinander verglichen. In diesen Studien wurde zumeist mithilfe eines Wiedererkennenstests (Ja/Nein - Test, siehe Abschnitt 1.2.2.3) beiden Versuchspersonengruppen eine bestimmte Anzahl an Gesichtern während einer Präsentationsphase dargeboten und nach einem bestimmten Intervall zusammen mit neuen, d.h. in der Präsentationsphase nicht gesehenen Gesichtern, wiederum präsentiert. Die Versuchspersonen sollten in der zweiten Phase angeben, ob sie das Gesicht, welches sie gerade sahen, wiedererkennen oder nicht. Dabei zeigte sich zumeist, dass Personen der

eigenen Gruppe besser wiedererkannt wurden. Studien mit anderen Gruppen sind seltener, zeigen aber ähnliche Resultate. Platz und Hosch (1988) untersuchten den Cross-Race Bias bei Schwarzen, Weißen und Latein-Amerikanern, die in El Paso, Texas, lebten. Weiße und Latein-Amerikaner konnten am besten Mitglieder der eigenen ethnischen Gruppe wiedererkennen, bei Schwarzen zeigten sich keine signifikanten Unterschiede. Allerdings war die Anzahl der Teilnehmer nicht bei allen drei Gruppen gleich. Vielmehr bestand die Gruppe der Schwarzen aus lediglich elf Probanden. Sporer (1999) untersuchte den Cross-Race Bias mit türkischen und deutschen Versuchspersonen und fand dabei eine Interaktion der Faktoren 'Ethnie der Versuchspersonen' und 'Ethnie der Stimulusgesichter'. Die türkischen Versuchspersonen zeigten keinen Unterschied im Wiedererkennen zwischen den Gruppen der Stimulusgesichter im Gegensatz zu den deutschen Versuchspersonen, die bei den deutschen Gesichtern bessere Leistungen zeigten als bei den türkischen Gesichtern. Demnach ist der Cross-Race Bias ein Phänomen, das nicht auf eine bestimmte ethnische Gruppe beschränkt ist, sondern zumindest die untersuchten ethnischen Gruppen betrifft.

1.1.3 Exkurs: Der Own-Age Bias und der Own-Sex Bias

Es läßt sich zeigen, dass der Cross-Race Bias kein einzelnes, eingegrenztes Phänomen ist, wenn man Leistungsunterschiede zwischen Gruppen betrachtet. Es gibt andere Biases, die ebenfalls unterschiedliche Wiedererkennensleistungen zum Inhalt haben und durch das In-Group/Out-Group Modell von Sporer (2001a) vorhergesagt werden (siehe Abschnitt 1.1.5.6). So beschreiben beispielsweise Wright und Stroud (2002) den *Own-Age Bias*, der besagt, dass Versuchspersonen in einem Wiedererkennensexperiment besser in der Lage sind, Personen der eigenen Altersgruppe wiederzuerkennen als Personen einer anderen Altersgruppe. Sie ließen Versuchspersonen, die entweder zwischen 18 und 25 Jahren oder zwischen 35 und 55 Jahren alt waren, Videos betrachten, in denen ein Auto bzw. ein Fernsehgerät gestohlen wurde. Die Täter waren Männer, die 21 bzw. 23 Jahre ('junge' Tätergruppe) oder 48 und 51 Jahre alt waren ('alte' Tätergruppe). Insgesamt sah jede Versuchsperson vier Videos (2 (Tat) x 2 (Tätergruppe)). Nach einem Intervall von einem Tag bzw. einer Woche wurden den Versuchspersonen vier Gegenüberstellungen à sieben Fotos simultan dargeboten, von denen eines das Bild des Täters zeigte. Bei dem Wiedererkennentest nach einem Tag zeigte sich eine signifikante Interaktion zwischen dem Alter der Versuchspersonen und dem Alter der Täter. Jüngere Versuchspersonen zeigten bessere Leistungen bei jüngeren Tätern als ältere Versuchspersonen. Die

Wiedererkennensleistungen von älteren Versuchspersonen bei älteren Tätern waren etwas besser als bei jüngeren Versuchspersonen, jedoch war dieser Unterschied nicht signifikant. Nach dem Intervall von einer Woche zeigten sich keine signifikanten Unterschiede. Rodin (1987) fand heraus, dass ältere Personen sowohl jüngere als auch Personen der eigenen Altersgruppe gleich gut wiedererkennen konnten, während jüngere Teilnehmer bei älteren Personen schlechter waren als bei jungen.

Ähnlich verhält es sich mit dem *Own-Sex Bias* (Jalbert & Getting, 1992; Wright & Sladden, 2003; Sporer, 1992, Slone, Brigham & Meissner, 2000), der die analogen Wiedererkennens-Defizite postuliert wie der Own-Age Bias und der Cross-Race Bias. Demnach erkennen Männer männliche Gesichter besser als Frauen, und Frauen können weibliche Gesichter besser wiedererkennen als Männer. Wright und Sladden (2003) untersuchten den Own-Sex Bias mit 20 Männern und 20 Frauen, denen 24 Gesichter (12 männliche und 12 weibliche Gesichter) dargeboten wurden. Nach einem Intervall von fünf Minuten wurden die bekannten Gesichter und weitere 24 Bilder von Männern und Frauen in zufälliger Reihenfolge gezeigt. Als zusätzlicher Faktor wurden bei der Hälfte der Gesichter die Haare der Person auf dem Bild geschwärzt, so dass die Frisur keinen weiteren Hinweisreiz auf das Geschlecht darstellte. In dieser Studie wurden männliche Gesichter von Männern und weibliche Gesichter von Frauen besser wiedererkannt. Ebenso zeigten sich bessere Leistungen bei Gesichtern mit Haaren im Gegensatz zu Gesichtern ohne Haare. Allerdings benutzten Wright und Sladden in ihrer Untersuchung lediglich 24 Gesichter, deren Fotos z.T. aus dem Internet heruntergeladen worden waren. Inwiefern die Qualität der Bilder dabei konstant war ist unklar. Wenn die Fotos von unterschiedlicher Qualität waren, kann dies ein ausreichender Cue sein, um Gesichter zu differenzieren.

1.1.4 Expertise in der Gesichterwahrnehmung

Allgemein betrachtet bedeutet der Begriff *Expertise*, dass Experten mehr Erfahrung mit dem Gegenstand bzw. der Domäne ihres Expertenwissens haben als Novizen (Eysenck & Keane, 1992). Diese Erfahrung drückt sich aus in einer größeren Anzahl von Domänen- oder gegenstandsrelevanten Repräsentationen im Gedächtnis von Experten gegenüber Novizen sowie einer schnelleren Verfügbarkeit dieser Informationen. Der Abruf und die Anwendung von Expertenwissen geschieht dabei größtenteils automatisch. Für die Art und Weise, wie Informationen erworben und gespeichert werden, gibt es mehrere Ansätze aus der kognitiven Psychologie. Zu den bekanntesten Theorien gehören die Speicherung von

Informationen im Kurzzeitgedächtnis in Form von *Chunks* (Miller, 1956), die Speicherung in Form von *Instanzen* (Logan, 1988) sowie die Bildung von *Produktionsregeln* während des Erlernens einer Tätigkeit (ACT-R², Anderson, 1993). Im Kontext der Speicherung von Gesichtern ist die Theorie von Logan (1988) interessant, da die Ergebnisse in seinen Experimenten u.a. zu alphabet-arithmetischen Aufgaben zeigen, dass eine verbesserte Leistung durch die alleinige Wiederholung der einzelnen Aufgaben entsteht. Wie später in Abschnitt 1.1.5.4 beschrieben wird, hat die gegenwärtige Annahme der Speicherung von Gesichtern als einzelne Exemplare Ähnlichkeit zur Theorie der Speicherung einzelner Instanzen bei der Bearbeitung von z. B. alphabet-arithmetischen Aufgaben.

Menschen entwickeln sich im Laufe ihres Lebens zu Experten in der Gesichterwahrnehmung, und Gesichter per se sind anscheinend bereits von Geburt an eine besondere Stimulusklasse, wie Untersuchungen mit Gesichtern über Augenbewegungen bei Neugeborenen und anderen Stimuli zeigen konnten (Carmel & Bentin, 2002). Eine Reihe von Studien (z.B. Diamond & Carey, 1977; Chung & Thomson, 1995) fand heraus, dass die Fähigkeit zur Gesichterwiedererkennung sich vor allem zwischen den Altersstufen 6-7 und 10 Jahren verbessert. Schwarzer (2000) untersuchte die Hypothese, dass sich die Art der Informationsverarbeitung bei Gesichtern in den oben genannten Altersbereichen verändert und zwar von einer eher analytisch geprägten hin zu einer holistischen Informationsverarbeitung. Die Autorin fand für aufrechte Gesichter in Abhängigkeit vom Alter eine Verschiebung der Informationsverarbeitung von analytischer Verarbeitung hin zu holistischer Verarbeitung. Dass die Grundlagen für die Gesichterverarbeitung bereits in den ersten Wochen nach der Geburt gelegt werden, konnten Geldart, Mondloch, Maurer, de Schonen und Brent (2002) zeigen. Die Autoren untersuchten 10 bis 38-jährige Probanden, die in den ersten sieben Wochen nach der Geburt keine visuellen Stimuli aufnehmen konnten, und die im Vergleich zu einer Kontrollgruppe schlechtere Wiedererkennungslleistungen bei veränderter Gesichtshaltung der Stimuli aufwiesen.

Eine Studie von Myles-Worsley, Johnston und Simons (1988) zeigte, dass Radiologen mit längerer Berufserfahrung anomale Röntgenbilder besser erinnern konnten als Kollegen mit geringerer Berufserfahrung und Studierende. Bei der Wiedererkennungslleistung mit ebenfalls in dieser Studie untersuchten Gesichtern waren alle Gruppen gleich gut. Die Leistung bei Gesichtern und anomalen Röntgenbildern war bei denjenigen Radiologen mit der längsten Berufserfahrung sogar gleich. Dagegen sank die

² ACT-R ist die Abkürzung für *Adaptive Control of Thought – Rational*. Diese Theorie wird hier jedoch nicht weiter ausgeführt.

Wiedererkennensleistung von 'normalen' Röntgenbildern, d.h. Röntgenbilder ohne pathologischen Befund, bei Radiologen mit der längsten Berufserfahrung auf Zufallsniveau. Myles-Worsley et al. (1988) ziehen daraus die Schlussfolgerung, dass ähnlich wie bei Gesichtern ein atypisches Merkmal, in diesem Falle ein pathologischer Befund auf einem Röntgenbild, ein Bild besser memorierbar macht, weil es von normalen Bildern abweicht. Jedoch scheint diese bessere Memorierbarkeit davon abzuhängen, ob ein atypisches Merkmal für die Domäne der Expertise von Bedeutung ist. Dies zeigt die sinkende Leistung bei normalen Röntgenbildern mit der Länge der Berufserfahrung. Auch normale Röntgenbilder unterscheiden sich in bestimmten Merkmalen, die für Experten der Radiologie aber von untergeordneter Bedeutung sind. Anfänger, d.h. Radiologen mit weniger als einem Jahr Berufserfahrung, waren hier weitaus besser als ihre älteren Kollegen.

Diamond und Carey (1986) zeigten in einem Experiment mit Hunden als Stimulusmaterial und Hundezüchtern als Experimental- sowie Nicht-Züchtern als Kontrollgruppe als Versuchspersonen, dass Hundezüchter bei invertierten Bildern größere Schwierigkeiten hatten als die Kontrollgruppe. Es wurde allerdings von Bruyer und Crispeels (1992) kritisiert, dass die Kontrollgruppe jünger war als die Experimentalgruppe, und dass sowohl in der Enkodierungs- als auch in der Testphase die gleichen Bilder verwendet wurden, d.h. die Ansicht der Hunde wurde zwischen Lern- und Testphase nicht variiert. In der gegenwärtigen Forschung mit Wiedererkennensexperimenten werden in der Testphase Bilder verwendet, die im Gegensatz zu dem Bild in der Lernphase leicht verändert wurden. Meistens wird dabei eine andere Ansicht gewählt als in der Lernphase, um auszuschließen, dass die Teilnehmer den Stimulus nicht nur an einzelnen Merkmalen erkennen, sondern eine gewissermaßen ganzheitliche Repräsentation des Stimulus entwickelt haben.

1.1.4.1 Trainingsstudien

Ob gezieltes Training der Gesichterwiedererkennung zu Verbesserungen in der Wiedererkennensleistung bzw. zu einer Reduktion des Cross-Race Bias führt, wurde von Malpass, Lavigneur und Weldon (1973) überprüft. Die Autoren testeten zunächst die Effekte verbalen Trainings auf die Wiedererkennensleistung. Die Teilnehmer sollten entweder Gesichter der eigenen bzw. der Fremdgruppe beschreiben und sich bei der Beschreibung auf die Merkmale des Gesichts beschränken, d.h. keine Beschreibung z.B.

von der Kleidung der Stimulus-Person abgeben. In einem späteren Test zeigte sich kein Einfluss des verbalen Trainings auf die Leistung. In einem zweiten Experiment wurde mit den Teilnehmern ein visuelles Trainingsparadigma durchgeführt, in dem sie aus einem Set mit vier Gesichtern dasjenige bestimmen sollten, welches ihnen zuvor dargeboten wurde. Eine Sitzung bestand aus 25 solcher Durchgänge, in denen den Teilnehmern entweder keine Rückmeldung, verbale Rückmeldung oder Elektroschocks bei falschen Antworten gegeben wurde. Ein anfänglicher Vorteil der weißen Gesichter gegenüber den schwarzen Gesichtern verschwand im Laufe des Trainings und kehrte sich zum Ende sogar um. Zusätzlich wurde die Leistung bei allen Gesichtern während des Trainings schlechter und stieg zum Ende wieder leicht an. Die Elektroschock-Bedingung zeigte bessere Leistungen als die Bedingung ohne Rückmeldung. Die Teilnehmer mit verbalem Training zeigten keine signifikanten Unterschiede zu den anderen beiden Trainingsbedingungen.

Elliot, Wills und Goldstein (1973) trainierten weiße Teilnehmer mittels eines Paar-Assoziationstests, bei dem entweder 30 weiße oder 30 orientalische Gesichter jeweils mit einer Zahl gepaart wurden, die von den Teilnehmern gelernt werden sollte. Einer Kontrollgruppe wurden 20 weiße und 20 orientalische Gesichter für jeweils 3-4 Sekunden gezeigt. In einem anschließenden Wiedererkennenstest wurde den Teilnehmern 80 Gesichter (20 'alte' und 60 'neue') lediglich einer der beiden Ethnien dargeboten, die als 'alt' oder 'neu' erkannt werden sollten. Die weißen Gesichter wurden in allen Bedingungen besser wiedererkannt, egal welche ethnische Gruppe im Paar-Assoziationstest gelernt wurde, allerdings zeigten sich für die in orientalischen Gesichtern trainierten Versuchspersonen bessere Leistungen bei den orientalischen Gesichtern im Vergleich zu den Teilnehmern, die in der Trainingsphase weiße Gesichter sahen, bzw. überhaupt kein Training erfahren hatten.

1.1.4.2 Stufen der Abstraktion

Objekte und Gesichter können auf unterschiedlichen Stufen der Abstraktion beschrieben werden. Die *grundlegende Stufe*, auf der Objekte als einer Kategorie zugehörig angesehen werden, wird definiert als die Stufe auf der ein einzelnes mentales Abbild der Objekte gebildet werden kann, und die eine gemeinsame Form bilden (Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976). Auf einer *untergeordneten Stufe* werden die einzelnen Objekte einer Kategorie differenziert. So wird z.B. eine Amsel auf der grundlegenden Stufe als Vogel und auf der untergeordneten Stufe als Amsel erkannt und

kategorisiert. Experten sind in der Lage, Objekte auf beiden Stufen gleich gut und schnell zu identifizieren (Tanaka, 2001). Tanaka untersuchte ebenfalls, ob sich die Verarbeitung von Gesichtern auf der grundlegenden Ebene oder einer untergeordneten Ebene unterscheidet. Dazu verglich er menschlicher Gesichter mit Hunden als Stimulusmaterial. Während auf der grundlegenden Stufe menschliche Gesichter und Hunde gleich gut und schnell kategorisiert werden konnten, stieg die Reaktionszeit bei Hunden, und die Korrektheit der Kategorisierung sank, wenn die Zuordnung auf der untergeordneten Ebene vollzogen wurde. Bei menschlichen Gesichtern wurde keine Veränderung in der Reaktionszeit und der Korrektheit auf der untergeordneten Ebene beobachtet. Maurer, Le Grand und Mondloch (2002) unterscheiden analog zu Rosch et al. (1976) und in Anlehnung an Rhodes, Brake und Atkinson (1989) zwei Ordnungen der konfiguralen Verarbeitung von Gesichtern: *Relationen erster Ordnung*, z.B. die Anordnung der Augen über einer Nase und einem Mund, sowie *Relationen zweiter Ordnung*, z.B. die Distanzen zwischen Gesichtsmerkmalen. *Konfigurale Information* bedeutet die Information über Abstände zwischen Gesichtsmerkmalen (Nase, Augen, Ohren etc.) sowie über die Gesichtsform (Fallshore & Schooler, 1995).

Ein anderer Aspekt, der von Schooler und Engstler-Schooler (1990) untersucht wurde, ist der Effekt der *verbalen Überlagerung* (*verbal overshadowing*). Wenn die Teilnehmer an einem Versuch gebeten wurden, verbale Beschreibungen von Gesichtern abzugeben, können diese Beschreibungen eine Verschlechterung beim Wiedererkennen dieser Gesichter zur Folge haben. Fallshore und Schooler (1995) vermuten, dass dieser Effekt darauf begründet ist, dass die Teilnehmer konfigurale Informationen der Gesichter (holistische Verarbeitung) vernachlässigen und Informationen über Gesichtsmerkmale überbewerten, deren Beschreibung ja letztendlich instruiert wird. Fallshore und Schooler (1995) untersuchten ihre Hypothese unter Berücksichtigung des Cross-Race Bias, indem sie schwarzen und weißen Probanden, die jeweiligen Eigen- bzw. Fremd-Gruppen Gesichter zeigten, wobei jeweils die Hälfte der Probanden die Gesichter verbal beschreiben sollten. Die Autoren nahmen an, dass Eigen-Gruppen Gesichter konfigural verarbeitet werden und dass nach einer Verbalisierung der Merkmale der Eigen-Gruppen Gesichter, diese Gesichter, ebenso wie die Fremd-Gruppen Gesichter merkmalsbasiert verarbeitet werden und dementsprechend die Wiedererkennungslleistung sinken muss. Hypothesenkonform zeigte sich, dass Personen, die Eigen-Gruppen Gesichter verbalisieren mussten, bei diesen Gesichtern ähnliche Leistungen zeigten, wie bei den Fremd-Gruppen Gesichtern.

Dehon und Brédart (2001) fanden bei weißen und schwarzen Probanden, die das Alter von schwarzen und weißen Gesichtern einschätzen sollten, ebenfalls einen Cross-Race Bias. Weiße Probanden waren in ihrer Einschätzung des Alters von weißen Personen genauer als bei schwarzen Personen. Die schwarzen Probanden zeigten bei beiden Ethnien die gleiche Genauigkeit.

1.1.4.3 Neuronale Grundlagen (Fusiform Face Area (FFA))

In diesem Abschnitt sollen kurz die neuronalen Grundlagen der Erkennung und Differenzierung von Gesichtern erläutert werden. Sie werden im folgenden nicht explizit untersucht, sollen aber besprochen werden, um einen möglichst vollständigen Überblick über das Thema zu bieten.

Der Frage nach der Lokalisation der neuronalen Repräsentation von Gesichtern im menschlichen Gehirn wurde u.a. von Kanwisher, Chun, McDermott, und Ledden (1996), Kanwisher, McDermott und Chun (1997) sowie von Schweinberger, Pickering, Jentsch, Burton und Kaufmann (2002) nachgegangen. Der Schwerpunkt liegt dabei auf dem *fusiformen Gyrus*, der vor allem dann eine erhöhte neuronale Aktivität aufweist, wenn den Probanden Gesichter dargeboten werden. In Gehirnen von Makaken-Affen wurden Neuronen entdeckt, die selektiv auf Gesichter reagieren und im inferioren temporalen Gyrus liegen (Bruce, Desimone, & Gross 1981; Desimone, 1991). Die fehlende Fähigkeit der Differenzierung von Gesichtern (Prosopagnosie) hat nach de Renzi, Perani, Carlesimo, Silveri und Fazio (1994) ihre Lokalisation in der rechten Hirnhemisphäre. Patienten mit Schädigungen in der rechten Hemisphäre können Schwierigkeiten in der Gesichterwiedererkennung und der Zuordnung von Gesichtsausdrücken haben, sind jedoch in ihrer Fähigkeit zum Lippenlesen nicht beeinträchtigt. Bei vergleichbaren Schädigungen der linken Hemisphäre ist das Verhältnis umgekehrt. Die Fähigkeit, Gesichter wiederzuerkennen ist dann nicht beeinträchtigt, dafür jedoch die Fähigkeit, von den Lippen zu lesen.

Einen weiteren Hinweis auf die Verarbeitung von Gesichtern im menschlichen Gehirn ist die sogenannte N170, eine elektrisch negative Welle, die mittels EEG in occipito-temporalen Gebieten des Gehirns messbar ist (siehe Abbildung 2). Dies bedeutet, dass nach einem Intervall von 170 ms nach Darbietung eines Gesichtes eine Erregung in einem lateralen occipito-temporalen Ort außerhalb des fusiformen Gyrus auftritt. Sie kann ebenfalls gemessen werden, wenn den Probanden Gesichter von Schimpansen oder

gesichtsähnliche Stimuli präsentiert werden. Die N170 hat jedoch eine größere Amplitude für Gesichter im Vergleich zu anderen Objekten, die ebenfalls eine N170 auslösen können (Bentin, Allison, Puce, Perez & McCarthy, 1996). Tanaka und Curran (2001) zeigten, dass die Amplitude des N170 durch Expertise erhöht werden kann und zwar auch bei nicht-menschlichen Stimuli. Die Autoren fanden bei Vogel- und Hundezüchtern jeweils erhöhte Amplituden im N170, wenn den Vogelzüchtern Vögel bzw. den Hundezüchtern Hunde als Stimuli präsentiert wurden.

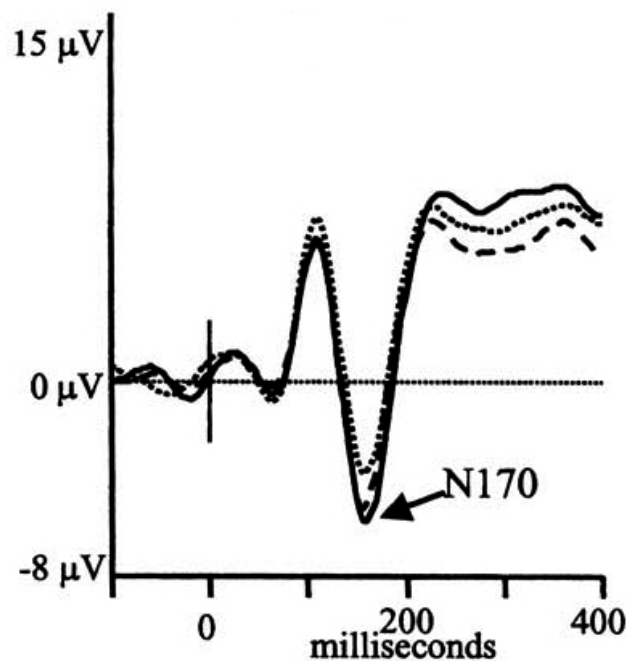


Abbildung 2. Drei N170 Wellen. Die gepunktete Linie zeigt eine N170 bei menschlichen Gesichtern und die durchgezogene Linie bei Autos nach einem Training (aus Carmel & Bentin, 2002).

Rossion, Gauthier, Delvenne, Tarr, Bruyer und Crommelinck (1999) postulierten ebenfalls eine Nicht-Spezifität der N170 für Gesichter, da die N170 auch bei Objekten ausgelöst werden kann und mit steigender Expertise bei Objekten eine größere Amplitude zeigt. Carmel und Bentin (2002) ziehen letztendlich den Schluss, dass der N170 gesichtsspezifisch ist, jedoch nicht spezies-spezifisch. Bentin, Sagiv, Mecklinger, Friederici und Von Cramon (2002) stellten darüber hinaus fest, dass einzelne Merkmale eines Gesichts (z.B. die Augen) eine N170 auslösen können, und dass selbst eine rein schematische Darstellung der Augen (zwei Punkte auf dem Bildschirm) eine N170

auslösen kann, wenn sie durch den Kontext als Schema eines Gesichts identifiziert werden kann. Eine rein schematische Darstellung von zwei Punkten erzeugt hingegen keine N170.

Die äußere Region des fusiformen Gyrus dient demnach nicht der Differenzierung einzelner Gesichter, sondern ist der eigentlichen Region für die Unterscheidung von Gesichtern vorgeschaltet. Rossion et al. (2002) fanden bei der Wahrnehmungsexpertise bei künstlich geschaffenen Objekten, die keine Merkmale von Gesichtern beinhalten ('Greebles'), dass die gleichen kognitiven Prozesse und Hirnareale involviert sind wie für die Wahrnehmungsexpertise bei Gesichtern. Eine N170-Amplitude ist ebenfalls existent bei Stimuli wie z.B. Vögeln (Rossion et al., 2000). Nach Untersuchungen von Pfütze, Sommer und Schweinberger (2002) sowie Schweinberger, Pickering, Jentsch, Burton und Kaufmann (2002), ist für die Differenzierung individueller Gesichter ein späterer Prozess verantwortlich, der sogenannte N250r (N250 repetition effect), der im fusiformen Gyrus lokalisiert werden kann. Die Autoren fanden bei wiederholt dargebotenen, bekannten Gesichtern eine N250r die ihrer Meinung nach auf den Zugang zu enkodierten Gesichtsrepräsentationen hindeutet. Golby, Gabrieli, Chiao und Eberhardt (2001) fanden bei einer Untersuchung des Cross-Race Bias mit der funktionellen Magnet-Resonanz Tomographie, dass Eigen-Gruppen Gesichter in bestimmten Regionen des fusiformen Gyrus eine stärkere Aktivierung auslösen als Fremd-Gruppen Gesichter.

Zusammenfassend lässt sich sagen, dass auf neuronaler Ebene nach dem heutigen Erkenntnisstand, die Wahrnehmung von Gesichtern wahrscheinlich im fusiformen Gyrus lokalisiert ist und die Differenzierung einzelner Gesichter ca. 250ms nach Wahrnehmung eines Gesichtes stattfindet.

1.1.4.4 Der Inversionseffekt beim Wiedererkennen von Gesichtern

Ob das Wiedererkennen von Gesichtern auf Prozessen beruht, die speziell für Gesichter ausgelegt sind, ist auch Gegenstand der Forschung zu invertierten Gesichtern. Invertierte Gesichter sind Gesichter, die im Gegensatz zu normalen, aufrechten Gesichtern um 180° rotiert sind. Der erste Befund zur Wiedererkennensleistung bei invertierten Gesichtern stammt von Yin (1969), der den Probanden Bilder von Gesichtern und anderen Objekten (z.B. von Häusern und Flugzeugen) entweder aufrecht oder invertiert dargeboten und anschließend in der gleichen Ausrichtung die Wiedererkennensleistung getestet hatte. Gesichter wurden in der aufrechten Bedingung besser wiedererkannt als die anderen Objekte, in der invertierten Bedingung jedoch wurde das Verhältnis umgekehrt, d.h. die

Gesichter wurden schlechter wiedererkannt als die anderen Objekte. In einer späteren Studie verglich Yin (1970) bei Probanden, die nach einer Hirnverletzung eine Schwäche hatten, Gesichter wiederzuerkennen, die Fähigkeit, invertierte und aufrechte Gesichter wiederzuerkennen. Die Probanden hatten Schwierigkeiten bei aufrechten Gesichtern, zeigten aber normale Leistungen bei invertierten Gesichtern. Ob allerdings schwächere Leistungen bei invertierten Gesichtern im Gegensatz zu aufrechten Gesichtern dafür sprechen, dass die Verarbeitung von Gesichtern eine speziellen Fähigkeit darstellt, wird von Valentine (1988) angezweifelt. Valentine spricht sich dafür aus, dass Faktoren wie die Vertrautheit einer Stimulus-Klasse sowie deren Homogenität eine größere Rolle spielen als spezielle Verarbeitungswege für Gesichter. Valentine und Bruce (1986) untersuchten den Cross-Race Bias mit invertierten Gesichtern und fanden bei Eigen-Gruppen Gesichtern einen größeren Inversionseffekt als bei Fremd-Gruppen Gesichtern. Sie argumentieren, dass sich Eigen-Gruppen und Fremd-Gruppen Gesichter, von Oberflächenmerkmalen wie der Hautfarbe einmal abgesehen, lediglich bezüglich ihrer Vertrautheit voneinander unterscheiden. Dies steht im Kontrast zu Yin, der vorschlägt, dass invertierte Gesichter schlechter wiedererkannt werden, weil die Enkodierung und das Gedächtnis für Gesichter eine spezielle Form der Verarbeitung von visuellen Stimuli darstellen. Demnach sollten zwischen Eigen- und Fremd-Gruppen Gesichtern keine Unterschiede bei der Wiedererkennung invertierter Gesichter auftreten. Dies kann aber mittels der Studie von Valentine und Bruce nicht bestätigt werden.

1.1.5 Stand der Forschung zum CR-Bias

Grundlage dieses Abschnittes sind vor allem die Meta-Analyse von Meissner und Brigham (2001) sowie ein Überblicksartikel von Sporer (2001a). In die Meta-Analyse von Meissner und Brigham (2001) gingen die Daten von 91 unabhängigen Stichproben aus 39 Forschungs-Artikeln mit 4996 Teilnehmern ein.

1.1.5.1 Theorien

In der sozialpsychologischen Literatur ist ein wichtiger Fokus auf die Forschung zu Stereotypen, Vorurteilen und Einstellungen von Personen gegenüber anderen Personen und Gruppen sowie deren Entstehung gelegt worden. Der *Out-Group-Homogenitäts-Effekt*

bezeichnet z.B. das Phänomen, dass Mitglieder einer Fremd-Gruppe als untereinander ähnlicher wahrgenommen werden als die Mitglieder der Eigen-Gruppe (Judd & Park, 1988; Quattrone & Jones, 1980). Dieser Befund wird u.a. dadurch erklärt, dass eine Akzentuierung der Unterschiede zwischen den Gruppen stattfindet. Die Zugehörigkeit zu einer bestimmten Gruppe dient ebenfalls der Generierung einer sozialen Identität und, indem andere Gruppen als homogen wahrgenommen und mit negativen Attributen versehen werden, ebenfalls einer Erhöhung des eigenen Selbstwertes (Tajfel & Turner, 1979). Die Kategorisierung anderer Personen aufgrund von Merkmalen dieser Person, die eine Kategorisierung erlauben (dazu gehören fast alle Eigenschaften oder äußeren Merkmale, die eine Differenzierung zwischen Menschen ermöglichen) ist eine sehr einfache Strategie, um Handlungen und Einstellungen anderer Personen vorherzusagen. Indem eine andere Person kategorisiert wird, kann es dazu kommen, dass dieser Person ebenfalls Eigenschaften und Fähigkeiten, die dem Stereotyp dieser Kategorie entsprechen, zugeschrieben werden (MacLin & Malpass, 2003). Zum Beispiel wird Afro-Amerikanern, zumindest in Basketballerkreisen, zugeschrieben, exzellent Basketball spielen zu können. Ein Stereotyp, das nicht immer zutreffend ist.

Eine solche Stereotypisierung kann aber nicht alle Facetten einer menschlichen Persönlichkeit abdecken (siehe u.a. *actor-observer bias*, Jones & Nisbett, 1972) und ist damit zwangsläufig mit Fehlern behaftet. Die Kategorisierung und Stereotypisierung von Personen erleichtert zunächst das tägliche Leben und ist ein Prozess, der täglich stattfindet. Die Nutzung von Stereotypen bedeutet nicht zwangsläufig, dass diese Stereotypen negativ sein müssen, wie man an dem Basketballbeispiel sehen kann. Bei der Frage nach der Entstehung des Cross-Race Bias taucht jedoch die Frage auf, ob negative Stereotypen bzw. Vorurteile gegenüber anderen Ethnien einen Einfluss auf den Cross-Race Bias haben. Personen mit Vorurteilen gegenüber anderen Ethnien grenzen die eigene Gruppe stärker von der anderen Gruppe ab als Personen ohne Vorurteile (Blascovich, Wyer, Swart & Kibler, 1997), bzw. sind eher darauf bedacht, nach Merkmalen zu suchen, die Personen voneinander unterscheiden. Blascovich et al. (1997) fanden bei einem Kategorisierungsexperiment, bei dem weiße Teilnehmer Bilder von schwarzen, weißen und uneindeutigen Gesichtern kategorisieren sollten, dass vorurteilsbehaftete Teilnehmer länger brauchten, die uneindeutigen Gesichter einzuordnen als vorurteilsfreie Teilnehmer. Dieser Befund ist insofern interpretierbar, als dass die vorurteilsbehafteten Teilnehmer länger nach relevanten Merkmalen suchten, die eine eindeutige Einordnung der gesehenen Person ermöglichten. Eine solche längere Suche bei Personen mit Vorurteilen gegenüber anderen Ethnien scheint jedoch nicht dazu zu führen, dass diese Personen Mitglieder

anderer Ethnien besser oder schlechter wiedererkennen können als vorurteilsfreie Personen.

Neuere Studien konnten zeigen, dass es keine Zusammenhänge zwischen den Einstellungen gegenüber anderen Ethnien und der Wiedererkennensleistung gibt (Slone, Brigham & Meissner, 2000). So zeigten Ferguson, Rhodes, Lee und Sriram (2001), dass sowohl für implizite Messung von Vorurteilen als auch für selbst-berichtete Einstellung gegenüber Fremden kein Zusammenhang zwischen der Vorurteilsneigung und dem Cross-Race Bias besteht.

1.1.5.1.1 Physiognomische Unterschiede

Die oft gehörte Aussage, dass sich Menschen anderer Ethnien sehr ähnlich sehen, lässt zunächst vermuten, dass sich Gesichter anderer Ethnien bezüglich ihrer Physiognomie stark ähneln. Zu dieser wahrgenommenen Ähnlichkeit gehören beispielsweise die Form der Nase, des Mundes oder auch das Fehlen von Lidfalten bei asiatischen Gesichtern. Wenn diese Annahme zuträfe, müssten Gesichter einer bestimmten ethnischen Gruppe (z.B. Asiaten) von allen anderen Gruppen entweder schlechter (bei physiognomischer Homogenität) oder besser (bei physiognomischer Heterogenität) wiedererkannt werden. Dies trifft aber nicht zu. Vielmehr ist der Cross-Race Bias ein Phänomen, das bei den meisten untersuchten ethnischen Gruppen auftritt. Goldstein und Chance (1976) zeigten weißen Versuchsteilnehmern jeweils zwei Gesichter von entweder Japanern oder Weißen und ließen die Teilnehmer entscheiden, ob die beiden Bilder von ein und derselben Person sind oder nicht. Für beide Stimuluse ethnien waren die Reaktionszeiten und die Fehlerraten nahezu gleich. Dies deutet darauf hin, dass sich Differenzen in der Physiognomie innerhalb der Gruppen nicht unterscheiden und damit nicht kausal für den Cross-Race Bias verantwortlich sind. Allerdings wurden keine japanischen Versuchspersonen erhoben, womit die Ergebnisse von Goldstein und Chance (1976) lediglich auf die Wahrnehmung von weißen Teilnehmern bezogen werden können. Entsprechend sind die Ergebnisse von Chance, Goldstein und McBride (1975) zu interpretieren. Sie boten schwarzen und weißen Versuchspersonen schwarze, weiße sowie orientalische Gesichter dar und untersuchten später die Wiedererkennensleistung. Dabei fanden sie bei beiden Teilnehmergruppen die schlechteste Wiedererkennensleistung bei den orientalischen Gesichtern, während die

Gesichter der eigenen Gruppe von den Versuchspersonen jeweils am besten wiedererkannt werden konnten.

Nichtsdestotrotz gibt es Unterschiede, zumindest bei schwarzen und weißen Personen in der Beschreibung von Gesichtsmerkmalen (Ellis, Deregowski & Shepherd, 1975). Die Autoren baten schwarze und weiße Personen, Beschreibungen von schwarzen und weißen Gesichtern vorzunehmen, und fanden dabei, dass Weiße häufiger Merkmale wie Haarfarbe und Augenfarbe berichteten, während Schwarze häufiger die Ohren, Kinn oder Augenbrauen erwähnten. Anscheinend werden vor allem jene Merkmale benutzt, die es der jeweiligen Person am besten ermöglichen, Gesichter zu differenzieren. Wenn Betrachter und Stimulusperson dabei der gleichen ethnischen Gruppe angehören, hat der Betrachter dabei einen Vorteil gegenüber einer Person, die einer anderen ethnischen Gruppe angehört, weil er bereits gelernt hat, welche Merkmale am besten dazu geeignet sind, zwischen den Mitgliedern der Eigen-Gruppe zu differenzieren. Palmer und Masling (1969) zeigten beispielsweise, dass schwarze Kinder einen größeren Wortschatz hatten, um die Hautfarbe einer Person zu beschreiben, als weiße Kinder.

1.1.5.2 Gesichts-Schema bzw. Gesichts-Prototyp

Goldstein und Chance (1980) führen den CR-Bias auf die Entwicklung von *Gesichtsschemata* während der Entwicklung eines Kindes zurück. In frühen Jahren der Kindesentwicklung entsteht nach Meinung der Autoren ein Gesichtsschema bzw. ein Gesichts-Prototyp³ für die Eigen-Gruppe des Kindes, das für Mitglieder einer Fremd-Gruppe nicht adäquat ist. Solch ein Gesichtsprototyp entsteht durch wiederholte Konfrontation des Kindes mit Gesichtern der Eigen-Gruppe, in den frühen Jahren vor allem durch die Familienmitglieder, in späteren Jahren durch Kontakte in der Schule und anderen Orten, wo sich das Kind aufhält.

³ Da die Bedeutung des Begriffs *Schema*, vor allem in der neueren Literatur zum verbalen Gedächtnis, von der von Goldstein und Chance (1980) abweicht, soll hier der Begriff *Prototyp* gebraucht werden.

1.1.5.3 Kontakthypothese

Weitere Erklärungsansätze für den Cross-Race Bias geben die Annahmen der *Kontakthypothese* (Chance, Goldstein & McBride, 1975). Es wird vermutet, dass häufigerer Kontakt mit anderen ethnischen Gruppen zu verbesserten Wiedererkennensleistungen führt. Ob dabei alleine die Quantität des Kontaktes eine Rolle spielt, oder ebenfalls die Qualität des Kontaktes, ist schwierig zu entscheiden, da Quantität und Qualität konfundiert sind. Häufiger Kontakt kann dazu führen, dass auch die Qualität des Kontaktes höher ist. Die Einstellungen gegenüber anderen Gruppen können ebenfalls die Quantität des Kontaktes beeinflussen. Personen mit negativen Einstellungen werden die Gegenwart von anderen Personen einer Gruppe mit negativ besetzten Attributen eher meiden und umgekehrt. Die Erhebung der Kontakthäufigkeit sowie von Einstellungen gegenüber anderen Personen mittels Fragebogen ist immer unter Berücksichtigung der Sozialen Erwünschtheit von Fragebogenitems zu sehen.

Eine weitere Möglichkeit, die Kontakthypothese zu untersuchen, besteht darin, Personengruppen zu erfassen, die in Gebieten mit entweder heterogener bzw. homogener Bevölkerungszusammensetzung leben. Wright, Boyd und Tredoux (2003) untersuchten schwarze und weiße Studenten in Südafrika und weiße Studenten in England. Sie fanden sowohl für südafrikanische als auch für englische Studenten, dass weiße Probanden weiße Gesichter besser wiedererkennen konnten als schwarze Gesichter. Überraschenderweise waren die schwarzen südafrikanischen Studenten ebenfalls bei den weißen Gesichtern besser als bei den schwarzen. Wright et al. (2003) erklären dies mit der überproportionalen Häufigkeit weißer Studenten an südafrikanischen Hochschulen im Vergleich zur Gesamtbevölkerung. Die Autoren untersuchten in einer früheren Studie (Wright, Boyd & Tredoux, 2001) Nicht-Studenten und fanden dabei, dass schwarze Südafrikaner schwarze Gesichter besser wiedererkennen konnten als weiße Gesichter. Weimann, Fishman und Rattner (1986) verglichen die Wiedererkennensleistung ashkenasischer Juden, sephardischer Juden, israelischer Araber und Kanadier bei ashkenasischen, sephardischen und arabischen Gesichtern. Die kanadischen Teilnehmer waren sowohl bei Falschidentifikationen als auch bei Nichtidentifikationen schlechter als die anderen drei Gruppen. Die Autoren verwenden allerdings den Begriff *nationale Distanz* anstelle des Begriffs *Kontakthypothese*. Die beiden Begriffe sind jedoch nicht klar voneinander zu trennen. Da aber der Versuchsplan einem Test der Kontakthypothese entspricht, wird diese Untersuchung als Indiz für die Gültigkeit der Kontakthypothese angesehen.

Dunning, Li und Malpass (1998) untersuchten die Qualität des Kontaktes, indem sie annahmen, dass Basketballfans aufgrund ihrer Vorliebe für den Sport ein großes Interesse an den Spielern haben und daher bessere Wiedererkennensleistungen zeigen als Nicht-Basketballfans. Sie unterteilten weiße 'Experten' (Fans) und schwarze Teilnehmer sowie weiße 'Laien' anhand eines Fragebogens und präsentierten ihnen Bilder mit weißen und schwarzen männlichen Gesichtern. Es zeigten sich keine Unterschiede zwischen den weißen Experten und den schwarzen Teilnehmern bezüglich der schwarzen Gesichter. Hingegen waren die weißen Laien bei schwarzen Gesichtern signifikant schlechter. Bezüglich der weißen Gesichter zeigten sich keine Unterschiede. Eine weitere Theorie über die Verarbeitung von Gesichtern wird im folgenden Abschnitt beschrieben.

1.1.5.4 Kognitive Missachtung und unterschiedliche Verarbeitungstiefen

Die Kontakthypothese geht davon aus, dass Quantität und Qualität des Kontakts eine wichtige Rolle bei der Fähigkeit zum Wiedererkennen von Gesichtern spielen. Einen Schritt weiter gehen Theorien, die die kognitive Verarbeitung von Gesichtern zum Inhalt haben und Fragen zur Aufmerksamkeit bzw. Beachtung externer Stimuli thematisieren. So hat die psychologische Forschung gezeigt, dass Menschen bei der Informationsverarbeitung, sei es nun bei der sozialen Informationsverarbeitung oder in der reinen Wahrnehmung externer Stimuli, Heuristiken anwenden, die zum einen die Gegebenheiten vereinfachen und zum anderen nicht alle Aspekte einer Gegebenheit beachten (z.B. *Repräsentativitäts-Heuristik*, Kahnemann & Tversky, 1973).

Rodin (1987) beschreibt ein Phänomen, das sie *kognitive Missachtung* nennt. Sie bezeichnet damit die Aufmerksamkeitsfokussierung einer Person auf Merkmale einer anderen Person, die es ermöglichen, die andere Person zu kategorisieren, um dann abschätzen zu können, inwieweit es sich lohnt (im Sinne einer Kosten-Nutzen-Rechnung) Kontakt zu dieser Person aufzunehmen. Auf dieser Stufe der Informationsverarbeitung, der Kategorisierung, erscheint die wahrgenommene Person noch nicht als Individuum, sondern als Angehöriger einer bestimmten Gruppe, die sich je nach Art der Kategorisierung (Geschlecht, Alter etc.) zusammensetzt.

Der Ansatz von Malpass (1990) einer utilitaristischen Analyse möglicher Kontakte, d. h. nützt die Aufnahme eines Kontaktes der Person, ist dem Ansatz von Rodin (1987) ähnlich. Malpass (1990) beschreibt in diesem Ansatz, dass der Kontakt mit Personen der

eigenen Gruppe vielfältiger ist und häufiger auftritt als mit Personen anderer Gruppen, daher ist auch der Erfahrungsschatz mit der eigenen Gruppe größer als mit anderen Gruppen. Leyens und Codol (1992) beschreiben eine unterschiedliche Verarbeitung der Eigenschaften von Personen der Eigen-Gruppe und der Fremd-Gruppe. Die Autoren behaupten, dass die Individuen der eigenen Gruppe eine differenziertere Vorstellung der Mitglieder der eigenen Gruppe haben gegenüber Mitgliedern anderer Gruppen. Dieser Ansatz soll hier nur kurz erwähnt sein, um zu verdeutlichen, dass Intergruppenprozesse in vielen Bereichen auftreten.

Der Ansatz unterschiedlicher Verarbeitungstiefen (Craik & Lockhart, 1972) beschäftigt sich mit der Frage, inwieweit eine 'tiefere' Verarbeitung von Gesichtern zu einer verbesserten Wiedererkennensleistung führt. Unterschiedliche Verarbeitungstiefen werden zumeist mittels der verwendeten Instruktionen induziert. So instruierten z. B. Devine und Malpass (1985) die Teilnehmer, entweder die ethnische Gruppenzugehörigkeit von Eigen-Gruppen und Fremd-Gruppen Gesichtern zu bestimmen, die Person als sympathisch bzw. unsympathisch einzuschätzen, oder zu versuchen, die Person im Gedächtnis zu behalten. Die Autoren fanden einen Effekt für die Versuchsinstruktion insofern, dass die Teilnehmer, die lediglich die Gruppenzugehörigkeit bestimmen sollten in einem Wiedererkennenstest am schlechtesten abschnitten, verglichen mit den anderen zwei Instruktionsbedingungen. Sporer (1991) kritisiert bei diesem Ansatz, dass auch die Menge an enkodierter Information mit größerer Verarbeitungstiefe zunimmt und somit für eine bessere Wiedererkennensleistung verantwortlich sein kann. Nichtsdestotrotz betrachtet Sporer (2001a) die in diesem Kapitel kurz vorgestellten Ansätze als interessante Erklärungsmodelle, die näher untersucht werden sollten.

1.1.5.5 Norm-basierte und Exemplar-basierte Modelle

Es wurde bereits in den 70er Jahren angenommen, dass die *Typikalität* eines Gesichtes dessen Wiedererkennbarkeit beeinflusst, in dem Sinne, dass Gesichter, die von einer Norm abweichen, besser wiedererkennbar sind als für eine Norm typische Gesichter (Light, Kayra-Start & Hollander, 1979). Das Gegenteil von Typikalität ist die *Distinktivität* eines Gesichtes und beschreibt den Grad der Abweichung von einer Norm. Ein typisches Gesicht besitzt wenig bzw. sogar keine distinkten Merkmale. So konnte z.B. Endo (Endo, 1986; Endo, 1982, zit. nach Valentine, 1988) zeigen, dass distinkte Merkmale der äußeren Gesichtsteile, z.B. Veränderungen von Haaren und Augen besser erkannt bzw. entdeckt

wurden als Veränderungen von Nase, Mund und Kinn. Ebenso wurde von Myles-Worsley, Johnston und Simons (1988) gezeigt, dass atypische Merkmale von Röntgenbildern besser erkannt werden konnten.

Neuere Studien fanden einen klaren Einfluss der Typikalität von Gesichtern auf deren Wiedererkennbarkeit, d.h. das Ausmaß, in dem ein Gesicht als typisch für eine bestimmte Norm angesehen wird (Valentine, 1991; Brigham, 1990). Vokey und Read (1992) beschreiben eine Zwei-Faktoren-Struktur, mit denen die Typikalität beschrieben werden kann und deren Faktoren mit *kontext-freier Vertrautheit* und *Einprägsamkeit* benannt wurden. Die Autoren verwendeten fünf Skalen, auf denen Gesichter von Highschool-Schülern bewertet wurden: *Typikalität*, *Einprägsamkeit*, *Vertrautheit*, *Attraktivität* und *Sympathie*. Die Skalen *Vertrautheit*, *Attraktivität* und *Sympathie* laden auf dem Faktor *Kontextfreie Vertrautheit*, der von Vokey und Read als Faktor beschrieben wird, der die Gefühle und Emotionen, die mit einem Gesicht verbunden werden, umfasst. Der zweite Faktor *Einprägsamkeit*, zu dem vor allem die Skalen *Einprägsamkeit* und *Typikalität* bzw. *Distinktivität* beitragen, umfasst Einschätzungen der Probanden darüber, wie leicht bzw. schwer es ist, sich an ein Gesicht zu erinnern. Ein ungewöhnliches Gesicht ist nach Vokey und Read besser zu erinnern, weil es hoch auf den Items des Faktors *Einprägsamkeit* lädt und gering auf den Items des Faktors *Vertrautheit*.

Sporer (2001a) stellt in Anlehnung an Valentine (1991a, 1991b) die Frage, ob die tägliche Konfrontation mit einer Vielzahl von Gesichtern zu der Entwicklung einer Norm führt, und ob neue Gesichter in Bezug zu dieser Norm gespeichert werden. Da Gesichter bereits seit frühester Kindheit wahrgenommen werden, bildet sich die Fähigkeit zur Unterscheidung von Gesichtern heraus. Die Wahrnehmung unterschiedlicher Gesichter führt zu der Bildung von Expertenwissen bezüglich dieser Klasse von Stimuli. Diese Gesichter bilden letztendlich Prototypen, die eine bestimmte Gruppe hinreichend charakterisieren sollen. Zu dieser Frage gehören Studien, die die Wiedererkennungsleistung bei Gesichtern untersucht haben und als Stimulusklasse ebenfalls invertierte Gesichter benutzten. Zu einem Gesichts-Prototyp gehört u.a., dass ein Gesicht aufrecht ist, d.h. der Mund liegt unterhalb der Nase und diese wiederum mittig unterhalb der Augen. Bei invertiert dargebotenen Gesichtern sinkt die Wiedererkennenswahrscheinlichkeit gegenüber aufrechten Gesichtern (Yin, 1970), was für die Annahme von Gesichts-Prototypen spricht.

Valentine (1991a) schlug einen *multidimensionalen Raum* für die Speicherung von Gesichtern vor. Nach den Annahmen der *Norm-basierten* und *Exemplar-basierten* Modelle sind Gesichter als Punkte in einem multidimensionalen Raum gespeichert (siehe

Abbildung 3). Die Dimensionen des Raumes ergeben sich aus solchen Merkmalen, die zwischen Gesichtern variieren können (z.B. Haarlänge, Haarfarbe, Nasenlänge etc.) und sind nicht durch das Modell vordefiniert. Demnach kann ein multidimensionaler Raum von Person zu Person unterschiedlich sein, gleichzeitig ist aber die Anzahl der Dimensionen durch die Anzahl der Eigenschaften von Gesichtern begrenzt. Norm-basierte und Exemplar-basierte Modelle unterscheiden sich darin, ob sie eine Norm als Basis des Raumes postulieren oder die einzelnen Exemplare in einem Raum ohne Norm liegen.

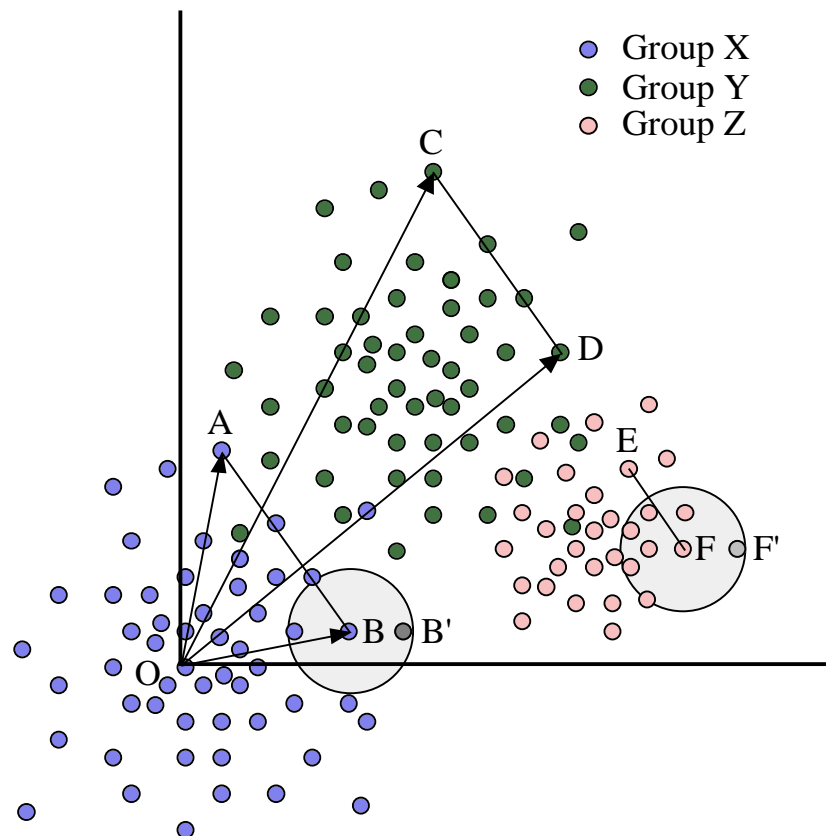


Abbildung 3. Schematische Darstellung der Modellannahmen: In einem multidimensionalen Raum verteilen sich die Gesichter der eigenen ethnischen Gruppe um die Basis (nach Sporer, 2001a).

In Abbildung 3 ist ersichtlich, wie Gesichter in einem multidimensionalen Raum gespeichert werden. In dieser vereinfachten Abbildung ist jedem Gesicht ein Punkt in einem zweidimensionalen Raum zugewiesen, der durch seine Ausprägung auf der jeweiligen Dimension bestimmt ist. Im Norm-basierten Modell liegen die Gesichter der Eigen-Gruppe um den Ursprung des Raumes verteilt, wobei durch die Entwicklung einer

Norm diese gewissermaßen den Prototyp eines Gesichtes bildet. Gesichter anderer Gruppen liegen weiter vom Ursprung entfernt, weil deren Merkmale sich von Merkmalen der Gesichter der Eigen-Gruppe bezüglich der gewählten Dimension unterscheiden können. Auch im Exemplar-basierten Modell liegen Gesichter von Eigen-Gruppe und Fremd-Gruppe voneinander entfernt. Hier gibt es jedoch keine Norm, die gebildet wird, sondern es wird eine *Zentrale Tendenz* postuliert, um die sich die einzelnen Punkte des Raumes analog zu einer Normalverteilung verteilen. Die einzelnen Exemplare, die sich an den Rändern der Normalverteilung befinden (distinkte Gesichter), liegen nach dem Exemplar-basierten Modell und dem Norm-basierten Modell in Bereichen geringerer Dichte (Valentine, 1991a). In Abhängigkeit von der Erfahrung mit einer bestimmten Gruppe, unterscheidet sich ebenfalls die Anzahl der Exemplare für jede Gruppe, so kann z.B. die Eigen-Gruppe mehr Exemplare in diesem multidimensionalen Raum als die Fremd-Gruppe besitzen, Eigen-Gruppen und Fremd-Gruppen können sich aber auch in der Anzahl der Exemplare nicht unterscheiden (Gruppe X und Y).

Im Norm-basierten Modell sind distinkte Merkmale eines Gesichts relativ zur Norm gespeichert, d.h. je distinkter ein Merkmal ist, desto stärker unterscheidet es sich von der Norm, und desto weiter ist die Entfernung eines Gesichtes bezüglich dieses Merkmales vom Ursprung in einem multidimensionalen Raum. Nach Valentine (1991a) bestimmt sich die Ähnlichkeit von Gesichtern durch den Winkel zwischen den Vektoren der Gesichter. Je geringer der Winkel ist, desto ähnlicher sind die Gesichter. In einem Exemplar-basierten Modell wird angenommen, dass Gesichter um einen normal-verteilten Bereich gespeichert werden, wobei distinkte Gesichter eher an der Rändern dieses Bereiches liegen, während typische Gesichter eher im mittleren Bereich repräsentiert sind und somit auch dichter beieinander liegen (Valentine, 1991a). In diesem Falle wären unähnliche Gesichter weit voneinander entfernt, ohne Beachtung einer Norm oder des Winkels ihrer Vektoren. Das Maß für die Ähnlichkeit wäre die euklidische Distanz zwischen den Gesichtern.

Bruce, Burton und Dench (1994) maßen an 89 männlichen und 86 weiblichen Gesichtern verschiedene Distanzen (Breite des Mundes, Länge der Nase) und ließen diese Gesichter bezüglich der Distinktivität beurteilen. Von den Gesichtern waren entweder die Haare verdeckt oder die Gesichter wurden mit den Haaren präsentiert. Anschließend wurden Korrelationen zwischen den Gesichtsmaßen und den Distinktivitätsurteilen berechnet, wobei sich Korrelationen zwischen $r = .16$ und $r = .56$ zeigten. Die größte Korrelation von $r = .56$ wurde bei männlichen Gesichtern, deren Haare verdeckt wurden, gefunden. Die Korrelationen bei weiblichen Gesichtern lagen zwischen $r = .16$ und $r = .38$, wobei die höchste Korrelation ebenfalls bei Gesichtern gefunden wurde, deren Haare

verdeckt wurden. In dieser Studie war eine Korrelation auf einem Alpha-Niveau von .01 signifikant, wenn sie den Wert $r = .28$ übersteigt. Bruce et al. (1994) sehen diese Ergebnisse als Beleg für den Zusammenhang von objektiven Maßen der Bestimmung von Distinktivität von Gesichtern und der wahrgenommenen Distinktivität der Gesichter durch die Probanden.

In Tabelle 1 sind die wichtigsten Annahmen und Vorhersagen des Norm-basierten und des reinen Exemplar-basierten Modells zusammengefaßt. In beiden Modellen ist die Anzahl der Dimensionen nicht spezifiziert. Eine weitere zusätzliche Annahme ist die, dass im Norm-basierten Modell Eigen-Gruppen und Fremd-Gruppen Gesichter die gleiche Dichteverteilung aufweisen, während im Exemplar-basierten Modell die Fremd-Gruppen Exemplare dichter verteilt sind als die Eigen-Gruppen Exemplare. In Abbildung 3 ist ebendalls ersichtlich, wie das Maß der Ähnlichkeit zwischen den Gesichtern festgestellt werden kann. Im Norm-basierten Modell ist der Winkel zwischen zwei Vektoren maßgeblich für die Ähnlichkeit zweier Gesichter, während im Exemplar-basierten Modell die Distanz zwischen zwei Gesichtern entscheidend ist.

Tabelle 1

Vergleich der Annahmen des Norm-basierten und des exemplar-basierten Modells eines multidimensionalen Gesichterraums (nach Sporer, 2001a)

	Norm-basiertes Modell	Exemplar-basiertes Modell
<i>Annahmen</i>		
Dimensionen	<u>n</u> : nicht spezifiziert	<u>n</u> : nicht spezifiziert
Speicherung	Exemplare + Norm	Exemplare
Gruppen-Norm?		
In-Group	Ja: In-group Norm	Nein: 'zentrale Tendenz'
Out-Group	Nein	Nein: 'zentrale Tendenz'
Dichte der Verteilung	Gleich für In- und Out-Group	Out-Group dichter
Maß der Ähnlichkeit	Winkel/Vektor, abhängig von Norm	Distanz zwischen den Exemplaren
<i>Vorhersagen</i>		
Out-Group Homogenität	Ja	bei dichterem Verteilung
Inversions Effekt	Stärker bei In-Group	Stärker bei Out-Group
Distinktivität der Gesichter		
Rec-task	Drei-fach Interaktion	Haupteffekt 'Distinktivität'
Own-sex bias	Bei own-sex Norm	Bei mehr Kontakt mit eigenem Geschlecht

Es wird ebenfalls angenommen, dass die Speicherung von Gesichtern von *Rauschen*⁴ beeinflusst wird und auch durch *Rauschen* überlagert werden kann (Valentine, 1999; Sporer, 2001a).

⁴ Der Begriff ‚Rauschen‘ stammt aus der Signal-Entdeckungs-Theorie (siehe Abschnitt 1.2.1) und bedeutet in diesem Zusammenhang, dass ein Gesicht nicht eindeutig gespeichert werden kann, da bestimmte Merkmale des Gesichtes anderen Gesichtern sehr ähnlich sind.

Levin (1996, 2000) postuliert, dass die ethnische Gruppenzugehörigkeit als visuelles Merkmal als Teil der Erklärung des Cross-Race Bias dient. Seiner Meinung nach betonen Personen die Informationen, die eine andere Person als Mitglied einer Fremd-Gruppe kategorisieren lassen und legen dabei weniger Wert auf Informationen, die einer Differenzierung von Personen dienen. Die visuelle Suche nach einem 'positiven' Merkmal, d.h. nach einem Merkmal, das einer Kategorisierung dient, ist nach Levin (1996) schneller als die Suche nach einem 'negativen' Merkmal. Der Autor untersuchte diese Annahme, indem er u.a. weißen Versuchspersonen ein schwarzes Gesicht mit mehreren weißen Gesichtern darbot ('positives' Merkmal) bzw. ein weißes Gesicht mit mehreren schwarzen Gesichtern ('negatives' Merkmal). Die weißen Teilnehmer fanden das schwarze Gesicht ('positives' Merkmal) schneller als das weiße Gesicht ('negatives' Merkmal). Diese *Search Asymmetry* (Treisman & Gormican, 1988, zit. Nach Levin, 2000) besagt, dass die Suche nach 'herausstechenden' Merkmalen einfach ist, weshalb Fremd-Gruppen Gesichter zunächst aufgrund ihrer Gruppenzugehörigkeit wahrgenommen werden. Dieser Ansatz wird im folgenden Abschnitt genauer beschrieben.

1.1.5.6 Das In-Group/Out-Group Modell

Sporer (2001a) nimmt an, dass der erste Schritt in der Wahrnehmung eines neuen Gesichts in der sozialen Wahrnehmung dieses Gesichts liegt. In Abhängigkeit von dieser Wahrnehmung postuliert das In-Group/Out-Group Modell der Gesichterverarbeitung von Sporer (2001a) zwei unterschiedliche Verarbeitungswege, je nachdem ob das zu verarbeitende Gesicht der Eigen-Gruppe oder der Fremd-Gruppe zugehörig eingestuft wird (siehe Abbildung 4). Gehört das wahrgenommene Gesicht der Eigen-Gruppe an, erfolgt der normale Verarbeitungsweg. Die wahrnehmende Person verarbeitet das Gesicht ganzheitlich, automatisch und enkodiert das Gesehene konfigural. Wird hingegen das Stimulusgesicht als einer Fremd-Gruppe zugehörig eingestuft, so werden andere Prozesse angestoßen, die die Verarbeitung des Gesichtes beeinflussen. Dazu gehört beispielsweise die in Abschnitt 1.1.2.3 beschriebene Stereotypisierung der Person.

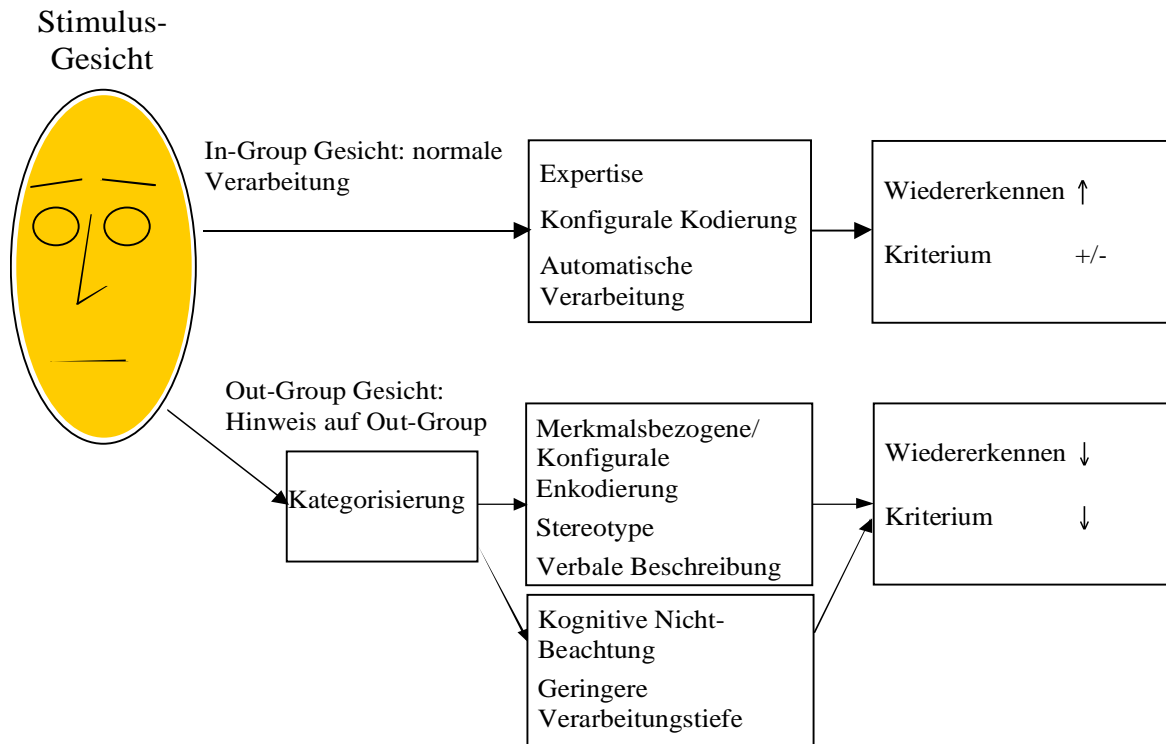


Abbildung 4. Verarbeitungswege von Eigen-Gruppen und Fremd-Gruppen Gesichtern (nach Sporer, 2001a)

Bei Eigen-Gruppen Gesichtern fällt der erste Schritt weg, d.h. es findet keine Kategorisierung des Gesichtes statt, da es ja zur eigenen Gruppe gehört. Die Kategorisierung kann durch jedes Merkmal geschehen, das ein Fremd-Gruppen Gesicht von einem Eigen-Gruppen Gesicht unterscheidet. Unterstützung erfährt dieses Modell durch Untersuchungen von MacLin und Malpass (2001, 2003), die bei der Entwicklung eines Stimulus-Sets auf einen Wahrnehmungseffekt stießen. Die Autoren generierten mit Hilfe des Programms 'Faces' (InterQuest inc. 1998) künstliche Gesichter, deren Merkmale bis auf die Frisur identisch waren. Diese Gesichter wurden dann mit Frisuren versehen, die entweder 'typisch' latein-amerikanisch oder 'typisch' afro-amerikanisch waren. Die Versuchspersonen wurden aufgefordert, diese Gesichter einer der beiden Ethnien zuzuordnen. Bei freier Antwortmöglichkeit (MacLin & Malpass, 2003), wurden 90 % der Gesichter mit afro-amerikanischen Frisuren als afro-amerikanisch kategorisiert; hatten dieselben Gesichter eine latein-amerikanische Frisur, wurden 80 % der Gesichter als latein-amerikanisch eingestuft. Die Autoren berichten ebenfalls, dass die Gesichter, je nach

Kategorisierung als latein- bzw. afro-amerikanisch anders wahrgenommen und den Gesichtern auch andere Eigenschaften zugeschrieben wurden. So wurden die als latein-amerikanisch eingestuften Gesichter als unterwürfiger beschrieben als die afro-amerikanischen, und obwohl alle Gesichter bis auf die Frisur identisch waren, wurde die Hautfarbe der als afro-amerikanisch eingestuften Gesichter als dunkler wahrgenommen. MacLin und Malpass (2001) führten mit diesen Gesichtern ein Wiedererkennungsexperiment durch, allerdings nur mit latein-amerikanischen Teilnehmern, und fanden dabei einen Cross-Race Bias insofern, dass latein-amerikanische Gesichter einen höheren Wert für den Leistungsindex A' aufwiesen als die afro-amerikanischen Gesichter. Der Index A' ist Maß für die Wiedererkennensleistung und wird in Abschnitt 1.2.1.1 ausführlicher besprochen.

1.2 Methoden in der Forschung zur Gesichtererkennung

Die theoretische Grundlage zur Auswertung von Experimenten in der Gesichterwiedererkennung bildet die Signal-Entdeckungstheorie, die im folgenden Kapitel näher beschrieben wird. Dabei wird zunächst ein Überblick über die Grundlagen der SDT gegeben, anschließend werden die verwendeten Maße und deren Berechnung dargestellt.

1.2.1 Signal-Entdeckungstheorie

Ursprünglich stammt die *Signal-Entdeckungstheorie* (engl. *signal detection theory*, im folgenden SDT genannt) aus der Entwicklung von Radar-Systemen während des zweiten Weltkrieges (Swets, 1964). Es war damals noch sehr diffizil, zu erkennen, ob Gegenstände, die von einem Radargerät erfasst wurden, als gefährlich (wie z.B. feindliche Flugzeuge), oder ungefährlich eingestuft werden konnten. Es war und ist nötig, ein Signal, wie im Falle des Radars ein Flugzeug, von einem Nicht-Signal zu unterscheiden.

Bezogen auf die Wahrnehmungspsychologie bedeutet dies, dass die Wahrnehmung eines Stimulus (Signal) durch sogenanntes *Rauschen* überlagert sein kann. Unter dem Begriff *Rauschen* werden alle anderen Umgebungseinflüsse bezeichnet, die auf das Individuum einwirken und vom Individuum als Signal fehlinterpretiert werden können (Goldstein, 1989). Als Beispiel sei hier die Wahrnehmung von Tönen genannt. So ist es möglich, einen Ton nicht zu erkennen, obwohl er präsent ist, weil er von anderen Tönen

überlagert ist, oder auch die physikalische Reizschwelle noch nicht überschritten hat, um vom menschlichen Ohr identifiziert zu werden. Andererseits kann sehr wohl ein Ton als vorhanden 'wahrgenommen' werden, obwohl er physikalisch nicht existiert. Im ersten Fall spricht man in der SDT von *Miss*, d.h. ein Signal wird nicht entdeckt trotz seines Vorhandenseins, im zweiten Fall ist die Rede von einem *Fehlalarm (FA)*, da ein Signal 'entdeckt' wurde, welches jedoch nicht existiert. Tabelle 2 gibt einen Überblick über mögliche Ergebnisse in der SDT.

Tabelle 2

Antwortmöglichkeiten nach der Signal-Entdeckungstheorie

Stimulus	Reaktion	
	Ja	Nein
Signal präsent	Treffer (Hit)	Falsche Zurückweisung (Miss)
Signal nicht Präsent	Falscher Alarm (False Alarm)	Korrekte Zurückweisung (Correct Rejection)

Es gibt zwei richtige Entscheidungen. Eine 'Ja'-Antwort auf einen dargebotenen Ton hin ist ein Treffer (Hit), und eine 'Nein'-Antwort auf einen nicht-existenten Ton eine *korrekte Zurückweisung (Correct Rejection)*. In diesem Zusammenhang ist es wichtig zu wissen, ab welchem Punkt von dem Probanden eine 'Ja'-Antwort gegeben wird, bzw. ab wann der Proband sagt, dass kein Signal vorliegt. Frühe Theorien gingen davon aus, dass es eine sog. *absolute Schwelle* gibt, bei der jede Person sagt, ein Signal sei vorhanden (Fechner, 1860). Diese Theorien wurden jedoch später in Frage gestellt. Heutzutage wird davon ausgegangen, dass jede Person mit einer bestimmten Wahrscheinlichkeit in Abhängigkeit von der wirklichen Signalintensität diese Signale wahrnimmt. Abbildung 5 zeigt zwei Verteilungen. Die linke Verteilung gibt die Wahrscheinlichkeit an, dass eine 'Ja' - Antwort erfolgt, unter der Voraussetzung, dass kein Signal, sondern lediglich 'Rauschen' präsent war. Die rechte Verteilung hingegen gibt die Wahrscheinlichkeit einer 'Ja' - Antwort unter der Voraussetzung eines präsenten Signals an (mit gleichzeitiger Präsenz des Rauschens).

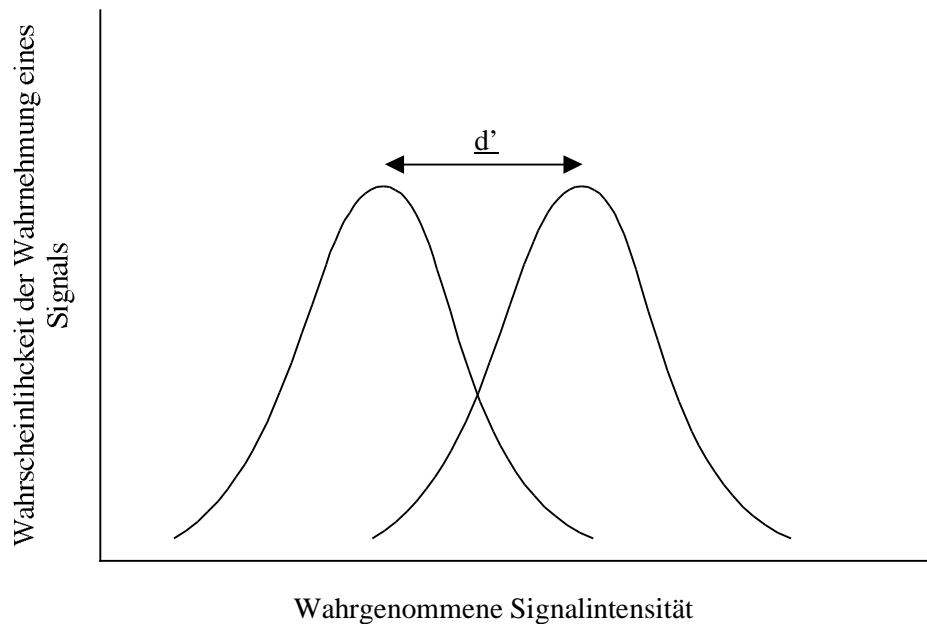


Abbildung 5. Die linke Verteilung ist die Verteilung einer Ja-Antwort unter Rauschen, die rechte Verteilung gibt die Verteilung einer Ja-Antwort unter Rauschen + Signal an.

Bei einer Person, die in der Lage ist, gut zu differenzieren, ob ein Signal vorliegt (Signal + Rauschen) oder nicht (Rauschen), wären die beiden Verteilungen weit auseinander. In diesem Falle wäre der sich überschneidende Bereich zwischen den Verteilungen relativ klein. Das Gegenteil wäre der Fall bei einer Person, deren Differenzierungsfähigkeit weniger stark ausgeprägt ist. Bei ihr lägen die beiden Verteilungen näher beieinander und die Wahrscheinlichkeit falscher Antworten, d.h. eines Fehlalarms oder einer falschen Zurückweisung wäre dementsprechend größer. Die Antworten eines Probanden hängen aber nicht nur von der Fähigkeit des Probanden ab, die An- bzw. Abwesenheit von Signalen zu erkennen (es wird an dieser Stelle nicht näher auf die verschiedenen Arten von Stimuli oder andere Faktoren eingegangen, die einen Einfluss auf die Leistungen haben und experimentell variiert bzw. kontrolliert werden). Zusätzlich zu dieser Fähigkeit muss auch die Reaktionsneigung eines Probanden betrachtet werden. Damit ist die Neigung von Personen gemeint, in der Tendenz der Antworten eher 'Ja' bzw. 'Nein' zu sagen. Auf diesen Aspekt und die Leistungsmessung in der Signal-Entdeckungstheorie wird im folgenden Abschnitt näher eingegangen.

1.2.1.1 Leistungsmaß A' und Reaktionsneigung B''

Beim Wiedererkennenstest und bei der Delayed-Matching Aufgabe (siehe Abschnitt 1.2.2) ist es möglich, die Daten nach der Signal-Entdeckungs-Theorie auszuwerten. Je nachdem, ob für die verwendeten alten und neuen Items (Wiedererkennenstest) bzw. identischen und nicht-identischen Items (Delayed-Matching Aufgabe) Normalverteilungen mit gleicher Varianz angenommen werden oder nicht, können parametrische bzw. nicht-parametrische Maße des Leistungsindex berechnet werden (Sporer, 1992). Bei einer Normalverteilungsannahme wird üblicherweise d' als Maß herangezogen. Ein Wert von $d' = 0$ entspricht einer Leistung auf Zufallsniveau. Sporer (1992) gibt an, dass unter optimalen Lernbedingungen d' auf Werte von 3 und mehr steigen kann. Da die Annahme der Normalverteilung der Items mit gleichen Varianzen nicht immer gegeben ist und die manchmal geringe Anzahl an Items nicht ausreicht, um hinreichende Verteilungen zu produzieren, werden in Wiedererkennensexperimenten meistens nicht-parametrische Maße berechnet, die keine bestimmte Verteilung der Items voraussetzen. Dabei können zwei Maße berechnet werden. Zum einen der Leistungsindex A' (Rae, 1976), zum anderen die Reaktionsneigung B'' (Donaldson, 1992; Grier, 1971).⁵ Zur Berechnung des Leistungsindex benötigt man die Anzahl der Treffer, d.h. die Anzahl der richtigen Identifikationen sowie die Anzahl der Fehllalarme, und somit die Anzahl der falschen Identifikationen. Ein hoher Wert von A' ist ein Indikator für eine gute Diskriminationsleistung. Dies bedeutet, ein Proband hätte in einem Test eine hohe Anzahl von Treffern bei einer geringen Anzahl an Fehllarmen. A' variiert dabei zwischen den Werten 0 und 1. Ein Wert für A' von .5 bedeutet hierbei eine Diskriminationsleistung auf dem Zufallsniveau. Zu beachten ist dabei, dass sich die Formeln für die Berechnung von A' unterscheiden, je nachdem, ob die Anzahl der Treffer höher ist als die der falschen Alarme oder umgekehrt.

Das Maß B'' ist das Maß für die Reaktionsneigung eines Probanden. Die Teilnehmer an einem Wiedererkennenstest, bzw. an einer Delayed-Matching Aufgabe können sich in der Art und Weise unterscheiden, in der sie ihre Antworten auf die Stimuli

⁵
$$A' = \frac{(H^2 + F^2 + 3H - F - 4FH)}{(4H(1 - F))} \quad \text{falls } \# \text{ Treffer} > \# \text{ falsche Alarme}$$

$$A' = \frac{(H - H^2 + F - F^2)}{(4H(1 - F))} \quad \text{falls } \# \text{ Treffer} < \# \text{ falsche Alarme}$$

$$B'' = \frac{((1 - H)(1 - F) - HF)}{((1 - H)(1 - F) + HF)} \quad \text{Anmerkung: 'H' Hits; 'F': Falsche Alarme}$$

geben. Dabei kann es vorkommen, dass eine Person im Falle der Unsicherheit über eine Antwort dazu neigt, eine Antwortoption zu bevorzugen. Probanden, die bei Unsicherheit dazu neigen, mit 'Ja' zu antworten, werden zum einen eine hohe Trefferrate erhalten, zum anderen wird aber auch die Anzahl der Fehllarme ansteigen. Solch ein Antwortverhalten bezeichnet man auch als *liberales Antwortkriterium*. Wenn ein Proband hingegen zu 'Nein'-Antworten neigt, sinkt dementsprechend die Anzahl der Fehllarme, die Anzahl der Treffer wird jedoch ebenfalls sinken. In diesem Fall spricht man von einem *konservativen Antwortkriterium*. Positive Werte von \underline{B}'' sind Indikatoren für ein konservatives Antwortkriterium, negative Werte bedeuten, dass der Proband ein liberales Antwortkriterium zeigt. \underline{B}'' variiert zwischen -1 und 1 (siehe Abbildung 6). Auf den beiden Achsen sind jeweils die Wahrscheinlichkeiten abgetragen, mit der eine Person positiv, d.h. mit einer Ja-Antwort, in einem Durchgang reagiert. Auf der x-Achse ist die Wahrscheinlichkeit einer Ja-Antwort bei nicht existentem Stimulus⁶ abgetragen ($P(S/N)$, Falscher Alarm), auf der y-Achse die Wahrscheinlichkeit einer positiven Antwort bei einem existenten Stimulus ($P(S/S)$, Treffer). Je besser die Differenzierungsleistung ist, d.h. je mehr Treffer (hohe Wahrscheinlichkeit auf der y-Achse, $P(S/S)$), und je weniger falsche Alarmer (geringe Wahrscheinlichkeit auf der x-Achse, $P(S/N)$), desto größer wird die Fläche unterhalb der beiden Linien, die sich in Punkt x treffen. Bei einer perfekten Leistung von $\underline{A}' = 1$, würde diese Fläche dementsprechend die gesamte Fläche des Diagramms einnehmen. Zu beachten ist hierbei die Reaktionsneigung. Liegt der Punkt x auf der Nebendiagonalen von Punkt $P(S/S) = 1; P(S/N) = 0$ zu Punkt $P(S,S) = 0; P(S/N) = 1$, ist von einer Reaktionsneigung zu sprechen, die $\underline{B}'' = 0$ beträgt, da weder 'Ja'- noch 'Nein'-Antworten übermäßig gegeben werden. Ist jedoch eine Antworttendenz in Richtung mehr 'Ja'- als 'Nein'-Antworten gegeben, verschiebt sich die Linie, die vorher durch die Diagonale gebildet wurde, nach rechts oben.

⁶ In einem Gesichterwiedererkennungsexperiment wäre dies eine positive Antwort des Probanden auf ein Gesicht, das er nicht vorher gesehen hat. Für A' wurde hier die Formel von Rae (1971), für B'' die von Donaldson (1992) verwendet.

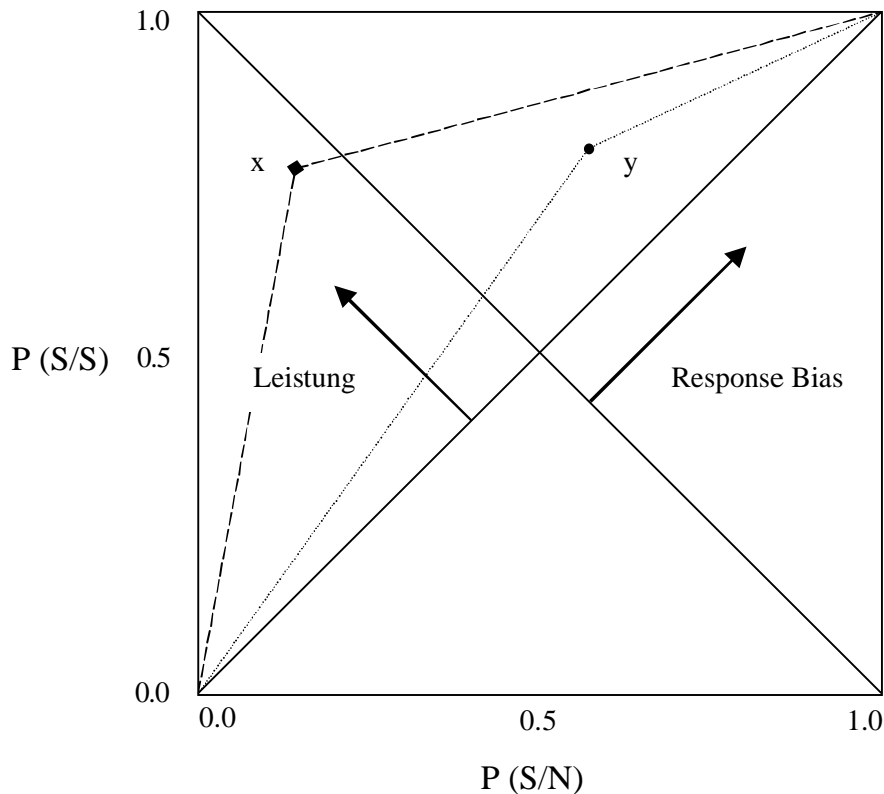


Abbildung 6. Graphische Darstellung der Parameter der Signal-Entdeckungs-Theorie (nach Sporer, 2001a).

In der Forschung zum Cross-Race Bias zeigt sich idealerweise eine Interaktion zwischen Stimulusethnie und der Ethnie der Probanden beim Leistungsmaß A' (Bothwell et al., 1989; Lindsay & Wells, 1983). In Abbildung 6 ist dies durch die beiden Punkte x und y gekennzeichnet. Der Punkt x ist in diesem Fall die Leistung einer Person bei der Eigen-Gruppe, mit einem hohen Wert von A' und einer Reaktionsneigung, die nahe an der Nebendiagonalen liegt. Zusätzlich konnten einige Autoren (Sporer, 1992; Slone, Brigham & Meissner, 2000) zeigen, dass bei Gesichtern anderer ethnischer Gruppen ein eher liberales Antwortkriterium angelegt wird. Dies zeigt sich darin, dass B'' bei Fremdgruppengesichtern zwischen 0 und -1 liegt. Dieser Umstand ist durch den Punkt y dargestellt. Die Leistung einer Person bei der Fremd-Gruppe ist gekennzeichnet durch einen geringeren Wert von A' (die Fläche, die unter dem Schnittpunkt der beiden Linien liegt, ist kleiner als bei Punkt x) und einer Reaktionsneigung, die weit von der Diagonalen entfernt liegt mit einer Tendenz der Person, die Existenz eines Stimulus zu bejahen, bzw. in einem Wiedererkennensexperiment zu sagen, dass ein dargebotenes Gesicht bereits vorher gesehen wurde.

Durch die systematische Veränderung der Reaktionsneigung lassen sich sog. *Receiver Operating Characteristics (ROC)* bilden, die bei konstantem \underline{A}' eine Kurve in der zwei-dimensionalen Fläche analog zur Abbildung 6 formen (siehe Abbildung 7).

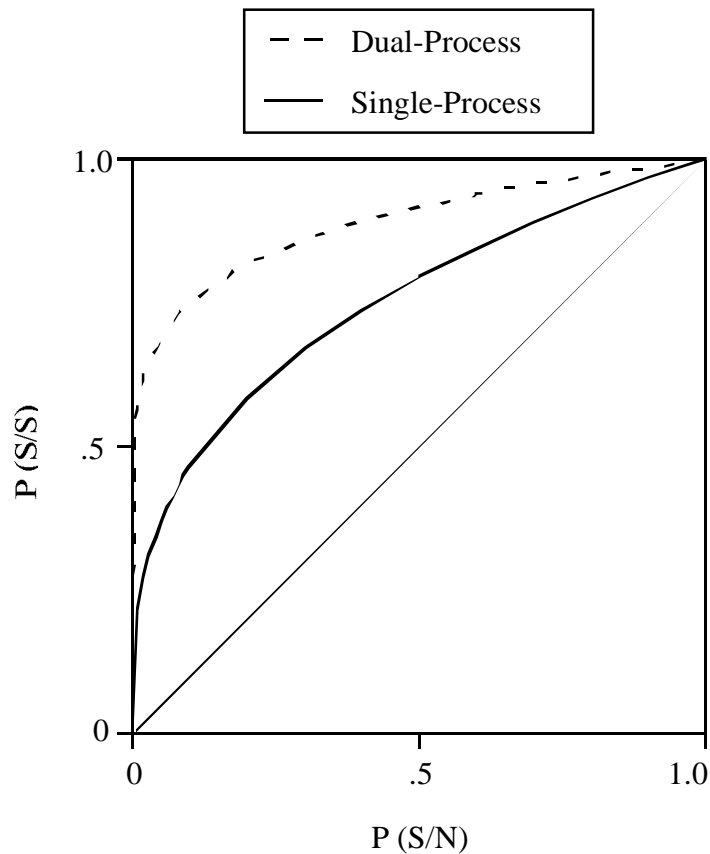


Abbildung 7. Schematische Darstellung zweier Receiver Operating Characteristics (ROCs).

Solche Kurven dienen der Veranschaulichung der Wiedererkennensleistung von Probanden, legen aber auch gleichzeitig ein Problem offen, das im folgenden Abschnitt diskutiert wird. Es geht dabei um die Frage, ob ein einzelner Faktor für das Wiedererkennen verantwortlich ist, oder ob zwei Faktoren beim Wiedererkennen eine Rolle spielen und dabei interagieren.

1.2.1.2 Single-process und Dual-process Ansätze in Wiedererkennungs-experimenten

Es gibt zwei Ansätze, die sich mit der Frage auseinandersetzen, welche Gedächtnisprozesse bei Wiedererkennungs-experimenten eine Rolle spielen. Sie unterscheiden sich in ihren Annahmen, ob ein Faktor (single-process, Vertrautheit mit einem Item) hinreichend ist, um die Effekte in Wiedererkennungs-experimenten zu erklären, oder ob die Wiedererkennung von zwei Faktoren (dual-process, Vertrautheit mit einem Item und Erinnerbarkeit des Items) beeinflusst wird, die unabhängig voneinander sind. Beide Modelle werden herangezogen, um den sogenannten *mirror effect* (Glanzer & Adams, 1985) zu erklären. Die Autoren fanden bei einer Durchsicht von 80 Wiedererkennungs-experimenten bei verschiedenen Klassen von Stimuli das gleiche Muster bezüglich der Treffer und der falschen Alarme. Dabei wird bei einer Stimulusklasse (im Zusammenhang mit dieser Arbeit wären dies Eigen-Gruppen Gesichter) eine höhere Anzahl an Treffern gefunden bei einer geringen Anzahl an falschen Alarmen, während bei einer anderen Stimulusklasse (Fremd-Gruppen Gesichter) dieses Verhältnis 'gespiegelt' wird. Demnach ist bei Fremd-Gruppen Gesichtern die Anzahl der Treffer gegenüber den Eigen-Gruppen Gesichtern geringer bei gleichzeitiger größerer Anzahl an falschen Alarmen gegenüber der Eigen-Gruppe. Meissner und Brigham (2001) schlagen vor, dass der mirror-effect durch einzelne single-process Modelle ausreichend erklärt werden kann. Die Vertrautheit von Personen mit Gesichtern der eigenen ethnische Gruppe führt danach zu einer effizienteren Speicherung von Eigen-Gruppen Gesichtern. Gesichter anderer ethnischer Gruppen werden weniger effizient gespeichert, da weniger adäquate Merkmale für die Speicherung genutzt werden.

Der single-process Ansatz wird durch die Signal-Entdeckungs-Theorie hinreichend beschrieben (Yonelinas, 2002), indem die Vertrautheit mit alten und neuen Items zwei Normalverteilungen der alten und neuen Items bildet, die sich an den Grenzen überlappen, und bei denen ein Antwortkriterium gesetzt wird, das möglichst gut alte von neuen Items trennen soll. Wenn dieses Antwortkriterium so gesetzt wird, dass die Person bei der Aufgabenbeantwortung rät, erhielte man die Diagonale von Punkt 0/0 zu Punkt 1/1 in Abbildung 7. Bei jeder Leistung, die besser als die Ratewahrscheinlichkeit ist, entsteht eine Kurve in der ROC. Die Kritik von Yonelinas basiert auf den empirischen Ergebnissen einiger Studien, die zu einer anderen ROC führen, als dies nach single-process Modellen der Fall sein sollte. Nach single-process Modellen müsste eine ROC Kurve symmetrisch

und kurvilinear sein, wie in Abbildung 7 dargestellt (untere Kurve). Yonelinas berichtet hingegen von ROC-Kurven, deren Form zwar kurvilinear ist, jedoch asymmetrisch zur Diagonalen in Abbildung 7 verläuft (obere Kurve). Er plädiert deshalb für einen dual-process Ansatz, in dem neben der Vertrautheit noch eine weitere Komponente aufgenommen wird. Andere Autoren ziehen ebenfalls dual-process Modelle für die Erklärung des mirror-effects heran (Joordens & Hockley, 2000; Hockley, Hemsworth & Consoli, 1999; Reder, Angstadt, Cary, Erickson & Ayers, 2002). Yonelinas nennt diese Komponente in Anlehnung an Tulving (1985) 'Recollection' (Erinnerung) und meint damit Prozesse, die unabhängig von der Vertrautheit mit einer Stimulusklasse ablaufen. Eine Person kann sich z.B. an Informationen erinnern, die mit dem gelernten Item während der Lernphase assoziiert wurden (Gegenstände im Raum, ein bestimmtes Wort etc.). Wenn es gelingt, diese Informationen abzurufen und durch eine Assoziation das gelernte Item ebenfalls, hat dieser Prozess nichts mit der Vertrautheit mit der Stimulusklasse zu tun. Ebenso ist dieser Prozess mit einer 'alles oder nichts' Antwort verbunden, d.h. mit einer Schwelle, ab der ein Item erinnert wird oder nicht. Gelingt eine Verbindung von anderen Informationen mit dem gelernten Item, wird das Item auch erinnert, gelingt dies nicht, kann eine Erinnerung und damit eine erfolgreiche Identifizierung nicht erfolgen. Die ROC der reinen Erinnerung ist keine Kurve, sondern eine Gerade, da ein festes Antwortkriterium gegeben ist (Erinnerung an das Item vs. keine Erinnerung an das Item). Yonelinas (1997) testete diese Annahme in einem Experiment, in dem er Wortpaare lernen ließ und anschließend die Teilnehmer zwischen den gelernten Wortpaaren und neu zusammengestellten Wortpaaren diskriminieren ließ. Seiner Ansicht nach ist bei dieser Aufgabe der Einfluss der Vertrautheit sehr gering, da keine neuen Wörter im Wiedererkennentest vorkamen und daher die Wiedererkennensleistung auf die reine Erinnerung an die Wortpaare zurückzuführen sein muss. Tatsächlich fand Yonelinas eine lineare ROC, wie auch in anderen Experimenten von Yonelinas (1999), Rotello, Macmillan und Van Tassel (2000) sowie Meissner und Brigham (in Druck).

1.2.2 Testparadigmen

In der Forschung zum Wiedererkennen bzw. zur Verarbeitung von Gesichtern werden vor allem drei Paradigmen verwendet: Die *Delayed-Matching-Aufgabe*, der *Alternative-Forced-Choice Test (AFC)* und der *Ja/Nein -Test*. Der übliche Vorgang ist, dass den Teilnehmern eine bestimmte Anzahl an Gesichtern der eigenen Gruppe und einer

oder mehrerer anderer Gruppen dargeboten werden. Dies kann in geblockter oder gemischter Reihenfolge geschehen. Der Teilnehmer erhält vorher eine Instruktion, worauf er in der Präsentationsphase achten soll. Jedes einzelne Bild wird für einen bestimmten Zeitraum (500 ms, 2000 ms etc.) dargeboten. Im Anschluss an die Darbietung der einzelnen Gesichter wird eine Aufgabe gestellt, die die Aufmerksamkeit von den Gesichtern ablenken und einen möglichen Recency Effekt vermeiden soll. Normalerweise werden die Teilnehmer vor dem Experiment nicht darüber informiert, dass der Darbietung eine Testphase folgt. Das in den Pilotstudien in den Abschnitten 3 und 5 verwendete Design diente der Erfassung von Daten zur weiteren Planung der Experimente in den Abschnitten 4 und 6 und bezieht sich nicht auf eines der drei im folgenden beschriebenen Paradigmen.

Die meisten Studien benutzten in ihren Experimenten zum Cross-Race Bias einen Wiedererkennentest im Sinne des Ja/Nein - Tests (Meissner & Brigham, 2001). Der Cross-Race Bias konnte aber auch bei Matching-Aufgaben gefunden werden (Lindsay, Jack & Christian, 1991; Malpass, Erskine & Vaughn, 1988, zit. nach Meissner & Brigham, 2001, Sporer, Trinkl & Guberova, im Druck).

Die drei Paradigmen beanspruchen ebenfalls andere Gedächtnisprozesse. Bei der Delayed-Matching Aufgabe wird von dem Teilnehmer ein Abgleich zweier kurz hintereinander dargebotener Bilder verlangt und somit eher das Kurzzeitgedächtnis beansprucht. Bei der AFC-Aufgabe hingegen müssen Bilder mit einer internen Repräsentation des richtigen Stimulus verglichen und anschließend entschieden werden, welches der Bilder mit dieser internen Repräsentation übereinstimmt. Diese Aufgabe erfordert neben Prozessen der Wahrnehmung eines Stimulus, eine Beanspruchung des Langzeitgedächtnisses und darüber hinaus eine Beanspruchung des Arbeitsgedächtnisses.

1.2.2.1 Delayed-Matching-Aufgabe

Bei der Delayed-Matching-Aufgabe wird der Versuchsperson zunächst ein Stimulus für einen bestimmten Zeitraum, beispielsweise 2000 ms, präsentiert. Anschließend folgt ein sogenannter *Filler-Screen*, der den Stimulus so lange verdeckt, bis das Netzhaut-Bild vollständig verschwunden ist. Im dritten Schritt wird entweder der gleiche Stimulus oder ein anderer Stimulus präsentiert, der von der Versuchsperson mit dem ersten Stimulus verglichen werden soll. Die Eingabe der Versuchsperson ist dabei im Allgemeinen in der Form 'Stimulus ist identisch' und 'Stimulus ist nicht identisch'. Das

Abgleichen (Matching) von Gesichtern, insbesondere von Fotos ist nicht auf Experimente im Labor beschränkt. Auch im Alltagsleben werden Bilder und Personen abgeglichen, so z.B. Bilder auf Führerscheinen, Studiausweisen und eben auch bei grenzüberschreitendem Verkehr. Dass auch ein direkter Abgleich von Personen mit Fotos nicht zu eindeutigen und fehlerfreien Identifikationen führt, zeigt eine Studie von Kemp, Towell und Pike (1997). Die Autoren untersuchten die Frage, inwieweit Fotos auf Kreditkarten mit den Überbringern durch Andere abgeglichen werden können. Dazu ließen sie für jeden der 'Überbringer' Kreditkarten drucken, die sich in vier experimentell variierten Merkmalen unterschieden: 1) die Karte zeigt das Gesicht des Trägers so wie er am Tag des Experimentes erscheint, 2) das Erscheinungsbild ist leicht verändert (Brille etc.), 3) ein anderes Gesicht ist auf der Karte, welches dem Träger jedoch ähnlich sieht und 4) ein völlig unähnliches Gesicht im Vergleich zum Träger ist auf der Karte abgedruckt. Anschließend 'kauften' die Träger der Kreditkarten in einem Supermarkt nach Ladenschluss an sechs Kassen mit den Kreditkarten ein, wobei die Kassierer die Anweisung hatten, die Kreditkarte aufgrund des Vergleichs des Fotos mit dem Überbringer anzunehmen oder zurückzuweisen. Kemp et al. (1997) berichten den durchschnittlichen Anteil korrekter Antworten der einzelnen Kassierer für die vier Kreditkarten. Insgesamt lag der Anteil richtiger Antworten bei 67.4 % (richtige Antworten bedeutet, dass die Karten 1) und 2) akzeptiert und die Karten 3) und 4) zurückgewiesen wurden). Das im Kontext dieser Dissertation jedoch wichtigere Ergebnis ist, dass 63.6 % der Karten, die ein dem Träger ähnliches Gesicht (Bedingung 3) und 34.1 % der Karten, die ein dem Träger unähnliches Gesicht zeigten (Bedingung 4), von den Kassierern akzeptiert wurden. Der Anteil der akzeptierten Kreditkarten in der Bedingung 1 (Gesicht und Foto sind unverändert) beträgt 93.3 %, und der Anteil der akzeptierten Kreditkarten in der Bedingung 2 (Erscheinungsbild des Trägers ist leicht verändert) beträgt 86.2 %.

1.2.2.2 Alternative-Forced-Choice Test (AFC)

Der Alternative-Forced-Choice Test (AFC) folgt in Studien zur Gesichter-Erkennung einer Lernphase. Er ist so aufgebaut, dass jeweils ein 'altes' gelerntes Gesicht zusammen mit einem oder mehreren neuen Gesichtern dargeboten wird. Die Aufgabe der Versuchsperson ist dann, zu entscheiden, welches der Gesichter im vorangegangenen Lernset gesehen wurde. Die Wahrscheinlichkeit, das alte Gesicht dabei zufällig zu wählen, variiert dabei in Abhängigkeit von der Anzahl der neuen Gesichter.

1.2.2.3 Wiedererkennenstest

Der Wiedererkennenstest folgt ebenso wie der AFC Test einer Lernphase. Dem Teilnehmer werden in der Lernphase eine Reihe von Stimuli dargeboten. Dabei kann selbstverständlich die Anzahl der Stimuli und die Darbietungsdauer der einzelnen Stimuli variiert werden. Der Lernphase schließt sich eine Testphase an, in der die Stimuli der Lernphase gemischt mit neuen, nicht gelernten Stimuli sukzessive dargeboten werden. Im Normalfall werden gleich viele alte, gelernte Stimuli wie neue, ungelernte dem Probanden präsentiert. Normalerweise wird zwischen Lern- und Testphase, sofern nicht eine beträchtliche Zeitdauer zwischen den beiden Phasen liegt, dem Probanden eine Aufgabe gegeben (Fragebogen, Rückwärtszählen), um die Aufmerksamkeit von den Stimuli abzulenken. Im Gegensatz zum AFC-Test werden die alten Bilder jedoch nicht gleichzeitig mit einem oder mehreren Distraktoren präsentiert, sondern einzeln danach beurteilt, ob das gesehene Bild in der vorangegangenen Lernphase präsentiert worden ist oder nicht.

1.3 Integration der dargestellten Theorien und Phänomene

In diesem Abschnitt sollen die beschriebenen Phänomene noch einmal kurz erwähnt werden und mit dem In-Group/Out-Group Modell von Sporer (2001a) sowie der Annahme der Speicherung von Gesichtern in einem multidimensionalen Raum in einen breiteren Zusammenhang gestellt werden.

Dass Gesichter eine besondere Rolle und Funktion haben, ist u.a. Gegenstand in der Entwicklungspsychologie, wie die Forschung zur Entwicklung der Fähigkeit zur Gesichterwiedererkennung zeigt. Menschen sind Experten in ihrer Fähigkeit, Gesichter zu verarbeiten, sie zu enkodieren und wiederzuerkennen. Die Faktoren, die bei den einzelnen Prozessen involviert sind, können dazu führen, dass die Wiedererkennbarkeit einzelner Gesichter und von Gruppen von Gesichtern abnimmt.

Ziel dieser Arbeit ist die Überprüfung einiger Annahmen des In-Group/Out-Group Modells, das in seiner Theorie auf den erwähnten Erklärungsmodellen aufbaut. Eine zentrale Annahme des In-Group/Out-Group Modells ist die Kategorisierung bzw. die fehlende Kategorisierung nach dem erstmaligen Zusammentreffen mit einer Person, je nachdem, ob an der Person Merkmale festgestellt werden, die sie als Mitglied einer Fremd-

Gruppe identifizieren können. Die Kategorisierung von Personen aufgrund einzelner Merkmale und die daraus resultierende Zuschreibung von Persönlichkeitsattributen ist ein bekanntes Phänomen in der Sozialpsychologie. Sie findet sich u.a. im Out-Group Homogenitäts Effekt und in den Experimenten zu äußeren Gesichtsmerkmalen als ethnische Marker (McLin & Malpass, 2001, 2003). Ebenso wurde von Levin (1996, 2000) die Bedeutung der ethnischen Kategorisierung bei der Erklärung des Cross-Race Bias hervorgehoben. Fremd-Gruppen Gesichter werden anhand eines schnell enkodierten Gruppenmerkmals als Fremd-Gruppen Gesichter identifiziert und dementsprechend einer Gruppe zugeordnet.

Wenn keine ethnische Kategorisierung einer Person stattfindet, d.h. wenn eine Person keine relevanten Merkmale besitzt, die sie als nicht der Eigen-Gruppe zugehörig einstufen lassen, wird sie dem In-Group/Out-Group Modell nach als Angehöriger der Eigen-Gruppe verarbeitet, d.h. die Verarbeitung des Gesichts dieser Person geschieht automatisch und holistisch, wie es in Domänen, mit denen die Person eine große Erfahrung und Expertise hat, ebenso der Fall ist. Die Aufmerksamkeit der Person ist nicht auf einzelne Merkmale gelenkt, um ein Gesicht von anderen Gesichtern differenzieren zu können, sondern das Gesicht wird 'als Ganzes' wahrgenommen und enkodiert. Es müssen hierbei keine internen Verbalisierungen des Gesehenen vorgenommen werden, die eine spätere Identifizierung des Gesichts erschweren könnten (verbal overshadowing, Fallshore & Schooler, 1995). In dem Fall, dass eine Kategorisierung einer Person als Angehöriger einer Fremd-Gruppe stattfindet, wird das Gesicht nicht holistisch verarbeitet, sondern es kommen andere der vorher beschriebenen Prinzipien zu tragen. Eine Hypothese ist, dass Fremd-Gruppen Gesichter eher merkmalsbasiert verarbeitet werden und damit auch die Mechanismen des verbal overshadowing eine spätere Wiedererkennensleistung negativ beeinflussen können. Die Klassifizierung einer Person als Angehöriger der Fremd-Gruppe führt zu einer gesteigerten Verwendung von Out-Group Schemata (Levin, 1996).

Die Speicherung der wahrgenommenen Gesichter, seien es Eigen-Gruppen oder Fremd-Gruppen Gesichter, wird nach heutigem Erkenntnisstand mit der Speicherung als Exemplare in einem multidimensionalen Raum am Besten beschrieben. Für die Speicherung von Gesichtern in einem multidimensionalen Raum bedeutet der häufigere Kontakt mit Mitgliedern der eigenen ethnischen Gruppe, dass die Verteilung der Exemplare durch die Mitglieder der eigenen ethnischen Gruppe weitgehend bestimmt wird und andere Gesichter zunächst in Relation zu dieser Verteilung abgespeichert werden. Dies heißt nicht, dass die Verteilung der Exemplare in einem multidimensionalen Raum unveränderbar ist. Neue Exemplare kommen hinzu und besetzen bestimmte Punkte in

solch einem Raum. Vielmehr ist durch die Enkodierung zusätzlicher Gesichter die Speicherung der Exemplare dynamisch, d.h. eine grundlegende Verteilung um eine 'zentrale Tendenz' kann sich nach Maßgabe der enkodierten Exemplare in Richtung der neuen Exemplare verschieben und neue Exemplare können in einen multidimensionalen Raum integriert werden.

Ein Faktor bei der Erinnerbarkeit von Gesichtern betrifft deren Distinktivität. Je distinkter ein Gesicht ist, d.h. je mehr 'hervorstechende' Merkmale es besitzt, die es von anderen Gesichtern unterscheiden, desto besser wird es erinnert, weil vor allem diese Merkmale als Unterscheidungskriterium dienen können. Bezogen auf die Speicherung von Gesichtern in einem multidimensionalen Raum bedeutet dies, dass distinkte Gesichter in diesem Raum Punkte besetzen, die von einer Norm bzw. einer zentralen Tendenz weiter entfernt sind als typische Gesichter. Fremd-Gruppen Gesichter werden verglichen mit der Eigen-Gruppe als distinkt betrachtet, weil sie in einem oder mehreren Merkmalen von der Eigen-Gruppe abweichen. Demnach sind Fremd-Gruppen Gesichter per se distinkt von der Eigen-Gruppe und sollten besser wiedererkannt werden. Wenn aber alle Fremd-Gruppen Gesichter als distinkte Gesichter eingeordnet werden, ist eine Differenzierung innerhalb der Fremd-Gruppe schwerer möglich, da die distinkten Merkmale als Differenzierungskriterium nicht mehr ausreichend sind. Anders verhält es sich bei distinkten Eigen-Gruppen Gesichtern. Sie weichen in einem oder mehreren Merkmalen von den anderen Exemplaren der Eigen-Gruppe ab und sind somit innerhalb der Eigen-Gruppe besser von den anderen Exemplaren der Eigen-Gruppe zu differenzieren.

Die gleichen Mechanismen können für den Own-Age und den Own-Sex Bias zum Tragen kommen. Wenn eine Person vorrangig Kontakt mit Personen hat, die ihr ähnlich sind (eigene Altersgruppe bzw. eigenes Geschlecht) so können die mit dem Geschlecht und /oder dem Alter von Personen verbundenen Differenzierungsmerkmale ebenfalls Dimensionen in einem multidimensionalen Raum bilden, auf denen Gesichtsmerkmale gespeichert werden.

Ein Cross-Race Bias drückt sich nicht nur in einer besseren Leistung bei Eigen-Gruppen Gesichtern, gemessen durch A' , aus. Ein mirror effect in den Treffern und Fehlalarmen ist ebenfalls Zeichen für ein differentiell schlechteres Wiedererkennen von Fremd-Gruppen Gesichtern. Bei distinkten Gesichtern gibt es mehr Treffer und weniger Fehlalarme in einem Wiedererkennensexperiment verglichen mit typischen Gesichtern. Daher sollte bei Fremd-Gruppen Gesichtern, die zwar als distinkt im Vergleich mit der Eigen-Gruppe angesehen werden, innerhalb der Fremd-Gruppen aber als nicht distinkt, weniger Treffer und mehr falsche Alarme gefunden werden als bei der Eigen-Gruppe.

1.4 Fragestellung

In diesem Abschnitt sollen die inhaltlichen Fragestellungen, die sich aus den bisherigen Erläuterungen ergeben, besprochen werden. Die drei Experimente zu einem möglichen Cross-Currency Bias, sowie zum Cross-Race Bias werden separat kurz erläutert und die dazugehörigen Hypothesen vorgestellt.

1.4.1 Hypothesen für die Experimente zu einem möglichen Cross-Currency Effekt

Die Annahmen der Speicherung von Gesichtern in einem multidimensionalen Raum finden weitgehende Zustimmung und experimentelle Bestätigung (Johnston, Milne, Williams & Hosie, 1997; Sporer, 2001a; Valentine, 1999). Gesichter werden als Exemplare enkodiert, deren Ort in einem multidimensionalen Raum von der Ausprägung verschiedener Merkmale auf verschiedenen Dimensionen bestimmt wird. Wie aber werden einzelne Objekte, deren Struktur keine Gesichtsmerkmale beinhaltet, enkodiert und abgerufen? Wie in Abschnitt 1.2.1.1 über die neuronalen Grundlagen der Gesichtererkennung beschrieben, existiert im menschlichen fusiformen Gyrus eine neuronale Struktur, die zum einen auf Merkmale von Gesichtern (N170) und zum anderen auf individuelle Gesichter (N250r) reagiert. Darüber hinaus fanden Rossion et al. (2002) bei Objekten, deren Struktur keinen bekannten Objekten ähnelt und auch keine Ähnlichkeit mit Gesichtern aufweist, dass die gleichen kognitiven Prozesse und Hirnareale involviert sind wie bei der Verarbeitung von Gesichtern. Caldara, Thut, Servois, Michel, Povet und Renault (2003) fanden in einer Studie mit weißen Teilnehmern und weißen sowie asiatischen Gesichtern eine N170 für beide Stimulus-Gruppen. Die Autoren fanden allerdings bei Eigen-Gruppen Gesichtern eine stärkere frühe Erregung in medial-occipitalen Strukturen als bei Fremd-Gruppen Gesichtern. Sie bringen dies mit den Experimenten von Phelps et al. (2000) sowie von Hart, Whalen, Shin, McInerney, Fischer und Rauch (2000) in Zusammenhang. Beide Studien fanden bei schwarzen und weißen Versuchsteilnehmern Unterschiede zwischen Fremd- und Eigen-Gruppen Gesichtern in der Amygdala insofern, dass diese Gehirnstruktur bei Fremd-Gruppen Gesichtern eine stärkere Erregung zeigt als bei Eigen-Gruppen Gesichtern.

Daraus leitet sich die Frage ab, ob Objekte ebenfalls als Exemplare in einem multidimensionalen Raum gespeichert werden, analog zu Gesichtern. Wenn dies der Fall ist, müssten für einen Wiedererkennenstest mit Objekten die gleichen Annahmen getroffen werden wie bei Gesichtern, nämlich, dass mit zunehmender Distanz eines Exemplars auf einer oder mehreren Dimensionen von den anderen Exemplaren einer Gruppe, die Wiedererkennensleistung bei diesem Exemplar sinkt.

Erweitert man die Frage nach der Speicherung von Objekten in einem multidimensionalen Raum um die Frage der Kategorisierung von Objekten anhand eines bestimmten Merkmals, so greifen zusätzlich die Annahmen des In-Group/Out-Group Modells von Sporer (2001a), nach dem bei der Wahrnehmung von Gesichtern zunächst eine Kategorisierung der Gesichter als Mitglied einer Eigen-Gruppe bzw. Fremd-Gruppe stattfindet und anschließend die Gesichter in Abhängigkeit der Kategorisierung weiterverarbeitet werden. Auf das Wiedererkennen von Objekten bezogen, hieße dies, für den Fall, dass Objekte bezüglich eines Merkmals als Mitglied einer Eigen-Gruppe bzw. Fremd-Gruppe kategorisierbar sind, sollten diese Objekte ebenfalls in Abhängigkeit dieser Kategorisierung verarbeitet und gespeichert werden. Es sollte sich ebenfalls beim Wiedererkennen von Objekten ein mirror-effect zeigen, d.h. eine höhere Anzahl an Treffern bei bekannten Objekten als bei unbekanntem und eine geringere Anzahl an falschen Alarmen im Vergleich zu unbekanntem Objekten.

Dieser Frage soll in den drei Experimenten zum *Cross-Currency Effekt* nachgegangen werden. Konkret soll untersucht werden, ob Objekte analog zu Gesichtern verarbeitet werden und damit auch die Merkmale der Speicherung in einem multidimensionalen Raum zeigen. Damit eine Kategorisierung durch die Teilnehmer stattfinden kann, müssen die Objekte so gestaltet sein, dass diese ein Merkmal besitzen, welches sie unterscheidet und für die Teilnehmer auch von Relevanz ist. Dazu sollen Geldscheine dreier verschiedener Währungen benutzt werden. Die Geldscheine mit hoher Relevanz, die als der 'Eigen-Gruppe' zugehörig betrachtet werden sollen, sind für eine deutsche Stichprobe Geldscheine der DM. Geldscheine der 'Fremd-Gruppe' werden durch das Britische Pfund und durch den Euro (vor dessen Einführung am 1.1. 2002) repräsentiert. Es wird in den drei Experimenten analog zum Cross-Race Bias ein Cross-Currency Effekt erwartet, d.h. bei Geldscheinen mit Zugehörigkeit zur Eigen-Gruppe sollten in allen drei verwendeten Paradigmen bessere Leistungen gezeigt werden als bei Geldscheinen mit Zugehörigkeit zur Fremd-Gruppe. Zusätzlich sollte sich in den Paradigmen, in denen eine Auswertung der Daten nach der Signal-Entdeckungs-Theorie möglich ist (Delayed-Matching Aufgabe und Wiedererkennenstest), ein mirror-effect

zeigen, wobei die DM auf der einen und das Britische Pfund und der Euro auf der anderen Seite gegenübergestellt werden sollen.

Ein weiteres Ziel der Experimente zum Cross-Currency Effekt ist, wie im vorherigen Absatz bereits angeklungen ist, der Vergleich mehrerer Testparadigmen. Meissner und Brigham (2001) bemängeln in ihrer Meta-Analyse, dass in vielen Studien ein 'Ja/Nein' Wiedererkennenstest benutzt wurde, während andere Testparadigmen seltener verwendet wurden. Die Frage, die sich dabei stellt, ist, inwiefern die Ergebnisse unterschiedlicher Testparadigmen miteinander vergleichbar sind. Für die Delayed-Matching Aufgabe und den Wiedererkennenstest konnte ein Cross-Race Bias bereits nachgewiesen werden. Ob dies auch für eine AFC-Aufgabe möglich ist, ist unklar. In Experiment I wurde eine Delayed-Matching-Aufgabe als Paradigma benutzt, in Experiment II eine AFC-Aufgabe und in Experiment III ein 'Ja/Nein' Wiedererkennenstest. Die Hypothese der Entstehung eines Cross-Currency Effekts durch die Verwendung von Geldscheinen unterschiedlicher Währungen und die Generierung eines mirror-effects gilt dabei für alle drei experimentellen Paradigmen.

1.4.2 Hypothesen zum Einfluss des Zeitintervalls und des Gesichtsauschnitts auf den Cross-Race Bias

Die Fähigkeit des Gedächtnisses, gelernte Inhalte wiederzugeben, ist u.a. von dem Zeitintervall zwischen dem Lernen neuer Inhalte und der Wiedergabe derselben abhängig, wie zuerst von Ebbinghaus (1885) nachgewiesen wurde. Chance und Goldstein (1987) untersuchten den Cross-Race Bias nach einem Zeitintervall von zwei bzw. sieben Tagen zwischen Lern- und Testphase und fanden keine Unterschiede bei der Anzahl richtiger Antworten, aber eine größere Anzahl an falschen Alarmen mit größerem Zeitintervall zwischen Lern- und Testphase. Eine Frage, die sich jedoch stellt, wenn man die unterschiedlichen Wiedererkennensleistungen bei Gesichtern anderer ethnischer Gruppen betrachtet, ist die Frage, welchen Einfluss das Zeitintervall auf den Cross-Race Bias ausübt. In den meisten Untersuchungen zum Cross-Race Bias wurden den Teilnehmern neue Gesichter dargeboten und ein Wiedererkennenstest nach einem relativ kurzen Intervall durchgeführt.

Eine eindeutige Antwort darauf, wie sich der Cross-Race Bias in einem experimentellen Setting verhält, wenn die Testung nach einem längeren Zeitintervall durchgeführt wird, ist bislang noch nicht gefunden worden. Shepherd, Gibling und

Ellis (1991) fanden in einer Untersuchung, dass es bei distinkten Gesichtern nach einem Zeitintervall von einem Monat zwischen Lern- und Testphase mehr Treffer gab als bei nicht-distinkten Gesichtern. Sie ziehen daraus die Schlussfolgerung, dass distinkte Gesichter über einen längeren Zeitraum besser erinnert werden können. Das In-Group/Out-Group Modell trifft zur Frage der Wiedererkennung von Gesichtern nach einem längeren Zeitintervall explizit keine Aussage, jedoch können sich Aussagen hierzu ableiten lassen. Im Gegensatz zu den Ergebnissen von Shepherd, Gibling und Ellis (1991) ist beim In-Group/Out-Group Modell bei einer merkmalsbasierten Enkodierung von Gesichtern anzunehmen, dass die Erinnerungsfähigkeit an einzelne Merkmale mit der Zeit in stärkerem Maße nachlässt als bei einer holistischen Enkodierung. Einen Hinweis darauf liefern u.a. Studien zum Verbal Overshadowing (Schooler & Engstler-Schooler, 1990; Fallshore & Schooler, 1995). Probanden wurden gebeten, Gesichter, die sie sich einprägen sollten, verbal zu beschreiben, mit dem Ziel, die Wiedererkennensleistung zu verbessern. Das Gegenteil war jedoch der Fall. Probanden, die Gesichter beschreiben sollten, zeigten schlechtere Leistungen als Probanden, denen diese Aufgabe nicht gestellt wurde. Es ist durchaus möglich, dass Prozesse des Verbal Overshadowing bei der Enkodierung von Fremd-Gruppen Gesichtern eine Rolle spielen. Die Möglichkeiten der Beschreibung eines Gesichts anhand einzelner Merkmale sind durch die endliche Menge an Dimensionen eingeschränkt, und eine Person ist dadurch nicht in der Lage, jedes Gesicht anhand von Beschreibungen eindeutig von anderen Gesichtern zu differenzieren. Es ist ebenfalls möglich, dass Dimensionen, welche eine Ethnie durchaus hinreichend beschreiben können, für eine andere Ethnie inadäquat sind. So wären beispielsweise mitteleuropäische Hauttypen für afro-amerikanische Gesichter ein schlechtes Unterscheidungsmerkmal, weil sich diese Hauttypen dort nicht finden lassen. Wenn nun Fremd-Gruppen Gesichter merkmalsbezogen enkodiert werden, treffen die gewählten Merkmale u.U. auf eine ganze Reihe von Gesichtern zu, die das gleiche Merkmal teilen, so dass nach einem längeren Zeitintervall diese Merkmale allein nicht mehr ausreichen, um eine größere Anzahl von Gesichtern zu differenzieren. Erfolgt die Enkodierung jedoch holistisch, ist die Person nicht darauf angewiesen, einzelne Merkmale aus dem Gedächtnis abzurufen, sondern erkennt das Gesicht 'als Ganzes'. Eine solche holistische Repräsentation sollte zeitlich überdauernder sein als eine merkmalsbezogene Repräsentation. Daher sollten sich bei den Eigen-Gruppen Gesichtern nach einem längeren Zeitintervall bessere Wiedererkennensleistungen zeigen als bei Fremd-Gruppen Gesichtern.

Eine experimentelle Variation der Gesichtsmerkmale, die bei der Enkodierung von Gesichtern eine Rolle spielen, ist schwierig zu realisieren, weil es zum einen eine relativ

große Zahl möglicher Merkmalsdimensionen gibt, und zum anderen eine Veränderung derselben dazu führen kann, dass ein Gesicht nicht mehr als natürliches Gesicht wahrgenommen wird. Merkmale, die häufig dazu dienen, Gesichter zu unterscheiden, sind die Frisur und die Haarfarbe, aber auch die Form der Nase, der Ohren und anderer individueller Merkmale. Dass die ethnische Kategorisierung bei der Enkodierung eine bedeutende Rolle spielt, wurde von Levin (2000) untersucht. MacLin und Malpass (2001) konnten zeigen, dass die Frisur als ethnischer Marker fungiert, anhand derer Personen einer bestimmten Ethnie zugeordnet werden. Wenn dieses Merkmal bei einem Gesicht fehlt, so fehlt damit auch eine Möglichkeit der Kategorisierung einer Person, aber es fehlt auch eine zusätzliche Möglichkeit, eine Person anhand dieses Merkmals wiederzuerkennen. Eine weitere Frage, die sich dabei stellt, ist, wie sich die Wiedererkennensleistung bei Gesichtern verhält, denen solche äußeren Merkmale fehlen? Wenn es Probanden nicht mehr möglich ist, eine Person mittels äußerer Merkmale zu enkodieren, müssen sie sich die inneren Merkmale, wie z.B. die Form der Nase oder die Distanz der Augen voneinander, einprägen. Nach dem In-Group/Out-Group Modell wären hierbei die Eigen-Gruppen Gesichter im Vorteil, da deren Enkodierung, im Gegensatz zu Fremd-Gruppen Gesichtern, holistisch ist, und weniger auf die Verarbeitung einzelner Merkmale angewiesen ist. Auf der anderen Seite würde ein Merkmal, das dazu dient, ein Gesicht als der Fremd-Gruppe zugehörig zu kategorisieren, fehlen. Der Prozess der Kategorisierung sollte aber dennoch stattfinden, solange im Gesicht andere Merkmale vorkommen, die in der Lage sind, Eigen-Gruppen von Fremd-Gruppen Gesichtern zu trennen. Dabei wird vermutet, dass bei fehlenden äußeren Merkmalen und dennoch ermöglichter Kategorisierung der Gesichter die Wiedererkennensleistung bei Fremd-Gruppen Gesichtern geringer ausfällt als bei Eigen-Gruppen Gesichtern, da diese immer noch holistisch verarbeitet werden, während für die Verarbeitung der Fremd-Gruppen Gesichter wichtige äußere Gesichtsmerkmale fehlen. Dieser Frage und der Frage nach der Form, d.h. der Größe des Effekts des Cross-Race Bias nach einem längeren Behaltensintervall wird in Abschnitt 4 zum Einfluss von Behaltensintervall und Gesichtsausschnitt auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern nachgegangen.

1.4.3 Hypothesen zum Einfluss der Kontakthäufigkeit auf den Cross-Race Bias

Die Leistung beim Wiedererkennen von Gesichtern der eigenen Ethnie und anderer Ethnien ist, neben dem Behaltensintervall und der Anzahl der sichtbaren Merkmale eines Gesichtes auch von der Quantität und Qualität des Kontaktes mit anderen Ethnien abhängig. Der Cross-Race Bias wird bei zunehmendem Kontakt mit anderen Ethnien geringer. In der Studie von Wright, Boyd und Tredoux (2001) wurden Probanden untersucht, die in Ländern mit unterschiedlicher Bevölkerungszusammensetzung leben, und bei größerer Kontakthäufigkeit mit Personen anderer Ethnien einen geringeren bzw. sogar einen entgegengesetzten Cross-Race Bias aufwiesen als bei geringer Kontakthäufigkeit. Personen mit einer größeren Kontakthäufigkeit in Südafrika waren vor allem Schwarze, die aus beruflichen oder akademischen Gründen mehr Kontakt mit Weißen hatten, bei denen also davon ausgegangen werden kann, dass bei ihnen auch eine besondere Qualität des Kontaktes bestand. Diese Personen zeigten sogar bessere Wiedererkennensleistungen bei weißen verglichen mit schwarzen Gesichtern. Es soll der Frage nachgegangen werden, ob auch eine größere Kontakthäufigkeit, deren Qualität nicht so groß ist, wie bei den Basketballfans von Dunning, Li und Malpass (1998) oder auch den südafrikanischen Studierenden eine Reduzierung des Cross-Race Bias zur Folge hat. Es bedarf dazu also einer Personengruppe als Teilnehmer, deren Kontakthäufigkeit mit Personen anderer Ethnien größer ist als die einer Kontrollgruppe, die Qualität der Kontakte sollte sich zwischen den Gruppen jedoch nicht unterscheiden.

Die Annahmen des In-Group/Out-Group Modells treffen dabei für beide Gruppen zu. Fremd-Gruppen Gesichter sollten von beiden Gruppen als Fremd-Gruppen kategorisiert werden und entsprechend den Annahmen verarbeitet werden. Dadurch, dass die Gruppe mit der größeren Kontakthäufigkeit mit Fremd-Gruppen Mitgliedern aber einen größeren Erfahrungsschatz mit der Fremd-Gruppe besitzt als die Gruppe mit der geringeren Kontakthäufigkeit, ist die Anzahl der Exemplare der Fremd-Gruppen im multidimensionalen Raum bei der Gruppe mit großer Kontakthäufigkeit größer. Eine weitere Vermutung, inwiefern sich die beiden Gruppen voneinander unterscheiden, ist, dass sich in der Gruppe mit der größeren Kontakthäufigkeit bessere Strategien zur Merkmalsbestimmung herausgebildet haben, verschiedene Gesichter innerhalb einer ethnischen Gruppe zu differenzieren. Um diese Frage zu untersuchen, werden in Abschnitt 6 in einem Experiment zum Einfluss der Kontakthäufigkeit auf die Wiedererkennens-

leistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern, Beamte des Bundesgrenzschutzes am Flughafen Frankfurt/Main untersucht und mit einer studentischen Kontrollgruppe verglichen. Aufgrund der größeren Kontakthäufigkeit der BGS-Beamten mit Personen anderer Ethnien als der eigenen sollten sie insgesamt im Sinne eines Haupteffektes der Berufsgruppe, bessere Leistungen in einem Wiedererkennenstest zeigen als die studentische Kontrollgruppe. Zusätzlich soll in Experiment III der Berufsgruppe der Bundesgrenzschutzbeamten Rechnung getragen werden und neben einem Wiedererkennenstest eine Delayed-Matching Aufgabe durchgeführt werden. Eine Delayed-Matching Aufgabe bildet in experimentell kontrollierter Weise die Aufgabe der Grenzschutzbeamten ab, in dem Sinne, dass bei dieser Aufgabe ein Abgleich von zwei Gesichtern verlangt wird, wie er im Berufsalltag der Beamten vorkommt. Daher sollen auch nur Beamte getestet werden, die an einer Dienststelle an der Passkontrolle am Flughafen Frankfurt/Main beschäftigt sind. Beamte, die beispielsweise für die Rückführung von Asylbewerbern zuständig sind, sollen nicht getestet werden, weil bei ihnen von einer anderen Qualität des Kontaktes mit Personen anderer Ethnien ausgegangen werden muss.

1.5 Schlußbemerkungen zur Fragestellung

Weiterhin sollen in den Experimenten zum Einfluss des Behaltensintervalls und des Gesichtsausschnitts auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern (siehe Abschnitt 4) und zum Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern (siehe Abschnitt 6), Gesichter als Stimulus-Material verwendet werden, die standardisiert aufgenommen und digital bearbeitet wurden, um mögliche Artefakte aufgrund des Materials zu minimieren, da die Qualität der Stimuli in vielen bisherigen Untersuchungen wahrscheinlich eher dürftig ist. Wright und Sladden (2003) haben z. B. ihre Stimulusbilder zum Teil aus dem Internet heruntergeladen, was den Nachteil in sich birgt, dass die Ansicht der Gesichter nicht gleich ist und auch der emotionale Ausdruck der Gesichter nicht ausreichend kontrolliert werden kann. Es sollen nicht nur die Leistungswerte A' und B'' in den Analysen betrachtet werden, sondern auch die Treffer und falschen Alarme, da der mirror-effect bezüglich dieser beiden Maße ebenfalls Unterschiede zwischen den untersuchten Gruppen postuliert.

Sofern nicht anders angegeben, wird für alle statistischen Tests ein Alpha-Niveau von $p = .05$ festgelegt, ab dem ein Test als signifikant angesehen wird. Wenn notwendig, wird eine Alpha-Adjustierung vorgenommen, um eine künstliche Inflation des Alpha-Fehlers zu vermeiden. Post-hoc Tests bei signifikanten Varianzanalysen werden nach den Formeln von Winer (1971) berechnet. Weiterhin sollen in Anlehnung an die APA-Richtlinien (Publication Manual of the American Psychological Association, 2001) und an Wright (2003) in den Darstellungen der Ergebnisse die exakten p -Werte und Effektgrößen (r ; siehe Mullen, 1989) berichtet werden. Nach Cohen (1988) wird ein Effekt als 'gering' bezeichnet, wenn $.2 \leq d \leq .49$ ist. Ergänzend ist ein Effekt von $.5 \leq d \leq .79$ ein mittlerer Effekt und ein Effekt von $d \geq .8$ ein starker Effekt. Diese Werte von d lassen sich in r -Werte transformieren⁷ (Cohen, 1988), die in den nachfolgenden Abschnitten über die Ergebnisse der Experimente benutzt werden sollen. Dementsprechend ist ein Effekt von $.1 \leq r \leq .29$ ein geringer Effekt, ein Effekt von $.3 \leq r \leq .49$ ein moderater Effekt und ein Effekt von $r \geq .5$ ein starker Effekt (Cohen, 1988). Bei Tests mit mehr als zwei Gruppen wird als Effektstärkemaß η^2 verwendet. Ein Effekt von $.01 \leq \eta^2 \leq .057$ ist dementsprechend ein geringer Effekt, ein Effekt von $.058 \leq \eta^2 \leq .137$ ein moderater Effekt und ein Effekt von $\eta^2 > .138$ ein starker Effekt (Cohen, 1988).

Meissner und Brigham (2001) berichten in ihrer Meta-Analyse zum Cross-Race Bias schwache bis moderate Effekte. Ausgehend von dieser Meta-Analyse werden in dieser Arbeit ebenfalls schwache bis moderate Effekte in den einzelnen Experimenten erwartet. Da die Auswertung der Daten varianzanalytisch erfolgt, wird in den einzelnen Abschnitten die erwartete Power nach Cohen (1988) in Abhängigkeit von Teilnehmeranzahl und Alpha-Niveau berichtet.

$${}^7 r = \frac{d}{\sqrt{d^2 + 4}}$$

Bei gleichen Zellbesetzungen

2 Die unterschiedliche Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Objekten (Cross-Currency Effekt)

Die Annahmen der Kontakthypothese (Chance, Goldstein & McBride, 1975) nach denen bessere Leistungen bei der Wiedererkennung von Gesichtern anderer ethnischer Gruppen als der eigenen darauf zurückzuführen sind, dass häufigerer Kontakt mit diesen Gruppen existiert, sollen in den nachfolgend beschriebenen ersten drei Telexperimenten auf die Wahrnehmung und Wiedererkennung von bekannten und unbekanntem Objekten angewandt werden. Rossion et al. (2002) fanden bei Objekten, deren Struktur keinen bekannten Objekten ähnelt und auch keine Ähnlichkeit mit Gesichtern aufweist, dass die gleichen kognitiven Prozesse und Hirnareale involviert sind wie bei der Verarbeitung von Gesichtern. Gauthier und Tarr (1996) fanden bei nicht-gesichtsähnlichen Stimuli einen Leistungszuwachs bei einem späteren Wiedererkennenstest. Es soll in diesem Experiment versucht werden, einen 'Cross-Race Bias' mittels Banknoten aus verschiedenen Ländern zu generieren. Dieser Effekt wird im folgenden in Anlehnung an den Cross-Race Bias mit *Cross-Currency Effekt* bezeichnet. Hierbei konnte durch die Einführung des Euro als allgemeines Zahlungsmittel in der Europäischen Union im Januar 2002 Stimulus-Material genutzt werden, das nicht künstlich generiert werden musste und trotzdem den Teilnehmern unbekannt war. Die Daten wurden als Within-Subjects Untersuchung zu zwei Zeitpunkten erhoben. Der erste Zeitpunkt (t_1) lag kurz vor Einführung des Euro im Dezember 2001, der zweite Messzeitpunkt (t_2) im April/Mai 2002. Dadurch, dass der Wiedererkennenstest des dritten Experiments von allen Teilnehmern als letztes bearbeitet wurde, konnten die in den ersten beiden Experimenten (Delayed-Matching Aufgabe und AFC-Aufgabe) benutzten Stimuli als Übungsstimuli für den Wiedererkennenstest herangezogen werden. Ein weiteres Ziel der Experimente zum Cross-Currency Effekt sollte sein, die einzelnen Testparadigmen, die in den Experimenten verwendet wurden, hinsichtlich ihrer Vergleichbarkeit zu untersuchen. Die Hypothesen, die für die einzelnen Experimente aufgestellt werden, ähneln sich daher im engeren Sinne.

2.1 Versuchsplan

Für die Erhebung der Daten fanden die in Abschnitt 1.3.2 erwähnten Testparadigmen Anwendung. Allen Teilnehmern wurden Geldscheine präsentiert, die nach Maßgabe einer Delayed-Matching-Aufgabe (siehe Abschnitt 2.2), einer AFC-Aufgabe (siehe Abschnitt 2.3) und eines Wiedererkennenstests (siehe Abschnitt 2.4) bearbeitet werden mussten. Jedes dieser drei Paradigmen wurde zu zwei Zeitpunkten in einem Messwiederholungsdesign bearbeitet, weil es vor allem darauf ankam, die Wiedererkennensleistung bei Banknoten nach einer natürlichen Lernphase zu testen, d.h. die Lernphase bestand aus dem täglichen Umgang mit dem neuen Zahlungsmittel und nicht, wie üblicherweise, aus einer experimentell kontrollierten Lernphase. Zwar dienten die Delayed-Matching Aufgabe und die AFC-Aufgabe auch als Lerndurchgänge für den Wiedererkennenstest, den jahrelangen Kontakt mit Geldscheinen der DM sollte eine solch vergleichsweise kurze Lernphase aber nicht ausgleichen können. Des weiteren wurde die Tastaturbelegung ausbalanciert, d.h. es wurde variiert, mit welchen Tasten auf der Tastatur während der Experimente die Antworten eingegeben werden sollte. Je nach Testparadigma lauteten die Antwortoptionen 'Geldschein ist identisch' vs. 'Geldschein ist nicht identisch' (Delayed-Matching-Aufgabe), 'echter Geldschein ist oben' vs. 'echter Geldschein ist unten' (AFC-Aufgabe) und 'Geldschein ist echt' vs. 'Geldschein ist falsch' (Wiedererkennenstest). Für die Antworten wurden die Tasten 'c' und 'm' belegt, die je nach Bedingung entweder für die erste oder die zweite Antwortoption Geltung besaßen.

Um die Wiedererkennensleistung bei Euro-Geldscheinen (€) als unbekanntem Objekten zu untersuchen, wurden als Kontrollstimuli Geldscheine der Deutschen Mark (DM) herangezogen, die zum Zeitpunkt der Versuchsdurchführung seit ca. zehn Jahren unverändert geblieben und offizielles Zahlungsmittel in der Bundesrepublik Deutschland waren. Als zweite Währung wurden Geldscheine des Britischen Pfunds (£) verwendet. Der Grund für die Hinzunahme des Britischen Pfunds war die Annahme, dass der überwiegende Teil der Teilnehmer weder zum Zeitpunkt t_1 noch zum Zeitpunkt t_2 häufigen Kontakt mit dieser Währung hatte. Außerdem bietet das Britische Pfund die Möglichkeit, die gleichen Experimente im Falle der Einführung des Euro in Großbritannien mit britischen Teilnehmern zusätzlich zu einem späteren Zeitpunkt durchzuführen. Tabelle 3 zeigt den Versuchsplan für die Experimente zum Cross-Currency Effekt.

Tabelle 3

Versuchsplan für die Zuordnung der Teilnehmer zu den Experimenten zum Cross-Currency Effekt

<u>Zeitpunkt t₁</u>		<u>Zeitpunkt t₂</u>	
Reihenfolge	Taste	Reihenfolge	Taste
DM - AFC - Rec	c (m)	DM - AFC - Rec	m (c)

Anmerkung: Der Faktor 'Geschlecht' wurde nicht systematisch variiert, weil durch die Wahl des ersten Messzeitpunktes kurz vor der Einführung des Euro eine gleiche Verteilung von Männern und Frauen nicht sichergestellt werden konnte. Innerhalb der einzelnen Untersuchungen wurden die folgenden Faktoren (siehe Tabelle 4) eingeführt.

Tabelle 4

Darstellung der Faktoren: Art der Fälschung, Seite des Geldscheins und Wertigkeit

Währung	Wertigkeit	<u>Farbe</u>		<u>Spiegelung</u>	
		Vorder	Rück	Vorder	Rück
DM/Pfund	(1) DM 10 /£ 1/€ 5	1	1	1	1
Euro	(2) DM 20 /£ 5/€ 10	1	1	1	1
	(3) DM 50 /£ 10/€ 20	1	1	1	1
	(4) DM 100 /£ 20/€ 50	1	1	1	1

Die dargestellten Faktoren dienen vor allem der Generierung des Stimulusmaterials (Art der Fälschung, Seite des Geldscheines). Zu diesen beiden Faktoren bestehen keine Hypothesen im Rahmen eines möglichen Cross-Currency Effektes. Molz und Hopf (2002) fanden in einem Experiment zur Größeneinschätzung von Geldmünzen, dass Münzen mit einer höheren Wertigkeit auch als physisch größer im Vergleich zu Münzen mit einer niedrigeren Wertigkeit eingeschätzt werden. Zur Kontrolle möglicher Effekte der Wertigkeit eines Geldscheines auf die Wahrnehmung, wurden die Geldscheine der Währungen bezüglich der verwendeten Wertigkeiten einander bestmöglich angepasst. Bei den Währungen DM und Euro ist dies auch sehr genau möglich, die Geldscheine des

Britischen Pfundes weichen jedoch in etwas stärkerem Maße von den Wertigkeiten der DM und des Euro ab⁸. Der derzeitige Wechselkurs beträgt ca. £ 1: € 1.5.

2.1.1 Stichprobe

An den drei Experimenten nahmen insgesamt 48 Personen teil (29 weiblich, 19 männlich), hauptsächlich Studierende der Psychologie an der Universität Giessen sowie Mitarbeiter am Fachbereich Psychologie und Sportwissenschaft der Universität Giessen. Von diesen 48 Probanden konnten 42 (24 weiblich, 18 männlich) in die Analyse aufgenommen werden. Vier Personen erschienen nicht zum zweiten Messzeitpunkt. Von weiteren zwei Personen waren die Daten-Dateien aufgrund wiederholten unsachgemäßen Antwortverhaltens unbrauchbar (die Teilnehmer drückten statt der vorgesehenen Tasten 'c' und 'm' andere Tasten auf der Tastatur, die jedoch nicht zuordenbar waren). Der Altersmedian betrug $\text{Mdn} = 22$ Jahre (männliche Teilnehmer $\text{Mdn} = 26$, weibliche Teilnehmer $\text{Mdn} = 21$). Der jüngste Teilnehmer war 19, der älteste 52 Jahre alt. Die Teilnehmer erhielten als Aufwandsentschädigung entweder zwei Versuchspersonenstunden oder € 15,-.

Die Anzahl der Teilnehmer ($N = 42$) ist ausreichend, um bei erwarteten mittleren Effektstärken ($.3 \leq r \leq .49$) beim Cross-Race Bias (Meissner & Brigham, 2001) und einem alpha-Niveau von $\alpha = .05$ je nach Anzahl der Zählerfreiheitsgrade eine Power von .63 bis .99 zu erreichen (siehe Cohen, 1988).

2.1.2 Material

In diesem Abschnitt wird beschrieben, wie die Geldscheine für die Delayed-Matching Aufgabe, die AFC-Aufgabe und den Wiedererkennenstest bearbeitet und für die Experimente vorbereitet wurden. Anschließend werden die drei Experimente in der Reihenfolge Delayed-Matching, AFC- Aufgabe und Wiedererkennenstest beschrieben und die jeweiligen Ergebnisse dargestellt.

⁸ Einige Teilnehmer an der Untersuchung und Leser dieser Arbeit merkten an, dass es keine £ 1 Banknote gäbe. Die £ 1 Banknote existiert in der Tat, sie wird jedoch durch die £ 1 Münze weitestgehend ersetzt.

2.1.3 Stimuli

Die Bilddateien der Geldscheine der drei Währungen (Deutsche Mark, Britisches Pfund und Euro) wurden von den jeweiligen Landesbanken (The Bank of England, n.d.; Deutsche Bundesbank, n.d.), die Muster der Geldscheine auf ihren Internetseiten bereitstellen, heruntergeladen. Bei jeder Währung wurden Geldscheine vier verschiedener Wertigkeiten, die zwischen den Währungen in etwa gleichwertig sind, verwendet. Die Geldscheine der DM waren die 10, 20, 50 und 100 DM Banknote, die Geldscheine des Britischen Pfunds die 1, 5, 10 und 20 £ Banknote und für den Euro die 5, 10, 20 und 50 € Banknote. Die Geldscheine des Britischen Pfunds weichen von der Maßgabe der Gleichwertigkeit leicht ab. Da jedoch zu einem späteren Zeitpunkt die gleiche Untersuchung mit Britischen Probanden durchgeführt werden sollte, wurde auf diese Noten zurückgegriffen. Somit bestand das Basismaterial aus zwölf verschiedenen Banknoten. Da die einzelnen Geldscheine je nach Herkunft und Wertigkeit von unterschiedlicher Größe waren, mussten sie noch standardisiert werden, damit sie nicht anhand dieses Unterscheidungsmerkmals für die Teilnehmer einfacher zu differenzieren waren. Zu diesem Zweck wurde die Breite und Höhe einer 10 DM Note als Standard festgelegt und alle anderen Geldscheine mittels des Programms 'Photoshop 5.5' in ihrer Größe an die 10 DM Note angeglichen. Die Originaldatei eines jeden Geldscheins wurde mit Photoshop geöffnet, die Größe des Bildes mit dem Befehl 'image size' auf (B x H) 400 x 211 Pixel festgelegt und unter der jeweiligen Kennung (Währung, Wert, Seite, Art der Fälschung) gespeichert.

Damit in den Experimenten die echten von den falschen Noten unterschieden werden konnten, mussten die Basisbanknoten verfälscht werden. Dazu wurden zwei Möglichkeiten realisiert, nämlich Spiegelungen und Farbfälschungen der Originalbanknoten.⁹ Bei Spiegelungsfälschungen wurde jeder einzelne Geldschein horizontal gespiegelt (siehe Abbildung 8). Die Bilder der Geldscheine wurden mit Photoshop geöffnet und der gesamte Bildbereich markiert. Anschließend wurde der Befehl 'Flip horizontal' ausgewählt und damit der gesamte Bereich horizontal gespiegelt. Dabei musste darauf geachtet werden, dass Text und Zahlen auf den Geldscheinen nicht gespiegelt erscheinen, da dies einen einfachen Hinweis zur Identifizierung falscher Geldscheine dargestellt hätte. Daher wurden die Ziffern der Geldscheine nach dem

⁹ Es existieren mehr Möglichkeiten, Banknoten zu fälschen, z.B. die Wertigkeiten verschiedener Banknoten zu vertauschen, dies war im Rahmen der Experimente aber nicht sinnvoll.

Spiegeln des gesamten Geldscheins noch einmal zusätzlich gespiegelt, was bedeutet, dass jeder Bereich auf den Geldscheinen mit Text oder Ziffern einzeln ausgewählt und gespiegelt werden musste. Für den Fall, dass sich Text oder Ziffern über zwei oder mehrere Kontrastebenen erstrecken, wie in Abbildung 8 die Seriennummer des Geldscheins, wurden entweder einzelne Teile des Textes separat gespiegelt, oder die Kontrastebenen wurden noch einmal separat vertauscht.



Abbildung 8. Beispiel für eine Spiegelungs- und eine Farbfälschung anhand einer 50 Euro Note. Die Farbfälschung der 50 Euro Banknote wurde der Grundfarbe der 5 Euro Banknote angepaßt. Die gefälschten 50 € Banknoten sind unten.

Aufgrund der Komplexität der Geldscheine des Britischen Pfunds (siehe Abbildung 9) wurden diese Geldscheine eingehender bearbeitet. Sie wurden wie die DM und Euro Geldscheine zunächst mit der beschriebenen Prozedur gespiegelt, die einzelnen Texte und Ziffern aber auf andere Art und Weise spiegelngefälscht. Die Texte und Ziffern wurden mithilfe des 'magic wand tools' einzeln markiert, sodass lediglich diese Bereiche markiert waren. Anschließend wurde mittels des Befehls 'new layer via copy' mit diesem Bereich eine neue Ebene geöffnet und diese Ebene horizontal gespiegelt. Die Ebenen wurden dann mit 'merge layers' wieder zusammengeführt und eventuell auffällige Ränder zwischen Text und Geldschein mit dem 'smudge tool' geglättet.

Für die Farbfälschungen gab es zwei Wege, die aus der unterschiedlichen farblichen Struktur der Geldscheine der drei Währungen resultieren. Die Geldscheine der DM und des Euro bestehen für jede Wertigkeit aus einer bestimmten Grundfarbe (z.B. braun bei der 50 DM Banknote, oder auch bei der 50 Euro Banknote, siehe Abbildung 8), während die Noten des Britischen Pfunds keine bestimmten Grundfarben besitzen (siehe Abbildung 9).



Abbildung 9. Die Vorderseiten von drei relativ gleichwertigen Banknoten der DM, des Euro und des Britischen Pfunds.

Daher war es bei den DM und Euro-Geldscheinen relativ einfach, Farbfälschungen herzustellen, d.h. es musste lediglich die Grundfarbe eines Geldscheins verändert werden. Eine solche Prozedur war für das Britische Pfund nicht möglich, weil diese Noten nicht aus einer genau definierten Grundfarbe bestehen. Bei diesen Geldscheinen wurde ein bestimmter Teil der Banknote, in dem eine bestimmte Farbe vorherrscht, separat von den anderen Teilen gefälscht. Es sollten bei der Herstellung der Farbfälschungen keine neuen Grundfarben verwendet werden, sondern die Grundfarben innerhalb einer Währung neu den Geldscheinen zugewiesen werden, um es den Teilnehmern nicht zu ermöglichen, anhand einer unbekanntenen Farbe, falsche von echten Geldscheinen voneinander zu unterscheiden. Von den neun möglichen Kombinationen der Zuweisung der 'neuen' Farben für die Farbfälschungen (siehe Tabelle 5) wurden drei zufällig ausgewählt und den drei Währungen zugeteilt.

Tabelle 5

Mögliche Kombinationen von Zuweisungen der Farben zu den Geldscheinen

	Wertigkeit			Kombination								
	DM	€	£	1	2	3	4	5	6	7	8	9
(1)	10	5	1	2	2	2	3	3	3	4	4	4
(2)	20	10	5	1	3	4	1	4	4	1	3	3
(3)	50	20	10	4	4	1	4	1	2	2	1	2
(4)	100	50	20	3	1	3	2	2	1	3	2	1

Anmerkung: Die Kombinationen 1-9 geben an, welche Grundfarbe mit der Grundfarbe eines anderen Geldscheins ersetzt wird.

Bei den DM- und Euro-Geldscheinen wurde der entsprechende Geldschein mit Photoshop geöffnet und die Grundfarbe mittels der Option 'Hue/Saturation' verändert. Dazu wird die originale Farbstruktur in Photoshop angezeigt, die durch Veränderung der unteren Leiste in dem Optionsmenü 'Hue/Saturation' der gewünschten Farbstruktur angeglichen werden kann. Die Geldscheine des Britischen Pfunds besaßen, wie bereits erwähnt, keine bestimmte Grundfarbe, sondern hatten eine zumeist sehr pastellfarbene Struktur. Bei diesen Geldscheinen wurde, je nach Geldschein, die Farbe bzw. die Farben extrahiert, die eine gewisse Grundstruktur darstellten und für die neue Kombination der Farbzusweisungen zu den Banknoten des Britischen Pfunds verwendet. Die Geldscheine wurden ebenfalls mit Photoshop geöffnet, und es wurde ähnlich mit ihnen verfahren, d.h. die Farbstruktur wurde ebenfalls mit der Option 'Hue/Saturation' verändert. Der einzige Unterschied bestand darin, dass bei Geldscheinen mit mehr als einer vorherrschenden Farbe die Farbbereiche auf dem Geldschein markiert und getrennt vom Rest des Geldscheins farblich verändert wurden. Die beschriebenen Prozeduren zur Erstellung von Farb- und Spiegelungsfälschungen wurden sowohl für die Vorder- als auch für die Rückseiten der Geldscheine verwendet.

Durch die Auswahl des Stimulus-Materials und dessen Eigenschaften (Vorder- und Rückseite, vier Wertigkeiten) sowie aufgrund der Spiegelungs- und Farbfälschungen resultieren für die Matching-Aufgabe und den Wiedererkennenstest insgesamt 72 verschiedene Stimuli (3 [Währung: DM, £, €] x 4 [Wertigkeit: 1, 2, 3, 4] x 3 [Fälschung: Normal, Farbfälschung, Spiegelung] x 2 [Seite: Vorderseite, Rückseite]). Für die AFC-Aufgabe, bei der zwei Geldscheine gleichzeitig übereinander dargeboten wurden (ein

'echter' und ein 'gefälschter'), wurden insgesamt 96 verschiedene Stimuli generiert (3 [Währung: DM, £, €] x 4 [Wertigkeit: 1, 2, 3, 4] x 2 [Fälschung: Farbfälschung, Spiegelung] x 2 [Seite: Vorderseite, Rückseite] x 2 [Darbietung 'echter' Geldschein: oben, unten]). Für die Rückmeldung über die Richtigkeit der Antworten wurden zwei Bilddateien erstellt: Bei falscher Antwort wurde das Wort "FALSCH" und bei richtiger Antwort das Wort "RICHTIG" eingeblendet. Beide Bildeinblendungen hatten eine Größe von (B x H) 640 x 480 Pixel. Der Hintergrund war schwarz bei weißer Schriftfarbe.

2.2 Experiment I: Delayed-Matching-Aufgabe

Das erste Experiment zum Cross-Currency Effekt benutzte eine Delayed-Matching Aufgabe als experimentelles Paradigma. Zunächst werden die Hypothesen erläutert und die Durchführung des Experiments dargestellt. Anschließend folgt die Darstellung der Ergebnisse und die Diskussion dieses Experiments.

2.2.1 Hypothesen

Nach den Annahmen der Kontakthypothese ist der Cross-Race Bias vor allem darauf zurückzuführen, dass häufiger Kontakt mit Personen anderer ethnischer Gruppen dazu führt, dass Gesichter anderer Gruppen besser wiedererkannt werden können. Wenn die Verarbeitungswege bei Objekten denen von Gesichtern ähneln, wie von Rossion et al. (2002) gezeigt wurde, müssten dann die gleichen Wiedererkennensdefizite bei Objekten auftauchen, wenn sie für den Betrachter eine ähnliche Relevanz wie Gesichter anderer Ethnien haben.

Es wird demnach erwartet, dass vermehrter Kontakt mit Objekten einer bestimmten Klasse, in diesem Falle mit Geldscheinen, zu einem ähnlichen Effekt wie dem Cross-Race Bias führt. Geldscheine der DM, mit denen die Teilnehmer sehr langen und häufigen Kontakt hatten, und die als 'eigene Währung' einen hohen Grad an Eigen-Gruppen Merkmalen besitzen, sollten besser wiedererkannt werden als Geldscheine anderer Währungen bzw. Geldscheine mit wenigen bzw. keinen Eigen-Gruppen Merkmalen. Indem an zwei Zeitpunkten getestet wurde (vor und nach der Einführung des Euro als gültiges Zahlungsmittel in der Bundesrepublik Deutschland), sollte es bei Geldscheinen des Euro zu einer Verbesserung der Wiedererkennensleistung von Zeitpunkt 1 zu

Zeitpunkt 2 kommen. Bei den Geldscheinen der DM könnte möglicherweise eine Verschlechterung der Wiedererkennensleistung von Zeitpunkt 1 zu Zeitpunkt 2 zu beobachten sein, weil der Kontakt mit der DM nach der Euro-Einführung abnimmt. Dem könnte aber eine zehnjährige Expertise mit den DM-Scheinen entgegenwirken. Beim Britischen Pfund wird keine Veränderung der Wiedererkennensleistung von Zeitpunkt 1 zu Zeitpunkt 2 erwartet. Es sollte sich daher ein Interaktionseffekt der Faktoren 'Währung' und 'Zeit' zeigen. Gleichzeitig bedeutet dies, dass für den Faktor 'Zeit' kein Haupteffekt erwartet wird, weil sich die Niveauunterschiede der einzelnen Währungen ausgleichen sollten. Bei den Währungen sollten sich Unterschiede zwischen dem Britischen Pfund und der DM auf der einen sowie dem Britischen Pfund und dem Euro auf der anderen Seite zeigen, weil für das Britische Pfund keine Veränderungen in der Kontakthäufigkeit von t_1 zu t_2 vermutet werden. Die Anzahl richtiger Identifizierungen sollte für das Britische Pfund geringer ausfallen als für DM und Euro. Die Faktoren 'Art der Fälschung', 'Wertigkeit' und 'Seite' sollten keine Bedeutung für die Identifizierungsleistung haben und damit keine signifikanten Haupteffekte aufweisen. Die postulierten Differenzen zwischen den Faktoren sollen mit Hilfe der Anzahl der Treffer, falschen Alarmen und des Leistungsindex 'A' überprüft werden.

Gleichzeitig lässt sich durch die Betrachtung der Treffer und der falschen Alarme eine Hypothese bezüglich des mirror-effects aufstellen: Bekannte und gut erinnerbare Stimuli weisen im Vergleich zu schlechter erinnerbaren Stimuli eine höhere Anzahl an Treffern sowie eine geringere Anzahl an falschen Alarmen auf. Dieser Effekt sollte sich auch bei den Geldscheinen zeigen. Die Geldscheine der DM sollten mehr Treffer und weniger falsche Alarme aufweisen als die Geldscheine des Euro und des Britischen Pfunds.

2.2.2 Stimulus-Darbietung

Die Delayed-Matching-Aufgabe wurde an zwei Apple 'iMac' Computern mit einer Taktfrequenz von 300 bzw. 350 Mhz, einem Arbeitsspeicher von jeweils 192 MB und einem 15"-Monitor erhoben. Die Farbtiefe betrug 16 Bit (65536 Farben) bei einer Bildschirmauflösung von 1024x768 Pixel und einer Bildwiederholffrequenz von 75 Hz. Die Programmierung erfolgte mit der Applikation 'Superlab' (Version 1.75) der Firma Cedrus (<http://www.cedrus.com>). Die Stimuli für die Delayed-Matching-Aufgabe hatten eine

Größe von (B x H) 211x400 Pixel bei einer Auflösung von 72x72 dpi. Die Farbtiefe der Stimuli betrug 256 Farben.

Der Bildschirmhintergrund war schwarz gehalten. Die Abfolge der Experimente war für alle Versuchspersonen gleich, d.h. zuerst musste die Matching-Aufgabe, dann die AFC-Aufgabe und letztendlich der Wiedererkennenstest bearbeitet werden. Zwischen den Aufgaben am Computer mussten noch sog. Fülleraktivitäten durchgeführt werden, nämlich die Beantwortung von Fragen zur Kontakthäufigkeit der Versuchsperson mit den einzelnen Währungen.

2.2.3 Versuchsdurchführung

Alle Teilnehmer wurden zunächst begrüßt und darüber informiert, dass das Experiment an zwei Zeitpunkten durchgeführt wird. Es wurde ihnen mitgeteilt, dass es dazu nötig ist, sie vor Beginn des zweiten Zeitpunktes per Email bzw. telefonisch davon in Kenntnis zu setzen, damit sie sich in eine Liste für die Belegung der Computer eintragen können. Anschließend begann die Bearbeitung der Experimente mit einer für alle Teilnehmer zu beiden Testzeitpunkten gleichen Reihenfolge (siehe Tabelle 3).

Vor dem Beginn des Experiments mußten die Teilnehmer den ersten Teil des Kontaktfragebogens ausfüllen (siehe Anhang B1, Fragen 1 – 5). Die Teilnehmer wurden instruiert, die Anweisungen auf dem Bildschirm zu lesen und danach mit dem Experiment zu beginnen. Die Instruktionen sind in Anhang A1 wiedergegeben.

Für die Darbietung der Stimuli mussten 72 Trials programmiert werden, d.h. 72 festgelegte Abfolgen der drei Phasen um eine vollständige Kreuzung der Faktoren zu gewährleisten. Ein Trial bestand aus fünf Phasen:

- 1) Präsentation eines 'echten' Geldscheines für 1000 ms;
- 2) Präsentation einer Füllergrafik (Random-Dots Grafik, den ganzen Bildschirm ausfüllend) für 1500 ms;
- 3) Präsentation des 'echten' Geldscheines aus Phase 1 oder der Spiegelungs- bzw. Farbfälschung des Geldscheines aus Phase 1 für 1000 ms;
- 4) Eingabe der Antwort ('Geldscheine sind identisch'; 'Geldscheine sind nicht identisch') durch Drücken der Tasten 'c' oder 'm';
- 5) Rückmeldung über die Richtigkeit der Antwort durch Darbietung der Wörter 'Richtig' bzw. 'Falsch'.

Nach Bearbeitung der zwei Beispieldurchgänge sollten die Teilnehmer nun die 72 Trials in randomisierter Reihenfolge bearbeiten.

Zum ersten Testzeitpunkt wurden die Teilnehmer nochmals darauf hingewiesen, dass sie kontaktiert werden, wenn die zweite Phase der Erhebung nach Einführung des Euro beginnt und sie sich in die Listen eintragen sollten. Nach Beendigung des Versuchsdurchganges zum zweiten Testzeitpunkt wurde den Teilnehmern für ihre Teilnahme gedankt und sie erhielten je nach Wunsch entweder Versuchspersonenstunden oder € 15,-.

2.2.4 Ergebnisse

Die Daten der Delayed-Matching Aufgabe wurden nach der Signal-Entdeckungstheorie ausgewertet. Ein Treffer ist in diesem Zusammenhang die Antwort 'Geldscheine sind identisch' bei Darbietung eines normalen Geldscheins in der dritten Phase eines Durchgangs, d. h. nach Präsentation des Füller- Bildschirms. Als falscher Alarm wird die Antwort 'Geldscheine sind identisch' nach Darbietung einer Farb- oder Spiegelungsfälschung gewertet. Die Maße der Signal-Entdeckungstheorie werden u.a. getrennt nach der Art der Fälschung ausgewertet, um mögliche Unterschiede zwischen den Fälschungsarten zu erkennen. Zunächst werden die Ergebnisse der Analyse der Treffer und der falschen Alarme dargestellt. Anschließend wird auf das Leistungsmaß A' und die Reaktionsneigung B'' näher eingegangen. Das Alpha-Niveau wird nach Bonferroni-Adjustierung bei sechs geplanten Kontrasten für die formulierten Hypothesen auf $\alpha = .008$ festgelegt.

Die Hypothese, nach der bei der DM die Anzahl der Treffer höher ist als bei den anderen beiden Währungen, kann teilweise bestätigt werden. Die DM unterscheidet sich vom Euro nicht, $F(1,38) = 1.39$, *ns*, während sich die DM ($M = .913$) vom Britischen Pfund signifikant unterscheidet ($M = .849$), $F(1,38) = 16.04$, $p = .001$, $r = .55$. Der Euro unterscheidet sich ebenfalls nicht signifikant vom Britischen Pfund, $F < 1$ *ns*. Zur Überprüfung der Hypothese, dass zwischen t_1 und t_2 beim Euro Leistungszuwächse zu beobachten sind, während die Leistung bei der DM abfällt und beim Britischen Pfund gleich bleibt, wurden die Differenzen der Treffer zwischen t_1 und t_2 für jede Währung berechnet und mittels t -Tests überprüft, ob diese Differenz signifikant von Null verschieden ist. Zwar zeigte sich deskriptiv bei allen drei Währungen ein Zuwachs von t_1 zu t_2 , dieser ist aber nur bei der DM signifikant, $t(41) = 3.53$, $p = .001$, $r = .263$,

($M_{Diff} = .125$). Daraus folgernd zeigte sich ebenfalls ein signifikanter Haupteffekt über den Faktor 'Zeit' mit , $F(1,38) = 6.40$, $p = .016$, $r = .38$, ($M_{t1} = .849$, $M_{t2} = .919$).

Die Treffer wurden zusätzlich mit einer Varianzanalyse mit Messwiederholung und den Between-Subjects Faktoren 'Geschlecht' (männlich, weiblich) und 'Taste' ('c', 'm') sowie den Within-Subjects Faktoren 'Zeit' (Zeitpunkt 1, Zeitpunkt 2), 'Währung' (DM, Euro, Pfund), 'Wertigkeit' (1,2,3,4) sowie 'Seite' (Vorderseite, Rückseite) analysiert. Der Faktor 'Geschlecht' wurde in dem Versuchsdesign zwar nicht systematisch kontrolliert, soll aber in die Analyse aufgenommen werden, um eventuelle Geschlechtsunterschiede aufzudecken.

Wie bei Betrachtung der Analysen der Differenzen zwischen t_1 und t_2 bereits erwartet, zeigen die Faktoren 'Währung' und 'Zeit' eine marginal signifikante Interaktion, $F(2,76) = 3.62$, $p = .032$, $\eta^2 = .09$, jedoch nicht in die erwartete Richtung. Zwar zeigt sich beim Euro ein Zuwachs der Treffer von Zeitpunkt 1 zu Zeitpunkt 2, dies trifft aber auch für die anderen beiden Währungen zu, sogar mit dem größten Zuwachs bei der DM (siehe Tabelle 6). Wider Erwarten hatte die Ansicht der Vorder- bzw. Rückseite einen Einfluss auf die Trefferrate, $F(1,38) = 4.47$, $p = .041$, $r = .32$.

Tabelle 6

Mittelwerte der Treffer, der falschen Alarme, des Leistungsindex \underline{A}' und der Reaktionsneigung \underline{B}'' in der Delayed-Matching Aufgabe getrennt nach Zeitpunkt und Währung

Währung	Zeitpunkt		\underline{M} :
	t1	t2	
Deutsche Mark			
Treffer	.850 ^a	.976 ^a	.913
Falsche Alarme (Farbe)	.048	.016	.032
Falsche Alarme (Spiegel)	.187	.154	.170
\underline{A}' (Farbe)	.921 ^a	.989 ^a	.955
\underline{A}' (Spiegel)	.885 ^a	.954 ^a	.919
\underline{B}'' (Farbe)	.341 ^a	.070 ^a	.206
\underline{B}'' (Spiegel)	-.251 ^a	-.447 ^a	-.349
Euro:			
Treffer	.873	.908	.891
Falsche Alarme (Farbe)	.148 ^a	.056 ^a	.102
Falsche Alarme (Spiegel)	.254 ^a	.165 ^a	.210
\underline{A}' (Farbe)	.904 ^a	.958 ^a	.931
\underline{A}' (Spiegel)	.865 ^a	.922 ^a	.893
\underline{B}'' (Farbe)	-.015	.106	.045
\underline{B}'' (Spiegel)	-.157	-.276	-.216
Britisches Pfund:			
Treffer	.825	.874	.849
Falsche Alarme (Farbe)	.182 ^a	.120 ^a	.151
Falsche Alarme (Spiegel)	.180	.132	.156
\underline{A}' (Farbe)	.881	.932	.907
\underline{A}' (Spiegel)	.877	.928	.903
\underline{B}'' (Farbe)	-.159	-.069	-.114
\underline{B}'' (Spiegel)	-.263	-.263	-.263
\underline{M} :			
Treffer	.849 ^a	.919 ^a	.884
Falsche Alarme (Farbe)	.126 ^a	.064 ^a	.095
Falsche Alarme (Spiegel)	.207 ^a	.150 ^a	.179
\underline{A}' (Farbe)	.902 ^a	.960 ^a	.931
\underline{A}' (Spiegel)	.877 ^a	.935 ^a	.906
\underline{B}'' (Farbe)	.056	.036	.046
\underline{B}'' (Spiegel)	-.224	-.329	-.276

Anmerkung: ^aDie Zellenmittelwerte innerhalb einer Zeile unterscheiden sich nach Bonferroni-Adjustierung signifikant (post-hoc Vergleiche, $p < .001$).

Die Anzahl der Treffer bei der Rückseite der Geldscheine war geringer ($\underline{M} = .870$) als bei der Vorderseite ($\underline{M} = .900$). Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 2.48$, alle $p_s > .065$).

Die Analyse der falschen Alarme geschah mittels einer Varianzanalyse mit Messwiederholung und den Between-Subjects Faktoren 'Geschlecht' (männlich, weiblich) und 'Taste' ('c', 'm') sowie den Within-Subjects Faktoren 'Zeit' (Zeitpunkt 1, Zeitpunkt 2), 'Währung' (DM, Euro, Pfund), 'Wertigkeit' (1,2,3,4), 'Art der Fälschung' (Farbfälschung, Spiegelungsfälschung) sowie 'Seite' (Vorderseite, Rückseite). Von Zeitpunkt 1 zu Zeitpunkt 2 sank die Zahl der falschen Alarme signifikant, $F(1,38) = 14.85$, $p = .001$, $r = .53$ ($M_{t1} = .167$; $M_{t2} = .107$). Die Anzahl der falschen Alarme bei der DM war am geringsten ($M_{DM} = .101$), gefolgt vom Britischen Pfund ($M_{BP} = .153$) und dem Euro ($M_{EU} = .156$), $F(2,76) = 9.64$, $p = .001$, $\eta^2 = .20$. Es zeigte sich keine Interaktion der Faktoren 'Währung' und 'Zeit', $F(2,76) = 2.63$, $p = .079$, $\eta^2 = .07$.

Die Wertigkeit der Geldscheine hatte ebenfalls einen Einfluss auf die Anzahl der falschen Alarme, $F(3,114) = 17.02$, $p = .001$, $\eta^2 = .31$. Die wenigsten Falschen Alarme gab es bei den zweithöchsten Werten der Geldscheine ($M_3 = .098$) gefolgt von den Geldscheinen mit der niedrigsten Wertigkeit ($M_1 = .130$) und den Geldscheinen mit der höchsten Wertigkeit ($M_4 = .134$). Die meisten falschen Alarme resultierten aus den Geldscheinen mit der zweitniedrigsten Wertigkeit ($M_2 = .186$). Bei anschließenden post-hoc Vergleichen zeigt sich, dass sich lediglich die Geldscheine mit der niedrigsten und der höchsten Wertigkeit nicht signifikant voneinander unterscheiden.

Einen großen Einfluss hatte die Art der Fälschung auf die falschen Alarme mit $F(1,38) = 52.67$, $p = .001$, $r = .76$. Die Spiegelungsgefälschten Geldscheine wurden häufiger als mit dem echten Geldschein identisch angesehen ($M_{SF} = .179$) als die farbgefälschten Geldscheine ($M_{FF} = .095$). Die Seite der Geldscheine zeigte einen signifikanten Haupteffekt $F(1,38) = 26.04$, $p = .001$, $r = .64$, mit mehr falschen Alarmen bei der Rückseite der Geldscheine ($M_{Rü} = .161$) als bei der Vorderseite ($M_{Vo} = .113$).

Die Interaktion der Faktoren 'Währung' und 'Art der Fälschung' war signifikant, $F(2,76) = 13.82$, $p = .001$, $\eta^2 = .27$. Während es beim Britischen Pfund keine Unterschiede zwischen den Fälschungsarten gab, $F < 1$, *ns*, zeigten sich bei der DM, $F(1,76) = 7.21$, $p = .009$, $r = .29$ ($M_{SF} = .170$, $M_{FF} = .032$), und dem Euro, $F(1,76) = 4.41$, $p = .039$, $r = .23$ ($M_{SF} = .210$, $M_{FF} = .102$) bei den Spiegelungsfälschungen mehr falsche Alarme als bei den Farbfälschungen. Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 2,02$, alle $p_s > .64$).

Die Hypothesen wurden ebenfalls bezüglich des Leistungsindex A' ausgewertet. A' sollte bei der DM höher sein als bei den anderen beiden Währungen. Diese Hypothese kann bestätigt werden. Einfache geplante Vergleiche zeigen signifikante Unterschiede zwischen der DM und dem Euro, $F(1,38) = 7.09$, $p = .011$, $r = .40$, ($A'_{DM} = .937$,

$\underline{A}'_{Eu} = .912$), und dem Britischen Pfund, $\underline{F}(1,38) = 17.76$, $p = .001$, $r = .56$, ($\underline{A}'_{DM} = .937$, $\underline{A}'_{BP} = .905$). Zur Überprüfung der Hypothese, dass zwischen t_1 und t_2 beim Euro Leistungszuwächse zu beobachten sind, während die Leistung bei der DM abfällt und beim Britischen Pfund gleich bleibt, wurden wie bei der Analyse der Treffer die Differenzen von \underline{A}' zwischen t_1 und t_2 für jede Währung berechnet und mittels t -Tests überprüft, ob diese Differenz signifikant von Null verschieden ist. Dabei zeigte sich für alle drei Währungen ein signifikanter Zuwachs von \underline{A}' zwischen t_1 und t_2 . Bei der DM beträgt dieser Zuwachs $\underline{A}' = .069$, $t(41) = 2.82$, $p = .007$, $r = .21$, beim Euro ist der Zuwachs $\underline{A}' = .056$, $t(41) = 2.62$, $p = .012$, $r = .20$, und beim Britischen Pfund beträgt der Zuwachs $\underline{A}' = .051$, $t(41) = 2.55$, $p = .014$, $r = .19$. Ebenfalls entgegen der Erwartung zeigte sich ein signifikanter Kontrast über den Faktor 'Zeit' mit, $\underline{F}(1,38) = 7.05$, $p = .011$, $r = .39$, ($\underline{M}_{t1} = .890$, $\underline{M}_{t2} = .948$).

Analog zu den Treffern und falschen Alarme wurde der Leistungsindex \underline{A}' mit einer Varianzanalyse mit Messwiederholung und den Between-Subjects Faktoren 'Geschlecht' (männlich, weiblich) und 'Taste' ('c', 'm') sowie den Within-Subjects Faktoren 'Zeit' (Zeitpunkt 1, Zeitpunkt 2), 'Währung' (DM, Euro, Pfund), 'Art der Fälschung' (Farbfälschung, Spiegelungsfälschung) durchgeführt. Da \underline{A}' Treffer und falsche Alarme zueinander in Beziehung setzt, erscheint es sinnvoll, \underline{A}' nicht auf jeder einzelnen Faktorstufe zu berechnen, sondern die Daten über mehrere Bedingungen zu akkumulieren. Es wurde darauf verzichtet, die Faktoren 'Wertigkeit' und 'Seite' in die Analyse mit aufzunehmen, da es für diese Faktoren keine Hypothesen gab.

Die vermutete Interaktion der Faktoren 'Währung' und 'Zeit' zeigte sich nicht, $\underline{F}(2,76) = .48$, $p = .622$, $\eta^2 = .01$. Die Art der Fälschung hatte einen signifikanten Einfluss auf den Leistungsindex mit $\underline{F}(1,38) = 30.48$, $p = .001$, $r = .67$ ($\underline{A}'_{FF} = .931$, $\underline{A}'_{SF} = .905$). Die Interaktion der Faktoren 'Währung' und 'Art der Fälschung' ist signifikant mit $\underline{F}(2,76) = 7.38$, $p = .001$, $\eta^2 = .16$. Beim Britischen Pfund zeigten sich keine Unterschiede zwischen den Fälschungsarten, $\underline{F} < 1$, ns , während sowohl bei der DM, $\underline{F}(1,38) = 29.58$, $p = .001$, $r = .662$ ($\underline{A}'_{FF} = .955$, $\underline{A}'_{SF} = .919$) als auch beim Euro mit $\underline{F}(1,38) = 32.96$, $p = .001$, $r = .682$ ($\underline{A}'_{FF} = .931$, $\underline{A}'_{SF} = .893$) bessere Leistungen bei den Farbfälschungen verglichen mit den Spiegelungsfälschungen gezeigt wurden. Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $\underline{F}s < 2.49$, alle $ps > .090$).

Schließlich wurde die Reaktionsneigung \underline{B}' mittels einer Varianzanalyse mit Messwiederholung und den Between-Subjects Faktoren 'Geschlecht' (männlich, weiblich) und 'Taste' ('c', 'm') sowie den Within-Subjects Faktoren 'Zeit' (Zeitpunkt 1, Zeitpunkt 2),

'Währung' (DM, Euro, Pfund), 'Art der Fälschung' (Farbfälschung, Spiegelungsfälschung) analysiert.

Die Teilnehmer legten in Abhängigkeit von der Fälschung der Geldscheine ein unterschiedliches Antwortverhalten an, $F(1,38) = 52.24$, $p = .001$, $r = .76$. Während es bei den Farbfälschungen keine Tendenz zu einem bestimmten Antwortverhalten gab ($B'' = .046$), zeichnet sich bei den Spiegelungsfälschungen ein eher liberales Antwortverhalten ab ($B'' = -.276$). Dies ist aber nicht für alle Währungen gleich, wie an der Interaktion von 'Währung' und 'Art der Fälschung' zu erkennen ist, $F(2,76) = 11.64$, $p = .001$, $\eta^2 = .23$. Die Spiegelungsfälschungen unterscheiden sich zwischen den Währungen nicht (alle $F_s < 3.95$, alle $p_s > .054$). Die Farbfälschungen hingegen unterscheiden sich signifikant zwischen den Währungen mit einem eher konservativen Antwortverhalten bei der DM von $B'' = .206$, einer fehlenden Tendenz beim Euro von $B'' = .045$ und einem eher liberalen Kriterium beim Britischen Pfund von $B'' = -.114$. Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 1.85$, alle $p_s > .164$).

2.2.5 Diskussion der Ergebnisse der Delayed-Matching-Aufgabe

Die postulierte Interaktion der Faktoren 'Zeit' und 'Währung' konnte bei keinem der verwendeten Maße gefunden werden. Zwar war die Leistung zum Zeitpunkt 1 bei den Eigen-Gruppen Geldscheinen besser als bei den Fremd-Gruppen Geldscheinen, was zunächst für einen Cross-Currency Effekt spricht. Im Vergleich zu Zeitpunkt 1 steigt zum Zeitpunkt 2 die Leistung jedoch bei allen drei Währungen an. Da nicht davon ausgegangen wird, dass alle Teilnehmer während der Phase zwischen den beiden Messzeitpunkten in gleichem Ausmaß Kontakt mit dem Britischen Pfund hatten, ist der Anstieg der Leistungen mit ziemlicher Sicherheit auf Übungseffekte der Teilnehmer zurückzuführen. Die Hypothese, nach der kein Haupteffekt für den Faktor 'Zeit' erwartet wurde, konnte daher nicht bestätigt werden, da bei allen drei Währungen signifikante Verbesserungen von Zeitpunkt 1 zu Zeitpunkt 2 beobachtet wurden. Dennoch kann bei Betrachtung der besseren Leistungen bei der DM zu beiden Messzeitpunkten im Vergleich zu den Geldscheinen des Euro und des Britischen Pfunds die Hypothese eines Cross-Currency Effekts noch aufrecht erhalten werden. Die DM als 'eigene Währung' sollte besser als der Euro und das Britische Pfund bearbeitet werden. Diese Hypothese wird durch den einen

starken Haupteffekt 'Währung', $r = .55$, bestätigt, der einen besseren Leistungsindex bei der DM verglichen mit dem Euro und dem Britischen Pfund zeigt. Eine Vermutung, die allerdings noch überprüft werden müsste, ist, dass die vier bis fünf Monate, die zwischen den beiden Messzeitpunkten lagen, ein zu kurzer Zeitraum war, um die anderen postulierten Effekte zu generieren. Die DM war zum zweiten Messzeitpunkt noch nicht lange durch den Euro abgelöst worden und dieser Messzeitpunkt fiel ebenfalls in eine Zeit, in der viel in den Medien über Verteuerungen durch den Euro gesprochen wurde. Es wäre in diesem Zusammenhang interessant zu wissen, wie die Leistungen zum jetzigen Zeitpunkt ausfallen würden, wenn das gleiche Experiment mit anderen Teilnehmern noch einmal durchgeführt werden würde.

Ein mirror-effect, wie er auch in Studien zum Cross-Race Bias berichtet wird, konnte ebenfalls beobachtet werden. Bei der DM gab es mehr Treffer als beim Euro oder beim Britischen Pfund. Ebenso werden bei der DM weniger falsche Alarmer gemeldet als bei den beiden anderen Währungen. Bei allen drei Währungen liess sich eine Verbesserung, d.h. mehr Treffer und weniger falsche Alarmer, von Zeitpunkt 1 zu Zeitpunkt 2 feststellen.

Die Art der Fälschung erzeugte überraschenderweise signifikante Unterschiede in der Leistung mit einem starken Effekt von $r = .67$. Die Teilnehmer waren bei den Farbfälschungen der DM und des Euro besser als bei den jeweiligen Spiegelungsfälschungen, während es beim Britischen Pfund keine Unterschiede zwischen den Fälschungsarten gab. Da dieser Effekt auch bei der AFC-Aufgabe und dem Wiedererkennenstest auftrat, wird er in der Gesamtdiskussion der Experimente zum Cross-Currency Effekt besprochen.

Ein interessantes Detail für die Interpretation der Daten bietet der Haupteffekt des Faktors 'Seite' in der Analyse der Treffer mit einer mittleren Effektstärke von $r = .32$. Für die Seite der Geldscheine wurden keine Hypothesen formuliert, weil nicht davon auszugehen war, dass bei diesem Faktor Unterschiede auftreten. Tatsächlich war es aber so, dass bei den Vorderseiten der Geldscheine ebenfalls weniger falsche Alarmer auftraten als bei den Rückseiten. Dieser Effekt ist für alle drei Währungen gleich und schwerlich auf ein bestimmtes Merkmal zurückzuführen. Auf den Vorderseiten der DM sind Gesichter abgebildet, auf den Vorderseiten des Euro Tore verschiedener Kulturen und auf den Vorderseiten sowie auf den Rückseiten des Britischen Pfundes sind Gesichter zu finden. Eine eindeutige Aussage, nach der dieser Effekt auf systematische Unterschiede zwischen den Vorder- und Rückseiten zurückzuführen ist, fehlt zur Zeit.

Warum die Wertigkeit der Geldscheine einen Einfluss auf die Anzahl der falschen Alarme hatte, ist ebenfalls schwer nachzuvollziehen. Eine mögliche Erklärung hierfür ist, dass zumindest die Geldscheine der DM, bei denen die falschen Alarme am geringsten waren, von den Teilnehmern am häufigsten benutzt wurden und es ihnen daher leichter gefallen ist, deren Fälschungen als Fälschungen zu identifizieren und sie dementsprechend zurückzuweisen. Im folgenden Abschnitt wird die AFC-Aufgabe als zweites Experiment zum Cross-Currency Effekt beschrieben.

2.3 Experiment II: AFC-Aufgabe

Das zweite Experiment zum Cross-Currency Effekt benutzte eine AFC-Aufgabe als experimentelles Paradigma. Zunächst werden die Hypothesen erläutert und die Durchführung des Experiments dargestellt. Die Ergebnisse und die Diskussion der AFC-Aufgabe bilden den Abschluss.

2.3.1 Hypothesen

Für die AFC-Aufgabe gelten die gleichen Annahmen wie für die Delayed-Matching-Aufgabe. Geldscheine der DM, mit denen die Teilnehmer sehr langen und häufigen Kontakt haben und die als 'eigene Währung' einen hohen Grad an Eigen-Gruppen Merkmalen besitzen, sollten besser wiedererkannt werden als Geldscheine anderer Währungen bzw. Geldscheine mit wenigen bzw. keinen Eigen-Gruppen Merkmalen. Es wird vermutet, dass vermehrter Kontakt mit Objekten der Klasse 'Geldscheine' zu einem Cross-Currency Effekt führt, nämlich einer Zunahme richtiger Identifizierungen von Euro-Geldscheinen bei sinkendem (DM) bzw. gleichbleibendem (Britisches Pfund) Niveau richtiger Identifizierungen von t_1 zu t_2 . Es wird ein Interaktionseffekt der Faktoren 'Währung' und 'Zeit' erwartet. Gleichzeitig bedeutet dies, dass für den Faktor 'Zeit' kein Haupteffekt erwartet wird, weil sich die Niveauunterschiede der einzelnen Währungen ausgleichen sollten. Bei den Währungen sollten sich Unterschiede zwischen dem Britischen Pfund auf der einen, und der DM sowie dem Euro auf der anderen Seite zeigen, weil für das Britische Pfund keine Veränderungen in der Kontakthäufigkeit von t_1 zu t_2 vermutet werden. Die Anzahl richtiger Identifizierungen sollte für das Britische Pfund geringer ausfallen als für die DM und den Euro. Die Faktoren 'Art der Fälschung',

'Wertigkeit' und 'Seite' sollten keine Bedeutung für die Identifikationsleistung haben und damit keine signifikanten Haupteffekte aufweisen.

2.3.2 Stimulus-Darbietung

Die AFC-Aufgabe wurde an zwei Apple 'iMac' Computern mit einer Taktfrequenz von 300 bzw. 350 Mhz, einem Arbeitsspeicher von jeweils 192 MB und einem 15"-Monitor erhoben. Die Farbtiefe betrug 16 Bit (65536 Farben) bei einer Bildschirmauflösung von 1024x768 Pixel und einer Bildwiederholfrequenz von 75 Hz. Die Programmierung erfolgte mit der Applikation 'SuperLab' (Version 1.75) der Firma Cedrus. Die Stimuli für die AFC-Aufgabe hatten eine Größe von 400x420 Pixel (bei Darbietung von zwei Geldscheinen) bzw. 400x200 Pixel (für die Darbietung der Geldscheine bei der Rückmeldung) bei einer Auflösung von 72x72 dpi. Die Farbtiefe der Stimuli betrug 256 Farben.

Die Stimuli mussten für die AFC-Aufgabe für die einzelnen Trials in einem weiteren Schritt bearbeitet werden, weil in jedem Trial zwei Geldscheine gleichzeitig dargeboten wurden. Dazu wurden in Photoshop ein echter Geldschein und, je nach Bedingung ein zweiter, in Wertigkeit und Seitenansicht gleichwertiger, farb- oder spiegelungsgefälschter Geldschein geladen und in eine Vorlage mit den Abmessungen (B x H) von 400 x 420 Pixel kopiert. Die Geldscheine wurden dann oben und unten in der Vorlage so justiert, dass die linken und rechten Kanten der Geldscheine genau übereinander lagen. Die beiden Scheine sollten dabei aber nicht exakt aneinander liegen, um sie optisch voneinander zu trennen. Dabei entstand ein weißer Bereich zwischen den Geldscheinen der mittels eines schwarz ausgefüllten Kastens in Photoshop gefärbt und damit dem Hintergrund in Superlab angepasst wurde.

2.3.3 Versuchsdurchführung

Die AFC-Aufgabe folgte dem zweiten Teil des Kontaktfragebogens (siehe Anhang B1, Fragen 6 - 15). Die Teilnehmer sollten auch hier den Instruktionen auf dem Bildschirm folgen (siehe Anhang A2). Nachdem die ersten fünf Instruktionen gelesen wurden, begann das Experiment mit der Darbietung von drei Zeichen (®, ∞, ×) für jeweils 2000 ms mit

einem Inter-Trial-Intervall (ITI) von 500 ms, um die Teilnehmer auf den Beginn der experimentellen Phase vorzubereiten.

Anschließend sollten in randomisierter Reihenfolge 24 Trials bearbeitet werden, wobei jeder Trial den folgenden Aufbau hatte:

- 1) Präsentation von zwei Geldscheinen (ein 'echter' und eine Spiegelungs- bzw. Farbfälschung) der gleichen Währung und gleicher Wertigkeit für 2000 ms;
 - 2) Eingabe der Antwort durch Drücken der Taste 'c' oder 'm';
 - 3) Rückmeldung über die Richtigkeit der Antwort ('Richtig' vs. 'Falsch') für 2000 ms;
 - 4) Präsentation des 'echten' Geldscheines für 2000 ms mit anschließendem ITI von 500 ms.
- Nach Bearbeitung der ersten 24 Trials folgte die Instruktion Nr. 6 (siehe Anhang A2). Anschließend sollten die letzten 24 Trials bearbeitet werden, die in ihrem Aufbau sich von den ersten 24 Trials lediglich darin unterschieden, dass die Phase 1) nicht mehr auf 2000 ms beschränkt war. Die Teilnehmer sollten bei diesen letzten 24 Trials so schnell wie möglich angeben, welcher der Geldscheine der echte Geldschein war. Nach Beendigung der 48 Trials endete das Experiment.

2.3.4 Ergebnisse

Die AFC-Aufgabe konnte im Gegensatz zu der Delayed-Matching Aufgabe und dem Wiedererkennenstest nicht nach der Signal-Entdeckungs Theorie ausgewertet werden, weil bei dieser Aufgabe in der Stimulus-Darbietung nicht zwischen 'target-absent' und 'target-present' Durchgängen unterschieden wird. Der richtige, bzw. alte Stimulus ist in jedem Durchgang vorhanden. Daher wird in diesem Experiment als abhängige Variable die korrekte Anzahl richtiger Identifikationen der echten Geldscheine analysiert. Das Alpha-Niveau wird nach Bonferroni-Adjustierung bei geplanten Kontrasten für die formulierten Hypothesen auf $\alpha = .008$ festgelegt.

Die Vermutung, dass die Leistung bei der DM besser ist als bei den anderen beiden Währungen konnte bestätigt werden. Beide einfachen Kontraste sind signifikant. Die DM konnte besser differenziert werden als der Euro, $F(1,38) = 21.28$, $p = .001$, $r = .60$, ($M_{DM} = .860$, $M_{eu} = .774$), und ebenfalls besser als das Britische Pfund, $F(1,38) = 23.88$, $p = .001$, $r = .62$, ($M_{DM} = .860$, $M_{BP} = .757$). Zwischen dem Euro und dem Britischen Pfund bestehen keine signifikanten Unterschiede, $F < 1$, ns. Zur Überprüfung der Hypothese, dass zwischen t_1 und t_2 beim Euro Leistungszuwächse zu beobachten sind, während die Leistung bei der DM abfällt und beim Britischen Pfund gleich bleibt, wurden

analog zur Delayed-Matching Aufgabe die Differenzen der korrekten Antworten zwischen t_1 und t_2 für jede Währung berechnet und mittels t -Tests überprüft, ob diese Differenz signifikant von Null verschieden ist. Dabei zeigten sich für das Britische Pfund und den Euro jeweils ein signifikanter Zuwachs t_1 und t_2 . Beim Euro beträgt der Zuwachs $\underline{M} = .098$, $\underline{t}(41) = 5.67$, $\underline{p} = .001$, $\underline{r} = .40$, und beim Britischen Pfund ist der Zuwachs $\underline{M} = .077$, $\underline{t}(41) = 3.16$, $\underline{p} = .003$, $\underline{r} = .24$. Bei der DM konnte kein Zuwachs festgestellt werden, $\underline{t}(41) = .84$, $\underline{p} = .408$, $\underline{r} = .06$. Entgegen der Hypothese gab es von Zeitpunkt 1 zu Zeitpunkt 2 insgesamt einen signifikanten Zuwachs an richtigen Antworten, $\underline{F}(1,38) = 26.03$, $\underline{p} = .001$, $\underline{r} = .64$, ($\underline{M}_{t1} = .764$, $\underline{M}_{t2} = .830$).

Die Daten für die AFC-Aufgabe wurden anschließend analog zur Delayed-Matching Aufgabe mit Hilfe einer Varianzanalyse mit Messwiederholung mit den Between-Subjects Faktoren 'Geschlecht' und 'Taste' sowie den Within-Subjects Faktoren 'Zeit' (Zeitpunkt 1, Zeitpunkt 2), 'Währung' (DM, Euro; Britisches Pfund), 'Wertigkeit' (1,2,3,4), und 'Art der Fälschung' (Farbfälschung, Spiegelungsfälschung) analysiert. Da für eine vollständige messwiederholte ANOVA mit allen Faktoren keine ausreichende Anzahl an Stimuli zur Verfügung stand, wurden die Within-Subjects Faktoren 'Seite' (Vorderseite, Rückseite) und 'Darbietung' (echter Geldschein oben/unten) mit zwei getrennten Analysen in die Berechnungen eingebunden.

Bezüglich des Faktors 'Taste' zeigte sich ein marginal signifikanter Haupteffekt, $\underline{F}(1,38) = 6.56$, $\underline{p} = .015$, $\underline{r} = .38$. Teilnehmer, die die Taste 'c' drücken mussten, um anzuzeigen, dass der obere Geldschein der 'echte' Geldschein ist, waren besser als die Teilnehmer, die die Taste 'm' für 'echte' Geldscheine in der oberen Hälfte des Bildschirms betätigen sollten ($\underline{M}_c = .824$ vs. $\underline{M}_m = .770$)

Entsprechend der Analyse der Differenzen der einzelnen Währungen von Zeitpunkt 1 zu Zeitpunkt 2 zeigte sich ein signifikanter Interaktionseffekt der Faktoren 'Währung' und 'Zeit' $\underline{F}(2,76) = 4.13$, $\underline{p} = .020$, $\eta^2 = .10$ (siehe Abbildung 10).

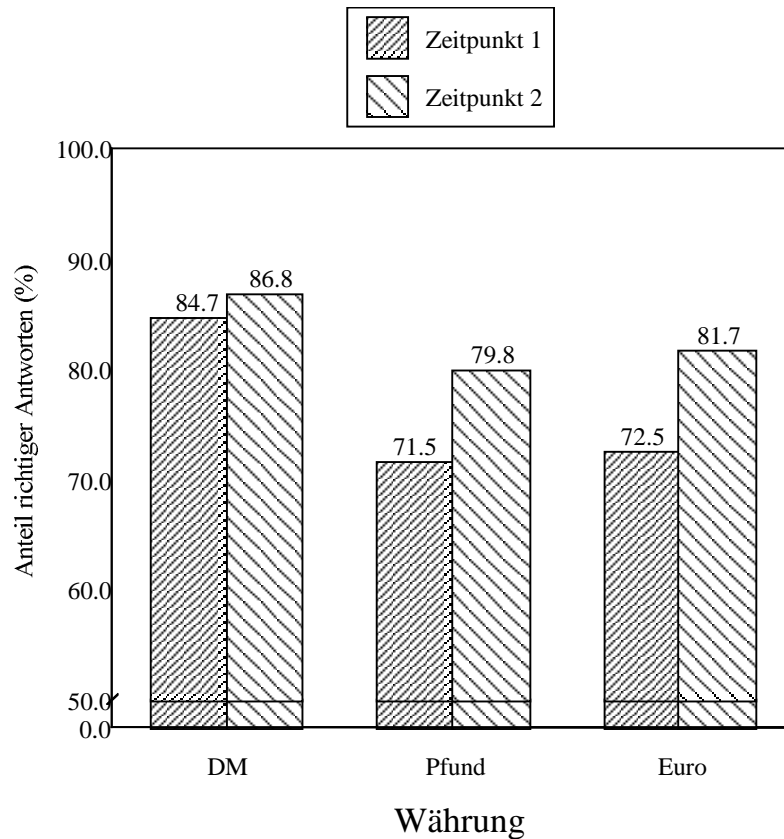


Abbildung 10. Anteil richtiger Antworten in der AFC-Aufgabe in Abhängigkeit von den Faktoren 'Währung' und 'Zeit'

Die Wertigkeiten der Geldscheine haben einen Einfluss auf die Leistung bei der AFC-Aufgabe. ($F(3,114) = 5.55$, $p = .001$, $\eta^2 = .13$). Dies liegt daran, dass die Leistungen zwischen Geldscheinen mit niedrigsten Wertigkeiten und den Geldscheinen mit den zweitniedrigsten Wertigkeiten, $F(1,38) = 5.71$, $p = .022$, $r = .22$, sowie zwischen Geldscheinen mit niedrigsten Wertigkeiten und den Geldscheinen mit den zweit-höchsten Wertigkeiten $F(1,114) = 15.28$, $p = .001$, $r = .34$, signifikant unterschiedliche sind, ($M_1 = .758$, $M_2 = .803$, $M_3 = .836$, $M_4 = .791$). Der Faktor 'Art der Fälschung' zeigt ebenfalls einen signifikanten Haupteffekt, $F(1,38) = 27.98$, $p = .001$, $r = .63$ ($M_{SF} = .757$, $M_{FF} = .837$). Der Faktor 'Seite' ist signifikant mit $F(1,38) = 18.90$, $p = .001$, $r = .58$, ($M_{Vo} = .824$, $M_{Rü} = .770$), ebenso zeigt sich wiederum eine signifikante Interaktion der Faktoren 'Währung' und 'Art der Fälschung', $F(2,76) = 33.53$, $p = .001$, $\eta^2 = .47$. Bei der DM, $F(1,76) = 8.70$, $p = .004$, $r = .32$, ($M_{SF} = .767$, $M_{FF} = .953$) werden die Farbfälschungen besser identifiziert als die Spiegelungsfälschungen. Dies ist deskriptiv auch beim Euro der Fall, ($M_{SF} = .713$, $M_{FF} = .835$), jedoch verfehlt der Vergleich ein signifikantes Niveau.

Beim Britischen Pfund besteht ebenfalls kein signifikanter Unterschied, ($\underline{M}_{SF} = .790$, $\underline{M}_{FF} = .723$). Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $\underline{F}_s < 2.67$, alle $p_s > .055$).

2.3.5 Diskussion der Ergebnisse der AFC-Aufgabe

Die Analyse der richtigen Antworten in der AFC-Aufgabe zeigt ein ähnliches Bild wie in der Delayed-Matching Aufgabe. Die DM als 'Eigen-Gruppen' Geldschein konnte besser differenziert werden als der Euro und das Britische Pfund. Es wurden bei den einzelnen Vergleichen der DM mit dem Euro und dem Britischen Pfund starke Effekte gefunden ($\underline{r} = .60$, bzw. $\underline{r} = .62$). Es kann daher von einem Cross-Currency Effekt gesprochen werden. Hier wäre es ebenfalls interessant zu wissen, wie die Leistungen nach einem längeren Zeitraum des Kontakts und der Erfahrung mit dem Euro als 'Eigen-Gruppen' Geldscheinen aussehen. Es zeigte sich hier im Gegensatz zur Delayed-Matching Aufgabe zwar eine signifikante Interaktion der Faktoren 'Zeit' und 'Währung', $\eta^2 = .10$, jedoch nicht in der Art und Weise, wie es in der Hypothese vermutet wurde. Die Leistung bei der DM ist zum Zeitpunkt 2 nicht gesunken, wie es postuliert wurde.

Interessanterweise ist die Leistung bei der DM jedoch von Zeitpunkt 1 zu Zeitpunkt 2 aber auch nicht signifikant angestiegen, während ein signifikanter Anstieg beim Euro und beim Britischen Pfund zu beobachten ist. Dieser Anstieg fällt beim Euro deutlicher aus ($\underline{r} = .40$) als beim Britischen Pfund ($\underline{r} = .24$). Der fehlende Anstieg bei der DM in der AFC-Aufgabe verglichen mit der Delayed-Matching Aufgabe ist auch als Folge eines Deckeneffektes in der AFC-Aufgabe zu erklären. Bei einer direkten Vergleichbarkeit der Delayed-Matching Aufgabe und der AFC-Aufgabe hieße dies, dass bei der AFC-Aufgabe die höchste Leistung nicht abgerufen werden konnte. Die beiden Paradigmen sind jedoch nicht in allen Punkten vergleichbar, da verschiedene Gedächtnisprozesse beteiligt sind. Bei der Delayed-Matching Aufgabe wird von dem Teilnehmer ein Abgleich zweier kurz hintereinander dargebotener Bilder verlangt, was in diesem Zusammenhang eine relativ leichte Aufgabe darstellt. Bei der AFC-Aufgabe hingegen müssen zwei Bilder mit einer vorher bereits auf irgendeine Art und Weise gelernten, internen Repräsentation des richtigen Stimulus verglichen und anschließend entschieden werden, welches der beiden Bilder mit dieser internen Repräsentation übereinstimmt. Diese Aufgabe erfordert neben den reinen Prozessen der Wahrnehmung eines Stimulus, eine Beanspruchung des Langzeitgedächtnisses und darüber hinaus eine Beanspruchung des Arbeitsgedächtnisses,

während bei der Delayed-Matching Aufgabe das Kurzzeitgedächtnis herangezogen wird. Bei der AFC-Aufgabe sind damit die Komponenten vielfältiger und fehleranfälliger als bei der Delayed-Matching Aufgabe. Nichtsdestotrotz gelten für die AFC-Aufgabe die gleichen Schlussfolgerungen bezüglich eines Cross-Currency Effekts.

Der Faktor 'Zeit' hatte entgegen der Vermutung einen signifikanten Einfluss auf die Leistung in der AFC-Aufgabe, $r = .64$. Auch hier können als Erklärung Übungsvorteile zum zweiten Messzeitpunkt verglichen mit dem ersten Messzeitpunkt herangezogen werden.

Die Art der Fälschung hatte bei der AFC-Aufgabe wieder einen signifikanten Einfluss auf die Leistung. Ebenso wie die Interaktion der Faktoren 'Währung' und 'Art der Fälschung'. Da dieser Effekt in allen drei experimentellen Paradigmen auftritt, wird dieser Effekt in der Gesamtdiskussion zum Cross-Currency Effekt diskutiert.

Überraschend ist der Haupteffekt des Faktors 'Wertigkeit'. Geldscheine mit mittleren Wertigkeiten werden besser differenziert als Geldscheine mit niedriger bzw. hoher Wertigkeit. Eine mögliche Erklärung sind, wie bei den falschen Alarmen in der Delayed-Matching Aufgabe, die unterschiedliche Frequenzen im Umgang mit verschiedenen Wertigkeiten. Man kann davon ausgehen, dass Geldscheine mit mittleren Wertigkeiten häufiger gebraucht werden als andere Geldscheine, was dazu führt, dass das Aussehen von Geldscheinen mit mittleren Wertigkeiten mental verfügbarer ist im Vergleich mit anderen Wertigkeiten. Dies träfe allerdings vor allem auf die DM und den Euro zu. Beim Britischen Pfund dürfte dies nicht der Fall sein, da es kein offizielles Zahlungsmittel in der Bundesrepublik Deutschland ist.

2.4 Experiment III: Wiedererkennenstest

In Experiment III zum Cross-Currency Effekt wurde ein Wiedererkennenstest als experimentelles Paradigma verwendet. Zunächst werden die Hypothesen erläutert und die Durchführung des Experiments dargestellt. Anschließend werden die Ergebnisse dargestellt und diskutiert.

2.4.1 Hypothesen

Für den Wiedererkennenstest gelten die gleichen Annahmen wie für die Delayed-Matching-Aufgabe und die AFC-Aufgabe. Es wird vermutet, dass vermehrter Kontakt mit Objekten der Klasse Geldscheine zu einem ähnlichen Effekt wie dem Cross-Race Bias führt. Es wird eine Zunahme richtiger Identifizierungen von Euro-Geldscheinen bei sinkendem (DM) bzw. gleichbleibendem Niveau (Britisches Pfund) richtiger Identifizierungen von t_1 zu t_2 vorhergesagt. Weiterhin wird ein Interaktionseffekt der Faktoren 'Währung' und 'Zeit' erwartet. Gleichzeitig bedeutet dies, dass für den Faktor 'Zeit' kein Haupteffekt erwartet wird, weil sich die Niveauunterschiede der einzelnen Währungen ausgleichen sollten. Bei den Währungen sollten sich Unterschiede zwischen dem Britischen Pfund auf der einen, und der DM sowie dem Euro auf der anderen Seite zeigen, weil für das Britische Pfund keine Veränderungen in der Kontakthäufigkeit von t_1 zu t_2 vermutet werden. Die Leistung bei der DM sollte besser ausfallen als beim Euro und dem Britischen Pfund. Die Faktoren 'Art der Fälschung', 'Wertigkeit' und 'Seite' sollten keine Bedeutung für die Identifikationsleistung haben und damit keine signifikanten Haupteffekte aufweisen. In diesem Wiedererkennenstest sollte sich ebenfalls ein mirror-effect zeigen, mit einer höheren Anzahl an Treffern und einer niedrigeren Anzahl an falschen Alarmen bei der DM verglichen mit dem Euro und dem Britischen Pfund.

2.4.2 Stimulus-Darbietung

Der Wiedererkennenstest wurde an zwei Apple 'iMac' Computern mit einer Taktfrequenz von 300 bzw. 350 Mhz, einem Arbeitsspeicher von jeweils 192 MB und einem 15"-Monitor erhoben. Die Farbtiefe betrug 16 Bit (65536 Farben) bei einer Bildschirmauflösung von 1024x768 Pixel und einer Bildwiederholfrequenz von 75 Hz. Die Programmierung erfolgte mit der Applikation 'Superlab' (Version 1.75) der Firma Cedrus (<http://www.cedrus.com>). Die Stimuli für die AFC-Aufgabe hatten eine Größe von 400x200 Pixel bei einer Auflösung von 72x72 dpi. Die Farbtiefe der Stimuli betrug 256 Farben.

2.4.3 Versuchsdurchführung

Nachdem der dritte Teil des Kontaktfragebogens beantwortet wurde (siehe Anhang B1, Fragen 16 - 25), folgte der Wiedererkennenstest. Die Instruktionen sind in Anhang A3 angefügt. Anschließend folgte die Darbietung der Geldscheine in randomisierter Reihenfolge. Ein einzelner Trial war einfach aufgebaut: Es wurde ein einzelner Geldschein präsentiert. Die Teilnehmer sollten nun durch Drücken der Tasten 'c' oder 'm' so schnell und genau wie möglich entscheiden, ob der präsentierte Geldschein echt oder falsch ist. Es gab keine Zeitbeschränkung für die einzelnen Trials und es wurde keine Rückmeldung über die Richtigkeit der Antwort gegeben. Alle Geldscheine wurden als normale, gespiegelte und farbgefälschte Geldscheine dargeboten, sowohl als Vorder- als auch Rückseite. Nach Bearbeitung der 72 Trials war das Experiment beendet.

2.4.4 Ergebnisse

Die Auswertung erfolgt anhand der Signal-Entdeckungs-Theorie mit Treffern, falschen Alarmen, dem Leistungsindex \underline{A}' und der Reaktionsneigung \underline{B}' . Ein Treffer bedeutet in diesem Experiment, dass ein 'echter' Geldschein auch als 'echter' Geldschein erkannt wird, während die Antwort 'echt' nach Darbietung eines falschen Geldscheines ein Fehlalarm wäre. Ein hoher Wert für \underline{A}' bedeutet eine gute Leistung in dem Sinne, dass viele 'echte' Geldscheine als 'echt' und viele falsche Geldscheine als unecht erkannt werden. Das Alpha-Niveau wird nach Bonferroni-Adjustierung bei geplanten Kontrasten für die formulierten Hypothesen auf $\alpha = .008$ festgelegt.

Die Hypothese nach der bei der DM die Anzahl der Treffer höher ist als bei den anderen beiden Währungen kann bestätigt werden. Die DM ($\underline{M} = .888$) unterscheidet sich marginal signifikant vom Euro ($\underline{M} = .841$), $\underline{F}(1,38) = 6.14$, $p = .018$, $\underline{r} = .37$, und vom Britischen Pfund ($\underline{M} = .800$), $\underline{F}(1,38) = 11.38$, $p = .002$, $\underline{r} = .48$. Zur Überprüfung der Hypothese, dass zwischen t_1 und t_2 beim Euro Leistungszuwächse zu beobachten sind, während die Leistung bei der DM abfällt und beim Britischen Pfund gleich bleibt, wurden die Differenzen der Treffer zwischen t_1 und t_2 für jede Währung berechnet und mittels t -Tests überprüft, ob diese Differenz signifikant von Null verschieden ist. Zwar zeigte sich deskriptiv bei allen drei Währungen ein Zuwachs von t_1 und t_2 , dieser ist aber nur beim Euro marginal signifikant, $\underline{t}(41) = 2.18$, $p = .035$, $\underline{r} = .17$, ($\underline{M}_{\text{Diff}} = .063$). Es zeigte sich

ebenfalls ein marginal signifikanter Haupteffekt über den Faktor 'Zeit' mit, $F(1,38) = 5.93$, $p = .020$, $r = .37$, ($M_{t1} = .821$, $M_{t2} = .865$).

Eine Varianzanalyse mit Messwiederholung und den Between-Subjects Faktoren 'Geschlecht' (männlich, weiblich) und 'Taste' ('c', 'm') sowie den Within-Subjects Faktoren 'Zeit' (Zeitpunkt1, Zeitpunkt2), 'Währung' (DM, Euro, Britisches Pfund), 'Wertigkeit' (1,2,3,4) und 'Seite' (Vorderseite, Rückseite) und der Anzahl der Treffer als abhängiger Variablen zeigte einen Haupteffekt für den Faktor 'Seite', $F(1,38) = 39.31$, $p = .001$, $r = .71$. Wurde die Rückseite eines Geldscheins gezeigt, war die mittlere Trefferanzahl $M = .789$, bei den Vorderseiten der Geldscheine lag sie dagegen bei $M = .897$. Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 2.33$, alle $p_s > .135$).

Die Falschen Alarme wurden mittels einer Varianzanalyse mit Messwiederholung und den Between-Subjects Faktoren 'Geschlecht' (männlich, weiblich) und 'Taste' ('c', 'm') sowie den Within-Subjects Faktoren 'Zeit' (Zeitpunkt1, Zeitpunkt2), 'Währung' (DM, Euro, Britisches Pfund), 'Art der Fälschung' (Farbfälschung, Spiegelungsfälschung), 'Wertigkeit' (1,2,3,4) und 'Seite' (Vorderseite, Rückseite) analysiert. Je nach untersuchter Währung wurden unterschiedlich viele falsche Alarme gegeben $F(2,76) = 35.05$, $p = .001$, $\eta^2 = .48$. Die wenigsten falschen Alarme gab es bei der DM ($M = .227$) gefolgt von dem Britischen Pfund ($M = .350$) und dem Euro ($M = .382$). Das Britische Pfund und der Euro unterscheiden sich nicht voneinander, $F < 1$, ns , während sich die DM und das Britische Pfund marginal unterscheiden, $F(1,76) = 3.84$, $p = .054$, $r = .22$, und zwischen der DM und dem Euro ein signifikanter Unterschied besteht, $F(1,76) = 6.52$, $p = .016$, $r = .27$. Die Wertigkeit der Geldscheine hat einen Einfluss auf die falschen Alarme, $F(3,114) = 8.31$, $p = .001$, $\eta^2 = .18$. Die wenigsten falschen Alarme gab es bei Geldscheinen mit der zweithöchsten Wertigkeit ($M = .283$) gefolgt von der höchsten Wertigkeit ($M = .310$), der niedrigsten Wertigkeit ($M = .317$) und der zweitniedrigsten Wertigkeit ($M = .369$), wobei lediglich die Differenz zwischen der zweithöchsten und der zweitniedrigsten Wertigkeit signifikant unterschiedlich ist $F(1,114) = 4.57$, $p = .035$, $r = .20$. Weiterhin gab es bei den Spiegelungsfälschungen mehr falsche Alarme als bei den Farbfälschungen $F(1,38) = 16.27$, $p = .001$, $r = .55$, ($M_{FF} = .269$, $M_{SF} = .37$) (siehe Tabelle 7). Die Interaktion der Faktoren 'Währung' und 'Art der Fälschung' ist ebenfalls signifikant, $F(2,76) = 63.61$, $p = .001$, $\eta^2 = .63$. Hier zeigte sich vor allem bei der DM ein großer Unterschied zwischen den falschen Alarmen bei Farbfälschungen ($M = .058$) und bei Spiegelungsfälschungen ($M = .397$), $F(1,76) = 52.09$, $p = .001$, $r = .64$. Beim Britischen Pfund zeigte sich ein signifikanter Unterschied zwischen Farbfälschungen ($M = .442$) und bei Spiegelungsfälschungen ($M = .258$), $F(1,76) = 10.23$, $p = .002$, $r = .34$, ebenso wie beim

Euro ($\underline{M}_{FF} = .311$, $\underline{M}_{SF} = .453$), $\underline{F}(1,76) = 8.88$, $\underline{p} = .004$, $\underline{r} = .32$. Alle anderen Haupteffekte und Interaktionen waren nicht signifikant, alle $\underline{F}s < 1.39$, alle $\underline{p}s > .254$.

Tabelle 7

Mittelwerte der Treffer, der falschen Alarme, des Leistungsindex \underline{A}' und der Reaktionsneigung \underline{B}'' im Wiedererkennentest getrennt nach Zeitpunkt und Wahrung

Wahrung	Zeitpunkt		\underline{M} :
	t1	t2	
Deutsche Mark			
Treffer	.873	.903	.888
Falsche Alarme (Farbe)	.045	.070	.058
Falsche Alarme (Spiegel)	.384	.410	.397
\underline{A}' (Farbe)	.952	.952	.952
\underline{A}' (Spiegel)	.818	.830	.824
\underline{B}'' (Farbe)	.317	.115	.243
\underline{B}'' (Spiegel)	-.286	-.517	-.395
Euro:			
Treffer	.811	.871	.841
Falsche Alarme (Farbe)	.373	.250	.312
Falsche Alarme (Spiegel)	.403	.504	.454
\underline{A}' (Farbe)	.794 ^a	.861 ^a	.823
\underline{A}' (Spiegel)	.775	.756	.766
\underline{B}'' (Farbe)	-.186	-.322	-.257
\underline{B}'' (Spiegel)	-.249	-.397	-.327
Britisches Pfund:			
Treffer	.779	.822	.801
Falsche Alarme (Farbe)	.424	.460	.442
Falsche Alarme (Spiegel)	.268	.249	.259
\underline{A}' (Farbe)	.757 ^a	.790 ^a	.774
\underline{A}' (Spiegel)	.809 ^a	.860 ^a	.835
\underline{B}'' (Farbe)	-.281	-.389	-.299
\underline{B}'' (Spiegel)	-.283	-.327	-.273
\underline{M} :			
Treffer	.821	.865	.843
Falsche Alarme (Farbe)	.281	.260	.271
Falsche Alarme (Spiegel)	.352	.399	.376
\underline{A}' (Farbe)	.834	.868	.851
\underline{A}' (Spiegel)	.800	.815	.808
\underline{B}'' (Farbe)	-.050 ^a	-.199 ^a	-.124
\underline{B}'' (Spiegel)	-.272 ^a	-.414 ^a	-.343

Anmerkung: ^aDie Zellenmittelwerte innerhalb einer Zeile unterscheiden sich nach

Bonferroni-Adjustierung signifikant (post-hoc Vergleiche, $\underline{p} < .001$).

Die Hypothese nach der \underline{A}' bei der DM höher ist als bei den anderen beiden Währungen kann bestätigt werden. Die DM ($\underline{A}' = .888$) unterscheidet sich signifikant vom Euro ($\underline{A}' = .797$), $\underline{F}(1,38) = 37.85$, $p = .001$, $r = .71$, und vom Britischen Pfund ($\underline{M} = .804$), $\underline{F}(1,38) = 35.87$, $p = .001$, $r = .70$. Zwischen dem Euro und dem Britischen Pfund besteht kein signifikanter Unterschied, $\underline{F} < 1$, ns. Zur Überprüfung der Hypothese, dass zwischen t_1 und t_2 beim Euro Leistungszuwächse zu beobachten sind, während die Leistung bei der DM abfällt und beim Britischen Pfund gleich bleibt, wurden die Differenzen von \underline{A}' zwischen t_1 und t_2 für jede Währung berechnet und mittels t -Tests überprüft, ob diese Differenz signifikant von Null verschieden ist. Es zeigte sich zwar beim Euro und beim Britischen Pfund deskriptiv ein Zuwachs von t_1 und t_2 , dieser ist aber bei keiner Währung signifikant von null verschieden ($\underline{A}'_{DiDM} = -.005$, $\underline{A}'_{DiEU} = .032$, $\underline{A}'_{DiBP} = .031$). Es zeigte sich kein signifikanter Haupteffekt über den Faktor 'Zeit', $\underline{F}(1,38) = 3.52$, ns.

Eine Varianzanalyse mit den Between-Subjects Faktoren 'Geschlecht' (männlich, weiblich) und 'Taste' ('c', 'm') sowie den Within-Subjects Faktoren 'Zeit' (Zeitpunkt1, Zeitpunkt2), 'Währung' (DM, Euro, Britisches Pfund) und 'Art der Fälschung' (Farbfälschung, Spiegelungsfälschung) zeigte keine signifikanten Effekte für die Faktoren 'Geschlecht', $\underline{F} < 1$, ns und 'Taste', $\underline{F} < 1$, ns).

Wiederum konnte keine Interaktion zwischen den Faktoren 'Zeit' und 'Währung' beobachtet werden, $\underline{F}(2,76) = .65$, $p = .524$, $\eta^2 = .02$. Der Leistungsindex war höher bei den Farbfälschungen verglichen mit den Spiegelungsfälschungen ($\underline{A}'_{FF} = .851$, $\underline{A}'_{SF} = .808$), $\underline{F}(1,38) = 12.42$, $p = .001$, $r = .49$. Der höhere Leistungsindex bei den Farbfälschungen zeigt sich nicht bei allen drei Währungen wie die signifikante Interaktion der Faktoren 'Währung' und 'Fälschung' zeigt, $\underline{F}(2,76) = 30.14$, $p = .001$, $\eta^2 = .44$ (siehe Abbildung 11). Der Leistungsindex \underline{A}' ist bei farbgefälschten DM-Banknoten höher als bei spiegelungsgefälschten ($\underline{A}'_{FF} = .952$, $\underline{A}'_{SF} = .842$). Dies ist auch bei Banknoten des Euro der Fall ($\underline{A}'_{FF} = .828$, $\underline{A}'_{SF} = .766$), während bei Banknoten des Britischen Pfunds das Verhältnis umgekehrt ist ($\underline{A}'_{FF} = .774$, $\underline{A}'_{SF} = .835$).

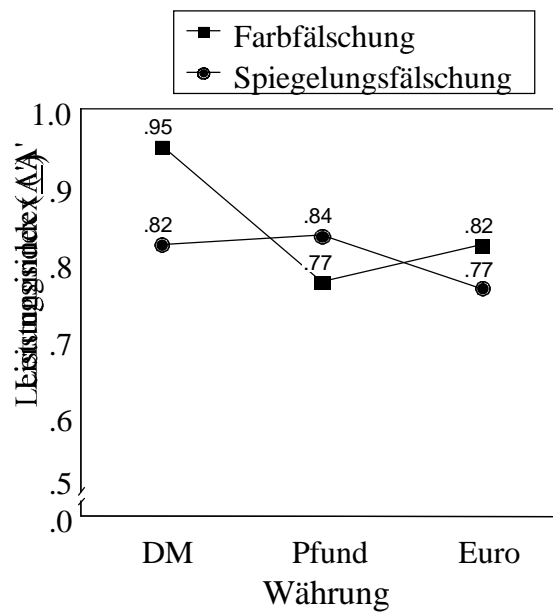


Abbildung 11: Interaktion der Faktoren 'Währung' und 'Fälschung' beim Leistungsindex \underline{A}

Bei den Spiegelungsfälschungen ist der Euro signifikant schlechter als die DM $F(1,76) = 7.48$, $p = .008$, $r = .30$, und das Britische Pfund, $F(1,76) = 10.83$, $p = .002$, $r = .35$. Die Paarvergleiche der Farbfälschungen hingegen sind für den Vergleich DM und Euro, $F(1,76) = 39.42$, $p = .001$, $r = .58$, der DM und dem Britischen Pfund, $F(1,76) = 81.24$, $p = .001$, $r = .72$, sowie dem Euro und dem Britischen Pfund, $F(1,76) = 7.48$, $p = .008$, $r = .30$, jeweils signifikant.

Wird der Faktor 'Zeit' mit einbezogen, erhält man eine signifikante Interaktion der Faktoren 'Währung', 'Fälschung' und 'Zeit', $F(2,76) = 3.60$, $p = .001$, $\eta^2 = .09$ (siehe Abbildung 12).

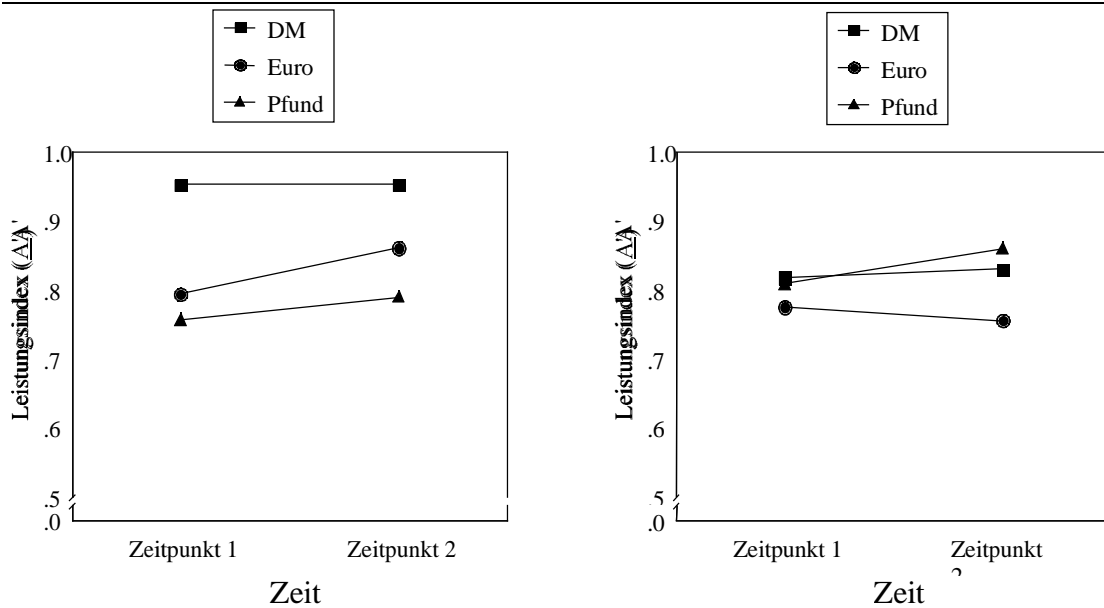


Abbildung 12: Interaktion der Faktoren 'Währung', 'Fälschung' und 'Zeit' bezüglich des Leistungsindex \underline{A}' . Links ist \underline{A}' bei den Farbfälschungen, rechts \underline{A}' bei den Spiegelungsfälschungen abgebildet.

Während \underline{A}' bei den Farbfälschungen der DM ($\underline{A}'_{1FF} = .952$, $\underline{A}'_{2FF} = .952$) auf der gleichen Stufe von Zeitpunkt 1 zu Zeitpunkt 2 bleibt, $F < 1$, ns, steigt \underline{A}' beim Euro ($\underline{A}'_{1FF} = .794$, $\underline{A}'_{2FF} = .861$) signifikant an, $F(1,76) = 55.22$, $p = .001$, $r = .66$, ebenso wie beim Britischen Pfund, $F(1,76) = 7.07$, $p = .010$, $r = .29$ ($\underline{A}'_{1FF} = .757$, $\underline{A}'_{2FF} = .790$). Bei den Spiegelungsfälschungen zeigen sich bis auf das Britische Pfund, $F(1,76) = 16.89$, $p = .001$, $r = .43$, ($\underline{A}'_{1SF} = .809$, $\underline{A}'_{2SF} = .860$) keine signifikanten Veränderungen der Währungen zwischen t_1 und t_2 . Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 1.78$, alle $p_s > .19$).

Zusätzlich zum Leistungsindex \underline{A}' wurde die Reaktionsneigung \underline{B}'' ausgewertet. Es wurde eine fünf-fache Varianzanalyse mit den Between-Subjects Faktoren 'Geschlecht' und 'Taste' sowie den Within-Subjects Faktoren 'Währung', 'Fälschung' und 'Zeit' durchgeführt. Die Faktoren 'Geschlecht', $F < 1$, ns und 'Taste', $F < 1$, ns hatten keinen signifikanten Einfluss auf die Reaktionsneigung. Es zeigte sich wieder ein signifikanter Haupteffekt für den Faktor 'Währung', $F(2,76) = 5.76$, $p = .004$, $\eta^2 = .13$ ($\underline{B}''_{DM} = -.093$, $\underline{B}''_{Eu} = -.288$, $\underline{B}''_{BP} = -.320$), der durch die signifikant niedrigeren Werte des Euro, verglichen mit der DM, $F(1,76) = 7.59$, $p = .007$, $r = .30$ und des Britischen Pfunds verglichen mit der DM, $F(1,76) = 5.61$, $p = .020$, $r = .26$ zu erklären ist. Der Euro und das Britische Pfund unterscheiden sich nicht, $F < 1$, ns. Die Reaktionsneigung war zum Zeitpunkt 1 ebenfalls

liberaler als zum Zeitpunkt 2, $F(1,38) = 8.97$, $p = .005$, $r = .44$, ($B''_{t1} = -.161$, $B''_{t2} = -.306$). Ein signifikanter Effekt zeigte sich für den Faktor 'Fälschung', $F(1,38) = 44.44$, $p = .001$, $r = .73$, ($B''_{FF} = -.124$, $B''_{SF} = -.343$). Interessanterweise zeigte sich eine signifikante Interaktion der Faktoren 'Währung' und 'Fälschung', $F(2,76) = 38.20$, $p = .001$, $\eta^2 = .50$. Während die Werte von B'' zwischen den Farbfälschungen und den Spiegelungsfälschungen für den Euro und das Britische Pfund nicht unterscheiden, $F < 1$, ns , ist die Differenz zwischen den beiden Fälschungsarten bei der DM signifikant unterschiedlich, $F(1,76) = 185.09$, $p = .001$, $r = .84$, ($B''_{FF} = .216$, $B''_{SF} = -.402$). Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 3.79$, alle $p_s > .059$).

2.4.5 Diskussion der Ergebnisse des Wiedererkennenstests

Die Leistung bei der DM war im Wiedererkennenstest besser als beim Britischen Pfund und beim Euro. Dieses Ergebnis bestätigt die Ergebnisse aus der Delayed-Matching Aufgabe und der AFC-Aufgabe. Die Teilnehmer hatten im Wiedererkennenstest bei der DM einen höheren Leistungsindex und zeigten aber ebenfalls ein liberales Antwortkriterium, zumindest bei den Spiegelungsfälschungen, verglichen mit dem Britischen Pfund und dem Euro. Bezüglich der Hypothese, einen Cross-Currency Effekt zu finden, bei dem die Leistungen beim Euro den Leistungen der DM angeglichen werden sollten, bedeutet dies, dass diese Hypothese nicht bestätigt werden konnte. Es gilt jedoch weiterhin wie bei der Delayed-Matching Aufgabe und der AFC-Aufgabe, dass eine bessere Leistung bei den 'Eigen-Gruppen' Geldscheinen im Vergleich zu den 'Fremd-Gruppen' Geldscheinen einen Hinweis auf einen Cross-Currency Effekt darstellen. Die Dominanz der DM wird im Wiedererkennenstest weiter durch das Fehlen einer Interaktion der Faktoren 'Zeit' und 'Währung' unterstützt. Bei keiner der drei Währungen kommt es zum zweiten Messzeitpunkt zu einer Verbesserung der Wiedererkennensleistung. Anscheinend beeinflussen zehn Jahre Erfahrung der Teilnehmer mit der DM immer noch signifikant die Wiedererkennensleistung und dieser Vorteil kann noch nicht innerhalb von vier Monaten ausgeglichen werden. Wie bei den vorherigen Experimenten wäre es in diesem Zusammenhang interessant zu wissen, wie sich das Verhältnis von DM zu Euro und zum Britischen Pfund nach einer Kontaktphase mit der neuen Währung von mehr als vier Monaten entwickelt, nachdem potentielle Versuchsteilnehmer einen längeren Zeitraum Kontakt und Umgang mit der neuen Währung hatten.

Obwohl von Zeitpunkt 1 zu Zeitpunkt 2 keine Steigerung im Leistungsindex A' gefunden werden konnte, stieg die Anzahl der Treffer von Zeitpunkt 1 zu Zeitpunkt 2 signifikant an, und die Reaktionsneigung B'' wechselte zu einem liberaleren Kriterium. Diese Befunde können erklären, warum ein signifikanter Zuwachs in der Leistung nicht gefunden werden konnte. Der Leistungsindex ist ein Maß für das Verhältnis zwischen Treffern und Fehlalarmen. Demnach sollte die Genauigkeit von Zeitpunkt 1 zu Zeitpunkt 2 steigen, wenn der Leistungsindex ansteigt. Gleichzeitig jedoch zeigt sich bei der Reaktionsneigung ein Wechsel von einem eher konservativen Antwortkriterium zu einem eher liberaleren Kriterium, was ein Hinweis darauf ist, dass die Teilnehmer häufiger mit einer 'Ja'-Antwort auf die präsentierten Stimuli reagieren.

Wenn man sich die Treffer und Fehlalarme näher betrachtet, zeigt sich, dass die Prozentwerte der Treffer um ca. sechs Prozent steigen und die Prozentwerte der Fehlalarme um ca. drei Prozent. Da für die Berechnung des Leistungsindex und der Reaktionsneigung Verhältnisse zwischen Treffern und Fehlalarmen berechnet werden, ist die absolute Anzahl von Treffern und Fehlalarmen nicht wichtig, bei der Berechnung der Genauigkeit jedoch sehr wohl vonnöten. Da weiterhin das Verhältnis zwischen gefälschten und normalen Banknoten zwei zu eins beträgt, bedeutet dies, dass Teilnehmer die mit einer 'Ja'-Antwort auf einen Stimulus antworten zweimal häufiger einen Fehlalarm produzieren als einen Treffer. Da nun zweimal mehr Stimuli in diesem Experiment vorkommen, die als Fehlalarm gewertet werden, wenn eine 'Ja'-Antwort gegeben wird (Farb- und Spiegelungsfälschungen) als Stimuli die als Treffer gewertet werden (normale Banknoten), bedeutet dies, dass ein Zuwachs an von drei Prozent Fehlalarmen eine Zuwachs von sechs Prozent Treffern ausgleicht und somit kein Zuwachs in der Genauigkeit von Zeitpunkt 1 zu Zeitpunkt 2 beobachtet werden kann, obwohl der Leistungsindex zwischen den beiden Zeitpunkten ansteigt.

Sowohl der Interaktion der Faktoren 'Währung' und 'Fälschung' als auch dem Haupteffekt für den Faktor 'Fälschung' bezüglich des Leistungsindex und der Reaktionsneigung, können unterschiedlichen Farbstrukturen der drei Währungen zugrunde liegen. Dies kann ebenfalls ein Grund dafür sein, dass die Spiegelungsfälschungen des Britischen Pfunds besser wiedererkannt werden konnten als die Spiegelungsfälschungen der DM und des Euro. Aufgrund der Komplexität der Banknoten des Britischen Pfunds, war es eventuell leichter für die Teilnehmer, bei diesen Banknoten Fehler zu entdecken, im Gegensatz zu den Banknoten der DM und des Euro.

Auch im Wiedererkennenstest zeigt sich ein mirror-effect. Die Anzahl der Treffer bei der DM ist am höchsten, gefolgt vom Euro und dem Britischen Pfund. Die Anzahl der

falschen Alarme ist bei der DM am geringsten gefolgt vom Britischen Pfund und dem Euro.

2.5 Diskussion der Experimente zum Cross-Currency Effekt

Wie bereits erwähnt, wurde in allen drei Experimenten durchgehend ein deutlicher Vorteil der DM gegenüber den anderen Währungen mit durchgehend starken Effekten gefunden. Es konnte in den drei Experimenten ein Cross-Currency Effekt aufgezeigt werden. Allerdings ist dies kein 'vollständiger' Bias, da per definitionem eine vollständige Interaktion von Versuchspersonen-Gruppe und Stimulus-Gruppe vorliegen muss. Es wurden jedoch nur deutsche Versuchspersonen in diesen Experimenten untersucht. Daher sind Rückschlüsse über die Leistungen anderer Nationalitäten nicht zulässig. Durch die Verwendung des Britischen Pfunds als Stimulusmaterial ist es aber möglich, Teilnehmer Britischer Nationalität zu testen und die Ergebnisse mit den deutschen Versuchspersonen zu vergleichen. Zumindest für die deutschen Teilnehmer ist jedoch eine bessere Leistung bei der eigenen Währung zu vermerken. Dieser Vorteil ist weniger deutlich, wenn man den Faktor 'Zeit' bei der Betrachtung mit heranzieht. Sowohl in der Delayed-Matching Aufgabe als auch in der AFC-Aufgabe zeigte sich der Vorteil der DM vor allem zum Zeitpunkt 1, während zum Zeitpunkt 2 sowohl das Britische Pfund als auch der Euro stark in der Leistung zunahen. Dies kann zum einen daran liegen, dass die Teilnehmer sich zum Zeitpunkt 2 an die Banknoten des Britischen Pfunds und des Euro erinnern konnten und daher bessere Leistungen zeigten. Dem widerspricht allerdings die fehlende Interaktion der Faktoren 'Zeit' und 'Währung' beim Wiedererkennenstest, da dieser Effekt auch hier hätte greifen müssen. Es ist daher eher zu vermuten, dass die Delayed-Matching-Aufgabe und die AFC-Aufgabe jedenfalls in dieser Beziehung andere Prozesse involvieren als der Wiedererkennenstest. Bei der Delayed-Matching Aufgabe und der AFC-Aufgabe werden die einzelnen Stimuli nicht aus dem Langzeit-Gedächtnis abgerufen, wie es beim Wiedererkennenstest der Fall ist. Die Stimuli werden vielmehr kurz nacheinander (Delayed-Matching Aufgabe) bzw. sogar gleichzeitig dargeboten und erfordern daher nicht den Abruf aus dem Langzeitgedächtnis bzw. sie müssen nicht mit einer internen Repräsentation des Gesehenen verglichen werden. Beim letzteren greift dann die jahrelange Erfahrung der Teilnehmer mit der DM besser als mit den neu gelernten Banknoten. Werden die Stimuli hingegen kurz hintereinander bzw. gleichzeitig

dargeboten, so erfordert die Aufgabe im ersten Falle lediglich eine kurze Speicherung des Stimulus im Kurzzeitgedächtnis und dann den Vergleich des Dargebotenen mit der inneren Repräsentation. Wenn nun die Teilnehmer die Stimuli bereits in einem früheren Teil des Experimentes (Zeitpunkt 1) kennen gelernt haben, haben sie bereits Erfahrung mit dieser Art der Testung, und weiterhin Erfahrung mit den benutzten Stimuli. Bei gleichzeitiger Darbietung genügt ein Vergleich der gesehenen Stimuli und ein Entdecken der Fehler auf einer gefälschten Banknote. Auch hier können die eben beschriebenen Effekt für eine bessere Leistung zum Zeitpunkt 2 als Erklärung herangezogen werden, ein Übungseffekt sowohl für die Aufgabe als auch für die Stimuli. Wenn es also um die Frage geht, welches der experimentellen Paradigmen bei Wiedererkennenstests vorteilhafter ist, hat der Wiedererkennenstest gegenüber der Delayed- Matching Aufgabe und der AFC-Aufgabe einen Vorsprung. Ein weiterer Nachteil der AFC-Aufgabe ist die fehlende Möglichkeit, die Daten nach der Signal-Entdeckungs-Theorie auszuwerten. Dies ist bei der Delayed-Matching Aufgabe und beim Wiedererkennenstest möglich und erlaubt eine vielschichtigere Analyse der anfallenden Daten, beispielweise eine Analyse des mirror effects. Bei einigen Aufgaben, vor allem bei der Frage nach Unterschieden zwischen Paradigmen bei der Gegenüberstellung im polizeilichen Kontext wäre eine AFC-Aufgabe angebracht. Die Unterscheidung zwischen simultaner und sequentieller Darbietung benötigt die Einführung einer AFC-Aufgabe als simultane Testung, während bei sequentieller Testung ein Wiedererkennenstest angebracht ist. Diese Frage wird aber in den nachfolgenden Experimenten nicht untersucht, weshalb die Einführung einer AFC-Aufgabe in den folgenden Experimenten nicht sinnvoll erscheint. Die drei Experimente zu einem Cross-Currency Effekt konnten zeigen, dass in allen drei Paradigmen deutliche Vorteile der DM gegenüber den anderen Währungen bestehen und dass damit Expertenwissen nach zehnjähriger Erfahrung mit der DM zum Tragen kommt. Ebenfalls wurde in der Delayed Matching Aufgabe und im Wiedererkennenstest ein mirror-effect beobachtet, mit einer größeren Zahl an Treffern und einer geringeren Zahl an falschen Alarmen bei der DM, verglichen mit dem Euro und dem Britischen Pfund. Beide Befunde sprechen für einen fundamentalen Vorteil der DM gegenüber den anderen verwendeten Währungen, der analog in der Forschung zum Cross-Race Bias bei Eigen-Gruppen Gesichtern beobachtet wird. Ein Problem, das bei der Interpretation der Daten entsteht, ist, dass bei allen Experimenten die gleichen Versuchspersonen mitwirkten und damit Übungseffekte der Teilnehmer wahrscheinlich sind. Eine mögliche Variation um diese Möglichkeit zu untersuchen wäre, die gleichen Experimente nochmals mit anderen Versuchspersonen durchzuführen und die Ergebnisse der zwei Stichproben zu vergleichen.

Bezüglich des Cross-Currency Effektes sind die Befunde zu den Fälschungsarten von Interesse und hierbei besonders die Farbfälschungen, da sich hier die größten Unterschiede zwischen den Währungen zeigen. Auffällig ist, dass die Farbfälschungen des Euro bei allen drei experimentellen Paradigmen von Zeitpunkt 1 zu Zeitpunkt 2 einen Zuwachs verzeichnen können. Dieser Effekt ist beim Britischen Pfund nicht erkennbar und auch bei der DM ist dies nicht der Fall. Hier muss allerdings gesagt werden, dass bei der DM auch kein besonders großer Zuwachs zu erwarten war, da bereits zu Zeitpunkt 1 die Wiedererkennensleistung bei der DM sehr gut gewesen ist. Es lässt sich jedoch erkennen, dass die Teilnehmer zum zweiten Messzeitpunkt besser mit den Farbfälschungen des Euro umgehen konnten als zum Messzeitpunkt 1. Dass die Farbe eine wesentliche Rolle bei der Enkodierung von Objekten spielt, zeigten Gegenfurtner und Rieger (2000). Die Autoren präsentierten den Versuchspersonen Bilder natürlicher Szenen entweder in Farbe oder in Schwarz-Weiß und testeten die Wiedererkennensleistung ebenfalls mit Farbbildern oder Schwarz-Weiß Bildern. Dabei zeigte sich, dass Bilder, die bei Darbietung und Test in Farbe dargeboten wurden, besser wiedererkannt werden konnten, als Bilder in den anderen Bedingungen, insbesondere bei Bildern, die sowohl in der Darbietungs- als auch in der Testphase in Schwarz-Weiß gezeigt wurden. Dieser Befund mag für die Euro-Experimente zunächst unwichtig erscheinen, der Schluss, den Gegenfurtner und Rieger (2000) aus ihrer Untersuchung ziehen, ist jedoch von Bedeutung. Ihrer Meinung nach trägt die Farbinformation eines Bildes wesentlich zur Wiedererkennung desselben bei, indem bei der Enkodierung und Repräsentation des Bildes diese Information weiter herangezogen wird. Der gleiche Prozess könnte bei den Farbfälschungen des Euro eine Rolle gespielt haben, indem der vermehrte Kontakt mit den echten Geldscheinen im täglichen Leben der Teilnehmer dazu geführt hat, dass die Teilnehmer eine Farbrepräsentation der Geldscheine erworben haben und damit besser in der Lage sind, echte von farbverfälschten Geldscheinen zu unterscheiden. Diese Repräsentationen sind für die DM noch vorhanden, bzw. für das Britische Pfund schwerer zu erwerben, weil sie zum einen nicht Gegenstand des täglichen Gebrauchs in Deutschland sind und zum anderen nicht so einheitlich oder auch klar strukturiert sind wie natürliche Szenen oder andere Banknoten.

In den folgenden Experimenten wird die Wiedererkennensleistung von Gesichtern untersucht und es werden einige Faktoren variiert, die einen Einfluss auf das Ausmaß des Cross-Race Bias haben sollten. Der Einfluss des Behaltensintervalls und des Gesichtsausschnitts werden in Abschnitt 4 untersucht. In Abschnitt 6 werden dann Berufsgruppen als Versuchsteilnehmer herangezogen, bei denen davon ausgegangen wird, dass sie unterschiedlich viel Kontakt mit anderen Ethnien haben. Um das nötige

Stimulusmaterial für die Experimente bereitzustellen, wird im folgenden Abschnitt zunächst eine Pilotstudie beschrieben, mit deren Hilfe Darbietungssets zusammengestellt wurden, um die Wiedererkennensleistung von Gesichtern zu testen.

3 Pilotstudie für die Auswahl der türkischen und deutschen Gesichter

In dieser Pilotstudie sollten 160 Fotos von 80 türkischen und 80 deutschen Männern auf fünf Skalen (siehe Abschnitt 3.1.3) beurteilt werden. Frühere Studien konnten zeigen, dass Personen, deren Gesichter als 'untypisch' bzw. 'einzigartig' eingestuft wurden, besser wiedererkannt werden konnten als Gesichter, die als 'typisch' oder 'normal' wahrgenommen wurden (Brigham, 1990; Valentine, 1991a). Ziel dieser Pilotstudie war die Erstellung von zwei Darbietungssets für die Wiedererkennstests in den Experimenten, die in Abschnitt 4 und 6 beschrieben werden, sowie für die Delayed-Matching Aufgabe in Abschnitt 6. Zusätzlich lässt sich in dieser Pilotstudie die Hypothese prüfen, ob Eigen-Gruppen Gesichter als distinkter (unterschiedlicher), angesehen werden als Fremd-Gruppen Gesichter. Die Eigen-Gruppen Gesichter sollten ebenfalls bei beiden Versuchspersonen-Gruppen als vertrauter angesehen werden als die Fremd-Gruppen Gesichter. Für die Wiedererkennstests und die Delayed-Matching Aufgabe der Experimente in den Abschnitten 4 und 6 standen zusätzlich jeweils 80 Fotos von Männern afro-amerikanischer und kaukasischer Herkunft zur Verfügung, für die bereits Ratings auf den in Abschnitt 3.1.3 beschriebenen Skalen durchgeführt wurden¹⁰ (Meissner & Brigham, in Druck).

3.1 Material

In diesem Abschnitt werden zunächst die Stichprobe, das Stimulusmaterial und die verwendeten Skalen beschrieben. Anschließend werden die Versuchsdurchführung und der Versuchsablauf besprochen.

¹⁰ Ich danke an dieser Stelle Christian Meissner von der Florida International University für die freundliche Überlassung der Bilder.

3.1.1 Stichprobe

An der Pilotstudie nahmen 64 Personen teil, größtenteils Studierende der Justus-Liebig-Universität Gießen. Jeweils die Hälfte der Teilnehmer waren türkischer bzw. deutscher Ethnizität. Das Verhältnis von Frauen zu Männern betrug 3:1, da in den Gebäuden, in denen die Teilnehmer rekrutiert wurden, mehr türkische Frauen studieren als türkische Männer. Um dieser Verteilung Rechnung zu tragen, wurde dieses Verhältnis auch bei den deutschen Versuchspersonen eingehalten. Die Teilnehmer erhielten nach Durchführung des Ratings entweder eine Versuchspersonenstunde oder acht Euro. Der Median der gesamten Stichprobe betrug $Mdn = 24$ Jahre. Die Teilnehmer waren zwischen 18 und 34 Jahren alt. Türkische und deutsche Teilnehmer unterschieden sich bezüglich des Altersmedians nicht (Türken: $Mdn = 24$; Deutsche: $Mdn = 24$). Die Teilnehmer und die abgebildeten Personen des Stimulusmaterials sind weiterhin in der gleichen Altersgruppe, was Effekte auf die Einstufung der Bilder in den verwendeten Skalen durch einen Cross-Age Bias ausschließt.

3.1.2 Stimulus-Material

Das Stimulus-Material bestand aus jeweils 80 Bildern in Frontalansicht von Männern türkischer und deutscher Herkunft. Die eine Hälfte der Gesichter stammt aus einer Untersuchung von Sporer (1999) bzw. den Diplomarbeiten von Guberova (2001) und Trinkl (2001). Die andere Hälfte wurde von zwei Diplomanden (Dogan, 2003, Henning, 2003) zum einen auf einer türkischen Kultur-Veranstaltung in Darmstadt (türkische Gesichter) und zum anderen an der Johann- Wolfgang von Goethe Universität in Frankfurt (deutsche Gesichter) aufgenommen. Die Aufnahmen wurden mit einer Digitalkamera angefertigt. Von jeder Person wurde sowohl ein Bild in der Frontalansicht als auch ein Bild im rechten Halbprofil (45° - Winkel) hergestellt. Für diese Pilotstudie wurde allerdings nur die Bilder in der Frontalansicht verwendet, weil die Bilder im rechten Halbprofil für die Testphase der Wiedererkennensexperimente und der Delayed-Matching Aufgabe vorgesehen waren, und auch für die Einstufung der Gesichter die Bilder in der Frontalansicht ausreichend sind. Das Alter der fotografierten Personen lag zwischen 20 und 30 Jahren. Sie wurden bei Aufnahme der Frontalansicht gebeten, direkt in die Kamera zu blicken und dabei einen 'neutralen' Gesichtsausdruck zu halten. Bei Bildern im rechten

Halbprofil wurden sie ebenfalls gebeten, einen neutralen Gesichtsausdruck zu halten, dabei jedoch einen Punkt rechts neben der Kamera zu fixieren, so dass der Winkel zwischen einer gedachten Linie von Kameralinse zum Gesicht und einer gedachten Linie vom Gesicht zu diesem Punkt 45° betrug. Alle Stimulus-Personen trugen während der Aufnahme einen schwarzen Schal, damit eine Wiedererkennung in den späteren experimentellen Phasen aufgrund der Kleidung ausgeschlossen werden konnte. Der Hintergrund der Aufnahmen bestand aus einem weißen Tuch (siehe Abbildung 13).

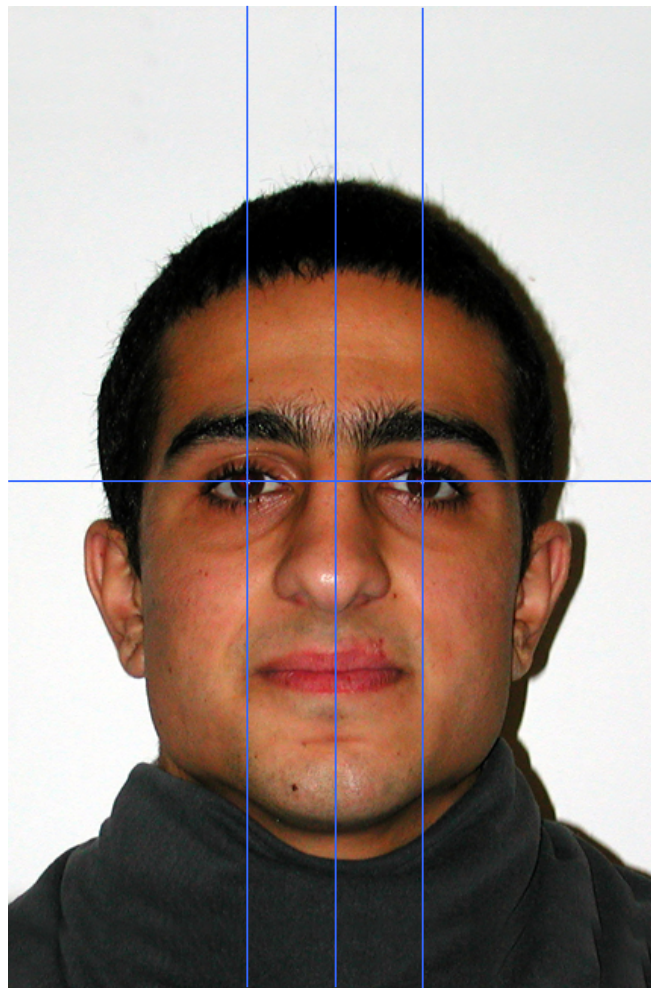


Abbildung 13. Beispiel-Gesicht mit Hilfslinien für die Standardisierung. Die Hilfslinien waren während der Pilotstudie und der Experimente nicht sichtbar.

Um eine unterschiedliche Einstufung der Gesichter auf den fünf genutzten Skalen aufgrund der Positionierung der Gesichter auf den Fotografien auszuschließen, wurden die Fotos anschließend in Anlehnung an die Diplomarbeiten von Guberova (2001) und Trinkl (2001) unter Verwendung des Programms 'Photoshop 5.5' standardisiert. Das Vorlagen-

Bild hatte eine Größe von (Breite x Höhe) 6 x 9 Inches mit einer Auflösung von 72 Pixel/Inch. Zur Orientierung für die Platzierung des Gesichtes während der Standardisierung wurden in das jeweilige Bild Hilfslinien eingefügt: eine horizontale Linie, bei 4 3/8 Inches und drei vertikale bei 2 1/8 Inches, 3 Inches, und 3 7/8 Inches (siehe Abbildung 13).

Die Standardisierungs-Prozedur war wie folgt: Das zu standardisierende Bild wurde in die Vorlage kopiert und das, vom Bearbeiter aus gesehene, linke Auge am Schnittpunkt der horizontalen mit der linken vertikalen (2 1/8 Inches) Hilfslinie positioniert und durch Drag & Drop der Zentrierungsmarke das gesamte Gesicht an diesem Punkt verankert. Mit dem Befehl 'Transform/Rotate' wurde, falls nötig, das Gesicht so mit bzw. gegen den Uhrzeigersinn gedreht, dass beide Pupillen auf der horizontalen Linie lagen. Das Ziel der Größenanpassung des neuen Bildes an die Vorlage war, dass das linke und das rechte Auge auf den Schnittpunkten der horizontalen Linie mit den beiden äußeren Linien lagen. Dazu musste das gesamte Gesicht mit dem Befehl 'Transform' sukzessive dergestalt verkleinert, bzw. vergrößert werden, bis das oben genannte Ziel erreicht war. Durch diese Prozedur kam es vor, dass Teile der Vorlage, d.h. der weiße Hintergrund, wieder sichtbar wurden, diese aber mit Hilfe des 'Clone Stamp Tool' mit dem Hintergrund der Fotografien überdeckt werden konnten. Dazu wurde das 'Clone Stamp Tool' ausgewählt und zunächst ein, dem weißen Hintergrund anliegender Teil durch gleichzeitiges Drücken der 'alt' sowie der Maustaste als zu kopierender Teil markiert und anschließend der zu überzeichnende Teil des Hintergrundes mit dem zu klonenden Teil überzeichnet. Als letzter Teil der Standardisierung wurden noch die beiden Befehle 'Adjust Auto Levels' und 'Adjust Auto Contrast' angewendet.

3.1.3 Skalen

Die Gestaltung der Rating-Skalen orientierte sich an Untersuchungen von Vokey und Read (1992) und O'Toole, Deffenbacher, Valentin und Abdi (1994) die in Untersuchungen zum Wiedererkennen von Gesichtern diese Skalen verwendeten und dabei zwei Faktoren extrahierten, auf denen die fünf Skalen laden. Vokey und Read (1992) postulieren, dass zwei Faktoren zur Fähigkeit der Gesichterwiedererkennung beitragen. Der erste Faktor wurde von den Autoren mit 'Kontext-freie Vertrautheit' und der zweite Faktor mit 'Einprägbarkeit' benannt. Der Faktor 'Kontext-freie Vertrautheit' bezieht sich auf Gefühle und Emotionen die mit einem Gesicht verbunden werden und beinhaltet hohe

positive Ladungen der Skalen 'Attraktivität', 'Sympathie' und 'Vertrautheit'. Der zweite Faktor 'Einprägsamkeit' bezieht sich dagegen auf die Beurteilung der Wiedererkennbarkeit eines Gesichts und beinhaltet die hohe positive Ladungen der Skala 'Einprägsamkeit' sowie hohe negative Ladungen der Skala 'Typikalität'. Mit Typikalität ist dabei immer die wahrgenommene Typikalität durch den Beobachter gemeint, da diese selbstverständlich zwischen den Personen variieren kann. Vokey und Read (1992) behaupten, ein distinktes Gesicht sei besser wiederzuerkennen, weil es zum einen in bezug auf die 'Kontext-freie Vertrautheit' niedrige Werte und zum anderen bezüglich der Einprägsamkeit hohe Werte besitzt, da sich solch ein Gesicht in starkem Maße von anderen Gesichtern unterscheidet. Die Items für die Messung der Skalen 'Typikalität' und 'Vertrautheit' wurden in Anlehnung an Meissner und Brigham (in Druck) von O'Toole et al. (1994) übernommen. Die anderen Items stammen von Vokey und Read (1992). Die Skalen wurden für diese Pilotstudie von der englischen Form ins Deutsche übersetzt. Folgende Skalen wurden beurteilt:

1. Typikalität (*typicality*): Wie schwierig wäre es, diese Person in einer Menschenmenge zu erkennen? (How difficult would it be to pick this person out of a crowd?)
(1 = "nicht schwierig" bis 7 = "sehr schwierig")
2. Sympathie (*likability*): Sieht diese Person sympathisch aus? (Does this person look as if he would be likable?)
(1 = "nicht sympathisch" bis 7 = "sehr sympathisch")
3. Attraktivität (*attractiveness*): Sieht diese Person attraktiv aus? (Is this person attractive looking?)
(1 = "nicht attraktiv" bis 7 = "sehr attraktiv")
4. Einprägsamkeit (*memorability*): Wäre es einfach, sich an diese Person zu erinnern? (Would this face be easy to remember?)
(1 = "nicht einfach" bis 7 = "sehr einfach")
5. Vertrautheit (*familiarity*): Ist dieses Gesicht verwechselbar mit jemandem, den sie kennen? (Is this face confusable with someone you know?)
(1 = "nicht verwechselbar" bis 7 = "sehr verwechselbar")

Für die Erstellung der Darbietungssets für die nachfolgenden Experimente ist der zweite Faktor 'Einprägsamkeit' von Bedeutung, da dieser diejenigen Skalen beinhaltet, die für die Wiedererkennbarkeit eines Gesichtes wichtig sind, nämlich 'Distinktivität' und 'Erinnerbarkeit'. Für die spätere Berechnung der Distinktivität eines Bildes wurde der Wert der Skala 'Typikalität' benutzt und von dem Wert acht subtrahiert, weil Typikalität und

Distinktivität die beiden Pole einer Dimension darstellen. Die anderen Skalen gingen gemäß ihrer Ratings in die spätere Berechnung ein.

3.1.4 Stimulus-Darbietung

Die Datenerhebung erfolgte an zwei Apple 'IMac' Computern mit einer Taktfrequenz von 300 bzw. 350 Mhz, einem Arbeitsspeicher von jeweils 192 MB und einem 15"-Monitor. Die Farbtiefe betrug 16 Bit (65536 Farben) bei einer Bildschirmauflösung von 1024x768 Pixel und einer Bildwiederholfrequenz von 75 Hz. Die Versuchssteuerung geschah durch das Programm 'Pyscope 1.2.5' (Cohen, MacWhinney, Flatt & Provost, 1993).

Jede der fünf Skalen sollte einmal pro Bild bearbeitet werden, wobei vermieden werden sollte, für jedes Bild die gleiche Reihenfolge der Skalenbearbeitung zu haben, um Habituationseffekte der Probanden bei der Skalenbearbeitung zu vermeiden. Schließlich sollte die Frage für jede Skala gewissenhaft beantwortet werden. Dazu wurden die fünf Skalen für die Darbietung in Form eines lateinischen Quadrates kombiniert (siehe Tabelle 8). Dieser Schritt war notwendig, weil die Programmierung der einzelnen Trials in Pyscope anders nicht möglich war.

Tabelle 8

Aufteilung der fünf Skalen in ein lateinisches Quadrat

Skala	Feld				
	1	2	3	4	5
Attraktivität (attr)	attr	typ	mem	lik	fam
Vertrautheit (fam)	fam	attr	typ	mem	lik
Sympathie (lik)	lik	fam	attr	typ	mem
Einprägsamkeit (mem)	mem	lik	fam	attr	typ
Typikalität (typ)	typ	mem	lik	fam	attr

Die standardisierten Fotos wurden für die Darbietung während der Pilotstudie bei ihrer Originalgröße von 6 x 9 Inches (432 x 648 Pixel) belassen.

3.1.5 Versuchsdurchführung

Die Pilotstudie gliederte sich in zwei Teile: 1) Instruktionsphase und 2) Versuchsphase. Der Versuchsablauf für jede einzelne Versuchsperson war wie folgt: Die Teilnehmer wurden instruiert, in der Instruktionsphase den Anweisungen auf dem Bildschirm zu folgen. Die einzelnen Instruktionen sind in Anhang A4 aufgeführt. Während der Instruktionsphase wurde nach der vierten Instruktion ein Beispieltrial bearbeitet, dessen Stimulusbild aus einem Foto der Fotoserie von Meissner genommen wurde. Nach der Instruktionsphase folgte die Versuchsphase mit dem Hinweis für die Teilnehmer, eine Taste zu drücken, wenn sie für den Versuch bereit sind. Anschließend folgten die ersten 80 Trials in randomisierter Reihenfolge mit, in Abhängigkeit von der Bedingung, entweder türkischen oder deutschen Gesichtern. Jeder Trial war gleichermaßen aufgebaut:

- 1) Präsentation des Stimulus in standardisierter Größe (432 x 648 Pixel) für 2000 ms,
- 2) Präsentation des Stimulus mit einer Größe von (Breite x Höhe) 266 x 400 Pixel in der oberen Hälfte des Bildschirms (horizontal zentriert, vertikal 32 %¹¹). Gleichzeitig wird nach randomisierter Auswahl einer der fünf Zeilen des lateinischen Quadrates die erste der Skalen in dieser Zeile in der unteren Hälfte des Bildschirms (700 x 158 Pixel, horizontal zentriert, vertikal 75 %) präsentiert. Die Teilnehmer sollten nun das Rating für den Stimulus für die präsentierte Skala durch Drücken einer der Tasten von 1 bis 7 abgeben. Diese Prozedur wiederholte sich für die verbleibenden vier Skalen. Das Stimulusgesicht blieb während der Bearbeitung der fünf Skalen auf der oben angegebenen Position präsent. Zwischen den einzelnen Trials lag ein ITI von 500 ms. Die Randomisierung der fünf Skalen war so ausbalanciert, dass jede der fünf Zeilen des lateinischen Quadrates 16 mal ausgewählt wurde. Nach Bearbeitung der ersten 80 Trials folgten die zweiten 80 Trials mit Fotos von Gesichtern der noch nicht bearbeiteten ethnischen Gruppe.

3.2 Ergebnisse

Die Werte der Ratings für die einzelnen Skalen wurden für jedes Stimulusfoto getrennt nach Ethnizität der Teilnehmer in Form von Mittelwerten zusammengefasst.

¹¹ Die Angabe der vertikalen Position in Prozentwerten bedeutet bei Psyscope 1.2.5 die relative Position des Mittelpunktes einer definierten Fläche zum vertikalen Mittelpunkt des Bildschirms. Ein Prozentwert von 0 heißt, dass sich der Flächenmittelpunkt am oberen Bildschirmrand befindet. Bei einem Prozentwert von 100 befindet sich der Flächenmittelpunkt am unteren Bildschirmrand.

Meissner und Brigham (in Druck) fanden bei den Ratings, die sie für Afro-Amerikaner und kaukasische Amerikaner durchführen ließ, analog zu Vokey und Read (1992) zwei Faktoren mit Eigenwerten > 1 , auf denen die fünf Skalen laden: 'Kontext-freie Vertrautheit' (Skalen: 'Sympathie', 'Vertrautheit' und 'Attraktivität') und 'Einprägsamkeit' (Skalen: 'Einprägsamkeit' und 'Distinktivität'). Diese Faktorenstruktur mit zwei Faktoren mit Eigenwerten > 1 konnte größtenteils auch bei den Ratings für deutsche und türkische Gesichter in der Pilotstudie gefunden werden. In den Tabelle 9 und 10 sind die Ergebnisse einer Faktorenanalyse mit Varimax-Rotation für die fünf Skalen getrennt nach türkischen und deutschen Versuchsteilnehmern dargestellt.

Tabelle 9

Ergebnisse der Faktorenanalyse für Gesichter von türkischen und deutschen Männern mit türkischen Teilnehmern

Skalen	Türkische Gesichter		Deutsche Gesichter	
	Faktor 1	Faktor 2	Faktor 1	Faktor 2
Einprägsamkeit (mem)	-.09	.94	.01	.96
Distinktivität (dis)	.02	.94	.01	.97
Sympathie (lik)	.89	-.03	.85	.04
Attraktivität (att)	.86	.06	.89	.11
Vertrautheit (fam)	.70	-.11	.77	-.11

Anmerkung: Die beiden Faktoren klären bei den türkischen Gesichtern 76.5 % der Varianz auf, bei den deutschen Gesichtern 80.2 %.

Tabelle 10

Ergebnisse der Faktorenanalyse für Gesichter von türkischen und deutschen Männern mit deutschen Teilnehmern

Skalen	türkische Gesichter		deutsche Gesichter	
	Faktor 1	Faktor 2	Faktor 1	Faktor 2
Einprägsamkeit (mem)	.17	.92	.12	.97
Distinktivität (dis)	-.01	.94	.01	.98
Sympathie (lik)	.90	.21	.88	.18
Attraktivität (att)	.94	-.01	.93	-.01
Vertrautheit (fam)	.60	-.55	.70	.01

Anmerkung: Die beiden Faktoren klären bei den türkischen Gesichtern 83.2 % der Varianz auf, bei den deutschen Gesichtern 81.1 %.

Tabelle 11 zeigt die Mittelwerte und Standardabweichungen der Ratings getrennt nach Skala, Ethnie der Teilnehmer und Stimulusethnie

Tabelle 11

Mittelwerte und Standardabweichungen der fünf Skalen getrennt nach Ethnie der Teilnehmer und Stimulusethnie

Teilnehmer	Türkische Gesichter		Deutsche Gesichter		<u>M_g</u>	<u>SD_g</u>
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>		
Türken						
Einprägsamkeit (mem)	3.86	1.82	3.47	1.83	3.67	1.83
Distinktivität (dis)	4.29	1.89	3.94	1.96	4.17	1.92
Sympathie (lik)	2.84	1.56	2.66	1.52	2.75	1.54
Attraktivität (att)	2.13	1.32	2.02	1.27	2.08	1.29
Vertrautheit (fam)	2.34	1.69	2.41	1.79	2.38	1.74
Deutsche						
Einprägsamkeit (mem)	3.68	1.62	3.98	1.61	3.83	1.61
Distinktivität (dis)	3.72	1.64	3.98	1.67	3.85	1.66
Sympathie (lik)	3.00	1.40	3.17	1.35	3.08	1.38
Attraktivität (att)	2.51	1.28	2.44	1.22	2.47	1.25
Vertrautheit (fam)	2.05	1.60	2.23	1.63	2.14	1.61
Gesamt						
Einprägsamkeit (mem)	3.77	1.72	3.72	1.72	3.75	1.72
Distinktivität (dis)	4.00	1.76	3.96	1.82	3.98	1.79
Sympathie (lik)	2.92	1.48	2.92	1.44	2.92	1.46
Attraktivität (att)	2.32	1.30	2.23	1.24	2.28	1.27
Vertrautheit (fam)	2.20	1.64	2.32	1.71	2.26	1.68

Zusätzlich wurden Varianzanalysen durchgeführt, um zu testen, ob die jeweiligen Fremd-Gruppen Gesichter von den beiden Teilnehmergruppen als signifikant abweichend auf den fünf Skalen angesehen werden, verglichen mit den Eigen-Gruppen Gesichtern. Das Alpha-Niveau wurde entsprechend der Anzahl der Tests angepasst¹². Eine zweifaktorielle Varianzanalyse mit dem Between-Subjects Faktor 'Ethnie der Versuchspersonen' (Türken, Deutsche) und dem Within-Subjects Faktor 'Stimulusethnie' (türkische Gesichter, deutsche Gesichter), sowie der Skala 'Einprägsamkeit' als abhängiger Variablen zeigte einen signifikanten Haupteffekt für die Ethnie der Versuchspersonen, $F(1,158) = 30.67$, $p = .001$,

¹² Das Alpha-Niveau beträgt bei 15 Tests nach Bonferroni-Adjustierung $\alpha = .0033$.

$r = .40$. Die deutschen Teilnehmer betrachten die Gesichter als einprägsamer als die türkischen Teilnehmer ($M_{de} = 3.83$, $M_{tü} = 3.67$). Die Interaktion der Faktoren 'Ethnie der Versuchspersonen' und 'Stimuluse ethnien', $F(1,158) = 142.80$, $p = .001$, $r = .69$ zeigt, dass die türkischen Teilnehmer die türkischen Gesichter und die deutschen Teilnehmer die deutschen Gesichter als erinnerbarer ansehen.

Die Varianzanalyse der Skala ‚Distinktivität‘ zeigt, dass die türkischen Teilnehmer die Gesichter als distinkter einstufen als die deutschen Teilnehmer, $F(1,158) = 74.67$, $p = .001$, $r = .57$ ($M_{tü} = 4.12$, $M_{de} = 3.85$). Ebenso fand sich eine signifikante Interaktion der Faktoren 'Ethnie der Versuchspersonen' und 'Stimuluse ethnien', $F(1,158) = 99.41$, $p = .001$, $r = .62$. Die türkischen Teilnehmer empfinden die türkischen Gesichter als distinkter ($M_{tü} = 4.28$, $M_{de} = 3.94$), während die deutschen Teilnehmer die deutschen Gesichter als distinkter wahrnehmen ($M_{tü} = 3.72$, $M_{de} = 3.98$).

Die Varianzanalyse der Skala ‚Sympathie‘ zeigt einen signifikanten Haupteffekt des Faktors 'Ethnie der Versuchspersonen', $F(1,158) = 164.54$, $p = .001$, $r = .71$. Die deutschen Teilnehmer betrachten die Gesichter insgesamt als sympathischer im Vergleich zu den türkischen Teilnehmern ($M_{de} = 3.08$, $M_{tü} = 2.75$). Die Interaktion der beiden Faktoren 'Ethnie der Versuchspersonen' und 'Stimuluse ethnien' war ebenfalls signifikant, $F(1,158) = 42.13$, $p = .001$, $r = .46$. Türkische Teilnehmer beurteilten die türkischen Gesichter als sympathischer als die deutschen Gesichter ($M_{tü} = 2.84$, $M_{de} = 2.66$), dies gilt ebenfalls für die deutschen Teilnehmer bei den deutschen Gesichtern ($M_{de} = 3.17$, $M_{tü} = 3.00$).

Die Varianzanalyse der Skala 'Attraktivität' zeigte einen signifikanten Haupteffekt für den Faktor 'Ethnie der Versuchspersonen', $F(1,158) = 246.62$, $p = .001$, $r = .78$, mit höheren Attraktivitätswerten der deutschen Teilnehmer verglichen mit den türkischen Teilnehmern ($M_{de} = 2.47$, $M_{tü} = 2.08$). Es zeigte sich kein Haupteffekt für den Faktor 'Stimuluse ethnien' und keine Interaktion der beiden Faktoren.

Zuletzt zeigt die Varianzanalyse der Skala 'Vertrautheit' einen Haupteffekt für die Ethnie der Versuchspersonen, $F(1,158) = 73.17$, $p = .001$, $r = .56$. Die türkischen Versuchspersonen beurteilten die Gesichter als vertrauter ($M = 2.38$) als die deutschen Versuchspersonen ($M = 2.139$). Die deutschen Gesichter werden wiederum als vertrauter beurteilt als die türkischen Gesichter wie der Haupteffekt 'Stimuluse ethnien' zeigt $F(1,158) = 9.36$, $p = .003$, $r = .24$ ($M_{de} = 2.32$, $M_{tü} = 2.20$).

Anhand der über die Teilnehmer gemittelten Ratings konnte das Stimulusmaterial in zwei Sets aufgeteilt werden. Die grundlegende Voraussetzung für die Erstellung der Sets

war, dass die Bilder zwischen den Sets sich in ihrer 'Schwierigkeit'¹³ nicht unterschieden. Da 'untypische' Gesichter besser wiedererkannt werden können als 'typische' Gesichter (Brigham, 1990; Valentine, 1991a), beruht die Aufteilung in die Sets auf dem Faktor 'Einprägsamkeit' mit den Skalen 'Distinktivität' sowie 'Einprägsamkeit'. Dazu wurden zunächst jeweils die Bilder von Türken und Deutschen anhand der Ratings der beiden Skalen 'Einzigartigkeit' und 'Einprägsamkeit' durch Mittelwertbildung zusammengefasst und anschließend anhand der Mittelwerte in aufsteigender Reihenfolge sortiert. Danach wurden 'Paare' gebildet, die durch Zufallszuweisung auf die beiden Sets verteilt wurden. Konkret bedeutet dies: Das Bild mit der geringsten Einstufung auf dem Faktor 'Einzigartigkeit' wurde mit dem Bild mit der zweit-niedrigsten Einstufung gepaart und per Zufall dem Set A bzw. dem Set B zugeordnet. Diese Prozedur wurde mit den weiteren Bildern wiederholt, sodass letztendlich die Sets A und B, für türkische und deutsche Stimuluspersonen, aus jeweils 40 Bildern bestanden. Um sicherzustellen, dass sich die beiden Sets bezüglich der Ratings nicht voneinander unterscheiden, wurde die Sets getrennt nach Stimuluse ethnien mittels t -tests für abhängige Stichproben auf signifikante Unterschiede getestet. Die beiden Sets der türkischen Stimuluspersonen unterscheiden sich nicht signifikant voneinander, $t(39) = -.442$, ns., ($M_{\text{setA}} = 3.88$, $M_{\text{setB}} = 3.89$), ebenso wie die beiden Sets der deutschen Stimuluspersonen, $t(39) = -.899$, ns., ($M_{\text{setA}} = 3.84$, $M_{\text{setB}} = 3.85$).

3.3 Diskussion

Die Faktorenstruktur von Vokey und Read (1992) und O'Toole et al. (1994), die auch von Meissner und Brigham (in Druck) repliziert werden konnte, wurde größtenteils auch in den Ratings von türkischen und deutschen Teilnehmern bei türkischen und deutschen Gesichtern gefunden. Beide beiden Teilnehmer-Gruppen zeigten sich in den Ratings zwei Faktoren mit Eigenwerten > 1 , und hohe Ladungen der Skalen 'Distinktivität' und 'Memorabilität' auf einem Faktor (Erinnerbarkeit) und zwar bei beiden Stimuluse ethnien. Bei türkischen Teilnehmern wurde die Faktorenstruktur bei beiden Stimuluse ethnien gefunden, mit sehr hohen Faktorladungen, vor allem der Skalen

¹³ Mit dem Begriff 'Schwierigkeit' ist in diesem Zusammenhang die bessere Wiedererkennbarkeit einzelner Bilder gegenüber anderen aufgrund spezifischer Charakteristika gemeint. Entlehnt ist diese Beschreibung der probabilistischen Testtheorie. (siehe Rost, 1995)

'Einprägsamkeit' und 'Distinktivität', auf dem zweiten Faktor. Die Ladungen der Skalen 'Sympathie', 'Attraktivität' und 'Vertrautheit' auf dem ersten Faktor sind zwar etwas geringer, aber dennoch ausreichend hoch. Die Ladungen der Skalen auf dem jeweils anderen Faktor liegen im Bereich einer Null-Korrelation. Ein ähnliches Bild zeigt sich bei der Faktorenanalyse der Daten der deutschen Teilnehmer. Auch hier sind die Faktorladungen der Skalen 'Einprägsamkeit' und 'Distinktivität' auf dem zweiten Faktor nahezu perfekt. Die Ladungen der Skalen 'Sympathie', und 'Attraktivität' sind auf dem ersten Faktor ebenfalls sehr hoch, mit einer moderaten Ladung der Skala 'Vertrautheit' bei deutschen Gesichtern. Auffällig ist in diesem Zusammenhang die Faktorenanalyse der Daten der deutschen Teilnehmer bei türkischen Stimulusgesichtern. Die Skala 'Vertrautheit' lädt bei dieser Analyse moderat positiv auf dem ersten und moderat negativ auf dem zweiten Faktor. Da diese Pilotstudie vor allem dazu dient, die Stimulusgesichter den Darbietungssets zuzuordnen, und die eindeutigen Faktorladungen der Skalen 'Einprägsamkeit' und 'Distinktivität' für eine Ordnung der Bilder mittels dieses Faktors sprechen, wird auf die Problematik der Skala 'Vertrautheit' nur kurz eingegangen. Morris und Wickham (2001), setzen sich kritisch mit den Analysen von Vokey und Read (1992) auseinander. In einem Experiment, in dem Morris und Wickham (2001) weibliche und männliche Gesichter bezüglich ihrer Attraktivität, Vertrautheit, Freundlichkeit (pleasantness), Erinnerbarkeit und Typikalität beurteilen liessen, zeigten sich zwei Faktoren, die sie jedoch wegen der hohen Ladungen der Skalen Distinktivität auf dem ersten Faktor und Attraktivität auf dem zweiten Faktor, ebenfalls mit *Distinktivität* und *Attraktivität* benannten. Eine derartige Interpretation ist auch mit den Daten aus dieser Pilotstudie möglich. Es soll hier jedoch nicht auf die Problematik der Faktorenbenennung weiter eingegangen werden, weil wie gesagt, diese Pilotstudie der Generierung von Darbietungssets für die Wiedererkennensexperimente in den Abschnitten 4 und 6 dienen sollte. Insgesamt läßt sich sagen, dass die deutsche Übersetzung der Skalen zu einer ähnlichen Faktorenstruktur führt wie bei Vokey und Read (1992) und in noch stärkerem Maße wie bei Morris und Wickham (2001).

Die Varianzanalysen der einzelnen Skalen zeigen weiterhin, dass Eigen-Gruppen und Fremd-Gruppen Gesichter von den beiden Versuchspersonen-Gruppen unterschiedlich wahrgenommen werden. Die Effektstärken weisen durchgängig auf mittlere bis starke Effekte hin. Die Eigen-Gruppen Gesichter werden von beiden Versuchspersonen-Gruppen als distinkter und erinnerbarer eingestuft als die Fremd-Gruppen Gesichter. Dieser Befund ist in Einklang mit der Annahme der Speicherung von Gesichtern in einem multidimensionalen Raum, nach der Fremd-Gruppen Gesichter per se zwar als distinkter

von der eigenen Gruppe wahrgenommen werden, sie sich aber in der Wahrnehmung innerhalb der Fremd-Gruppe nicht voneinander unterscheiden. Dementsprechend werden die Eigen-Gruppen Gesichter als distinkter und damit zugleich als erinnerbarer angesehen, da die Distinktivität eines Gesichts gleichbedeutend ist mit gut erinnerbaren Merkmalen, die nicht alle Exemplare teilen.

Die Eigen-Gruppen Gesichter werden zudem von beiden Versuchspersonen-Gruppen als vertrauter angesehen als die Fremd-Gruppen Gesichter. Auch dieser Befund ist in Einklang mit der Annahme der Speicherung von Gesichtern in einem multidimensionalen Raum. Da die Eigen-Gruppen Gesichter die größere Zahl der Exemplare darstellt und der Umgang mit Personen der eigenen Ethnie häufiger und eben vertrauter ist, stellen Fremd-Gruppen Gesichter eine Menge von Exemplaren dar, mit denen im allgemeinen noch nicht viele Erfahrungen gemacht wurden und dementsprechend als nicht vertraut eingestuft werden.

Für die nachfolgenden Experimente ist, der Faktor 'Einprägbarkeit' von Bedeutung, da dieser Faktor weitgehend die Wiedererkennung von Gesichtern beeinflusst (Vokey & Read, 1992; Meissner & Brigham, in Druck). Die Sets für türkische und deutsche Gesichter sowie die bereits vorhandenen Sets für afro-amerikanische und kaukasische Gesichter konnten nun in den folgenden Experimenten für die Wiedererkennstests in Abschnitt 4 und die Delayed-Matching Aufgabe in Abschnitt 6 verwendet werden.

4 Der Einfluss von Behaltensintervall und Gesichtsausschnitt auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern

Ein zentraler Punkt des In-Group/Out-Group Modells für die Erklärung des Cross-Race Bias ist die unterschiedliche Verarbeitung von Eigen-Gruppen und Fremd-Gruppen Gesichtern. Eigen-Gruppen Gesichter werden nach dem Modell holistisch und Fremd-Gruppen Gesichter merkmalsbasiert verarbeitet. Welchen Einfluss Gesichtsmerkmale wie Haare und Frisur auf die Kategorisierung von Gesichtern haben, wurde von MacLin und Malpass (2001) gezeigt. Die Autoren generierten mit dem Computer 'neutrale' Gesichter, und fügten entweder typisch latein-amerikanische bzw. afro-amerikanische Frisuren hinzu, sodass diese Gesichter in Abhängigkeit von dieser Frisur von den Versuchsteilnehmern als latein-amerikanisch bzw. afro-amerikanisch eingestuft wurden. Wright und Sladden (2003) fanden, je nachdem, ob auf den Bildern der Stimuluspersonen das Haupthaar sichtbar war, schlechtere Wiedererkennensleistungen bei fehlendem Haupthaar. Es ist bei der Untersuchung von Wright und Sladden allerdings etwas unklar, mit welcher Prozedur die Bilder bearbeitet wurden. Ebenfalls unklar ist, ob die Bilder standardisiert wurden, weil die Autoren einen Teil ihrer Stimuli aus dem Internet heruntergeladen hatten und einen anderen Teil aus einem 'Gesichter-Pool' des ansässigen Instituts für Psychologie heranzogen.

In diesem Experiment sollte die Bedeutsamkeit der Haare und der Frisur der Stimuluspersonen, und somit der wichtigsten äußeren Gesichtsmerkmale auf die Wiedererkennensleistung bei Eigen-Gruppen und Fremd-Gruppen Gesichtern untersucht werden. Es sollte u.a. die Hypothese getestet werden, nach der Gesichter anderer ethnischer Gruppen als der eigenen eher durch äußere (Oberflächen-) Merkmale, Gesichter der Eigengruppe jedoch eher durch innere Gesichtsmerkmale enkodiert werden. Um die Auswirkungen einer Darbietung von Gesichtern ohne äußere Merkmale auf den Cross-Race Bias zu untersuchen, sollten türkischen und deutschen Teilnehmern Gesichter dargeboten werden, bei denen entweder das gesamte Gesicht sichtbar war, oder die Gesichtsansicht auf das Gesichtsinere beschränkt wurde. Dass der Cross-Race Bias auch bei türkischen und deutschen Versuchsteilnehmern auftritt, konnte Sporer (1999, Juli) zeigen.

In den meisten Untersuchungen zum Cross-Race Bias wurden vor allem Gesichter von Schwarzen und Weißen als Stimulusmaterial verwendet und die Wiedererkennensleistung an schwarzen und/oder weißen Teilnehmern getestet. Die Stimulusgesichter und Teilnehmer gehören demnach in den meisten Untersuchungen den gleichen Ethnien an (Dieser Umstand liegt selbstverständlich in der Natur des Cross-Race Bias). In der Literatur finden sich jedoch keine Untersuchungen, die sich der Frage widmen, wie gut Gesichter wiedererkannt werden können, deren Ethnie zu keiner der untersuchten Teilnehmergruppen gehören. Bei solchen Gesichtern ist nach dem In-Group/Out-Group Modell anzunehmen, dass sie von beiden Teilnehmergruppen gleichermaßen enkodiert werden, d.h. merkmalsbasiert, und sie damit auch zu ähnlichen (schlechteren) Wiedererkennensleistungen führen. Zur Überprüfung dieser Hypothese wurden zusätzlich zu den Gesichtern von Türken und Deutschen, Gesichter von Afro-Amerikanern und kaukasischen Amerikanern als Stimulusgesichter verwendet.

Es gibt ebenfalls wenig Untersuchungen über die Form des Cross-Race Bias nach einem längeren Behaltensintervall. Die meisten Tests der Wiedererkennensleistung finden nach einem kurzen Intervall von fünf bis zehn Minuten statt. Shepherd, Gibling, und Ellis (1991) fanden nach einem Behaltensintervall von einem Monat eine höhere Anzahl an Treffern bei distinkten Gesichtern verglichen mit typischen Gesichtern. Nach den Annahmen des In-Group/Out-Group Modells und der Speicherung von Gesichtern in einem multidimensionalen Raum sollte ein einzelnes Fremd-Gruppen Gesicht als distinkter verglichen mit den Eigen-Gruppen Gesichtern wahrgenommen werden. Wenn aber alle bzw. viele Fremd-Gruppen Gesichter ein distinktes Merkmal teilen, fällt dieses Merkmal als Unterscheidungskriterium weg. Die Fremd-Gruppen Gesichter sollten demnach als homogener/typischer als die Eigen-Gruppen Gesichter wahrgenommen werden, solange die beobachtende Person keine weiteren Merkmale identifizieren kann, die es ihr ermöglichen, diese Gesichter voneinander zu unterscheiden. In der Pilotstudie in Abschnitt 3 wurde gezeigt, dass türkische und deutsche Teilnehmer die Gesichter der eigenen Ethnie als distinkter einstufen als die Gesichter der anderen Ethnie. Die Teilnehmer finden also bei den Gesichtern der Eigen-Gruppe eher Merkmale, die nach Einordnung der Teilnehmer dazu geeignet sind, diese Gesichter voneinander zu unterscheiden. Dies hieße für einen Wiedererkennentest nach einem längeren Zeitintervall unter Berücksichtigung der Ergebnisse von Shepherd, Gibling und Ellis (1991), dass es bei Gesichtern der eigenen ethnische Gruppe zu mehr Treffern kommen müsste, als bei Gesichtern anderer ethnischer Gruppen. Ebenso ist nach dem In-Group/Out-Group Modell anzunehmen, dass bei einer merkmalsbasierten Enkodierung von Fremdgruppen-Gesichtern, deren Distinktheit in der

Wahrnehmung der Teilnehmer geringer ist als bei Eigengruppen-Gesichtern, die Wiedererkennensleistung von Fremdgruppen-Gesichtern nach einem längeren Zeitintervall geringer ist als bei Eigengruppen-Gesichtern, deren Enkodierung holistisch erfolgt. Eine mögliche Erklärung hierfür könnte der Effekt des Verbal Overshadowing liefern.

Damit ergibt sich der in Tabelle 12 dargestellte Versuchsplan. Der Faktor 'Taste' wurde mit einbezogen, um Antworttendenzen aufgrund der Orientierung der Antworttasten zu kontrollieren.

Tabelle 12

Versuchsplan für das Experiment zum Einfluss des Behaltensintervalls und des Gesichtsausschnitts auf den Cross-Race Bias, unterteilt nach Set, Ausschnitt und Taste

BI	Vp-Eth	Sex	Sequenz	A				B			
				In		Out		In		Out	
				c	m	c	m	c	m	c	m
Unm.	Türken	Männl.	T-G/S-W	1	1	1	1	1	1	1	1
			S-W/T-G	1	1	1	1	1	1	1	1
		Weibl.	T-G/S-W	1	1	1	1	1	1	1	1
			S-W/T-G	1	1	1	1	1	1	1	1
	Deut.	Männl.	T-G/S-W	1	1	1	1	1	1	1	1
			S-W/T-G	1	1	1	1	1	1	1	1
		Weibl.	T-G/S-W	1	1	1	1	1	1	1	1
			S-W/T-G	1	1	1	1	1	1	1	1
3 Wn.	Türken	Männl.	T-G/S-W	1	1	1	1	1	1	1	1
			S-W/T-G	1	1	1	1	1	1	1	1
		Weibl.	T-G/S-W	1	1	1	1	1	1	1	1
			S-W/T-G	1	1	1	1	1	1	1	1
	Deut.	Männl.	T-G/S-W	1	1	1	1	1	1	1	1
			S-W/T-G	1	1	1	1	1	1	1	1
		Weibl.	T-G/S-W	1	1	1	1	1	1	1	1
			S-W/T-G	1	1	1	1	1	1	1	1

Anmerkung: A bzw. B = Set A, B; In bzw. Out = inneres / vollständiges Gesicht; c bzw. m = Taste „Gesicht alt“; BI = Behaltensintervall; Vp-Eth = Versuchspersonen-Ethnie; Sex = Geschlecht; Unm = unmittelbare Testung; 3 Wn = Testung nach drei Wochen; Deut = Deutsche Teilnehmer

Die Anzahl der Teilnehmer ($N = 128$) ist ausreichend, um bei erwarteten mittleren Effektstärken ($.3 \leq r \leq .49$) beim Cross-Race Bias (Meissner & Brigham, 2001) und einem alpha-Niveau von .05 je nach Anzahl der Zählerfreiheitsgrade eine Power von .97 bis .99 zu erreichen (siehe Cohen, 1988).

4.1 Hypothesen

Es wird vermutet, dass das Behaltensintervall einen Einfluss auf die Wiedererkennensleistung hat. Betrachtet man die Vergessenskurve nach Ebbinghaus (1885) sollten selbstverständlich die Teilnehmer, die nach drei Wochen getestet werden, insgesamt schlechtere Wiedererkennensleistungen aufweisen als die Teilnehmer, deren Testung sich direkt an die Darbietungsphase anschließt.

Wenn Eigen-Gruppen Gesichter holistisch verarbeitet werden und Fremd-Gruppen Gesichter merkmalsbasiert, dann sollte zunächst ein Cross-Race Bias zu beobachten sein, unabhängig vom Zeitintervall, d.h. Türken sollten türkische Gesichter besser wiedererkennen und Deutsche die deutschen Gesichter, verglichen mit den Fremdgruppen. Dieser Effekt sollte bei einer Testung nach drei Wochen noch stärker zum Tragen kommen, weil Prozesse des Verbal Overshadowing bei Fremd-Gruppen Gesichtern zu einer stärkeren Verschlechterung der Wiedererkennensleistung führen sollten. Shepherd, Gibling und Ellis (1991) berichten von einer höheren Trefferrate bei distinkten Gesichtern nach einem Zeitintervall von einem Monat. Bei einer höheren Intra-Gruppen Distinktheit hieße dies, dass Eigen-Gruppen nach einem Intervall von drei Wochen eine höhere Trefferrate aufweisen sollten als Fremd-Gruppen Gesichter.

Weiterhin ist anzunehmen, dass die Präsentation eines vollständigen Gesichts bzw. eines Gesichts, bei dem nur das Gesichtsinere sichtbar ist, einen Einfluss auf die Wiedererkennensleistung hat. Wie bei Wright und Sladden (2003) sollten vollständige Gesichter besser wiedererkannt werden als unvollständige, weil äußere Merkmale als zusätzliche Enkodierungsmöglichkeit fehlen. Dies sollte auch hier der Fall sein. Zwar sollte es bei Eigen-Gruppen Gesichtern, wegen der holistischen Verarbeitung nicht zu einem Effekt des Faktors 'Maske' kommen, insgesamt sollte sich dieser Faktor aber auf die Wiedererkennensleistung auswirken, mit schlechteren Leistungen für die maskierten Gesichter. Wenn weiterhin nach dem In-Group/Out-Group Modell von Sporer (2001a) Fremd-Gruppen Gesichter zunächst kategorisiert und anschließend merkmalsbasiert verarbeitet werden, Eigen-Gruppen Gesichter hingegen holistisch enkodiert werden, sollten

die Fremd-Gruppen Gesichter, bei denen lediglich das Gesichtsinere sichtbar ist, am schlechtesten wiedererkannt werden. Es wird demnach eine Interaktion der Faktoren 'Maske', 'Stimulusethnie' und 'Vpn-Ethnie' erwartet. Die besten Leistungen sollten sich bei den vollständigen Gesichtern der Eigen-Gruppe zeigen.

In den meisten Untersuchungen werden den Teilnehmern Gesichter dargeboten, die sowohl der eigenen ethnischen Gruppe entstammen als auch einer anderen. In diesem Experiment sollte die Wiedererkennensleistung sowohl bei der eigenen ethnischen Gruppe als auch bei mehreren anderen Gruppen getestet werden. Die Frage, wie sich die Leistung bei Fremd-Gruppen Gesichtern verhält, ist nach multi-dimensionalen Modellen der Gesichterrepräsentation relativ einfach zu beantworten. Mit zunehmender Distanz eines Gesichtes von der Eigen-Gruppe und damit von den häufigsten Repräsentationen in diesem Raum sollte die Wiedererkennensleistung abnehmen. Für beide Teilnehmergruppen hieße dies, dass die Gesichter der Eigen-Gruppe am besten wiedererkannt werden. Je nach der Dimension in dem subjektiven multidimensionalen Raum, die für die Teilnehmer relevant ist, sollten sich die Wiedererkennensleistungen für die weiteren Ethnien staffeln. Aufgrund der Hautfarbe und relativen Seltenheit von afro-amerikanischen Gesichtern in der Umgebung der meisten Teilnehmer ist anzunehmen, dass Gesichter afro-amerikanischer Personen am schlechtesten wiedererkannt werden können. Da die Gesichter von Deutschen und von kaukasischen Amerikanern viele Merkmale teilen, sollten für beide Teilnehmergruppen diese beiden Gruppen ähnlich gut wiederzuerkennen sein. Die Teilnehmergruppen sollten sich lediglich bezüglich der Reihenfolge unterscheiden.

4.2 Material

In diesem Abschnitt werden die für die Untersuchung verwendeten Materialien beschrieben. Anschließend werden die Ergebnisse dargestellt und diskutiert.

4.2.1 Stimuli

Als Stimulusmaterial wurden die in der Pilotstudie eingestufteten Bilder von Türken und Deutschen sowie die Bilder von afro-amerikanischen und kaukasischen Gesichtern von Meissner und Brigham (in Druck) verwendet. Beispielgesichter sind in Anhang D1 abgebildet. Für die Versuchsbedingung 'Maske' (d.h. lediglich das innere Gesicht ist

sichtbar) wurden zusätzlich alle Bilder maskiert, d.h. von den Bildern wurden die Gesichter derart ausgeschnitten, dass lediglich das Innere eines Gesichtes und ein Teil des Haaransatzes sowie des Halses sichtbar waren. Diese Teile waren zusätzlich 'weichgezeichnet', d.h., die Kontraste in diesem Bereich wurden reduziert (siehe Abbildung 14).

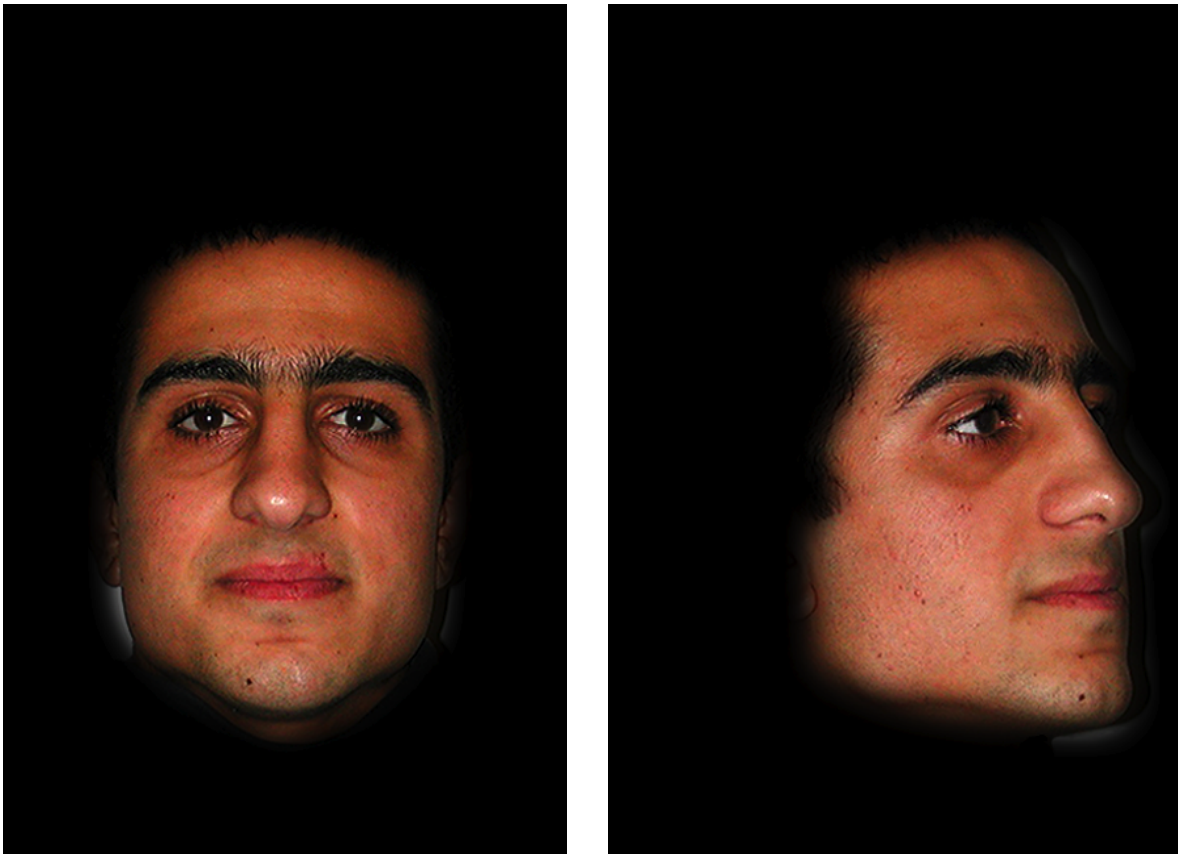


Abbildung 14. Beispiel für die Maskierung eines Gesichtes.

Die Prozedur für die Retuschierung äußerer Gesichtsmarkmale (Haare, Kinn, Ohren etc.) war folgende. Mit dem Programm Photoshop 5.5 wurden die vollständigen Originalbilder, sowohl Frontalansicht als auch im rechten Halbprofil, geöffnet. Das Gesichtsinere wurde mit der Option 'magnetic lasso tool' dergestalt markiert, dass die Markierung entlang der Grenze zwischen Gesichtsinerem und den äußeren Merkmalen (Kinn, Ohren und Haare) verlief. Das tool wurde dabei an einem Punkt dieser Grenze angesetzt und entlang der Grenze mittels Setzen von sog. Ankern an Stellen an denen keine gerade Führung der Markierung möglich ist, entlanggeführt. Dadurch wurde eine Fläche erhalten, die das Gesichtsinere umschließt. Dieser markierte Bereich konnte dann durch

die 'Layer' Leiste als eigene Ebene definiert werden (Dadurch ist es bei Photoshop möglich, einzelne Teile eines Bildes separat zu bearbeiten, ohne dabei andere Ebenen oder Bereiche, die unangetastet bleiben sollen, zu verändern). Dies geschah durch die Option 'Layer via copy' aus der Leiste 'Layer'. Durch den Befehl 'feather' unter der Leiste 'Select' wurde der Grenzbereich zwischen Gesichtsinnerem und Umgebung weichgezeichnet, um dem Bild ein möglichst natürliches Aussehen zu geben. Damit ist der Übergang zwischen dem Grenzbereich des Gesichtes und dem Hintergrund fließend. Anschließend wurde unter dem Menüpunkt 'Select' die Option 'Select all' angewählt, um das gesamte Bild einschließlich des Hintergrundes zu markieren. Der Hintergrund des Gesichtes sollte für alle Bilder einheitlich sein und bei der Darbietung der Gesichter auf dem Computer nicht vom Gesicht ablenken. Dazu musste noch der Hintergrund schwarz eingefärbt werden, was mit der Option 'paint bucket tool' durchgeführt wurde. Als Hintergrund wurde die Farbe 'Schwarz' mit einer Deckkraft von 100 % und dem Modus normal ausgewählt. Damit war das Gesichtsinne ausgeschnitten und konnte als separate Datei gespeichert werden.

4.2.2 Aufteilung der Stimuli zu den Darbietungssets

Die jeweils 80 Gesichter von Afro-Amerikanern, kaukasischen Amerikanern, Türken und Deutschen mussten anhand der Ratings aus der Pilotstudie für die türkischen und deutschen Gesichter bzw. von Meissner und Brigham (in Druck) für die afro-amerikanischen und kaukasisch-amerikanischen Gesichter in die beiden Darbietungssets aufgeteilt werden. Die Ratings der Skalen *Erinnerbarkeit* und *Distinktivität* wurden zu einem Mittelwert zusammengefasst und die Bilder getrennt für jede Ethnie anhand dieses Mittelwerts sortiert. Im Anschluss daran wurden Paare gebildet, wobei jeweils zwei bezüglich des Mittelwerts aufeinanderfolgende Bilder ein Paar bilden, und diese Paare wurden zufällig auf die zwei Darbietungssets aufgeteilt. Die Mittelwerte und die Zugehörigkeit jedes einzelnen Bildes zu den Darbietungssets sind in Anhang C1 – C4 aufgeführt.

4.2.3 Stimulus-Darbietung

Die Daten wurden an zwei Apple 'iMac' Computern mit einer Taktfrequenz von 300 bzw. 350 Mhz, einem Arbeitsspeicher von jeweils 192 MB und einem 15"-Monitor

erhoben. Die Farbtiefe betrug 16 Bit (65536 Farben) bei einer Bildschirmauflösung von 1024x768 Pixel und einer Bildwiederholfrequenz von 75 Hz. Die Programmierung erfolgte mit der Applikation 'Superlab' (Version 1.75) der Firma Cedrus (<http://www.cedrus.com>). Die Stimuli hatten eine Größe von 432x648 Pixel bei einer Auflösung von 72x72 dpi. Die Farbtiefe der Stimuli betrug 16,7 Mio. Farben.

Da ein einzelnes Bild eine Größe zwischen 600 kb und 900 kb hatte, und der Arbeitsspeicher der Computer 192 MB betrug, konnten nicht alle Gesichter gleichzeitig in den Arbeitsspeicher geladen werden.¹⁴ Daher wurden die Gesichter in zwei Sequenzen dargeboten. Die erste Sequenz umfasste jeweils 40 Gesichter von Türken und Deutschen, die zweite Sequenz jeweils 40 Gesichter von Afro-Amerikanern und kaukasischen Amerikanern. In der Testphase wurden dann die gleichen Sequenzen mit der doppelten Anzahl an Stimuli (Sets A und B) geladen. Die Reihenfolge der Sequenzen wurde experimentell ausbalanciert (siehe Abschnitt 4).

4.2.4 Versuchsdurchführung

Die Daten wurden am Institut für Psychologie der Philipps-Universität Marburg in der Zeit von Mitte März 2003 bis Ende Mai 2003 erhoben.¹⁵ Dies geschah aus dem Grund, dass die Wahrscheinlichkeit reduziert werden sollte, dass Teilnehmer an diesem Experiment die Personen auf den Bildern kannten, bzw. früher bereits einmal gesehen hatten. Die Zuordnung der Teilnehmer zu den Versuchsbedingungen sollte randomisiert erfolgen. Dies war für die Bedingung 'Gesichtsausschnitt' unproblematisch, bei der Bedingung 'Zeitintervall' war die Zuordnung jedoch von der Bereitschaft bzw. der Zeitplanung der Probanden abhängig, je nachdem, ob sie nach drei Wochen wiederkommen konnten. Sie wurden deshalb ggf. entsprechend ihren Wünschen in dieser Bedingung eingeteilt.

¹⁴ Bei Apple Macintosh Computern mit dem Betriebssystem 9.2.2 wird jedem Programm ein begrenzter Umfang an Arbeitsspeicher zugewiesen. Dieser wurde bei den benutzten Computern manuell auf 120 MB festgelegt, um zu gewährleisten, dass in der Testphase 160 Bilder problemlos in den Arbeitsspeicher geladen werden können. Das Betriebssystem benutzt je nach Erweiterungen zwischen 20 und 30 MB des Arbeitsspeichers.

¹⁵ Ich danke an dieser Stelle Prof. Dr. Ulrich Wagner von der Philipps-Universität Marburg für die Bereitstellung der Räume für die Untersuchung.

4.2.5 Versuchsablauf

Die Teilnehmer wurden begrüßt und nahmen vor einem der beiden Computer Platz. Zunächst wurde den Teilnehmern die Instruktionen auf dem Bildschirm präsentiert (siehe Anhang A5). Anschließend folgte je nach Sequenz die Darbietungsphase mit jeweils 40 Gesichtern türkischer und deutscher Gesichter bzw. jeweils 40 Gesichtern afro-amerikanischer und kaukasisch-amerikanischer Gesichter, die in randomisierter Reihenfolge jeweils 5000 ms dargeboten wurden (insgesamt 160 verschiedene Gesichter). Die Gesichter waren, je nach Bedingung, entweder vollständig oder auf das Gesichtsinere beschränkt. Das Inter-Trial-Intervall betrug 500 ms. Bevor das erste Gesicht auf dem Bildschirm erschien, wurden ein weißes Quadrat (Ⓜ), ein weißer Kreis (⊙) und ein liegendes weißes Kreuz (×), ebenfalls jeweils für 5000 ms und einem Inter-Trial-Intervall von 500 ms dargeboten, um den Blick der Teilnehmer auf die Bildschirmmitte zu lenken, wo dann im Anschluss an das weiße Kreuz die Gesichter erschienen. Die drei Zeichen hatten die Abmessung (B x H) von 2,8 Inch x 2,7 Inch und waren, wie die Stimulus-Gesichter, auf dem Bildschirm zentriert.

Nach der Darbietungsphase sollten die ersten dreizehn Items des Kontaktfragebogens beantwortet werden. Dieser Teil der Sitzung nahm, je nach Bearbeitungsgeschwindigkeit ca. fünf bis zehn Minuten in Anspruch. Bei Teilnehmern in der Bedingung der unmittelbaren Testung folgte der Wiedererkennenstest im Anschluss an die Bearbeitung des Kontaktfragebogens. Die letzten Items des Fragebogens erfragten die subjektive Wahrscheinlichkeitseinschätzung der Teilnehmer, wie viele der Gesichter sie glaubten, wiedererkennen zu können. Teilnehmer in der Bedingung der Testung nach einem Intervall von drei Wochen beantworteten diese Fragen bevor sie den Wiedererkennenstest durchführten.

In der Testphase wurden die Teilnehmer wiederum gebeten, den Instruktionen auf dem Bildschirm zu folgen. Je nach Sequenz wurden dann jeweils 160 Gesichter von Türken und Deutschen bzw. jeweils 160 Gesichter von Afro-Amerikanern und kaukasischen Amerikanern in randomisierter Reihenfolge dargeboten. Die Platzierung und Größe des Bildes auf dem Bildschirm erfolgte wie in der Darbietungsphase. Die Teilnehmer sollten durch Drücken der Taste 'c' bzw. 'm' angeben, ob sie das jeweilige Gesicht in der Darbietungsphase gesehen hatten oder nicht. Es gab keine Zeitbegrenzung für die Beurteilung der einzelnen Bilder. Nachdem beide Sequenzen bearbeitet wurden, erfolgte nochmals eine Befragung der Teilnehmer, wie sie den Prozentsatz richtig

wiedererkannter Gesichter einschätzen. Im Anschluss an diese letzten Fragen endete das Experiment und der Teilnehmer wurde verabschiedet.

4.2.6 Stichprobe

Teilnehmer waren $N = 128$ Personen, hauptsächlich Studierende an der Philipps-Universität Marburg. Davon waren je $n = 64$ türkischer bzw. deutscher Herkunft. Die Hälfte der Teilnehmer war männlichen, die andere Hälfte weiblichen Geschlechts. Der Altersmedian betrug $Mdn = 23$ (männliche Teilnehmer $Mdn = 24$, weibliche Teilnehmer $Mdn = 21$). Der jüngste Teilnehmer war 18, der älteste 34 Jahre alt. Die beiden untersuchten Gruppen unterscheiden sich in ihrem Alter nicht ($Mdn_{tü} = 23$, $Mdn_{de} = 23$). Die Teilnehmer erhielten entweder, je nach Versuchsbedingung eine bis zwei Versuchspersonenstunden oder acht bzw. 15 Euro nach Beendigung des Versuches ausbezahlt.

4.3 Auswertung

Die Auswertung erfolgte nach der Signal-Entdeckungs-Theorie mit der Berechnung des Anzahl der Treffer, der falschen Alarme, des Leistungsindex A' und der Reaktionsneigung B'' (siehe Abschnitt 1.2.1.1) Zunächst werden die Resultate der Analyse der Treffer und der falschen Alarme präsentiert, anschließend werden die Ergebnisse des Leistungsindex A' und der Reaktionsneigung B'' dargestellt. Das Alpha-Niveau wird nach Bonferroni-Adjustierung bei sechs geplanten Vergleichen auf $\alpha = .008$ festgelegt.

4.4 Ergebnisse

Zunächst wird die Analyse der Treffer dargestellt. Die Hypothese, nach der bei einem Zeitintervall von drei Wochen zwischen Lern- und Testphase mehr Treffer bei Eigen-Gruppen Gesichtern zu beobachten sind, wurde getestet, indem die Differenz der Treffer zwischen Eigen-Gruppen und Fremd-Gruppen Gesichtern an beiden Zeitintervallen berechnet und mithilfe dieser Werte ein Kontrast aufgestellt wurde. Dieser Kontrast zeigte sich als nicht signifikant von Null verschieden, $t(126) = -.671$, ns. Bei den Eigen-Gruppen

Gesichtern gibt es nicht signifikant mehr Treffer als bei Fremd-Gruppen Gesichtern nach einem Zeitintervall von drei Wochen, verglichen mit der sofortigen Testung.

Anschließend wurde die Anzahl der Treffer mittels einer Varianzanalyse mit den Between-Subjects Faktoren 'Ethnie der Versuchsperson' (Türken, Deutsche), 'Geschlecht' (weiblich, männlich), 'Behaltensintervall' (unmittelbar, drei Wochen), 'Gesichtsausschnitt' (Maske, vollständig) sowie dem Within-Subjects Faktor 'Stimuluse ethnien' (Afro-Amerikaner, kaukasische Amerikaner, Türken, Deutsche) analysiert. Die Anzahl der Treffer ist je nach Stimuluse ethnien signifikant unterschiedlich, $F(3,336) = 15.81$, $p = .001$, $\eta^2 = .29$. Die meisten Treffer wurden bei den türkischen Gesichtern erzielt ($M = .522$), gefolgt von den afro-amerikanischen ($M = .469$), den deutschen ($M = .450$) und den kaukasisch-amerikanischen Gesichtern ($M = .429$), (siehe Tabelle 13). Die türkischen Gesichter unterscheiden sich signifikant von den afro-amerikanischen Gesichtern, $F(1,336) = 27.55$, $p = .001$, $r = .28$, den deutschen Gesichtern, $F(1,336) = 50.84$, $p = .001$, $r = -.36^{16}$, und den kaukasisch-amerikanischen Gesichtern, $F(1,336) = 84.83$, $p = .001$, $r = -.45$). Die afro-amerikanischen und die kaukasisch-amerikanischen Gesichter unterscheiden sich ebenfalls signifikant, $F(1,336) = 15.96$, $p = .001$, $r = -.21$. Türkische und deutsche Versuchspersonen unterscheiden sich in der Gesamtzahl an Treffern nicht, $F < 1$, *ns*.

Die beiden Versuchspersonen-Ethnien haben je nach Stimuluse ethnien unterschiedlich viele Treffer, wie die Interaktion der Faktoren 'Ethnie der Versuchsperson' und 'Stimuluse ethnien' zeigt, $F(3,336) = 3.06$, $p = .029$, $\eta^2 = .08$. Die deutschen Versuchspersonen haben bei den Paarvergleichen zwischen den Stimuluse ethnien keine signifikanten Differenzen (alle $F_s < 2.45$, alle $p_s > .123$), während bei den türkischen Versuchspersonen signifikante Differenzen zwischen den kaukasisch-amerikanischen und den türkischen Gesichtern, $F(1,336) = 13.96$, $p = .001$, $r = .20$, und zwischen den türkischen und deutschen Gesichtern bestehen, $F(1,336) = 8.15$, $p = .005$, $r = .15$.

Die Maskierung der Gesichter hat keinen signifikanten Einfluss auf die Zahl der Treffer, $F(1,112) = 1.46$, $p = .230$, $r = .11$, ebensowenig wie das Behaltensintervall, bei dem jedoch ein Trend in Richtung mehr Treffer bei unmittelbarer Testung zu erkennen ist, $F(1,112) = 3.68$, $p = .058$, $r = .18$. ($M_{\text{unm}} = .492$, $M_{\text{3wo}} = .443$). Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 2.57$, alle $p_s > .054$).

¹⁶ Vergleiche, bei denen eine Ethnie, die weniger Treffer bzw. eine höhere Leistung A' aufweisen als die Vergleichsethnie, erhalten bei der Effektstärke r ein negatives Vorzeichen.

In der Analyse der falschen Alarme mittels einer Varianzanalyse mit den Between-Subjects Faktoren 'Ethnie der Versuchsperson' (Türken, Deutsche), 'Geschlecht' (weiblich, männlich), 'Behaltensintervall' (unmittelbar, drei Wochen), 'Gesichtsausschnitt' (Maske, vollständig) sowie dem Within-Subjects Faktor 'Stimuluse ethnien' (Afro-Amerikaner, kaukasische Amerikaner, Türken, Deutsche) zeigte sich lediglich ein signifikanter Haupteffekt für den Faktor 'Stimuluse ethnien' mit $F(3,336) = 13.50$, $p = .001$, $\eta^2 = .26$. Die wenigsten falschen Alarme gab es bei den deutschen Gesichtern ($M = .321$), gefolgt von den kaukasisch-amerikanischen ($M = .331$), den türkischen ($M = .383$) und den afro-amerikanischen ($M = .399$) Gesichtern. Dabei unterscheiden sich die Werte zwischen afro-amerikanischen und kaukasisch-amerikanischen Gesichtern, $F(1,336) = 42.89$, $p = .001$, $r = .34$, zwischen afro-amerikanischen und deutschen Gesichtern, $F(1,336) = 56.43$, $p = .001$, $r = .38$, zwischen kaukasisch-amerikanischen und türkischen, $F(1,336) = 25.08$, $p = .001$, $r = .26$, sowie zwischen türkischen und deutschen Gesichtern signifikant voneinander, $F(1,336) = 35.65$, $p = .001$, $r = .31$ (siehe Tabelle 13). Es zeigte sich keine Interaktion der Faktoren 'Ethnie der Versuchsperson' und 'Stimuluse ethnien', $F(3,336) = .10$, $p = .961$, $\eta^2 = .001$. Alle anderen Haupteffekte und Interaktionen sind nicht signifikant (alle $F_s < 2.59$, alle $p_s > .110$).

Tabelle 13

Mittelwerte der Treffer, der falschen Alarme, des Leistungsindex A' und der Reaktionsneigung B'' nach Vp-Gruppe und Stimulusethnie

Gesichter von	Türken	Deutsche	<u>M</u>
Afro-Amerikanern:			
Treffer	.477	.461	.469
Falsche Alarme	.384 ^a	.415 ^a	.399
<u>A'</u>	.586 ^a	.546 ^a	.566
<u>B''</u>	.235	.168	.202
Türken:			
Treffer	.539 ^a	.505 ^a	.522
Falsche Alarme	.374	.391	.383
<u>A'</u>	.635 ^a	.599 ^a	.617
<u>B''</u>	.159	.168	.164
Kauk.-Amerikanern:			
Treffer	.407 ^a	.450 ^a	.429
Falsche Alarme	.320	.343	.331
<u>A'</u>	.586	.603	.595
<u>B''</u>	.412 ^a	.313 ^a	.363
Deutschen:			
Treffer	.438	.462	.450
Falsche Alarme	.306 ^a	.336 ^a	.321
<u>A'</u>	.627	.612	.620
<u>B''</u>	.395	.320	.358
<u>M:</u>			
Treffer	.465	.469	.467
Falsche Alarme	.346	.371	.359
<u>A'</u>	.609	.590	.600
<u>B''</u>	.300	.240	.270

Anmerkung: ^a Die Zellenmittelwerte innerhalb einer Zeile unterscheiden sich nach Bonferroni-Adjustierung signifikant (post-hoc Vergleiche, $p < .001$).

Tabelle 14

Mittelwerte der Treffer, der falschen Alarme, des Leistungsindex \underline{A}' und der Reaktionsneigung \underline{B}'' getrennt nach Maskierung und Stimulusehnie

Ansicht	Afro- Am.	Türken	Kauk.- Am	Deutsche	<u>M:</u>
Inneres Gesicht:					
Treffer	.452 ^{ab}	.500 ^{acd}	.424 ^{abc}	.434 ^d	.453
Falsche Alarme	.426 ^{abc}	.393 ^{adf}	.364 ^{bde}	.337 ^{cef}	.380
\underline{A}'	.524 ^{abc}	.588 ^a	.563 ^{bd}	.594 ^{cd}	.567
\underline{B}''	.185 ^{ab}	.166 ^{cd}	.304 ^{ac}	.349 ^{bd}	.251
Inneres + äußeres Gesicht:					
Treffer	.487 ^{abc}	.545 ^{ade}	.433 ^{bd}	.464 ^{ce}	.482
Falsche Alarme	.373 ^{ab}	.372 ^{cd}	.298 ^{bc}	.305 ^{bd}	.337
\underline{A}'	.608 ^{ab}	.645 ^a	.626	.646 ^b	.631
\underline{B}''	.219 ^{abc}	.162 ^{ade}	.421 ^{ad}	.367 ^{ce}	.292
<u>M:</u>					
Treffer	.469 ^{ab}	.522 ^{acd}	.429 ^{bc}	.450 ^d	.468
Falsche Alarme	.399 ^{ab}	.383 ^{cd}	.331 ^{bc}	.321 ^{bd}	.359
\underline{A}'	.566 ^{ab}	.617 ^a	.595	.620 ^b	.600
\underline{B}''	.202 ^{ab}	.164 ^{cd}	.363 ^{bc}	.358 ^{bd}	.272

Anmerkung: ^{abcd} Die Zellenmittelwerte innerhalb einer Zeile mit gleichen Subskripta unterscheiden sich nach Bonferroni-Adjustierung signifikant (post-hoc Vergleiche, $p < .001$).

Im folgenden werden Tests der Hypothesen zum Leistungsindex \underline{A}' dargestellt. Zur Überprüfung der Hypothese schlechterer Wiedererkennensleistungen nach einem Zeitintervall von drei Wochen wurde ein einfacher Kontrast, d.h. ein Haupteffekt 'Behaltensintervall,' berechnet. Dieser wurde marginal signifikant mit $F(1,112) = 5.14$, $p = .025$, $r = .21$. \underline{A}' beträgt bei sofortiger Testung $\underline{A}' = .614$ und bei Testung nach drei Wochen $\underline{A}' = .585$. Der Faktor 'Maske' sollte mit besseren Leistungen bei vollständigen Gesichtern im Gegensatz zu maskierten Gesichtern einen Einfluß auf die Wiedererkennensleistung haben. Zum Testen dieser Hypothese wurde ein einfacher Kontrast berechnet, der signifikant von null verschieden ist, $F(1,112) = 25.59$, $p = .001$, $r = .43$. Bei vollständigen Gesichtern beträgt $\underline{A}' = .631$, bei maskierten Gesichter ist $\underline{A}' = .567$. Für den Einfluß der Maskierung auf die Wiedererkennensleistung bei Fremd-Gruppen und Eigen-Gruppen Gesichtern wurde eine Interaktion der Faktoren 'Maske' 'Ethnie der Teilnehmer' und 'Stimulusehnie' postuliert, die aber nicht signifikant war, $F < 1$, *ns.*

Nach einem Zeitintervall von drei Wochen zwischen Lern- und Testphase sollte ebenfalls die Wiedererkennensleistung bei Eigen-Gruppen Gesichtern besser sein als bei Fremd-Gruppen Gesichtern. Diese Hypothese wurde so getestet, wie die Hypothese über die Treffer nach einem Zeitintervall von drei Wochen. Die Differenz von \underline{A}' zwischen

Eigen-Gruppen und Fremd-Gruppen Gesichtern an beiden Zeitintervallen wurde berechnet und mithilfe dieser Werte ein Kontrast aufgestellt. Dieser Kontrast wurde marginal signifikant, $t(126) = -1.924$, $p = .057$, allerdings mit umgekehrten Vorzeichen. Bei sofortiger Testung beträgt die Differenz von A' zwischen Eigen-Gruppen und Fremd-Gruppen Gesichtern $A'_{Diff} = .05$, $SD_{Diff} = .099$, bei Testung nach drei Wochen ist diese Differenz geringer mit $A'_{Diff} = .014$, $SD_{Diff} = .12$.

Die Hypothese, nach der die Wiedererkennensleistung der vier Stimulusethnien je nach Ethnie der Teilnehmer linear in Abhängigkeit der Distanz von der eigenen Ethnie ansteigt wurde mithilfe linearer Kontraste getrennt für die beiden Teilnehmerethnien getestet. Für die türkischen Teilnehmer konnte diese Hypothese bestätigt werden, $F(1,56) = 12.15$, $p = .001$, $r = .42$, und für die deutschen Teilnehmer ebenfalls $F(1,56) = 12.73$, $p = .001$, $r = .43$. Beide Teilnehmer-Ethnien waren bei der Eigen-Gruppe am besten (siehe Abbildung 16) und bei afro-amerikanischen Gesichtern am schlechtesten.

Der Leistungsindex A' wurde im Anschluß an die Überprüfung der Hypothesen ebenfalls mit einer Varianzanalyse mit den Between-Subjects Faktoren 'Ethnie der Versuchsperson' (Türken., Deutsche), 'Geschlecht' (weiblich, männlich), 'Behaltensintervall' (unmittelbar, 3 Wochen), 'Gesichtsausschnitt' (maskiert, nicht maskiert) sowie dem Within-Subjects Faktor 'Stimulusethnie' (Afro-Amerikaner, kaukasische Amerikaner, Türken, Deutsche) analysiert. Dabei zeigte sich ein Haupteffekt für den Faktor 'Stimulusethnie', $F(3,336) = 8.44$, $p = .001$, $\eta^2 = .18$. Die Teilnehmer hatten bei afro-amerikanischen Gesichtern die schlechtesten Leistungen verglichen mit den anderen Ethnien ($M_{A-A} = .566$, $M_{K-A} = .595$, $M_{Tü} = .614$, $M_{DE} = .620$) (siehe Abbildung 15).

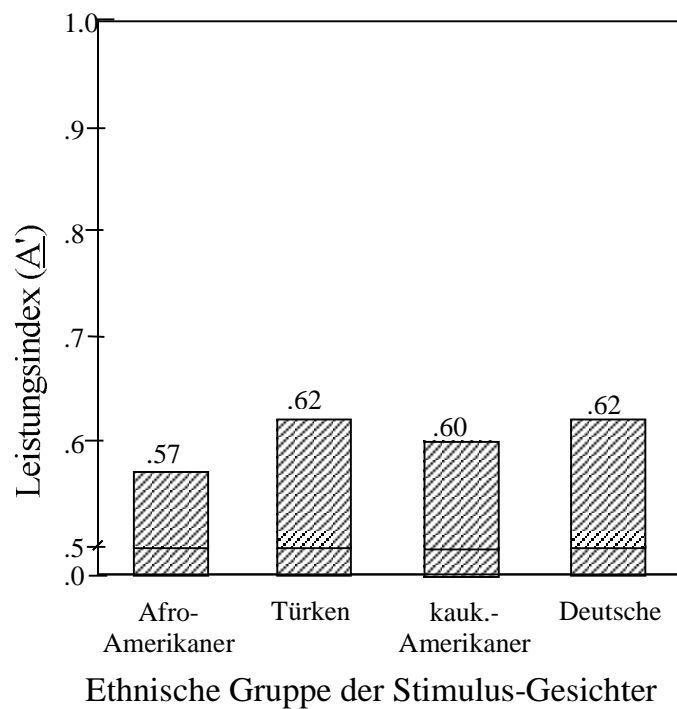


Abbildung 15. Leistungsindex \underline{A}' bei den verschiedenen Stimulusethnien

Der Wert von \underline{A}' der afro-amerikanischen Gesichter unterscheidet sich signifikant von den türkischen, $F(1,336) = 7.13$, $p = .008$, $r = .14$, und deutschen Gesichtern, $F(1,336) = 8.02$, $p = .005$, $r = .15$. Türkische und Deutsche Teilnehmer unterscheiden sich nicht voneinander, $F(1,112) = 2.12$, $p = .149$, $r = .14$.

Die Interaktion der Faktoren 'Ethnie der Versuchsperson' und 'Stimulusethnie' wurde marginal signifikant mit $F(3,336) = 2.27$, $p = .080$, $\eta^2 = .02$ (siehe Abbildung 16).

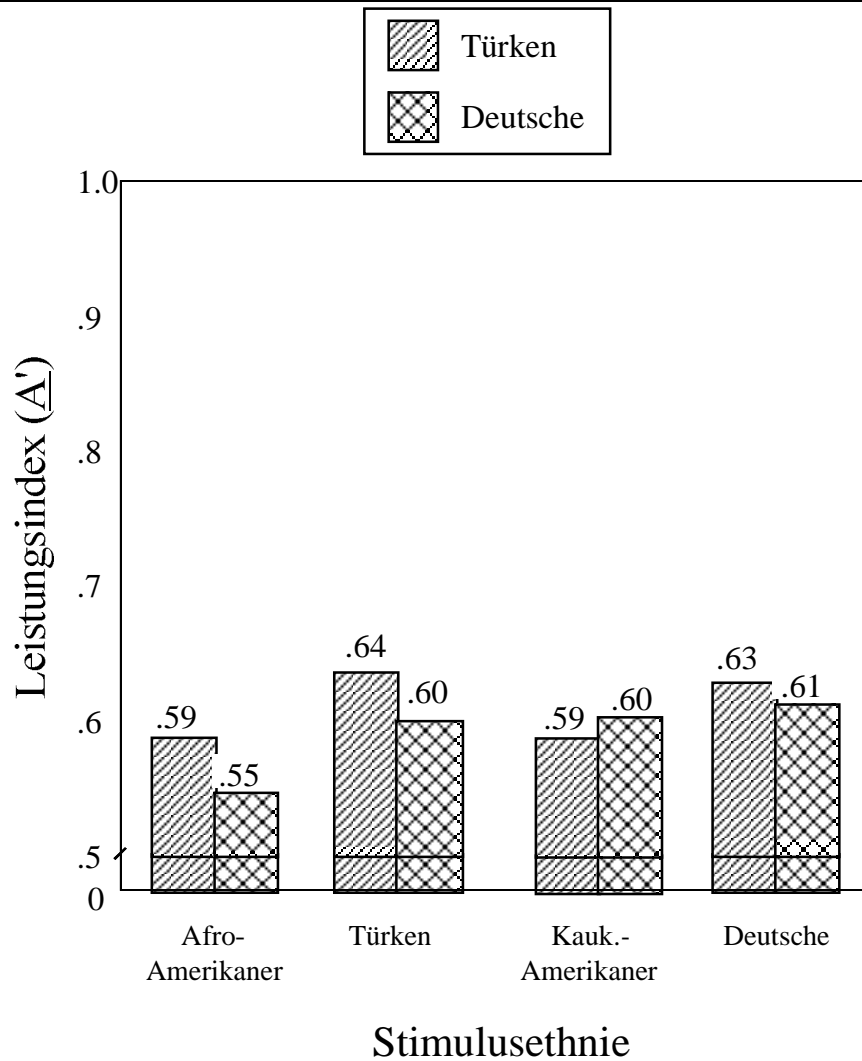


Abbildung 16. Interaktion der Faktoren 'Stimulusehnie' und 'Ethnie der Versuchspersonen' beim Leistungsindex $\underline{A'}$

Der Wert von $\underline{A'}$ der afro-amerikanischen Gesichter unterscheidet sich signifikant zwischen den deutschen und türkischen Teilnehmern, $F(1,336) = 19.67$, $p = .001$, $r = .24$, ebenso wie $\underline{A'}$ bei den türkischen Gesichtern $F(1,336) = 16.88$, $p = .001$, $r = .22$. $\underline{A'}$ der kaukasisch-amerikanischen, $F(1,336) = 3.52$, $p = .062$, $r = .10$ und deutschen Gesichter, $F(1,336) = 2.76$, $p = .098$, $r = .09$, unterscheiden sich hingegen zwischen den beiden Teilnehmergruppen nicht. Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle F s < 2.57, alle p s > .054).

Die Reaktionsneigung $\underline{B''}$ wurde ebenfalls mithilfe einer Varianzanalyse mit den Between-Subjects Faktoren 'Ethnie der Versuchsperson' (Türken, Deutsche), 'Geschlecht' (weiblich, männlich), 'Behaltensintervall' (unmittelbar, drei Wochen), 'Gesichtsausschnitt' (Maske, vollständig) sowie dem Within-Subjects Faktor 'Stimulusehnie' (Afro-

Amerikaner, kaukasische Amerikaner, Türken, Deutsche), inferenzstatistisch getestet. Dabei zeigten sich bis auf den Haupteffekt 'Stimulusethnie', $F(3,336) = 13.26$, $p = .001$, $\eta^2 = .11$, keine signifikanten Haupteffekte und Interaktionen. Die Teilnehmer zeigten bei den kaukasischen Amerikanern und den deutschen Stimulusgesichtern ein eher striktes Antwortkriterium, verglichen mit den afro-amerikanischen und türkischen Gesichtern. Dabei unterscheiden sich die afro-amerikanischen Gesichter von den kaukasisch-amerikanischen, $F(1,336) = 31.96$, $p = .001$, $r = .30$, und den deutschen Gesichtern, $F(1,336) = 30.00$, $p = .001$, $r = .29$ signifikant. Ebenso unterscheiden sich die türkischen Gesichter von den kaukasisch-amerikanischen, $F(1,336) = 48.83$, $p = .001$, $r = .35$, und den deutschen Gesichtern, $F(1,336) = 46.41$, $p = .001$, $r = .35$ signifikant (siehe Tabelle 14). Andere Haupteffekte oder Interaktionen zeigten sich nicht, alle $F_s < 2.44$, alle $p_s > .064$.

Im Anschluß an die Analyse der einzelnen Stimulusethnien wurden die Daten für die jeweiligen Fremd-Gruppen der beiden Teilnehmergruppe zusammengefaßt, d.h. für die türkischen Teilnehmer wurden die Daten der afro-amerikanischen, der kaukasisch-amerikanischen und der deutschen Gesichter zusammengefaßt. Für die deutschen Teilnehmer bedeutete dies analog eine Akkumulation der Daten der Fremd-Gruppen Gesichter über die afro-amerikanischen, kaukasisch-amerikanischen und türkischen Gesichter. Dieser Schritt wurde vollzogen, um zu testen, ob sich ein Cross-Race Bias finden läßt, wenn erstens die Eigen-Gruppe gegen die Fremd-Gruppen zusammen verglichen wird und zweitens, um zu testen ob sich ein Cross-Race Bias je nach Maskierung und Zeitintervall finden läßt. Dazu wurden getrennt nach Testzeitpunkt und Maskierung der Gesichter separate Varianzanalysen mit den Between-Subjects Faktoren 'Ethnie der Teilnehmer' (Türken, Deutsche) und 'Geschlecht' (weiblich, männlich) sowie dem Within-Subjects Faktor 'Stimulusethnie' (Fremd-Gruppen, Eigen-Gruppe) berechnet. Die Mittelwerte der Treffer, falschen Alarme, A' und B'' sind in Tabelle 15 angegeben.

Tabelle 15

Mittelwerte der Treffer, der falschen Alarme, des Leistungsindex A' und der Reaktionsneigung B'' getrennt nach Vp-Gruppe Zeitintervall, Ausschnitt und Stimulusethnie

Teilnehmer	Zeitintervall												
	unmittelbar						3 Wochen						
	Eigen		Fremd		Fremd		Eigen		Maske		Vollständig		
	Maske	Vollständig	Maske	Vollständig	Maske	Vollständig	Maske	Vollständig	Maske	Vollständig	Maske	Vollständig	M
Türken													
Treffer	.539	.567	.423	.458	.505	.547	.445	.437	.490				
Falsche Alarme	.361	.369	.362	.290	.413	.353	.373	.322	.355				
A'	.643	.664	.559	.652	.576	.656	.573	.615	.617				
B''	.012	.076	.047	.124	.031	.098	.074	.117	.072				
Deutsche													
Treffer	.484	.534	.505	.527	.380	.450	.403	.453	.467				
Falsche Alarme	.344	.317	.428	.377	.340	.344	.378	.348	.359				
A'	.632	.669	.566	.626	.547	.602	.529	.610	.598				
B''	.131	.046	.010	.027	.042	.130	.034	.103	.065				
M:													
Treffer	.512	.551	.464	.492	.442	.498	.424	.445	.479				
Falsche Alarme	.352	.343	.395	.333	.376	.349	.376	.335	.357				
A'	.637	.667	.563	.639	.561	.629	.551	.612	.607				
B''	.072	.061	.029	.076	.037	.114	.054	.110	.069				

Es zeigten sich bei unmittelbarer Testung und maskierten Gesichtern in der Testphase signifikant bessere Leistungen bei Gesichtern der Eigen-Gruppe verglichen mit Gesichtern der Fremd-Gruppen dadurch, dass der Faktor 'Stimulusethnie' signifikant war, $F(1,28) = 22.13$, $p = .001$, $r = .66$ ($A'_{E-G} = .637$, $A'_{F-G} = .563$). Die bessere Leistung bei Eigen-Gruppen Gesichtern gegenüber Fremd-Gruppen Gesichtern trifft in dieser Bedingung für beide Teilnehmergruppen zu ($A'_{E-G/Tü} = .643$, $A'_{F-G/tü} = .559$; $A'_{E-G/D} = .632$, $A'_{F-G/D} = .566$). Bei sofortiger Testung und vollständigen Gesichtern, sowie bei Testung nach drei Wochen und maskierten Gesichtern und Testung nach drei Wochen und vollständigen Gesichtern zeigten sich keine signifikanten Haupteffekte des Faktors 'Stimulusethnie', alle $F_s < 2.33$, ns. Die Interaktion der Faktoren 'Geschlecht' und 'Stimulusethnie' bei Testung nach drei Wochen sowie vollständigen Gesichtern wurde signifikant mit $F(1,28) = 5.38$, $p = .028$, $r = .40$. Während sich bei Fremd-Gruppen Gesichtern weibliche und männliche Teilnehmer nicht unterscheiden ($A'_{weibl} = .610$, $A'_{männl} = .615$), haben die weiblichen Versuchspersonen bei Eigen-Gruppen Gesichtern eine signifikant schlechtere Leistung ($A' = .577$) als die männlichen Versuchspersonen, ($A' = .681$), $F(1,28) = 93.60$, $p = .001$, $r = .88$. Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 3.24$, alle $p_s > .083$).

4.5 Diskussion

Die Zielsetzung dieses Experiments war, potentielle Einflussfaktoren auf den Cross-Race Bias zu untersuchen. Ein Cross-Race Bias wurde gefunden, beide Teilnehmer-Gruppen konnten die Eigen-Gruppen Gesichter am besten wiedererkennen, mit mittleren Effektstärken von, $r = .42$ bei türkischen Teilnehmern bzw. $r = .43$ bei deutschen Teilnehmern. Wie vermutet zeigte sich ein mittlerer Effekt des Zeitintervalls zwischen Lern- und Testphase, $r = .25$. Nach einem Intervall von drei Wochen waren die Versuchsteilnehmer schlechter als bei sofortiger Testung. Dies ist ein Befund, der in vielen Bereichen der Psychologie bereits erforscht wurde und daher nicht überraschend ist. Von Bedeutung ist allerdings eine fehlende dreifach-Interaktion der Faktoren 'Behaltensintervall', 'Stimulusethnie' und 'Ethnie der Teilnehmer'. Eigentlich sollte nach einem dreiwöchigen Zeitintervall, wegen einer holistischen Verarbeitung der Eigen-Gruppen Gesichter der Cross-Race Bias noch ausgeprägter sein als bei sofortiger Testung. Dass dies nicht der Fall ist, könnte mehrere Gründe haben. Zum einen besteht die Möglichkeit, dass die große Anzahl der Gesichter dafür ausschlaggebend ist, dass die

Wiedererkennensleistung insgesamt sehr niedrig ist und nahe dem Zufallsniveau liegt, schließlich wurden in der Darbietungsphase 160 Gesichter und in der Testphase 320 Gesichter gezeigt. Eventuell waren die Teilnehmer nicht in Lage, diese hohe Zahl an Gesichtern angemessen zu enkodieren. Wie bei Wright und Sladden (2003) wurden Gesichter, bei denen lediglich das Gesichtsinere sichtbar war, schlechter wiedererkannt als vollständige Gesichter. Hier wurde eine mittlere Effektstärke von $r = .43$ gefunden. Dies ist als Hinweis darauf zu werten, dass für das Wiedererkennen von Gesichtern, zumindest von unbekanntem vs. bekannten Personen (vgl. Bruce, 1988; Sporer, 1992), äußere Merkmale von besonderer Bedeutung sind. Die Speicherung von Gesichtern anhand einzelner Merkmale scheint bei neuen Gesichtern eine einfachere Strategie zu sein, als sich diese Gesichter holistisch einzuprägen. Selbstverständlich leidet die Fähigkeit zum Wiedererkennen, wenn diese Merkmale fehlen. Entgegen der Vermutung, dass Bilder mit lediglich dem Gesichtsineren vor allem bei Gesichtern der Fremd-Gruppen zu schlechteren Leistungen führen, fanden sich auch bei Gesichtern der Eigen-Gruppe solche Nachteile im Vergleich zu vollständigen Gesichtern. Bei genauerer Betrachtung der Interaktionen zwischen 'Stimulusethnie' und 'Ethnie der Teilnehmergruppen' je nach Testzeitpunkt, zeigt sich jedoch ein interessanteres Bild, das diese Befunde vereinbaren kann. Der Cross-Race Bias tritt in diesem Experiment nur dann auf, wenn die Testung unmittelbar geschieht und die Bilder maskiert sind und zwar mit einem starken Effekt von $r = .66$. Bei den anderen drei Faktorkombinationen (drei Wochen – Maske, drei Wochen – vollständig, sofort – vollständig) ist kein Cross-Race Bias zu beobachten (siehe Tabelle 15). Allerdings könnte es hierbei problematisch sein, bei deutschen Teilnehmern Gesichter von kaukasischen Amerikanern (Weiße) als Stimulusmaterialien zu verwenden bzw. zu den Fremd-Gruppen Gesichtern zu zählen, weil sich beide Ethnien sehr ähnlich sind. Bei sofortiger Testung und vollständigem Gesicht sind die Leistungen bei Eigen-Gruppen und Fremd-Gruppen Gesichtern ungefähr gleich und entsprechen ungefähr der Leistung bei maskierten Eigen-Gruppen Gesichtern. Bei der Testung nach drei Wochen gleichen sich die Leistungen bei den Stimulusgesichtern für beide Teilnehmergruppen und bei beiden Gesichtsausschnitten einander an. Was bedeutet dies für die Verarbeitung von Gesichtern? Zum einen könnte dies als Beleg für eine eher ganzheitliche Verarbeitung der Eigen-Gruppen Gesichter zu deuten sein, da bei fehlenden äußeren Merkmalen die inneren Merkmale für die Wiedererkennung ausreichen müssen. Eigen-Gruppen Gesichter sind dabei im Vorteil, da eine Differenzierung der Eigen-Gruppen Gesichter häufiger gelernt wurde als von Fremd-Gruppen Gesichtern. Dass dieser Effekt nach einem langen Intervall zwischen Lern- und Testphase nicht mehr auftritt, kann zum einen daran liegen, dass

holistische Verarbeitung nicht zu einer länger überdauernden Gedächtnisleistung führt, zum anderen aber auch, wie bereits beschrieben, an der hohen Anzahl an Stimuli, die in diesem Experiment enkodiert werden mussten, sowie einem Bodeneffekt nach drei Wochen. Eine interessante Ausweitung dieses Experiments wäre die Untersuchung der Wiedererkennensleistung nach mehreren Zeitintervallen, mit Testphasen, die beispielsweise im einwöchigen Abstand stattfinden. Auf diese Weise wäre man besser in der Lage festzustellen, wie stark zeitüberdauernd der Cross-Race Bias ist.

Weiterhin wurde die Hypothese getestet, nach der die vier Stimulusethnien in Abhängigkeit von ihrer Distanz von den Eigen-Gruppen Gesichtern in einem multidimensionalen Raum unterschiedlich gut wiedererkannt werden können. Diese Hypothese konnte sowohl insgesamt als auch für die beiden Teilnehmergruppen getrennt bestätigt werden. Die afro-amerikanischen Gesichter wurden, konform zur Hypothese, von beiden Gruppen am schlechtesten wiedererkannt. Die weitere Staffelung der Stimulusethnien ist für beide Gruppen entsprechend der Distanz von der Eigen-Gruppe (Türken: Afro-Amerikaner, Kaukasische Amerikaner, Deutsche, Türken; Deutsche: Afro-Amerikaner, Türken, Kaukasische Amerikaner, Deutsche). Dass gleichzeitig die Teilnehmer bei kaukasischen Amerikanern deskriptiv etwas schlechtere Leistungen als bei türkischen und deutschen Gesichtern zeigen mag zunächst etwas verwundern. Eine mögliche Erklärung könnte sein, dass die afro-amerikanischen und die kaukasisch-amerikanischen Gesichter aus einem Stimulus-Pool (Meissner & Brigham, in Druck) kommen und die türkischen und deutschen Gesichter aus einem anderen Stimulus-Pool (Sporer, 1999; Dogan, 2003; Henning, 2003). Die Qualität der Bilder von Meissner und Brigham ist etwas schlechter als die der Bilder von Sporer (1999), Dogan (2003) und Henning (2003), was bedeutet, dass die Auflösung der Bilder etwas niedriger und die Helligkeit etwas höher ist. Dies kann aber nicht als ursächlich dafür herangezogen werden, dass der Haupteffekt des Faktors 'Stimulusethnie' gefunden wurde. Die Gesichter selbst waren eindeutig differenzierbar, und die Gesichter der kaukasischen Amerikaner konnten von den deutschen Teilnehmern besser wiedererkannt werden als von den türkischen Teilnehmern, was gegen ein Artefakt der Ergebnisse durch das Stimulusmaterial spricht.

Die Analyse zeigte einen Haupteffekt für die Stimulusethnie für die Reaktionsneigung $\underline{B''}$, $\eta^2 = .26$. Die Werte bei allen vier Ethnien sind $\underline{B''} > 0$. Das bedeutet, bei allen vier Ethnien wurde ein konservatives Antwortkriterium angelegt. Dabei ist die Reaktionsneigung vor allem bei den deutschen und kaukasisch-amerikanischen Gesichtern konservativer als bei den afro-amerikanischen und türkischen Gesichtern. Nichtsdestotrotz waren die Teilnehmer insgesamt eher vorsichtig in ihrem Antwortverhalten. Entsprechend

dem von einigen Autoren (Sporer, 1992; Slone, Brigham & Meissner, 2000) postulierten liberaleren Antwortkriterium bei Fremd-Gruppen Gesichtern konnte in diesem Experiment zumindest teilweise eine liberale Reaktionsneigung bei Fremd-Gruppen Gesichtern gefunden werden. Beide Teilnehmergruppen zeigten bei kaukasisch-amerikanischen und deutschen Gesichtern ein konservativeres Kriterium als bei den afro-amerikanischen und türkischen Gesichtern. Insofern kann man im Vergleich sagen, dass zumindest für die deutschen Teilnehmer bei den Fremd-Gruppen Gesichtern liberaler geantwortet wurde als bei den Eigen-Gruppen Gesichtern. Lediglich die türkischen Teilnehmer bildeten hier bei türkischen Gesichtern eine Ausnahme.

5 Pilotstudie für die Auswahl der latein-amerikanischen Gesichter

In den beiden Experimenten zum Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung (siehe Abschnitt 6) sollten, ähnlich wie im vorangegangenen Experiment, den Teilnehmern Gesichter mehrerer Ethnien vorgelegt werden. In dem Experiment zum Einfluss des Behaltensintervalls und des Gesichtsausschnittes auf die Wiedererkennensleistung wurden den Teilnehmern Gesichter von afro-amerikanischen, kaukasisch-amerikanischen, türkischen und deutschen Personen dargeboten. Das Problem, das bei dieser Zusammensetzung der Stimuluse ethnien entstand, war, dass sich die deutschen Stimulus-Personen und die kaukasischen Amerikaner bezüglich ihrer ethnischen Abstammung und ihres Aussehens sehr ähnlich sind. Es konnten zwar in dem vorherigen Experiment leichte Unterschiede in der Wiedererkennensleistung zwischen den kaukasischen und den deutschen Stimuluspersonen gefunden werden, doch um diesen Kritikpunkt zu entschärfen, wurden in den Experimenten zum Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung die Bilder der kaukasischen Amerikaner gegen eine andere Stimuluse ethnien ausgetauscht. Dazu standen Bilder von latein-amerikanischen Männern zur Verfügung.¹⁷ Es wurde zunächst wiederum eine Pilotstudie durchgeführt, um die Bilder der latein-amerikanischen Stimuluspersonen auszuwählen. Ob sich die Ratings für die Gesichter zwischen Versuchspersonen-Gruppen unterscheiden, läßt sich in dieser Pilotstudie im Gegensatz zur Pilotstudie in Abschnitt 3 nicht klären, weil in dieser Pilotstudie nur deutsche Versuchspersonen befragt wurden. Die Rekrutierung von Latein-Amerikanern hat sich als sehr problematisch herausgestellt, weil in der Gegend in und um Gießen deren Anteil an der Bevölkerung sehr gering ist. Diese Studie sollte demnach lediglich der Generierung von zwei Darbietungssets dienen und verzichtet daher auf weitergehende Analysen der verwendeten Skalen.

In dieser Pilotstudie wurden 70 Fotos latein-amerikanischer Männer in bezug auf die fünf Skalen von Vokey und Read (1992) beurteilt, weil es für diese Gesichter noch keine Beurteilungen bezüglich dieser Skalen von Vokey und Read (1992) gab. Ursprünglich standen mehr als die verwendeten 70 unterschiedlichen Gesichter zur

¹⁷ Ich danke Prof. Roy S. Malpass von der University of Texas at El Paso, der die Bilder zur Verfügung gestellt hat.

Verfügung, es stellte sich jedoch heraus, dass nicht für jedes Gesicht in der Frontalansicht auch das gleiche Gesicht im rechten Halbprofil vorhanden war, was für Wiedererkennungstests unerlässlich ist.

5.1 Material

In diesem Abschnitt werden zunächst die Stichprobe, das Stimulusmaterial und die verwendeten Skalen beschrieben. Anschließend wird die Versuchsdurchführung und der Versuchsablauf besprochen.

5.1.1 Stichprobe

An der Pilotstudie nahmen 32 Personen deutscher Herkunft teil, größtenteils Studierende der Justus-Liebig-Universität Gießen. Das Verhältnis von Frauen zu Männern war 3:1, um das gleiche Verhältnis von Frauen zu Männern zu erhalten wie in der Pilotstudie zum Experiment zum Behaltensintervall und dem Gesichtsausschnitt. Der Median der gesamten Stichprobe betrug $Mdn = 24$ Jahre. Der jüngste Teilnehmer war 19, der älteste Teilnehmer 43 Jahre alt. Der Altersmedian der weiblichen Teilnehmer betrug $Mdn = 22$, der Altersmedian der männlichen Teilnehmer $Mdn = 27.5$. Die Teilnehmer erhielten nach Durchführung des Ratings entweder eine Versuchspersonenstunde oder fünf Euro.

5.1.2 Stimulus-Material

Das Stimulus-Material bestand aus 70 Bildern in Frontalansicht von Männern latein-amerikanischer Herkunft. Für diese Gesichter waren ebenfalls die Bilder im rechten Halbprofil vorhanden, diese wurden jedoch erst für die Experimente zum Einfluss der Kontakthäufigkeit benötigt. Deren Bearbeitung wird aber dennoch in diesem Abschnitt beschrieben, da die Bearbeitung beider Ansichten zeitgleich durchgeführt wurde. Auf den Bildern der Latein-Amerikaner war bei einigen Gesichtern ein Schattenwurf im Hintergrund zu erkennen, der es Probanden erleichtert hätte, diese Bilder wiederzuerkennen. Aus diesem Grund wurden Schatten, sofern sie vorhanden waren,

wegretuschiert. Die Gesichter der anderen Ethnien hatten keinen Schattenwurf und wurden daher nicht noch einmal extra behandelt. Die Bilder der Latein-Amerikaner wurden zunächst standardisiert und vorhandene Schatten mittels Photoshop entfernt.

Als Vorlage diente das gleiche Bild wie in der ersten Pilotstudie. Das Originalfoto wurde in die Vorlage kopiert und die Vorlage auf 300 % vergrößert, um das kopierte Bild besser und genauer bearbeiten zu können. Der Fotohintergrund war bei allen Bildern gleich, sodass eine Hintergrundvorlage angefertigt werden konnte, mit deren Hilfe die Schatten auf den Bildern wegretuschiert werden konnten. Dazu wurde das 'Pattern Stamp Tool' mit diesem Hintergrund und dem Pinsel 'Hard Rock' mit einer Stärke von 19 Pixel und einer Härte von 100 % ausgewählt. Damit konnte dann der Schatten durch Entlangziehen des Pinsels an der äußeren Haarkante entfernt werden. Das Bild wurde wieder auf 100 % verkleinert und die Pupille des linken Auges an dem Schnittpunkt der linken vertikalen mit der horizontalen Linie durch Drag & Drop der Zentrierungsmarke an diesem Punkt verankert (siehe Abbildung 17). Dann konnte, falls nötig, das Bild mit Hilfe des Befehls 'Transform rotate' gegen oder mit dem Uhrzeigersinn so gedreht werden, dass die Pupille des rechten Auges auf der horizontalen Linie lag. Durch den Befehl 'Free Transform' konnte das Bild anschließend so gedehnt bzw. gestaucht werden, dass sich die Pupille des rechten Auges ebenfalls auf dem Schnittpunkt der rechten vertikalen mit der horizontalen Linie befand. Es war bei den latein-amerikanischen Gesichtern ebenfalls möglich, dass durch diese Prozedur Teile des weißen Hintergrundes sichtbar wurden, die dann mit Hilfe des 'Pattern Stamp Tool' und dem generierten Hintergrund retuschiert werden konnten. Letztendlich musste mit dem Tool 'magic wand' das Hemd auf den Fotos ausgewählt werden, damit es vergleichbar mit den anderen Fotos geschwärzt werden konnte. Mit Hilfe des 'Pattern Stamp Tool' und eines schwarzen 'Musters' wurde dann der markierte Bereich schwarz eingefärbt. Mit dem 'Smudge Tool' wurde der Rand zwischen Hemd und Hintergrund anschließend noch begradigt und das Bild mit den Befehlen 'Auto-Levels' und 'Auto-Contrast' den anderen Bildern angeglichen.



Abbildung 17. Beispiel-Gesicht in Frontal und Halbprofilansicht mit Hilfslinien für die Standardisierung. Die Hilfslinien waren in der Darbietungs- und Testphase nicht sichtbar.

Bei den Fotos im rechten Halbprofil (45° - Winkel) wurde ähnlich verfahren (siehe Abbildung 17). Hier zeigte sich jedoch, dass es nicht möglich war, die einzelnen Bilder anhand einer Bildvorlage zu standardisieren. Der Grund hierfür ist zum einen die unterschiedliche Kopfform eines jeden Gesichts, was dazu führen würde, dass bei einigen Bildern der Hinterkopf über den Bildrand hinausragen würde. Daher wurde als 'Standard' für jedes Bild im Halbprofil das bereits bearbeitete, dazugehörige Bild in der Frontalansicht herangezogen. In dieses Bild wurden zusätzlich zwei horizontale Linien eingezeichnet; die erste Linie wurde am oberen Haaransatz eingefügt, die zweite Linie am unteren Ende des Kinns. Auf diese Weise wurden Anhaltspunkte für die vertikale Größe des Gesichtes geschaffen, an die das Gesicht im Halbprofil angepasst werden konnte. Das Gesicht in der Frontalansicht wurde mitsamt dem Hintergrund entfernt (Befehl: 'Delete Layer') und das Gesicht im Halbprofil in die neu entstandene Vorlage kopiert, sodass die linke Pupille auf der mittleren horizontalen Linie lag. Falls nötig, konnte das Bild so gedreht werden, dass die Pupille des rechten Auges ebenfalls auf der mittleren horizontalen Linie lag. Anschließend wurde das Gesicht mit Hilfe des Befehls 'Free Transform' so

vergrößert, bzw. verkleinert, dass oberer Haaransatz und unteres Kinn an den neu eingefügten horizontalen Hilfslinien lagen und somit das Gesicht die gleiche Größe wie in der Frontalansicht hatte. Falls nun der Hinterkopf über den Bildrand hinausragte, wurde das Bild horizontal in die Bildmitte verschoben. Anschließend wurde wie bei den Frontalansichten die Schatten entfernt und die Farbe sowie die Ränder der T-Shirts an die Vorlage angepasst. Die Gesichter im Halbprofil wurden, wie bereits erwähnt, nicht in der Pilotstudie verwendet, sondern als Stimuli in der Testphase der Wiedererkennenstests eingesetzt.

5.1.3 Skalen

Wie in der Pilotstudie zum Experiment in Abschnitt 4 orientierte sich die Gestaltung der Rating-Skalen an Untersuchungen von Vokey und Read (1992) und O'Toole, Deffenbacher, Valentin und Abdi (1994), die zwei Faktoren extrahierten, auf denen die fünf Skalen laden. Die beiden Faktoren sind in Abschnitt 3.1.3 bereits beschrieben worden, deshalb werden sie hier nicht wiederholt.

Für die Erstellung der Darbietungssets der latein-amerikanischen Gesichter im Experiment zum Einfluss der Kontakthäufigkeit auf den Cross-Race Bias wird wiederum der zweite Faktor 'Einprägsamkeit' verwendet, da dieser die Skalen beinhaltet, die für die Wiedererkennbarkeit eines Gesichtes wichtig sind. Für die spätere Berechnung der Distinktheit eines Bildes wurde der Wert der Skala Typikalität benutzt und umgepolt, weil Typikalität und Distinktivität die beiden Pole einer Dimension darstellen. Die anderen Skalen gingen gemäß ihrer Ratings in die spätere Berechnung ein.

5.1.4 Stimulus-Darbietung

Die Datenerhebung erfolgte an zwei Apple 'iMac' Computern mit einer Taktfrequenz von 300 bzw. 350 Mhz, einem Arbeitsspeicher von jeweils 192 MB und einem 15"-Monitor. Die Farbtiefe betrug 16 Bit (65536 Farben) bei einer Bildschirmauflösung von 1024x768 Pixel und einer Bildwiederholfrequenz von 75 Hz. Die Versuchssteuerung geschah durch das Programm 'PsyScope 1.2.5' (Cohen, MacWhinney, Flatt & Provost, 1993).

Jede der fünf Skalen sollte einmal pro Bild bearbeitet werden, wobei vermieden werden sollte, für jedes Bild die gleiche Reihenfolge der Skalenbearbeitung zu haben, um Reihenfolgeeffekte der Probanden bei der Skalenbearbeitung zu vermeiden. Schließlich sollte die Frage für jede Skala gewissenhaft beantwortet werden. Dazu wurden die fünf Skalen für die Darbietung in Form eines lateinischen Quadrates kombiniert. Dieser Schritt war notwendig, weil die Programmierung der einzelnen Trials in Psyscope anders nicht möglich war.

5.1.5 Versuchsdurchführung

Die Prozedur war die gleiche wie in der Pilotstudie zum Experiment zum Einfluss des Behaltensintervalls und des Gesichtsausschnitts. Nach der Begrüßung durch die Versuchsleitung setzte sich der Teilnehmer vor den Computer. Diesem Vorgang schloss sich zunächst (1) die Instruktionsphase und (2) die Versuchsphase an. Der Versuchsablauf für jede einzelne Versuchsperson war wie folgt: Die Teilnehmer wurden instruiert, in der Instruktionsphase den Anweisungen auf dem Bildschirm zu folgen. Die einzelnen Instruktionen sind in Anhang A5 aufgeführt. Während der Instruktionsphase wurde nach der vierten Instruktion ein Beispieltrial bearbeitet, dessen Stimulusbild aus einem Foto der Fotoserie von Malpass bestand. Nach der Instruktionsphase folgte die Versuchsphase mit dem Hinweis für die Teilnehmer, eine Taste zu drücken, wenn sie für den Versuch bereit sind. Anschließend folgten die 70 Trials in randomisierter Reihenfolge mit den latein-amerikanischen Gesichtern. Jeder Trial war gleichermaßen aufgebaut: 1) Präsentation des Stimulus in standardisierter Größe (432 x 648 Pixel) für 2000 ms, 2) Präsentation des Stimulus mit einer Größe von (Breite x Höhe) 266 x 400 Pixel in der oberen Hälfte des Bildschirms (horizontal zentriert, vertikal 32%¹⁸). Gleichzeitig wird nach randomisierter Auswahl einer der fünf Zeilen des lateinischen Quadrates die erste der Skalen in dieser Zeile in der unteren Hälfte des Bildschirms (700 x 158 Pixel, horizontal zentriert, vertikal 75 %) präsentiert. Die Teilnehmer sollten nun das Rating für den Stimulus für die präsentierte Skala durch Drücken einer der Tasten von 1 bis 7 abgeben. Diese Prozedur wiederholte sich für die verbleibenden vier Skalen. Das Stimulusgesicht blieb während der

¹⁸ Die Angabe der vertikalen Position in Prozentwerten bedeutet bei Psyscope 1.2.5 die relative Position des Mittelpunktes einer definierten Fläche zum vertikalen Mittelpunkt des Bildschirms. Ein Prozentwert von 0 heißt, dass sich der Flächenmittelpunkt am oberen Bildschirmrand befindet. Bei einem Prozentwert von 100 befindet sich der Flächenmittelpunkt am unteren Bildschirmrand.

Bearbeitung der fünf Skalen auf der oben angegebenen Position präsent. Zwischen den einzelnen Trials lag ein ITI von 500 ms. Die Randomisierung der fünf Skalen war so ausbalanciert, dass jede der fünf Zeilen des lateinischen Quadrates 14 mal pro Ethnie ausgewählt wurde. Nach Bearbeitung der 70 Trials endet der Versuch. Dem Teilnehmer wurde gedankt, eine Versuchspersonenstunde gutgeschrieben bzw. fünf € ausbezahlt und er wurde verabschiedet.

5.2 Ergebnisse

Die Werte der Ratings für die einzelnen Skalen wurden für jedes Stimulusfoto in Form von Mittelwerten über die Versuchspersonen zusammengefasst. Die Mittelwerte der fünf Skalen für die einzelnen Bilder sind in Anhang C2 dargestellt. Analog wurde für die Ratings der latein-amerikanischen Gesichter wie in der Pilotstudie für das Experiment zum Einfluss von Behaltensintervall und Gesichtsausschnitt eine Faktorenanalyse durchgeführt, um zu testen, ob die Faktorenstruktur der latein-amerikanischen Gesichter mit der von Vokey und Read (1992) übereinstimmt. Tabelle 16 zeigt die Mittelwerte und Standardabweichungen der verwendeten Skalen. Die Faktorenstruktur konnte auch bei den Ratings für latein-amerikanische Gesichter in dieser Pilotstudie gefunden werden. In Tabelle 17 sind die Ergebnisse einer Faktorenanalyse mit Varimax-Rotation für die fünf Skalen dargestellt.

Tabelle 16

Mittelwerte und Standardabweichungen für die einzelnen Skalen

Skalen	<u>M</u>	<u>SD</u>
Einprägsamkeit (memorability)	3.74	.56
Distinktivität (distinctiveness)	3.85	.55
Sympathie (likability)	3.05	.71
Attraktivität (attractiveness)	2.21	.58
Vertrautheit (familiarity)	1.85	.40

Tabelle 17

Ergebnisse der Faktorenanalyse für Gesichter von latein-amerikanischen Männern mit deutschen Teilnehmern

Skalen	Faktor 1	Faktor 2
Einprägsamkeit (memorability)	.13	.95
Distinktivität (distinctiveness)	.05	.94
Sympathie (likability)	.85	.05
Attraktivität (attractiveness)	.92	.01
Vertrautheit (familiarity)	.55	.38

Anmerkung: Die beiden Faktoren klären bei den latein-amerikanischen Gesichtern 76.6 % der Varianz auf.

Damit können auch die latein-amerikanischen Gesichter anhand der Skalen bezüglich des relevanten Faktors 'Erinnerbarkeit' in die Stimulus-Sets aufgeteilt werden. Auch hier galt wieder, dass sich die Bilder zwischen den Sets nicht in ihrer Schwierigkeit unterscheiden. Dazu wurden zunächst die Bilder anhand der Ratings der beiden Skalen 'Distinktivität' und 'Einprägsamkeit' durch Mittelwertbildung zusammengefasst und anschließend anhand der zusammengefaßten Mittelwerte in aufsteigender Reihenfolge sortiert. Anschließend wurden 'Paare' gebildet die durch Zufallszuweisung auf die beiden Sets verteilt werden, sodass in jedem Set letztendlich 35 Bilder zu finden waren. Die beiden Sets unterscheiden sich nicht signifikant voneinander, $t(34) = 1.79$, *ns*, ($M_{\text{setA}} = 3.78$, $M_{\text{setB}} = 3.76$).

5.3 Diskussion

Auch in dieser Pilotstudie konnte die Faktorenstruktur von Vokey und Read (1992) gefunden werden. Die interpretierbaren Faktorladungen liegen zwischen .38 und .95. Die Skala 'Vertrautheit' lädt auf beiden Faktoren mit einem Wert $\geq .38$. Es zeigten sich in den Ratings zwei Faktoren mit Eigenwerten > 1 , und hohe Ladungen der Skalen 'Distinktivität' und 'Memorabilität' auf einem Faktor (Erinnerbarkeit). Wie bereits in der ersten Pilotstudie bei deutschen Teilnehmern und türkischen Stimulusgesichtern ist die Skala 'Vertrautheit' nicht eindeutig einem der beiden Faktoren zuzuordnen. Bei dieser Pilotstudie lädt diese Skala positiv auf dem ersten und moderat positiv auf dem zweiten Faktor. In der ersten Pilotstudie ist die Ladung auf dem ersten Faktor moderat positiv, auf dem zweiten Faktor hingegen moderat negativ. Warum dies so ist, ist unklar. Eine Umpolung der Skala fand weder bei der ersten noch bei der zweiten Pilotstudie statt. Diese Ergebnisse sprechen eher für eine besondere Rolle der Skala 'Vertrautheit' bei der Beurteilung von Gesichtern. Warum Vokey und Read (1992) den ersten Faktor 'Kontextfreie Vertrautheit' nannten, ist angesichts dieser Ergebnisse schwer nachzuvollziehen, wie auch bereits von Morris und Wickham (2001) angesprochen wurde. Auch hier erscheint es wieder sinnvoll zu sein, die beiden Faktoren angesichts der hohen Ladungen der Skalen 'Distinktivität' auf dem ersten Faktor und 'Attraktivität' auf dem zweiten Faktor ebenfalls mit 'Distinktivität' und 'Attraktivität' zu bezeichnen. Eine derartige Interpretation ist auch mit den Daten aus dieser Pilotstudie möglich.

Da diese Pilotstudie aber vor allem dazu dient, die Stimulusgesichter den Darbietungssets zuzuordnen, und die eindeutigen Faktorladungen der Skalen 'Einprägsamkeit' und 'Distinktivität' für eine Ordnung der Bilder mittels dieses Faktors sprechen, wird auf die Problematik der Skala 'Vertrautheit' nicht weiter eingegangen. Auch hier läßt sich wieder sagen, dass die deutsche Übersetzung der Skalen zu annähernd den gleichen Ergebnissen und damit auch zu den gleichen Einschätzungen von Gesichtern führt wie in den englischen Originalartikeln.

6 Der Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern

Ein immer wieder diskutierter Erklärungsansatz für den Cross-Race Bias ist die Kontakthypothese (Chance, Goldstein & McBride, 1975) mit konträren Ergebnissen, bei denen entweder kein Einfluss des Kontakts auf die Wiedererkennensleistung gefunden wird (z. B. Ng & Lindsay, 1994) oder der Kontakt mit anderen Ethnien zu einer besseren Wiedererkennensleistung führt (z. B. Wright, Boyd & Tredoux, 2003; Weimann, Fishman & Rattner, 1986). Dabei wird ebenfalls diskutiert, ob die Qualität oder die Quantität des Kontakts der ausschlaggebende Faktor ist. Im Experiment zum Einfluss des Behaltensintervalls und des Gesichtsausschnittes auf den Cross-Race Bias wurden Personengruppen untersucht, bei denen eine Trennung von Qualität und Quantität des Kontaktes nicht vorgenommen wurde. Türkische und deutsche Teilnehmer sollten 'lediglich' einen Cross-Race Bias zeigen, ungeachtet möglicher Unterschiede zwischen den Teilnehmergruppen bezüglich der Qualität und Quantität des Kontakts mit anderen Ethnien. Das Ziel in diesem Experiment zum Einfluss der Kontakthäufigkeit auf den Cross-Race Bias war, die Wiedererkennensleistung von Personengruppen zu testen, bei denen eine unterschiedliche Ausprägung der Kontakthäufigkeit mit Personen anderer ethnischer Gruppen angenommen wird. Wie aber kann diese Unterschiedlichkeit bestimmt werden? Die beiden Gruppen müssen sich in der Kontakthäufigkeit unterscheiden, ohne jedoch Unterschiede in der Kontaktqualität aufzuweisen. Deshalb wurde im Gegensatz zum vorhergehenden Experiment, in dem die Versuchspersonen aus verschiedenen ethnischen Gruppen stammen (Türken und Deutsche), als Versuchspersonen ausschließlich Deutsche rekrutiert. Die postulierte Kontakthäufigkeit erschließt sich aus den Berufsfeldern, denen die Teilnehmer nachgehen. Dies sind zum einen Studierende an der JLU Gießen. Bei dieser Personengruppe wird angenommen, dass für sie während des Studiums bzw. vor dem Studium wenig Kontakt zu anderen ethnischen Gruppen besteht bzw. bestand. Grundlage dieser Annahme ist zum einen die Tatsache, dass der Anteil ausländischer Kinder und Jugendlicher an Gymnasien oder Gesamtschulen, die zum Abitur führen in Deutschland und auch in Hessen gering ist und unter dem Durchschnitt der ausländischen Schüler an weiterführenden Schulen liegt. So beträgt der Anteil ausländischer Schüler an Gymnasien in Hessen im Jahr 2002 ca. 7 %, während an

Hauptschulen dieser Anteil bei 38 % liegt (Tischler, Kisseler und Trabert, 2002). Zudem ist auch der Anteil ausländischer Studierender mit ca. 9,5 % relativ gering (JLU Gießen, 2003). Die zweite Gruppe besteht aus Beamten des Bundesgrenzschutzes (BGS-Beamten), die ihre Arbeitsstelle am Flughafen Frankfurt/Main haben. Als größter deutscher Flughafen mit starkem Verkehrsaufkommen und einer großen Zahl internationaler Passagiere¹⁹, ist es am Flughafen Frankfurt sehr wahrscheinlich, dass die Kontakthäufigkeit mit Personen anderer ethnischer Gruppen dort weitaus größer ist als im Vergleich mit der Gruppe der Studierenden. Ein Teil der Tätigkeit von BGS-Beamten besteht darin, die Pässe von einreisenden Passagieren zu kontrollieren. Eine weitere Tätigkeit der BGS-Beamten ist die Rückführung von ausländischen Staatsbürgern in ihre Heimatländer. Da dies jedoch eine qualitativ andere Aufgabe ist als die Kontrolle von Pässen, mit längerem und auch intensiverem Kontakt, sollten keine BGS-Beamten teilnehmen, die bei der Rückführung ausländischer Staatsangehöriger mitwirken. Auf diese Art und Weise sollte eine gute Abgrenzung von Kontakthäufigkeit und Kontaktqualität gelingen, mit zwei Teilnehmergruppen, die sich in ihrer Kontakthäufigkeit stark voneinander unterscheiden.

Zwar sind Polizeibeamte eine in der forensischen Psychologie häufig untersuchte Gruppe, über die Fähigkeiten von Polizeibeamten als Augenzeugen oder bei Wiedererkennensexperimenten wird in der Literatur wenig berichtet. Die Vermutung, dass Polizeibeamte per se besser in der Wiedererkennung von Gesichtern sind als Zivilpersonen kann angesichts der Studien von Clifford und Richards (1977) und Tickner und Poulton (1975) sowie Burton, Wilson, Cowan und Bruce (1999) nicht gehalten werden. In einer Untersuchung von Clifford und Richards (1977), über den Zusammenhang zwischen Kontaktdauer und Gedächtnisabruf von Merkmalen der Person wurden Polizeibeamte und Zivilpersonen unwissentlich mit einer gestellten Situation, in der eine Zielperson nach der Zeit (kurzer Kontakt) bzw. dem Weg (langer Kontakt) fragte, konfrontiert und anschließend nach zehn bestimmten Merkmalen der Zielperson befragt. Nach kurzem Kontakt konnten die Polizeibeamten gleichviel korrekte Angaben machen wie die Zivilpersonen, nach langem Kontakt waren die Polizeibeamten besser bei der Angabe richtiger Details. Auf der anderen Seite gaben die Polizeibeamten nach kurzem Kontakt mehr falsche Details als die Zivilpersonen. Die Autoren interpretieren diese Ergebnisse als Folge einer größeren Verarbeitungskapazität der Polizeibeamten, dadurch dass ihnen die Aufgabenstellung einer Wegauskunft vertrauter bzw. geübter ist, womit Aufmerksamkeit

¹⁹ Im Jahr 2002 betrug das Passagieraufkommen am Flughafen Frankfurt/Main 48,5 Mio. Passagiere (Fraport AG, 2002)

von der Aufgabe zur Personenwahrnehmung geleitet wird. Tickner und Poulton (1975) fanden hingegen bei einem Vergleich von 24 Polizeibeamten und 156 Zivilpersonen, die nach Darbietung eines Films eine bzw. drei handelnde Personen identifizieren sollten, keine Differenzen zwischen Polizeibeamten und Zivilpersonen. Ebenfalls zeigten sich in der Studie von Burton, Wilson, Cowan und Bruce (1999) keine Unterschiede zwischen Polizisten und Zivilisten beim Wiedererkennen von Personen mittels einer Überwachungskamera. Angesichts dieser Studien ist nicht davon auszugehen, dass Polizeibeamte per se besser in der Wiedererkennung von Gesichtern sind. Da BGS-Beamte ebenfalls Polizeibeamte sind, die jedoch dem Bund und nicht den einzelnen Bundesländern unterstellt sind, und polizeiliche Aufgaben wahrnehmen, können Untersuchungen zu 'normalen' Polizeibeamten mit einiger Berechtigung auch mit denen mit BGS-Beamten verglichen werden.

Da in den Experimenten zum Einfluss der Kontakthäufigkeit auf den Cross-Race Bias keine unterschiedlichen Ethnien als Teilnehmergruppen erhoben wurden, ist der Rückschluss auf einen Cross-Race Bias schwierig, weil per definitionem ein Cross-Race Bias eine Interaktion von mindestens zwei Teilnehmerethnien sowie mindestens zwei Stimulusethnien benötigt. In diesem Fall sollte jedoch hauptsächlich die Leistung einer bestimmten Berufsgruppe bei der Wiedererkennung verschiedener Ethnien getestet und mit einer Kontrollgruppe verglichen werden. Außerdem ist es sehr problematisch, wenn nicht gar unmöglich, beim Bundesgrenzschutz eine ausreichend große Anzahl von Personen zu rekrutieren, die anderer ethnischer Herkunft als der deutschen ist und selbst wenn, wäre diese Gruppe wahrscheinlich sehr heterogen bezüglich ihrer ethnischen Herkunft. Daher bezieht sich der Cross-Race Bias in diesen Experimenten auf die unterschiedliche Wiedererkennensleistung deutscher Teilnehmer bei Gesichtern der eigenen ethnischen Gruppe und Gesichtern anderer ethnischer Gruppen.

Das Wiedererkennen von Gesichtern wird normalerweise getestet, indem die Ansicht der Gesichter zwischen Lern- und Testphase unterschiedlich ist, um ein rein merkmalsbasiertes Wiedererkennen zu vermeiden. Argstatter, Gehrke, Hüweler, Kampschulte, Schraub und Sporer (2002) untersuchten in einem Wiedererkennenstest die Leistung bei türkischen und deutschen Gesichtern, die in der Testphase der Hälfte der Teilnehmer wie in der Lernphase in der Frontalansicht, der anderen Hälfte der Teilnehmer in einer 3/4-Ansicht dargeboten wurden. Dabei zeigte sich keine schlechtere Leistung bei deutschen Gesichtern in der 3/4-Ansicht verglichen mit der Frontalansicht, türkische Gesichter wurden allerdings in der 3/4-Ansicht schlechter wiedererkannt als in der Frontalansicht. Der Cross-race bias von deutschen Versuchsteilnehmern wurde nur bei

einem Wechsel der Pose, nicht aber bei gleicher Ansicht in Lern- und Testphase festgestellt.

Da in den meisten Untersuchungen ein Ja/Nein Wiedererkennenstest verwendet wurde, und eine Delayed-Matching Aufgabe weniger Verwendung fand, soll in dieser Untersuchung der Faktor der Ansicht eines Gesichtes zusätzlich getestet werden. Interessant ist dabei, zu erfahren, inwiefern die Ansicht eines Gesichtes, wenn, wie bei einer Delayed-Matching Aufgabe, zwischen Darbietung und Test nur wenige Sekunden liegen, einen Einfluss auf die Leistung hat.

6.1 Darstellung des Wiedererkennenstests und der Delayed-Matching Aufgabe

In dieser Untersuchung wurden, nachdem die latein-amerikanischen Gesichter aus der Pilotstudie in Abschnitt 5 in zwei Darbietungssets aufgeteilt wurden, zwei Experimente mit BGS-Beamten als Experimental- und Studierenden als Kontroll-Gruppe durchgeführt. Zum einen wurde die Wiedererkennensleistung mit einem Ja/Nein -Test als experimentellem Paradigma getestet, zum anderen wurde ebenfalls eine Delayed-Matching Aufgabe den Teilnehmern präsentiert, um der, für psychologische Untersuchungen, etwas ungewöhnlichen Gruppe der BGS-Beamten Rechnung zu tragen. Die Delayed-Matching Aufgabe als 'zeitverzögerter Abgleich von Gesichtern' ist der täglichen Aufgabe von Beamten am Frankfurter Flughafen relativ ähnlich, und erlaubt nicht nur Rückschlüsse über Gedächtnisleistungen, sondern auch über die Wahrnehmung von Gesichtern als Funktion unterschiedlicher Expertise. Zunächst werden die Hypothesen zu beiden Experimenten erläutert, anschließend der Versuchsplan, die Stichprobe und das verwendete Material sowie der Versuchsablauf beschrieben. Darauf folgend werden die Ergebnisse der beiden Experimente dargestellt und diskutiert.

6.2 Hypothesen

Es wird vermutet, dass aufgrund der Annahmen der Speicherung von Gesichtern in einem multidimensionalen Raum, die BGS-Beamten durch den häufigen Kontakt mit anderen Ethnien ein größeres Repertoire an Repräsentationen besitzen als die Gruppe der Studierenden. Es sollte ein Haupteffekt für den Faktor 'Berufsgruppe' (BGS-Beamte vs.

Studierende) gefunden werden. Die Gruppe der BGS-Beamten sollte insgesamt mehr Treffer und bessere Wiedererkennensleistungen aufweisen, gemessen durch \underline{A}' , als die Gruppe der Studierenden.

Es sollte sich ebenfalls ein Haupteffekt für den Faktor 'Stimuluse ethnien' (afro-amerikanische, latein-amerikanische, türkische und deutsche Gesichter) bei den Treffern und dem Leistungsindex \underline{A}' zeigen. Aufgrund der Tatsache, dass alle Versuchspersonen deutscher Nationalität sind, sollten deutsche Gesichter besser wiedererkannt werden als Gesichter anderer Ethnien. Geht man davon aus, dass Gesichter von Ethnien, die weniger häufig enkodiert werden und darüber hinaus durch bestimmte Merkmale der Gesichter andere Punkte im multidimensionalen Raum besetzen als die Gesichter der eigenen ethnischen Gruppe, so sollte sich bei den verwendeten Stimulusgesichtern eine lineare Beziehung zwischen Stimuluse ethnien und Wiedererkennensleistung zeigen. Diese Hypothese konnte bereits im Experiment in Abschnitt 4 bestätigt werden. Demnach wären afro-amerikanische Gesichter am schlechtesten wiederzuerkennen, weil sie zum einen relativ selten in Deutschland auftreten und zum anderen bereits durch die Hautfarbe einen anderen Punkt auf dieser Dimension im multidimensionalen Modell besetzen als die Ethnie der Probanden. Die Gesichter von Latein-Amerikanern sollten besser wiederzuerkennen sein als die Gesichter von Afro-Amerikanern, weil deren Auftretenshäufigkeit in Deutschland zwar ebenfalls nicht sehr groß ist, die ethnische Verwandtschaft mit Südeuropäern jedoch eine größere Nähe im multidimensionalen Raum zur Ethnie der Teilnehmer nahe legt als die afro-amerikanischen Gesichter. Die türkischen Gesichter sollten besser als die latein-amerikanischen Gesichter wiedererkannt werden, weil diese ethnische Gruppe in Deutschland weitaus häufiger anzutreffen ist als Afro-Amerikaner und Latein-Amerikaner und daher mehr Exemplare von den Probanden enkodiert wurden. Letztendlich sollten sich bei den deutschen Gesichtern als eigene ethnische Gruppe die besten Wiedererkennensleistungen zeigen.

Es wird ebenfalls davon ausgegangen, dass die Unterschiede zwischen den Ethnien bei BGS-Beamten sowohl bei den Treffern und falschen Alarmen als auch bei Leistungsindex \underline{A}' geringer ausfallen als bei Studierenden. Dies liegt daran, dass vermutet wird, dass durch den täglichen wiederholten Kontakt mit anderen Ethnien, sich bei den BGS-Beamten ein breiteres Spektrum an Exemplaren von Gesichtern vieler unterschiedlicher Ethnien herausgebildet hat, im Vergleich zu den Studierenden, und es damit den BGS-Beamten leichter fällt, neue Gesichter zu enkodieren und abzurufen. Demnach sollte sich ein Interaktionseffekt der Faktoren 'Berufsgruppe' und 'Stimuluse ethnien' zeigen. Die BGS-Beamten sollten sowohl bei den Treffern, als auch beim

Leistungsindex A' keine Unterschiede zwischen den Stimuluse ethnien zeigen, während bei den Studierenden eine lineare Beziehung zwischen den Stimuluse ethnien zu beobachten sein sollte. D.h. die Studierenden sollten wie beim Experiment in Abschnitt 4 bei den afro-amerikanischen Gesichtern die schlechtesten Leistungen zeigen, gefolgt von den latein-amerikanischen, den türkischen und den deutschen Gesichtern.

Um den Einfluss der Ansicht eines Gesichts auf die Leistung Beim Matching zu testen, sollten Gesichter, die bei der Delayed-Matching Aufgabe im Halbprofil dargeboten werden, schlechter wiedererkannt werden, als Gesichter, die in der Frontalansicht dargeboten werden, da ein rein merkmalsbasiertes Wiedererkennen nicht möglich ist. Da die BGS-Beamten mehr Kontakt mit anderen Ethnien haben sollten, müssten auch die Leistungen der BGS-Beamten bei Gesichtern im Halbprofil besser sein als bei den Studierenden. Dies gilt sowohl für die Treffer als auch für den Leistungsindex A' .

6.2.1 Versuchsplan

Aus den bisherigen Überlegungen resultierend, wurden in dieser Untersuchung die folgenden Faktoren eingeführt. Die Teilnehmer wurden zum einen aus den Beamten des Bundesgrenzschutzes am Flughafen Frankfurt zum anderen aus den Studierenden der Justus-Liebig-Universität Giessen rekrutiert. Jeweils die Hälfte waren Männer, respektive Frauen. Beide Gruppen führten sowohl den Wiedererkennenstest als auch die Delayed-Matching Aufgabe durch. Der Wiedererkennenstest war für alle Teilnehmer so aufgebaut, dass in der Darbietungsphase die Gesichter in der Frontalansicht und in der Testphase im rechten Halbprofil dargeboten wurden. Die Hälfte der Teilnehmer bearbeitete die Delayed-Matching Aufgabe in der Version, dass erstes und zweites Gesicht in der Frontalansicht dargeboten wurden, bei der anderen Hälfte wurde das erste Gesicht in der Frontalansicht und das zweite Gesicht im rechten Halbprofil dargeboten. Weiterhin wurde variiert, welches Darbietungsset dem Teilnehmer in der Darbietungsphase gezeigt wurde und welche Taste vom Teilnehmer gedrückt werden sollte, um anzuzeigen, welches Bild er bereits gesehen hatte, bzw. ob die beiden Gesichter in der Delayed-Matching Aufgabe identisch sind oder nicht. Ebenfalls wurde für die Delayed-Matching Aufgabe ein weiterer Faktor eingeführt ('Identitätspräsentation'), der mögliche Effekte des Vergleichsgesichts in Durchgängen mit nicht-identischen Gesichtern kontrollieren soll. Dazu wurden zwei Bedingungen eingeführt, die für das gleiche Stimulusgesicht zwei verschiedene Vergleichsgesichter darboten. Die genaue Aufteilung der Stimuli auf die Darbietungssets

und die Versuchsbedingungen ist in Abschnitt 6.3.2 aufgeführt. Dementsprechend ergibt sich der in Tabelle 18 abgebildete Versuchsplan.

Tabelle 18

Versuchsplan für die Experimente zum Einfluss der Kontakthäufigkeit auf den Cross-Race Bias, unterteilt nach Set (A,B), Identitätsbedingung(1,2) und Taste(c,m)

B-Gr	Sex	Rec-Test	DM	A				B			
				1		2		1		2	
				c	m	c	m	c	m	c	m
BGS	männl.	Fro/HP	Fro/Fro	1	1	1	1	1	1	1	1
		Fro/HP	Fro/HP	1	1	1	1	1	1	1	1
	Weibl.	Fro/HP	Fro/Fro	1	1	1	1	1	1	1	1
		Fro/HP	Fro/HP	1	1	1	1	1	1	1	1
Stud.	männl.	Fro/HP	Fro/Fro	1	1	1	1	1	1	1	1
		Fro/HP	Fro/HP	1	1	1	1	1	1	1	1
	Weibl.	Fro/HP	Fro/Fro	1	1	1	1	1	1	1	1
		Fro/HP	Fro/HP	1	1	1	1	1	1	1	1

Legende: B-Gr = Berufsgruppe; Rec-Test = Wiedererkennenstest; DM =Delayed-Matching

Die Anzahl der Teilnehmer ($N = 64$) ist ausreichend, um bei erwarteten mittleren Effektstärken ($.3 \leq r \leq .49$) beim Cross-Race Bias (Meissner & Brigham, 2001) und einem alpha-Niveau von .05 je nach Anzahl der Zählerfreiheitsgrade eine Power von .81 bis .99 zu erreichen (siehe Cohen, 1988).

6.2.2 Stichprobe

An dieser Untersuchung nahmen insgesamt 64 Versuchspersonen teil. Der Altersmedian der gesamten Stichprobe beträgt $Mdn = 27$. Der jüngste Teilnehmer war 19, der älteste 48 Jahre alt. Jede der zwei Gruppen bestand aus 32 Teilnehmern, dem Versuchsplan entsprechend ein Teilnehmer je Zelle. Der Altersmedian beträgt für die BGS-Beamten $Mdn = 30$, und für die Studierenden $Mdn = 22$. Die BGS-Beamten sind

älter als die Studierenden, dies ist jedoch in Anbetracht der bereits abgeschlossenen Berufsausbildung und einer durchschnittlichen Tätigkeitsdauer als BGS-Beamter von $M = 6.52$ Jahren nicht überraschend.

6.3 Material

Es wurden jene Bilder verwendet, die in den beiden Pilotstudien (Abschnitte 3 und 5) einem Rating unterzogen und anschließend aufgrund ihrer Distinktheit in zwei verschiedene Sets unterteilt wurden. Zusätzlich musste ein Kontaktfragebogen bearbeitet werden, der die Ausprägung und Häufigkeit des Kontaktes mit Personen anderer Ethnien erfasste (siehe Anhang B-3). Der Kontaktfragebogen diente sowohl der Erfassung einer subjektiven wahrgenommenen Kontakthäufigkeit der Teilnehmer mit anderen Ethnien, als auch als Füller-Aktivität zwischen den beiden Experimenten. Eine objektive Messung der Kontakthäufigkeit wäre bei den Studierenden insofern nicht möglich gewesen als dass die Studierenden lediglich zu den Experimenten im Labor erschienen. Bei den BGS-Beamten wäre eine objektive Erfassung der Kontakthäufigkeit nicht minder aufwändig, da eine solche Prozedur die Zählung der Passagiere und eine gleichzeitige Einschätzung bzw. Befragung ihrer Nationalität etc. bedeutet hätte.

6.3.1 Stimuli

Als Stimulusmaterial wurden die Fotos von Türken und Deutschen, die in der Pilotstudie zum Experiment zum Einfluss von Behaltensintervall und Gesichtsausschnitt beurteilt wurden, sowie die Bilder der Afroamerikaner von Meissner und Brigham (im Druck) als auch die Gesichter von Latein-Amerikanern von Malpass (University of Texas at El Paso) aus der Pilotstudie in Abschnitt 5 verwendet. Beispielgesichter sind in Anhang D2 abgebildet. Alle Bilder waren vollständig zu sehen. Es gab keine Einschränkung durch Retuschierung der Haare etc.

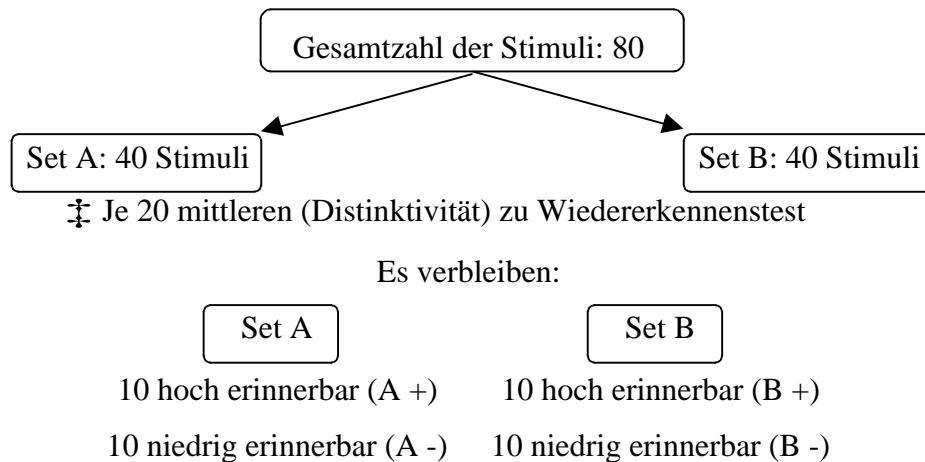
6.3.2 Aufteilung der Stimuli auf die Darbietungssets und die Experimente

Für diese Untersuchung mussten die vorhandenen Fotos auf die zwei Experimente aufgeteilt werden. Ausgehend von den Ratings auf dem Faktor 'Erinnerbarkeit' (siehe Abschnitt 3) wurden die Stimuli zunächst den beiden Experimenten zugeteilt. Dabei war vorgesehen, für den Wiedererkennenstest diejenigen Stimuli zu verwenden, deren Ratings auf der Skala 'Erinnerbarkeit' im mittleren Bereich lagen. Das heisst, von den Stimulusethnien wurden für Set A und Set B jene 20 Bilder für den Wiedererkennenstest verwendet, die im mittleren Bereich der Ratings lagen (siehe Anhang C1 – C5).

Die Stimuli für die Delayed-Matching Aufgabe wurden aus den verbleibenden Bildern gebildet. So konnten theoretisch aus den jeweils verbleibenden 40 Bildern von Türken, Afro-Amerikanern und Deutschen die Durchgänge der Delayed-Matching Aufgabe gebildet werden. Da aber lediglich 70 latein-amerikanische Gesichter zur Verfügung standen, musste die Anzahl der Durchgänge auf 16 herabgesetzt werden. Die verbliebenen Bilder bestanden bei den afro-amerikanischen, türkischen und deutschen Gesichtern aus den jeweils zehn als hoch- bzw. niedrig 'erinnerbar' eingestuften Bildern, bei den latein-amerikanischen Gesichtern aus den sieben als hoch- bzw. acht niedrig 'erinnerbar' eingestuften Bildern.²⁰ Für jede Ethnie mussten 16 Durchgänge programmiert werden, wobei in acht Durchgängen zweimal dasselbe Gesicht gezeigt wurde (Identitätsbedingung) und in weiteren acht Durchgängen die Gesichter nicht identisch waren (Nicht-Identitätsbedingung). Dafür wurden pro Set jeweils zwei hoch- und niedrig distinkte Bilder per Zufall ausgewählt. Diese waren zunächst als identische Darbietung vorgesehen. Da für diejenigen Durchgänge, in denen zwei verschiedenen Gesichter gezeigt werden sollten, auch zwei unterschiedliche Stimuli benötigt wurden, mussten vier weitere Bilder aus jedem Set jeweils aus den hoch- und niedrig- distinkten Bildern per Zufall ausgewählt werden. Diese Bilder wurden dann getrennt nach Darbietungsset und hoher bzw. niedriger Erinnerbarkeit wiederum in eine Zufallsreihenfolge gebracht, wobei das erste und das dritte Bild als erste Präsentation in einem Durchgang verwendet wurden, und das zweite bzw. vierte Bild als zweite Präsentation dem ersten und dritten Bild zugeordnet wurden. Abbildung 18 zeigt noch einmal zum besseren Verständnis die Prozedur der

²⁰ Die unterschiedliche Anzahl an Bildern entsteht durch die verschiedenen Gesamtzahlen an verfügbaren Bildern, 70 latein-amerikanische vs. 80 afro-amerikanische, türkische bzw. deutsche Gesichter. Bei zwei Sets à 35 Bildern (Latein-Amerikaner) werden die mittleren 20 für den Wiedererkennenstest verwendet. Die restlichen 15 Bilder setzen sich zusammen aus sieben hoch bzw. acht niedrig 'erinnerbaren' Bildern.

Stimuluszuweisung zu den Bedingungen in der Delayed-Matching Aufgabe. Zusätzlich wurden in einem zweiten Faktor die Durchgänge mit identischen Gesichtern mit den Durchgängen mit nichtidentischen Gesichtern ausbalanciert. Das bedeutet, dass die Gesichter in Durchgängen in der identischen Bedingung nun mit einem nicht-identischen Gesicht gepaart wurden und vice versa. Dazu wurde dem ersten Gesicht eines Durchgangs in der identischen Bedingung das zweite Gesicht eines Durchgangs in der Nicht-Identitätsbedingung zugewiesen, dem ersten Gesicht eines Durchgangs in der Nicht-Identitätsbedingung wurde das identische Gesicht wiederum zugewiesen.



⚡ Auswahl von jeweils 6 den mittleren Ratings am nächsten gelegenen Bildern aus jeder Kombination

⚡ Aufteilung der Stimuli zur Delayed-Matching-Aufgabe:

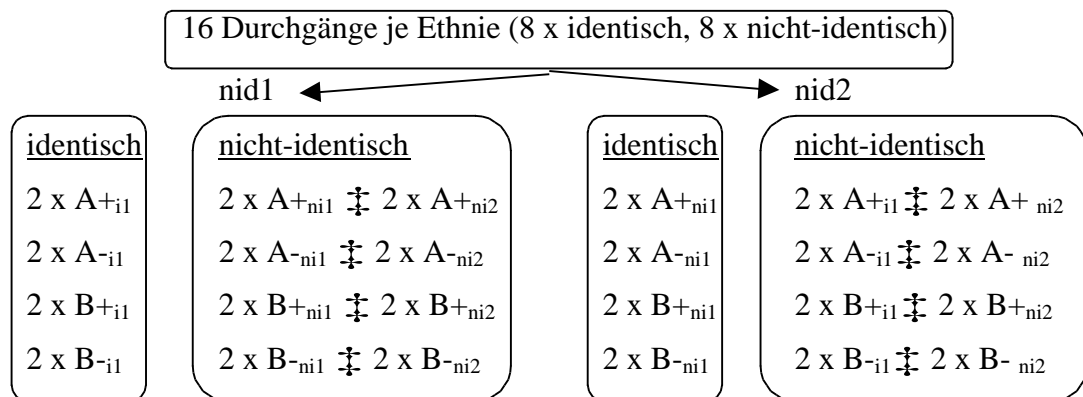


Abbildung 18. Stimuluszuweisung in der Delayed-Matching-Aufgabe bei jeder Ethnie. Bei den nicht-identischen Bedingungen geben die Zahlen in Klammern, die Kombination des ersten Bildes mit dem zweiten Bild an.

6.3.3 Stimulus-Darbietung

Die zwei Experimente wurden an zwei Apple 'iMac' Computern mit einer Taktfrequenz von 300 bzw. 350 Mhz, einem Arbeitsspeicher von jeweils 192 MB und einem 15"-Monitor erhoben. Die Farbtiefe betrug 16 Bit (65536 Farben) bei einer Bildschirmauflösung von 1024x768 Pixel und einer Bildwiederholfrequenz von 75 Hz. Die Programmierung erfolgte mit der Applikation 'Superlab' (Version 1.75) der Firma Cedrus (<http://www.cedrus.com>). Die Stimuli hatten eine Größe von 432x648 Pixel bei einer Auflösung von 72x72 dpi. Die Farbtiefe der Stimuli betrug 16,7 Mio. Farben. Der Bildschirmhintergrund war für alle Bilder schwarz.

6.3.4 Versuchsdurchführung

Die Daten der BGS-Beamten wurden direkt am Flughafen Frankfurt erhoben. Dazu konnten in einem ca. 3m x 4m großen Büro des Bundesgrenzschutzes im Terminal 1 des Flughafens die Computer aufgebaut werden. Die Beamten wurden freundlicherweise während der Dienstzeit freigestellt, um an der Untersuchung teilzunehmen. Die Erhebung erfolgte nach Absprache mit der Dienststellenleitung an drei Tagen.²¹

Es wurden immer zwei Beamte gleichzeitig getestet. Der Ablauf war wie folgt: Zunächst wurden die Beamten begrüßt und gebeten, sich an einen der zwei Plätze zu setzen. Anschließend wurden sie kurz über den Ablauf der Untersuchung aufgeklärt, und es wurde ihnen mitgeteilt, dass sie, falls Interesse für die Rückmeldung ihrer Leistungen in den Tests besteht, sie sich ihre Codenummer notieren sollten. Anschließend begann die Untersuchung. Nach Beendigung des zweiten Experimentes wurde den Beamten für die Teilnahme gedankt, und sie verließen den Raum. Die nächsten zwei Teilnehmer erschienen zumeist kurz im Anschluss. Bis auf die letzten zwei Beamten, die an der Untersuchung teilnahmen, wurden sie am Ende des zweiten Experimentes nicht über die Ziele der Untersuchung aufgeklärt. Sie wurden deshalb nicht aufgeklärt, um zu verhindern, dass Beamte, die bereits teilgenommen hatten, andere Beamte über die Untersuchung aufklären und somit die Ergebnisse verfälschen könnten. Stattdessen wurde mit der Rückmeldung an

²¹ Ich danke an dieser Stelle den teilnehmenden Beamten sowie Herrn Gabriel für den reibungslosen Ablauf der Untersuchung und die gute Zusammenarbeit. Ebenso danke ich Frau Dipl.-Psych. Andrea Schergung für die Hilfe bei der Datenerhebung

die Beamten und die Dienststellenleitung nach Auswertung aller Daten eine Aufklärung über die Ziele der Untersuchung durchgeführt.

Die Daten der Studierenden wurden in einem ebenfalls ca. 3m x 4m großen Raum am Fachbereich Psychologie und Sportwissenschaft an der Justus-Liebig-Universität Gießen erhoben. Auch bei dieser Gruppe wurden jeweils zwei Teilnehmer gleichzeitig getestet. Der Ablauf der Untersuchung bei dieser Gruppe war analog zum Versuchsablauf bei den Beamten vom BGS. Sie wurden ebenfalls nur kurz über den Ablauf der Untersuchung aufgeklärt.

6.3.5 Versuchsablauf

Jede Sitzung begann mit dem Wiedererkennenstest. Zunächst wurde den Teilnehmern die Instruktionen auf dem Bildschirm präsentiert (siehe Anhang A7). Anschließend folgte die Darbietungsphase mit 80 Gesichtern (20 Gesichter je Ethnie), die in randomisierter Reihenfolge jeweils 5000 ms dargeboten wurden. Das Inter-Trial-Intervall betrug 500 ms. Bevor das erste Gesicht auf dem Bildschirm erschien, wurden ein weißes Quadrat (Ⓢ), ein weißer Kreis (⊖) und ein liegendes weißes Kreuz (×), ebenfalls jeweils für 5000 ms und einem Inter-Trial-Intervall von 500 ms, dargeboten, um den Blick der Teilnehmer auf die Bildschirmmitte zu lenken, wo dann im Anschluss an das weiße Kreuz die Gesichter erschienen. Die drei Zeichen waren, wie die Stimulus-Gesichter, auf dem Bildschirm zentriert. Nach der Darbietungsphase sollten die ersten 13 Items des Kontaktfragebogens beantwortet werden. Dieser Teil der Sitzung nahm, je nach Bearbeitungsgeschwindigkeit, ca. fünf bis zehn Minuten in Anspruch. Der Wiedererkennenstest, in dem die 80 'alten' sowie 80 'neue' Gesichter in randomisierter Reihenfolge dargeboten wurden, folgte im Anschluss an die erste Bearbeitung des Kontaktfragebogens. Die Teilnehmer folgten wiederum den Instruktionen auf dem Bildschirm (siehe Anhang A7). Nach dem Wiedererkennenstest sollten die Items 14 bis 25 des Kontaktfragebogens bearbeitet werden, in denen es u.a. um die Einschätzung der Anzahl täglicher Kontakte mit Personen anderer ethnischer Gruppen ging. Während dieser Zeit wurde der Programmablauf für die Delayed-Matching Aufgabe geladen, und die Teilnehmer konnten sich nach Bearbeitung der Fragebogen-Items dieser Aufgabe zuwenden. Die Instruktionen für die Delayed-Matching Aufgabe sind in Anhang A8 aufgeführt. Nachdem die Instruktionen gelesen wurden, folgten zwei Beispieldurchgänge, anhand derer die Teilnehmer in die Prozedur eingewiesen wurden. Jeder Durchgang war

wie folgt aufgebaut: Ein Gesicht wurde für 1000 ms in der Frontalansicht dargeboten, danach folgte für 1500 ms ein 'Filler-Screen', der den gesamten Bereich des Bildschirms mit Zufallspunkten verdeckte. Nach dem Filler-Screen folgte entweder das gleiche oder ein anderes Gesicht, je nach Bedingung ebenfalls in Frontalansicht oder im rechten Halbprofil. Danach wurde ohne Zeitbeschränkung ein komplett schwarzes Bild eingeblendet, währenddessen die Teilnehmer die Taste 'c' bzw. 'm' drücken sollten, um anzuzeigen, ob die beiden Gesichter identisch waren oder nicht. Nach Drücken einer der beiden Tasten wurde durch Präsentation der Worte 'FALSCH' oder 'RICHTIG' für 2000 ms den Teilnehmern rückgemeldet, ob Ihre Wahl korrekt war oder nicht. Das Inter-Trial-Intervall betrug 500 ms. Die Delayed-Matching Aufgabe bestand insgesamt aus 64 trials (16 je Ethnie), die in randomisierter Reihenfolge dargeboten wurden.

6.4 Ergebnisse

In diesem Abschnitt wird zunächst die Annahme überprüft, nach der die beiden Berufsgruppen Unterschiede in der Kontakthäufigkeit mit verschiedenen Ethnien aufweisen. Anschließend werden die Ergebnisse des Wiedererkennenstests und der Delayed-Matching Aufgabe anhand der Maße der Signal-Entdeckungs-Theorie dargestellt.

6.4.1 Kontaktfragebogen

Da sich in diesem Experiment die beiden Gruppen hinsichtlich der Kontakthäufigkeit unterscheiden sollten, wurde zunächst der Kontaktfragebogen bezüglich der hierfür relevanten Items ausgewertet. Es zeigte sich, dass sich die beiden Gruppen hinsichtlich der geschätzten Kontakthäufigkeit mit anderen Personen und Personen anderer Ethnien voneinander unterscheiden.²² Eine Varianzanalyse mit den Between-Subjects-Faktoren 'Berufsgruppe' (BGS, Studierende) und 'Geschlecht' (männlich, weiblich) sowie dem Within-Subjects Faktor 'Stimulusethnie' (Deutsche, Weiße, nicht-deutsche weiße Europäer, ferner Osten, mittlerer Osten, naher Osten, Afrika), zeigte einen

²² Unglücklicherweise fehlen von sieben Teilnehmern zu dieser Frage die Daten, weil sie anstatt einer absoluten Schätzung der Personen eine relative Schätzung der Kontakthäufigkeit vornahmen. Diese Daten wurden für die Analyse der Kontakthäufigkeit aus der Berechnung ausgeschlossen.

signifikanten Haupteffekt für den Faktor Berufsgruppe, $F(1,53) = 27.77$, $p = .001$, $r = .59$. BGS-Beamte geben an, im Mittel $M = 288.3$ verschiedene Personen täglich zu sehen, während dieser Wert bei den Studierenden bei $M = 35.4$ liegt. Ein ebenfalls signifikanter Effekt zeigte sich für den Faktor 'Stimulusethnie', $F(6,318) = 18.34$, $p = .001$, $\eta^2 = .33$ (siehe Tabelle 19).

Tabelle 19.

Geschätzte Anzahl täglicher Begegnungen mit Personen verschiedener Ethnien für BGS-Beamte und Studierende

Kontakt mit Ethnie	BGS	Studierende	<u>M</u>	Ratio
Deutsche	489.6 ^a	150.1 ^a	319.9	3.3:1
Weißer	515.4 ^a	51.4 ^a	283.4	10:1
N/S/O Europa	265.0 ^a	16.9 ^a	141.0	15.7:1
Ferner Osten	370.0 ^a	5.2 ^a	187.6	71.2:1
mittl. Osten	95.4 ^a	5.3 ^a	50.3	18:1
Naher Osten	192.3 ^a	13.6 ^a	103.0	14.1:1
Afrika	90.2 ^a	5.2 ^a	47.7	17.3:1
Gesamt	288.3 ^a	35.4 ^a	161.8	8.1:1

Anmerkung: ^a Alle Zellenmittelwerte innerhalb einer Zeile unterscheiden sich nach Bonferroni-Adjustierung signifikant (post-hoc Vergleiche, $p < .001$).

Es gab keine Unterschiede zwischen den Geschlechtern und ebenfalls keine Interaktion der Faktoren 'Geschlecht' und 'Berufsgruppe' (für beide Vergleiche gilt: $F < 1$, ns). Damit kann die Auswahl der Berufsgruppen hinsichtlich der vermuteten Kontakthäufigkeit als erfolgreich betrachtet werden.

6.4.2 Wiedererkennenstest

Die Daten des Wiedererkennenstests wurden bezüglich der Maße der Signal-Entdeckungs-Theorie ausgewertet. Zunächst werden die Daten für die Treffer und falschen Alarme berichtet und die Hypothesen sowie anschließende Varianzanalysen dargestellt. Anschließend wird auf den Leistungsindex \underline{A}' sowie die Reaktionsneigung \underline{B}'' eingegangen. In Tabelle 20 sind die Mittelwerte der Treffer, der falschen Alarme, von \underline{A}' und \underline{B}'' zusammengefasst. Das Alpha-Niveau wird nach Bonferroni-Adjustierung bei fünf geplanten Kontrasten für die formulierten Hypothesen auf $\alpha = .010$ festgelegt.

Die Hypothese, nach der die BGS-Beamten insgesamt mehr Treffer haben sollten als die Studierenden wurde mit einem geplanten Kontrast getestet, bei dem die Differenz der Treffer der BGS-Beamten und der Studierenden größer als null sein sollte. Dieser Kontrast zeigte sich als nicht signifikant von Null verschieden, $F(1,48) = .14$, $p = .711$, $r = .054$. Die BGS-Beamten haben im Mittel eine Trefferrate von $\underline{M} = .561$, während es bei den Studierenden im Mittel $\underline{M} = .573$ Treffer sind. Die Hypothese, nach der die Anzahl der Treffer bei den vier Stimulusethnien je nach Ethnie der Teilnehmer linear in Abhängigkeit der Distanz von der eigenen Ethnie ansteigt wurde mithilfe linearer Kontraste getestet. Diese Hypothese konnte bestätigt werden, $F(1,48) = 40.25$, $p = .001$, $r = .68$. Die mittlere Trefferanzahl bei afro-amerikanischen Gesichtern beträgt $\underline{M}_{a-a} = .454$, bei latein-amerikanischen Gesichtern $\underline{M}_{l-a} = .593$, bei türkischen Gesichtern $\underline{M}_{t\ddot{u}} = .614$, und bei deutschen Gesichtern $\underline{M}_{de} = .608$. Lineare Kontraste werden bei separater Berechnung für beide Berufsgruppen signifikant. Bei den BGS-Beamten ist dieser Kontrast, $F(1,24) = 9.09$, $p = .006$, $r = .52$. Die mittlere Trefferanzahl der BGS-Beamten bei afro-amerikanischen Gesichtern beträgt $\underline{M}_{a-a} = .477$, bei latein-amerikanischen Gesichtern $\underline{M}_{l-a} = .588$, bei türkischen Gesichtern $\underline{M}_{t\ddot{u}} = .605$, und bei deutschen Gesichtern $\underline{M}_{de} = .575$. Bei den Studierenden betragen diese Werte entsprechend, $F(1,24) = 34.32$, $p = .001$, $r = .77$. Die mittlere Trefferanzahl liegt bei dieser Berufsgruppe bei afro-amerikanischen Gesichtern bei $\underline{M}_{a-a} = .431$, bei latein-amerikanischen Gesichtern $\underline{M}_{l-a} = .598$, bei türkischen Gesichtern $\underline{M}_{t\ddot{u}} = .623$, und bei deutschen Gesichtern $\underline{M}_{de} = .641$. Damit kann gleichzeitig die Hypothese verworfen werden, dass es bei den BGS-Beamten keine Unterschiede zwischen den Stimulusethnien bezüglich der Anzahl der Treffer gibt.

Tabelle 20

Mittelwerte der Treffer, der falschen Alarme, des Leistungsindex A' und der Reaktionsneigung B'' nach Berufsgruppe und Stimulusethnie im Wiedererkennenstest

Gesichter von	BGS-Beamte	Studierende	<u>M</u> :
Afro-Amerikanern:			
Treffer	.477	.431	.454
Falsche Alarme	.344	.380	.362
A'	.612 ^a	.544 ^a	.578
B''	.073	.022	.048
Latein-Amerikanern:			
Treffer	.588	.598	.593
Falsche Alarme	.405	.423	.414
A'	.651	.651	.651
B''	.053	-.027	.013
Türken:			
Treffer	.605	.623	.614
Falsche Alarme	.364	.363	.364
A'	.693	.706	.699
B''	.028	-.036	-.004
Deutschen:			
Treffer	.575 ^a	.641 ^a	.608
Falsche Alarme	.327 ^a	.281 ^a	.304
A'	.696 ^a	.768 ^a	.732
B''	.072	.074	.073
<u>M</u> :			
Treffer	.561	.573	.567
Falsche Alarme	.360	.362	.361
A'	.663	.667	.665
B''	.057	.033	.045

Anmerkung: ^a Die Zellenmittelwerte innerhalb einer Zeile unterscheiden sich nach Bonferroni-Adjustierung signifikant (post-hoc Vergleiche, $p < .001$).

Eine Varianzanalyse mit den Treffern als abhängige Variable und den Between-Subjects Faktoren 'Berufsgruppe' (BGS-Beamte, Studierende), 'Geschlecht' (weiblich, männlich), 'Set' (A,B) und 'Taste' (c,m) und dem Within-Subjects Faktor 'Stimulusethnie' (Afro-Amerikaner, Latein-Amerikaner, Türken, Deutsche) zeigte darüber hinaus eine marginal signifikante Interaktion der Faktoren 'Geschlecht' und 'Set', $F(1,48) = 4.1$, $p = .048$, $r = .28$. Während die männlichen Teilnehmer beim Set B mehr Treffer aufweisen, $M_B = .623$, als beim Set A, $M_A = .534$, ist dieses Verhältnis bei den weiblichen Teilnehmern umgekehrt, $M_B = .533$, $M_A = .580$. Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 2.51$, alle $p_s > .061$).

Analog zu den Treffern wurde auch für die falschen Alarme als abhängige Variable eine Varianzanalyse und den Between-Subjects Faktoren 'Berufsgruppe' (BGS-Beamte,

Studierende), 'Geschlecht' (weiblich, männlich), 'Set' (A,B) und 'Taste' (c,m) und dem Within-Subjects Faktor 'Stimulusethnie' (Afro-Amerikaner, Latein-Amerikaner, Türken, Deutsche) durchgeführt. Es zeigte sich ebenfalls ein signifikanter Haupteffekt für den Faktor 'Stimulusethnie', $F(3,144) = 8.30$, $p = .001$, $\eta^2 = .18$, mit einer mittleren Anzahl an Fehlalarmen für die afro-amerikanischen Gesichter von $M_{a-a} = .362$, bei latein-amerikanischen Gesichtern $M_{l-a} = .414$, bei türkischen Gesichtern $M_{t\ddot{u}} = .363$, und bei deutschen Gesichtern $M_{de} = .304$. Es zeigte sich kein Unterschied zwischen den Berufsgruppen, $F(1,48) = .00$, ns und keine Interaktion der Faktoren 'Berufsgruppe' und 'Stimulusethnie' $F(3,144) = 1.25$, ns . Alle anderen Haupteffekte und Interaktionen waren nicht signifikant, alle $F_s < 2.77$, alle $p_s > .102$.

Die Hypothesen bezüglich der Analyse der Treffer gelten ebenfalls für den Leistungsindex A' . Der Kontrast nach dem die Differenz der Treffer der BGS-Beamten und der Studierenden größer als null sein sollte, d.h. die BGS-Beamten sollten insgesamt besser sein als die Studierenden, war nicht signifikant von Null verschieden, $F < 1$, ns . Die BGS-Beamten haben im Mittel einen Leistungsindex von $A' = .663$ während die Studierenden einen Leistungsindex von $A' = .667$ aufweisen. Die Hypothese, nach der der Leistungsindex bei den vier Stimulusethnien linear in Abhängigkeit der Distanz von der eigenen Ethnie ansteigt wurde mithilfe linearer Kontraste getestet. Diese Hypothese konnte bestätigt werden, $F(1,48) = 82.73$, $p = .001$, $r = .80$. Der Leistungsindex bei afro-amerikanischen Gesichtern beträgt $A'_{a-a} = .578$, bei latein-amerikanischen Gesichtern $A'_{l-a} = .651$, bei türkischen Gesichtern $A'_{t\ddot{u}} = .699$, und bei deutschen Gesichtern $A'_{de} = .732$. Der lineare Kontrast ist bei beiden Berufsgruppen signifikant. Bei den BGS-Beamten ist $F(1,24) = 17.53$, $p = .001$, $r = .65$, mit $A'_{a-a} = .612$ bei den afro-amerikanischen, $A'_{l-a} = .651$ bei den latein-amerikanischen, $A'_{t\ddot{u}} = .693$ bei den türkischen und $A'_{de} = .695$ bei den deutschen Gesichtern. Für die Studierenden gilt $F(1,24) = 68.86$, $p = .001$, $r = .86$, mit $A'_{a-a} = .544$ bei den afro-amerikanischen, $A'_{l-a} = .651$ bei den latein-amerikanischen, $A'_{t\ddot{u}} = .706$ bei den türkischen und $A'_{de} = .768$ bei den deutschen Gesichtern. Beide Berufsgruppen waren bei der Eigen-Gruppe am besten und bei den afro-amerikanischen Gesichtern am schlechtesten.

Anschließend wurde noch eine multivariate Varianzanalyse des Leistungsindex A' mit den Between-Subjects Faktoren 'Berufsgruppe' (BGS-Beamte, Studierende), 'Geschlecht' (weiblich, männlich), 'Set' (A,B) und 'Taste' (c,m) und dem Within-Subjects Faktor 'Stimulusethnie' (Afro-Amerikaner, Latein-Amerikaner, Türken, Deutsche) durchgeführt. Der Faktor 'Stimulusethnie' interagiert mit dem Faktor 'Berufsgruppe', $F(3,144) = 5.06$, $p = .002$, $\eta^2 = .12$ (siehe Abbildung 19). Während sich die Werte für A' bei

den latein-amerikanischen, $F < 1$, ns und türkischen Gesichtern, $F < 1.03$, ns, wie auch über alle Gesichter zusammen nicht zwischen den Versuchspersonengruppen unterscheiden, zeigen sich post-hoc bei afro-amerikanischen, $F(1,144) = 14.08$, $p = .001$, $r = .30$, und deutschen Gesichtern, $F(1,144) = 16.22$, $p = .001$, $r = .32$, signifikante Differenzen zwischen den Berufsgruppen. Der Leistungsindex ist mit $A'_{BGS/a-a} = .612$ bei den BGS-Beamten für afro-amerikanische Gesichter besser als bei den Studierenden $A'_{St/a-a} = .544$, während für die deutschen Gesichter das Gegenteil zutrifft mit besseren Leistungen bei den Studierenden ($A'_{St/de} = .768$) gegenüber den BGS-Beamten ($A'_{BGS/de} = .695$).

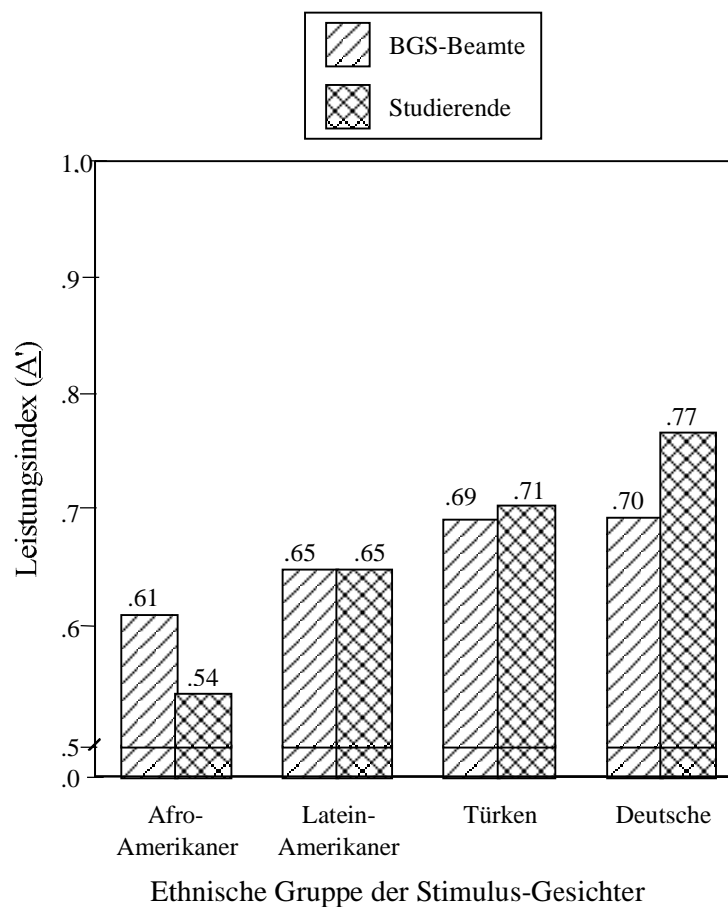


Abbildung 19. Leistungsindexwerte A' für BGS-Beamte und Studierende getrennt nach Stimulusethnie.

Es zeigte sich ebenfalls ein marginal signifikanter Haupteffekt für den Faktor 'Set', $F(1,48) = 6.18$, $p = .016$, $r = .34$. Der Leistungsindex war bei Set A mit $A' = .644$ schlechter als bei Set B mit $A' = .686$. Die Interaktion der Faktoren 'Stimulusethnie' und

'Set' war ebenfalls marginal signifikant, $F(3,144) = 3.15$, $p = .027$, $\eta^2 = .08$. Bis auf die türkischen Gesichter war bei den Stimulusethnien die Leistung bei Set B besser als bei Set A. Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 2.63$, alle $p_s > .111$).

Die Auswertung der Reaktionsneigung B'' zeigt, dass es keine systematischen Ausprägungen der Reaktionsneigung gibt. Eine Varianzanalyse der Reaktionsneigung B'' mit den Between-Subjects Faktoren 'Berufsgruppe' (BGS-Beamte, Studierende), 'Geschlecht' (weiblich, männlich), 'Set' (A,B) und 'Taste' (c,m) und dem Within-Subjects Faktor 'Stimulusethnie' (Afro-Amerikaner, Latein-Amerikaner, Türken, Deutsche) zeigte lediglich eine marginal signifikante Interaktion der Faktoren 'Geschlecht' und 'Set', $F(1,48) = 5.07$, $p = .029$, $r = .31$. Die männlichen Teilnehmer zeigten bei Set B eine Reaktionsneigung von $B'' = -.029$, während die weiblichen Teilnehmer eine Reaktionsneigung von $B'' = .086$ zeigten, $F(1,48) = 5.25$, $p = .026$, $r = .31$. Beim Set A zeigte sich kein signifikanter Unterschied, $F < 1$, *ns*. Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 2.35$, alle $p_s > .132$).

6.4.3 Diskussion des Wiedererkennenstests

Entgegen der Hypothese, dass BGS-Beamten sowohl mehr Treffer hätten als auch beim Leistungsindex A' insgesamt besser seien als die Studierenden, wurde im Wiedererkennenstest keine Bestätigung für diese Annahme gefunden. BGS-Beamte und Studierende sind insgesamt auf dem gleichen Leistungsniveau. Es scheint demnach nicht so zu sein, dass ein durchaus deutlich vermehrter Kontakt mit anderen Ethnien und auch der eigenen Ethnie zu einem besseren Gedächtnis für Gesichter führt. Interessant ist jedoch die Betrachtung der Leistungen bei den Stimulusethnien innerhalb der beiden untersuchten Gruppen. Zwar zeigen beide Berufsgruppen separat bei dem berechneten Kontrast über den Faktor 'Stimulusethnie' einen starken Effekt, $r = .65$ bei BGS-Beamten bzw. $r = .86$ bei den Studierenden, dieser ist bei den Studierenden aber höher ausgeprägt als bei den BGS-Beamten. Während sich bei den Studierenden ein klarer linearer Trend bei den Stimulusethnien abzeichnet (siehe Abbildung 19), ist dieser Trend bei den BGS-Beamten deutlich schwächer ausgeprägt. Auffallend ist hierbei außerdem noch die Differenz in der Leistung zwischen den beiden Berufsgruppen bei den afro-amerikanischen und den deutschen Gesichtern. Die BGS-Beamten sind bei den afro-amerikanischen Gesichtern deutlich besser als die Studierenden, $r = .30$. Bei den deutschen Gesichtern zeigen

wiederum die Studierenden bessere Leistungen als die BGS-Beamten, $\underline{r} = .32$. Was bedeuten diese Ergebnisse für die Annahmen des In-Group/Out-Group Modells und die Speicherung von Gesichtern in einem multidimensionalen Raum? Für die Theorie der Speicherung von Gesichtern in einem multidimensionalen Raum sind diese Ergebnisse als Bestätigung der Annahmen zu werten. Mit zunehmender Distanz der Ethnien zur eigenen Ethnie, in diesem Fall ausschließlich deutsche Teilnehmer, sinkt die Fähigkeit, Gesichter zu differenzieren und wiederzuerkennen. Für das In-Group/Out-Group Modell bedeuten diese Ergebnisse ebenfalls eine Bestätigung der Annahmen insofern, dass eine Verschlechterung der Leistung bei Fremd-Gruppen Gesichtern auftritt. Ob diese Verschlechterung mit einer merkmalsbasierten Enkodierung von Fremd-Gruppen Gesichtern zusammenhängt, kann im zweiten Experiment geklärt werden, in dem die Enkodierungsstrategie dadurch operationalisiert wird, dass die Vergleichsgesichter im Halbprofil bzw. Frontalansicht dargeboten werden. Bis hierhin sind die Annahmen der Speicherung in einem multidimensionalen Raum und die des In-Group/Out-Group Modells noch nicht trennbar. Weiterhin hat keine der beiden Berufsgruppen eine konservative bzw. liberale Reaktionsneigung gezeigt. Die Werte von \underline{B}'' für die Stimuluse ethnien variieren zwischen .074 und -.036.

Die Frage inwiefern die Kontakthäufigkeit einen Einfluss auf den Cross-Race Bias hat, kann mit den Ergebnissen des Wiedererkennentests zumindest teilweise geklärt werden. Wie postuliert wurde, unterscheiden sich die beiden Berufsgruppen hinsichtlich ihrer Kontakthäufigkeit mit anderen Ethnien stark voneinander. Bezüglich der Kontaktqualität sind keine Unterschiede zu vermuten, obwohl ein Beweis hierfür nicht erbracht werden kann. Diese Vermutung stützt sich auf die Art der Tätigkeit, die die BGS-Beamten ausüben, nämlich die Passkontrolle, bei der i. A. keine besondere Interaktion zwischen den Beamten und den Passagieren stattfindet. Wie bereits angesprochen, ist in diesem Zusammenhang die Betrachtung der Leistungen der beiden Berufsgruppen bei den untersuchten Stimuluse ethnien von Interesse. Bei den Studierenden zeigen sich zwischen den Stimuluse ethnien große Differenzen (zwischen $\underline{A}' = .544$ bei den afro-amerikanischen Gesichtern und $\underline{A}' = .768$ bei den deutschen Gesichtern). Auch zeigt sich bei den Studierenden ein starker 'mirror effect' mit vielen Treffern und wenigen falschen Alarmen bei deutschen Gesichtern, bzw. wenig Treffern und vielen falschen Alarmen bei den afro-amerikanischen Gesichtern. Dies ist bei den BGS-Beamten weniger stark ausgeprägt. Bei dieser Berufsgruppe ist die Differenz der Leistungen zwischen den Stimuluse ethnien geringer ($\underline{A}' = .612$ bei den afro-amerikanischen Gesichtern und $\underline{A}' = .696$ bei den deutschen Gesichtern), und ebenfalls ist bei dieser Berufsgruppe ein weniger stark

ausgeprägter 'mirror effect' zu beobachten. Dies bedeutet für die Diskussion der Frage, ob der Cross-Race Bias auf die Kontakthäufigkeit oder die Kontaktqualität zurückzuführen ist, zumindest eine Beeinflussung, d.h. eine Verminderung des Cross-Race Bias durch die Kontakthäufigkeit mit Personen anderer ethnischer Gruppen. Die Vermutung, dass eine solche geringere Ausprägung des Cross-Race Bias auf die größere Häufigkeit des Kontaktes der BGS-Beamten mit anderen Ethnien zurückzuführen ist, lässt sich durch die weitaus höheren berichteten Kontakthäufigkeiten der BGS-Beamten stützen. Wenn allerdings in Betracht gezogen wird, dass die BGS-Beamten im Durchschnitt ca. 8 Jahre älter sind als die Studierenden und ebenfalls im Durchschnitt ca. 6 1/2 Jahre am Flughafen Frankfurt bei der Passkontrolle tätig sind und darüber hinaus mit weitaus mehr Personen täglichen Kontakt haben, werden die Ergebnisse relativiert. Die BGS-Beamten zeigen zwar einen relativ geringen Cross-Race Bias, in Anbetracht der im Vergleich zu den Studierenden hohen Anzahl des Kontaktes mit anderen Ethnien wäre allerdings kein Cross-Race Bias vermutet worden. Bei Betrachtung der Werte bei den deutschen Stimulusgesichtern zwischen den beiden Berufsgruppen fällt zudem noch auf, dass die BGS-Beamten schlechtere Leistungen zeigen als die Studierenden. Diese Ergebnisse legen die Vermutung nahe, dass sich mit zunehmendem Kontakt mit anderen Ethnien, die Wiedererkennensleistungen bei den Ethnien einander angleichen. Ein Grund hierfür könnte sein, wenn die Speicherung von Gesichtern in einem multidimensionalen Raum angesprochen wird, dieser multidimensionale Raum nicht unendlich ist und die Möglichkeiten der Wiedererkennung von Gesichtern durch Gedächtnis und Enkodierungsprozesse begrenzt ist. Anscheinend wird eine Verbesserung der Wiedererkennensleistung bei Fremd-Gruppen Gesichtern durch eine Verschlechterung der Wiedererkennensleistung bei Eigen-Gruppen Gesichtern 'erkauft'. Die Prozesse, die dabei eine Rolle spielen, müssten weiter erforscht werden. Nichtsdestotrotz kann die Hypothese, dass eine höhere Kontakthäufigkeit zur Reduzierung eines Cross-Race Bias beiträgt, gestützt werden.

6.4.4 Delayed-Matching-Aufgabe

Die Daten der Delayed-Matching-Aufgabe wurden ebenfalls bezüglich der Maße der Signal-Entdeckungs-Theorie ausgewertet. Dabei gilt, dass eine richtige Identifizierung gleichbedeutend mit einer 'Ja'-Antwort bei identischen Gesichtern ist und ein falscher Alarm als eine 'Nein'-Antwort bei identischen Gesichtern klassifiziert wird. Zunächst

werden die Hypothesen für die Treffer berichtet und die Varianzanalysen für die Treffer und die falschen Alarme dargestellt. Anschließend wird auf den Leistungsindex \underline{A}' sowie die Reaktionsneigung \underline{B}'' eingegangen. Im Gegensatz zum Wiedererkennenstest wurden bei der Delayed-Matching Aufgabe die Faktoren 'Ansicht des Gesichtes' und 'Identitätspräsentation' eingeführt (siehe Abschnitt 6.2.1). Das Alpha-Niveau wird nach Bonferroni-Adjustierung bei sechs geplanten Kontrasten für die formulierten Hypothesen auf $\alpha = .008$ festgelegt

Die Hypothese, nach der die BGS-Beamten insgesamt mehr Treffer haben sollten als die Studierenden wurde mit einem geplanten Kontrast getestet, bei dem die Differenz der Treffer der BGS-Beamten und der Studierenden größer als null sein sollte. Dieser Kontrast zeigte sich als marginal signifikant von Null verschieden, $\underline{F}(1,32) = 5.43$, $p = .026$, $r = .38$. Die BGS-Beamten haben im Mittel eine Trefferrate von $\underline{M} = .956$ während es bei den Studierenden im Mittel $\underline{M} = .928$ Treffer sind. Die Hypothese, nach der die Anzahl der Treffer bei den vier Stimulusethnien je nach Ethnie der Teilnehmer linear in Abhängigkeit der Distanz von der eigenen Ethnie ansteigt wurde mithilfe linearer Kontraste getestet. Diese Hypothese konnte bestätigt werden, $\underline{F}(1,32) = 56.66$, $p = .001$, $r = .80$. Die mittlere Trefferanzahl bei afro-amerikanischen Gesichtern beträgt $\underline{M}_{a-a} = .857$, bei latein-amerikanischen Gesichtern $\underline{M}_{l-a} = .965$, bei türkischen Gesichtern $\underline{M}_{t-ü} = .971$, und bei deutschen Gesichtern $\underline{M}_{de} = .975$. Lineare Kontraste werden bei separater Berechnung für beide Berufsgruppen signifikant. Bei den BGS-Beamten ist dieser Kontrast $\underline{F}(1,16) = 15.35$, $p = .001$, $r = .70$. Die mittlere Trefferanzahl der BGS-Beamten bei afro-amerikanischen Gesichtern beträgt $\underline{M}_{a-a} = .906$, bei latein-amerikanischen Gesichtern $\underline{M}_{l-a} = .965$, bei türkischen Gesichtern $\underline{M}_{t-ü} = .980$, und bei deutschen Gesichtern $\underline{M}_{de} = .973$. Bei den Studierenden betragen diese Werte entsprechend $\underline{F}(1,16) = 41.58$, $p = .001$, $r = .85$. Die mittlere Trefferanzahl liegt bei dieser Berufsgruppe bei afro-amerikanischen Gesichtern $\underline{M}_{a-a} = .809$, bei latein-amerikanischen Gesichtern $\underline{M}_{l-a} = .965$, bei türkischen Gesichtern $\underline{M}_{t-ü} = .961$, und bei deutschen Gesichtern $\underline{M}_{de} = .977$.

Die Vermutung, dass die Ansicht der Gesichter einen Einfluss auf die Anzahl der Treffer hat wird mit einem signifikanten Haupteffekt, $\underline{F}(1,32) = 51.10$, $p = .001$, $r = .78$, bestätigt. Gesichter, die im Anschluss an den Filler-Screen ebenfalls in der Frontalansicht gezeigt wurden, hatten eine mittlere Trefferanzahl von $\underline{M} = .985$, verglichen mit Gesichtern im Halbprofil, $\underline{M} = .898$. Bei alleiniger Analyse der Gesichter im Halbprofil zeigt sich ein differenzierteres Bild. Zwar unterscheiden sich die beiden Berufsgruppen nur marginal voneinander, jedoch zeigt sich eine signifikante Interaktion der Faktoren 'Stimulusethnie' und 'Berufsgruppe', $\underline{F}(3,48) = 6.41$, $p = .001$, $\eta^2 = .29$. Dies ist alleine

darauf begründet, dass sich die Treffer der BGS-Beamten ($\underline{M} = .828$) bei den afro-amerikanischen Gesichtern von denen der Studierenden ($\underline{M} = .648$) signifikant unterscheiden $\underline{F}(1,48) = 48.46$, $\underline{p} = .001$, $\underline{r} = .71$.

Eine anschließende Varianzanalyse mit den Treffern als abhängige Variable und den Between-Subjects Faktoren 'Berufsgruppe' (BGS-Beamte, Studierende), 'Geschlecht' (weiblich, männlich), 'Ansicht' (Frontal, Halbprofil), 'Identitätspräsentation' (identisch, nicht identisch) und 'Taste' (c,m) und dem Within-Subjects Faktor 'Stimulusethnie' (Afro-Amerikaner, Latein-Amerikaner, Türken, Deutsche) zeigte insgesamt eine signifikante Interaktion der Faktoren 'Berufsgruppe' und 'Stimulusethnie', $\underline{F}(3,96) = 5.94$, $\underline{p} = .001$, $\eta^2 = .16$, wobei dieser Interaktionseffekt darauf zurückzuführen ist, dass bei den afro-amerikanischen Gesichtern die BGS-Beamten signifikant mit $\underline{M} = .906$ mehr Treffer haben als die Studierenden mit $\underline{M} = .809$, $\underline{F}(1,96) = 61.0$, $\underline{p} = .001$, $\underline{r} = .62$ (siehe Tabelle 21).

Tabelle 21

Mittelwerte der Treffer, der falschen Alarme, des Leistungsindex A' und der Reaktionsneigung B'' nach Berufsgruppe und Stimulusethnie in der Delayed-Matching Aufgabe

Gesichter von	BGS-Beamte	Studierende	<u>M</u> :
Afro-Amerikanern:			
Treffer	.906 ^a	.809 ^a	.858
Falsche Alarme	.082 ^a	.055 ^a	.069
<u>A'</u>	.945 ^a	.920 ^a	.933
<u>B''</u>	.080 ^a	.368 ^a	.224
Latein-Amerikanern:			
Treffer	.965	.965	.965
Falsche Alarme	.082	.098	.090
<u>A'</u>	.968	.961	.965
<u>B''</u>	-.212	-.375	-.294
Türken:			
Treffer	.981	.961	.971
Falsche Alarme	.031	.043	.037
<u>A'</u>	.985	.977	.981
<u>B''</u>	-.120	-.054	-.087
Deutschen:			
Treffer	.973	.977	.975
Falsche Alarme	.020	.016	.018
<u>A'</u>	.986	.990	.988
<u>B''</u>	.060	.060	.060
<u>M</u> :			
Treffer	.956 ^a	.928 ^a	.942
Falsche Alarme	.054	.053	.054
<u>A'</u>	.971	.962	.967
<u>B''</u>	-.048	-.001	-.025

Anmerkung: ^a Die Zellenmittelwerte innerhalb einer Zeile unterscheiden sich nach Bonferroni-Adjustierung signifikant (post-hoc Vergleiche, $p < .001$).

Die beiden Faktoren 'Stimulusethnie' und 'Ansicht' interagieren signifikant, $F(3,96) = 27.35$, $p = .001$, $\eta^2 = .46$. Während in der Frontalansicht die bekannten Gesichter aller vier Ethnien nahezu vollständig wiedererkannt werden ($\underline{M}_{a-a} = .977$, $\underline{M}_{l-a} = .988$, $\underline{M}_{Tü} = .992$, $\underline{M}_{de} = .984$) gehen die Leistungen in der Halbprofilansicht stark auseinander ($\underline{M}_{a-a} = .738$, $\underline{M}_{l-a} = .941$, $\underline{M}_{Tü} = .949$, $\underline{M}_{de} = .965$) (siehe Tabelle 22).

Tabelle 22

Mittelwerte der Treffer, der falschen Alarme, des Leistungsindex A' und der Reaktionsneigung B'' nach Ansicht und Stimulusethnie

Ansicht	Afr.- Am.	Lat.- Am.	Türken	Deutschen	M :
Frontal:					
Treffer	.977 ^a	.988 ^a	.992 ^a	.984	.985
Falsche Alarme	.008	.039 ^a	.023	.004 ^a	.019
A'	.992	.986	.992	.995	.991
B''	.092 ^a	-.186 ^{ab}	-.122 ^a	.092 ^b	.324
Halbprofil:					
Treffer	.738 ^a	.941 ^a	.949 ^a	.965 ^a	.898
Falsche Alarme	.129 ^a	.141 ^{ab}	.051 ^{ab}	.031 ^{ab}	.088
A'	.873 ^{abc}	.943 ^a	.970 ^b	.981 ^c	.942
B''	.356 ^a	-.401 ^{ab}	-.051 ^{ab}	.028 ^{ab}	.241
M :					
Treffer	.858 ^a	.965 ^a	.971 ^a	.975 ^a	.449
Falsche Alarme	.068 ^a	.090 ^{ab}	.037 ^{abc}	.018 ^{abc}	.053
A'	.932 ^{abc}	.965 ^a	.981 ^b	.988 ^c	.966
B''	.224 ^a	-.294 ^{ab}	-.087 ^{abc}	.060 ^{abc}	-.024

Anmerkung: ^{abc} Die Zellenmittelwerte innerhalb einer Zeile unterscheiden sich nach Bonferroni-Adjustierung signifikant (post-hoc Vergleiche, $p < .001$).

Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle F s < 4.03 , p s $> .053$).

Die Analyse der falschen Alarme mittels einer Varianzanalyse mit den Between-Subjects Faktoren 'Berufs -Gruppe' (BGS-Beamte, Studierende), 'Geschlecht' (weiblich, männlich), 'Ansicht' (Frontal, Halbprofil), 'Identitätspräsentation' (identisch, nicht identisch) und 'Taste' (c,m) und dem Within-Subjects Faktor 'Stimulusethnie' (Afro-Amerikaner, Latein-Amerikaner, Türken, Deutsche) zeigte keinen signifikanten Haupteffekt der Berufsgruppe, $F < 1$, ns. Allerdings gab es auch bei der fälschlichen Identifizierung identischer Gesichter wieder Unterschiede zwischen den Ethnien, $F(3,96) = 13.27$, $p = .001$, $\eta^2 = .29$ (siehe Tabelle 22). Die Teilnehmer hatten bei den latein-amerikanischen Gesichtern die meisten falschen Alarme ($M_{l-a} = .090$), gefolgt von den afro-amerikanischen Gesichtern ($M_{a-a} = .068$), den türkischen ($M_{tü} = .037$) und den deutschen Gesichtern ($M_{de} = .018$).

Die falschen Alarme waren bei Darbietung im Halbprofil signifikant häufiger ($M = .088$) als bei Darbietung in der Frontalansicht ($M = .019$), $F(1,32) = 26.39$, $p = .001$,

$r = .67$. Die Ansicht des zweiten Stimulus eines Durchgangs interagiert ebenfalls mit dem Faktor 'Stimuluse ethnien', $F(3,96) = 7.75$, $p = .001$, $\eta^2 = .20$ (siehe Tabelle 22).

Die Interaktion der Faktoren 'Identitätspräsentation' und 'Stimuluse ethnien' ist signifikant, $F(3,96) = 13.27$, $p = .001$, $\eta^2 = .29$. Bei den afro-amerikanischen, $F(1,96) = 51.88$, $p = .001$, $r = .59$ ($M_{\text{nid1}} = .023$, $M_{\text{nid2}} = .113$), den latein-amerikanischen, $F(1,96) = 25.08$, $p = .001$, $r = .46$ ($M_{\text{nid1}} = .121$, $M_{\text{nid2}} = .059$) und deutschen Gesichtern, $F(1,96) = 4.80$, $p = .031$, $r = .22$ ($M_{\text{nid1}} = .004$, $M_{\text{nid2}} = .03$) gibt es je nach Bedingung (nid1, nid2) unterschiedlich viele falsche Alarme.

Die Hypothese, nach der die BGS-Beamten insgesamt eine bessere Leistung zeigen sollten als die Studierenden wurde mit einem geplanten Kontrast getestet, bei dem die Differenz der Treffer der BGS-Beamten und der Studierenden größer als null sein sollte. Dieser Kontrast ist jedoch nicht signifikant von Null verschieden. $F(1,32) = 1.63$, $p = .211$, $r = .22$. Bei den BGS-Beamten ist $\underline{A}' = .971$, bei den Studierenden beträgt $\underline{A}' = .963$. Die Hypothese, nach der der Leistungsindex \underline{A}' bei den vier Stimuluse ethnien je nach Ethnie der Teilnehmer linear in Abhängigkeit der Distanz von der eigenen Ethnie ansteigt, wurde ebenfalls mithilfe linearer Kontraste getestet. Diese Hypothese konnte bestätigt werden, $F(1,32) = 55.50$, $p = .001$, $r = .796$. \underline{A}' beträgt bei afro-amerikanischen Gesichtern $\underline{A}'_{\text{a-a}} = .932$, bei latein-amerikanischen Gesichtern $\underline{A}'_{\text{l-a}} = .965$, bei türkischen Gesichtern $\underline{A}'_{\text{t}} = .981$, und bei deutschen Gesichtern $\underline{A}'_{\text{de}} = .988$. Lineare Kontraste werden bei separater Berechnung für beide Berufsgruppen signifikant. Bei den BGS-Beamten ist dieser Kontrast $F(1,16) = 27.84$, $p = .001$, $r = .80$. Bei afro-amerikanischen Gesichtern $\underline{A}'_{\text{a-a}} = .945$, bei latein-amerikanischen Gesichtern $\underline{A}'_{\text{l-a}} = .968$, bei türkischen Gesichtern $\underline{A}'_{\text{t}} = .985$, und bei deutschen Gesichtern $\underline{A}'_{\text{de}} = .986$. Bei den Studierenden betragen diese Werte entsprechend $F(1,16) = 29.76$, $p = .001$, $r = .81$, mit $\underline{A}'_{\text{a-a}} = .920$ bei afro-amerikanischen Gesichtern, bei latein-amerikanischen Gesichtern $\underline{A}'_{\text{l-a}} = .961$, bei türkischen Gesichtern $\underline{A}'_{\text{t}} = .977$, und bei deutschen Gesichtern $\underline{A}'_{\text{de}} = .989$. Es zeigte sich keine Interaktion der Faktoren 'Berufsgruppe' und 'Stimuluse ethnien', $F(3,96) = 1.83$, $p = .147$, $\eta^2 = .05$.

Die Vermutung, dass die Ansicht der Gesichter einen Einfluss auf den Leistungsindex \underline{A}' hat, wird mit einem signifikanten Kontrast, $F(1,32) = 50.93$, $p = .001$, $r = .78$, bestätigt. Gesichter, die im Anschluss an den Filler-Screen ebenfalls in der Frontalansicht gezeigt wurden, hatten ein $\underline{A}' = .991$, verglichen mit Gesichtern im Halbprofil, $\underline{A}' = .942$. Die Interaktion der Faktoren 'Ansicht' und 'Stimuluse ethnien', $F(3,96) = 32.01$, $p = .001$, $\eta^2 = .50$, zeigt, dass \underline{A}' zwischen den Stimuluse ethnien unterschiedlich ist, je nach Ansicht der Gesichter (siehe Abbildung 20).

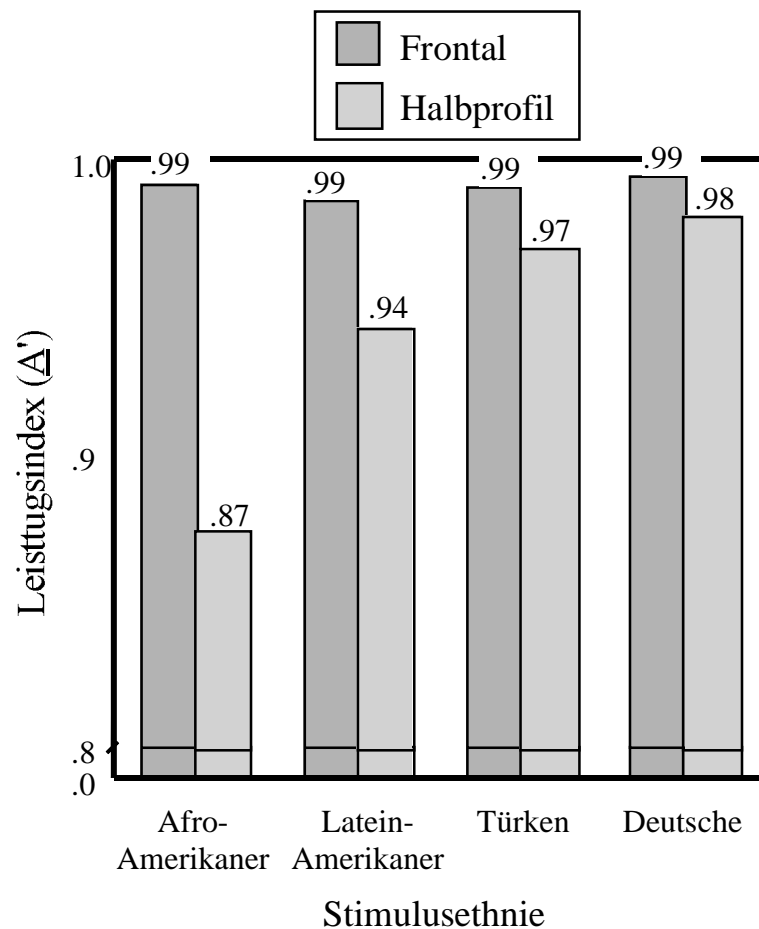


Abbildung 20. Leistungsindexwerte \underline{A}' für die Frontal und Halbprofilbedingung getrennt nach Stimuluse ethn ie.

Bei zwei separaten Analysen der Daten getrennt nach der Ansicht der Gesichter zeigt sich, dass bei Bildern in der Frontalansicht keine signifikanten Haupteffekte und Interaktionen gefunden werden, alle $\underline{F}_s < 2.86$, *ns*. Bei den Bildern in der Halbprofilansicht hingegen zeigt sich ein signifikanter Haupteffekt der Stimuluse ethn ie, $\underline{F}(3,48) = 35.02$, $p = .001$, $\eta^2 = .69$. \underline{A}' beträgt bei den afro-amerikanischen Gesichtern $\underline{A}'_{a-a} = .873$, bei den latein-amerikanischen Gesichtern $\underline{A}'_{l-a} = .943$, bei den türkischen Gesichtern $\underline{A}'_{t\ddot{u}} = .970$ und bei den deutschen Gesichtern ist $\underline{A}'_{de} = .981$. Die einzelnen Paarvergleiche sind alle signifikant (siehe Tabelle 23). Es zeigte sich in der Halbprofilansicht keine Interaktion der Faktoren 'Stimuluse ethn ie' und 'Berufsgruppe', $\underline{F}(3,48) = 2.16$, $p = .105$, $r = .21$.

Tabelle 23

Vergleich von \underline{A}' zwischen den Stimulusethnien in der Halbprofilansicht. ' $\underline{A}' 1$ ' und ' $\underline{A}' 2$ ' geben den Leistungsindex bei der Ethnie in der ersten bzw. zweiten Spalte an.

Vergleich Ethnie		$\underline{A}' 1$	$\underline{A}' 2$	\underline{F}	\underline{p}	\underline{r}
Afr.- Am.	Lat.- Am.	.873	.943	261.8	.001	.902
Afr.- Am.	Türken	.873	.970	504.5	.001	.945
Afr.- Am.	Deutsche	.873	.981	620.8	.001	.955
Lat.- Am.	Türken	.943	.970	39.4	.001	.630
Lat.- Am.	Deutsche	.943	.981	76.3	.001	.748
Türken	Deutsche	.970	.981	6.03	.017	.302

Legende: Afro-Am. Afro-Amerikaner; Lat-A: Latein-Amerikaner

Für das Leistungsmaß \underline{A}' wurde anschließend ebenfalls eine Varianzanalyse mit den Between-Subjects Faktoren 'Berufsgruppe' (BGS-Beamte, Studierende), 'Geschlecht' (weiblich, männlich), 'Ansicht' (Frontal, Halbprofil), 'Identitätspräsentation' (identisch, nicht identisch) und 'Taste' (c,m) und dem Within-Subjects Faktor 'Stimulusethnie' (Afro-Amerikaner, Latein-Amerikaner, Türken, Deutsche) durchgeführt. Neben den bereits erwähnten Effekten zeigte sich eine signifikante Interaktion der Faktoren 'Identitätspräsentation' und 'Stimulusethnie', $\underline{F}(3,96) = 9.58$, $\underline{p} = .001$, $\eta^2 = .23$. Während sich \underline{A}' bei den latein-amerikanischen, den türkischen und den deutschen Gesichtern nicht zwischen den Identitätsbedingungen unterscheidet, alle $\underline{F}_S < 3.07$, ns, ist \underline{A}' bei den afro-amerikanischen Gesichtern zwischen den Bedingungen signifikant unterschiedlich, $\underline{F}(1,96) = 53.76$, $\underline{p} = .001$, $\underline{r} = .60$, $\underline{A}'_{nid1} = .958$, $\underline{A}'_{nid2} = .907$. Weitere Haupteffekte oder Interaktionen wurden nicht gefunden, alle $\underline{F}_S < 2.28$, ns.

Letztendlich wurde die Reaktionsneigung \underline{B}'' mittels einer Varianzanalyse mit den Between-Subjects Faktoren 'Berufsgruppe' (BGS-Beamte, Studierende), 'Geschlecht' (weiblich, männlich), 'Ansicht' (Frontal, Halbprofil), 'Stimulusgleichheit' (identisch, nicht identisch) und 'Taste' (c,m) und dem Within-Subjects Faktor 'Stimulusethnie' (Afro-Amerikaner, Latein-Amerikaner, Türken, Deutsche) analysiert. Die Reaktionsneigung war nicht bei allen untersuchten Ethnien gleich, $\underline{F}(3,96) = 13.74$, $\underline{p} = .001$, $\eta^2 = .30$. \underline{B}'' beträgt bei den afro-amerikanischen Gesichtern $\underline{B}''_{a-a} = .224$, bei den latein-amerikanischen Gesichtern $\underline{B}''_{l-a} = -.293$, bei den türkischen Gesichtern $\underline{B}''_{tü} = -.087$ und bei den deutschen Gesichtern $\underline{B}''_{de} = .059$. Eine marginal signifikante Interaktion zeigte sich auch bei den Faktoren 'Stimulusethnie' und 'Ansicht', $\underline{F}(3,96) = 2.95$, $\underline{p} = .036$, $\eta^2 = .08$. Dabei

unterscheiden sich die Reaktionsneigungen bei afro-amerikanischen Gesichtern zwischen den Gesichtsansichten, $F(1,96) = 9.89$, $p = .002$, $r = .31$ ($B''_{fr} = .092$, $B''_{hp} = .356$), ebenso wie bei den latein-amerikanischen Gesichtern $F(1,96) = 6.62$, $p = .012$, $r = .25$ ($B''_{fr} = -.186$, $B''_{hp} = -.401$). Alle anderen Haupteffekte und Interaktionen waren nicht signifikant (alle $F_s < 2.27$, alle $p_s > .085$).

6.4.5 Diskussion der Ergebnisse der Delayed-Matching Aufgabe

Die Hypothese, nach der BGS-Beamte mehr Treffer zeigen sollten als die Studierenden, konnte bestätigt werden. Hingegen konnte die Vermutung, dass sich eine bessere Leistung auch beim Leistungsidex A' zeigt, nicht bestätigt werden. Der Grund hierfür liegt in der etwas höheren Anzahl der falschen Alarme bei den BGS-Beamten, die die bessere Leistung bei den Treffern wieder relativieren. Wie beim Wiedererkennenstest waren die Leistungen zwischen den verwendeten Stimulusethnien insgesamt sehr unterschiedlich, mit wiederum den schlechtesten Leistungen bei den afro-amerikanischen Gesichtern und den besten Werten bei deutschen Gesichtern. Dieses Ergebnis der Delayed-Matching Aufgabe wird ebenfalls als Beleg für die Speicherung von Gesichtern in einem multidimensionalen Raum interpretiert. Dass ein lineares Ansteigen der Wiedererkennensleistung in Abhängigkeit von der Ethnie der Stimulusperson wiederum bei beiden Berufsgruppen gefunden wird, kann zunächst dermaßen interpretiert werden, dass die größere Kontakthäufigkeit der BGS-Beamten keinen Einfluss auf die Leistung bei Fremd-Gruppen Gesichtern hat. Wiederum analog zum Wiedererkennenstest lässt sich aber beobachten, dass die BGS-Beamten bei den afro-amerikanischen Gesichtern besser sind als die Studierenden. Daraus lässt sich schließen, dass eine größere Kontakthäufigkeit dennoch zu einer Verbesserung der Wiedererkennensleistung bei Fremd-Gruppen Gesichtern führen kann, die aber letztendlich nicht zu gleichen Leistungen wie bei Eigen-Gruppen Gesichtern führt. Zusätzlich muss noch gesagt werden, dass über alle Gesichter hinweg, und vor allem bei den Eigen-Gruppen Gesichtern, die Gesamtleistung sehr gut ist. Solch ein Deckeneffekt führt dazu, dass Verbesserungen in der Wiedererkennensleistung nur marginal sein können. Die Ergebnisse der Delayed-Matching Aufgabe geben durch die Betrachtung der Bedingung 'Ansicht' Aufschluss über die Enkodierung von Eigen-Gruppen und Fremd-Gruppen Gesichtern. Hier zeigte sich ein starker Effekt mit $r = .78$, mit besseren Leistungen bei Gesichtern in der Frontalansicht. In der Literatur wird eine

merkmalsbasierte Enkodierung von Fremd-Gruppen Gesichtern und eine konfigurale, holistische Enkodierung von Eigen-Gruppen Gesichtern diskutiert. Eine holistische Verarbeitung postuliert gleichzeitig eine bessere Wiedererkennensleistung bei einem Posenwechsel des enkodierten Gesichts, d.h. ein Gesicht wird als ganzes wiedererkannt, und die Bedeutung einzelner Merkmale bei der Wiedererkennung wird geringer. Wie berichtet, zeigt sich zunächst bei allen Stimuluse ethnien in der Frontalbedingung ein beinahe perfektes Ergebnis, d.h. die Teilnehmer waren bei fast jedem Gesicht, egal welcher Ethnie, in der Lage, zu differenzieren, ob die beiden nacheinander dargebotenen Gesichter übereinstimmen oder nicht (vgl. Argstatter et al., 2002). In dieser Bedingung ist lediglich ein Abgleich der Merkmale eines Gesichts erforderlich. Stimmen diese Merkmale überein, sind die Gesichter identisch, stimmen sie nicht überein, ist die Antwort eindeutig, sie sind nicht identisch. Durch den Wechsel der Gesichtsansicht in der Halbprofilbedingung ist dieser Weg der Antwortfindung nicht mehr sinnvoll. Ein direkter Merkmalsabgleich ist nicht möglich, es sei denn, der Teilnehmer ist in der Lage, ein Gesicht mental zu rotieren und die Merkmale ebenfalls mental abzugleichen. Einer rein merkmalsbasierten Enkodierung werden hier schnell die Grenzen aufgezeigt, denn bei einem Wechsel der Pose werden wiederum andere Merkmale bzw. eine andere Merkmalszusammensetzung sichtbar, die vorher nicht in dieser Art und Weise enkodiert wurden und daher nicht mit den enkodierten Merkmalen verglichen werden können, auch wenn die beiden dargebotenen Gesichter eines Durchganges die gleiche Person zeigen. Eigen-Gruppen Gesichter werden in der Halbprofilbedingung eindeutig besser wiedererkannt als Fremd-Gruppen Gesichter. Die Reihenfolge der Stimuluse ethnien (Afro-Amerikaner < Latein-Amerikaner < Türken < Deutsche) zeigt sich nicht nur im Leistungsmaß A' , sondern auch bei den Treffern und bei den falschen Alarmen, wobei hier bei den Latein-Amerikanern mehr falsche Alarme gemacht werden als bei den Afro-Amerikanern. Die Vermutung, wonach Eigen-Gruppen Gesichter holistisch und Fremd-Gruppen Gesichter merkmalsbasiert enkodiert werden, wird durch den Vergleich der Ethnien in den beiden Ansichten unterstützt. Bezüglich der Annahmen des In-Group/Out-Group Modells bedeuten diese Ergebnisse eine Stützung der Vermutung einer merkmalsbasierten Enkodierung von Fremd-Gruppen Gesichtern.

Gleichzeitig wird in der Delayed-Matching Aufgabe ein 'mirror-effect' beobachtet. Gesichter der Eigen-Gruppe weisen in der Halbprofilbedingung eine höhere Zahl an Treffern bei gleichzeitiger geringerer Anzahl an falschen Alarmen auf als Gesichter der Fremd-Gruppen. Eine weitere Annahme in der Forschung zum Cross-Race Bias ist, dass die Reaktionsneigung bei Gesichtern der Fremd-Gruppe liberaler ist. Für die Delayed-

Matching Aufgabe bedeutet dies, die Teilnehmer neigen dazu, eher zwei Gesichter als identisch zu betrachten. Im Experiment in Abschnitt 4 konnte eine liberalere Reaktionsneigung bei Fremd-Gruppen Gesichtern nicht gefunden werden. Hier ist dieser Effekt bei Gesichtern im Halbprofil zu beobachten, jedoch nur bei den latein-amerikanischen Gesichtern ($B'' = -.401$). Bei den afro-amerikanischen Gesichtern zeigen die Teilnehmer hingegen ein eher konservatives Antwortkriterium ($B'' = .356$).

Weshalb sich die Leistungen bei den afro-amerikanischen Gesichtern zwischen den Identitätsbedingungen unterscheiden, bleibt noch zu klären. Denkbar ist, dass die Gesichter in der zweiten Identitätsbedingung sich stärker ähneln als in der ersten Identitätsbedingung, oder zumindest den Teilnehmer als einander ähnlicher vorkamen als in der ersten Identitätsbedingung. Solch ein Effekt sollte aber durch die Zuteilung der Gesichter zu den Sets ausgeschaltet sein. Eine genauere Analyse der Daten ist hier vonnöten.

6.5 Diskussion der Experimente zum Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung bei Eigen-Gruppen und Fremd-Gruppen Gesichtern

Ein Ziel in beiden Experimenten war, den Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung bei Eigen-Gruppen und Fremd-Gruppen Gesichtern zu untersuchen. Die Kontakthäufigkeit wurde dadurch variiert, indem zwei Berufsgruppen untersucht wurden, die sich bezüglich der Kontakthäufigkeit unterschieden. Da nur eine ethnische Gruppe teilnahm, ist, wie bereits beschrieben, eine vollständige Interaktion zwischen der Ethnie der Versuchsperson und der Ethnie der Stimulusgesichter nicht berechenbar. Daher wird ein Cross-Race Bias in diesem Fall einseitig definiert, d.h. die Teilnehmer sollten bei Eigen-Gruppen Gesichtern besser sein als bei Fremd-Gruppen Gesichtern. Beide Teilnehmergruppen zeigten demnach einen Cross-Race Bias in beiden Experimenten mit Effektstärken von $r = .65$ bis $r = .86$, was einem starken Effekt nach Cohen (1988) entspricht. Trotz dieser starken Effekte wäre es den Daten nach zu einfach, keinen Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung zu postulieren. Vergleicht man vor allem die Daten der beiden Teilnehmergruppen im Wiedererkennentest, so fällt auf, dass die Unterschiede in der Wiedererkennensleistung der vier untersuchten Ethnien bei den Studierenden größer sind als bei den BGS-Beamten. Letztgenannte zeigen in der Tendenz eine Angleichung der Leistungen zwischen den Ethnien, mit besseren Leistungen bei den afro-amerikanischen Gesichtern und schlechteren

Leistungen bei den deutschen Gesichtern, verglichen mit den Studierenden. Allerdings hätte es keine Unterschiede zwischen den Ethnien geben sollen, wenn man die durchschnittliche Dauer des Dienstes der BGS-Beamten am Flughafen Frankfurt betrachtet (ca. 6 1/2 Jahre) und damit die sehr große Anzahl an unterschiedlichen Gesichtern während dieser Zeitspanne. Auch die Leistung der BGS-Beamten bei der Delayed-Matching Aufgabe, vor allem bei den Gesichtern im Halbprofil, hätte unter Berücksichtigung der Kontakthäufigkeit besser ausfallen müssen. Bis auf die afro-amerikanischen Gesichter sind die BGS-Beamten bei dieser Aufgabe vergleichbar mit denen der Studierenden. Dies läßt vermuten, dass es keinen einzelnen Einflussfaktor zur Erklärung eines Phänomens gibt. Wahrscheinlich ließe sich mittels einer Interaktion von Kontakthäufigkeit und Kontaktqualität eine größere Reduktion des Cross-Race Bias realisieren. Interessant wäre es in diesem Zusammenhang auch zu wissen, wie die Wiedererkennensleistung bei BGS-Beamten aussieht, die vornehmlich im Bereich der Rückführung von Ausländern arbeiten, da bei dieser Personengruppe eine andere Qualität des Kontaktes als an der Passkontrolle vermutet wird.

Zieht man weiterhin die Befunde von Myles-Worsley, Johnston und Simons (1988) sowie die Aussagen einiger BGS-Beamten für die Erklärung der Ergebnisse heran, wird deutlich, dass dieser Befund gar nicht so überraschend ist. In der Studie von Myles-Worsley, Johnston und Simons zeigten Radiologen mit längerer Berufserfahrung die besten Leistungen bei anomalen Röntgenbildern, waren zugleich aber auch am schlechtesten bei normalen Röntgenbildern. Dieser Befund wurde damit erklärt, dass sich die Expertise bei erfahrenen Radiologen vor allem darauf konzentriert, pathologische Befunde zu erkennen, und andere Merkmale, die zwar von der Norm abweichen, jedoch nicht pathologisch sind, nicht zu beachten. Ähnlich könnte es sich mit den BGS-Beamten verhalten. Die meisten befragten BGS-Beamten teilten mit, dass es in ihrem Beruf nicht wichtig sei, sich Gesichter per se einprägen zu können. Für die Arbeit bei der Passkontrolle sei es vielmehr von Bedeutung, die Passagiere einer bestimmten Ethnie zuzuordnen zu können²³. Die Klassifizierung der Gesichter ist für BGS-Beamte von größerer Bedeutung als die Wiedererkennung.

²³ Ein Beamter sagte, er müsse sich eher darauf konzentrieren, ob jemand, der mit einem japanischen Pass einreist, auch Japaner ist und kein Chinese, was relativ häufig vorkomme. Ein weiterer Beamter berichtete während einer Demonstration der Experimente im Vorfeld der Untersuchung, er denke automatisch darüber nach, ob das Gesicht (in der Demonstration waren es türkische Gesichter) von einem Marokkaner oder Tunesier stamme.

Was bedeuten diese Ergebnisse für die Speicherung von Gesichtern in einem multidimensionalen Raum und bezüglich der Annahmen des Norm-basierten Modells? Gesichter von Afro-Amerikanern sind in diesem Modell weiter von der "Norm", die durch die eigene Ethnie geformt wird, entfernt als latein-amerikanische oder türkische Gesichter, die zumindest in einem Merkmal, nämlich der Hautfarbe, den deutschen Gesichtern ähnlicher sind. Sporer (1992, 2001a) zeigte u.a., dass (weiße) Deutsche einen stärkeren Cross-Race Bias für schwarze Gesichter aufweisen als beispielsweise bei Anthony, Copper und Mullen (1992) berichtet wird. Dass sich solche Normen erweitern können, zeigen die Ergebnisse der BGS-Beamten. Deren Werte, sowohl bezüglich der Anzahl richtiger Antworten und des Leistungsindex' \underline{A}' liegen für die vier Stimuluse ethnien relativ dicht beieinander, bzw. sie variieren nicht in dem Ausmaß, wie bei der studentischen Gruppe (siehe Abbildung 19). Anscheinend geht jedoch eine bessere Leistung für Gesichter anderer ethnischer Gruppen einher mit einer geringeren Leistung für die eigene Gruppe. Ähnlich wie in der Studie von Myles-Worsley, Johnston und Simons (1988) mit Radiologen als Versuchspersonen, wird eine zunehmende Differenzierungsleistung und Expertise im Wiedererkennen von Gesichtern einer ethnischen Gruppe dadurch erkaufte, dass die Leistung bei einer anderen ethnischen Gruppe abfällt. Die erfahrensten Radiologen in der Studie von Myles-Worsley, Johnston und Simons (1988) waren bei der Wiedererkennensleistung von Röntgenbildern ohne pathologischen Befund am schlechtesten, verglichen mit jüngeren Radiologen und der Kontrollgruppe. Ähnliche Interpretationen lassen auch die Befunde von Chiroro und Valentine (1995) zu. Afro-amerikanische Teilnehmer mit hohen Werten des Kontakts mit anderen Ethnien zeigten eine hohe Wiedererkennensleistung bei den Fremd-Gruppen Gesichtern aber ebenfalls eine Reduktion der Wiedererkennensleistung bei den Eigen-Gruppen Gesichtern. In diesem Zusammenhang muss eine dynamische Verschiebung der Verteilung der Exemplare diskutiert werden, die durch einen vermehrten Kontakt mit anderen Ethnien und dadurch eine erhöhte Anzahl an Exemplaren andere Ethnien im multidimensionalen Raum zustande kommt. Zur Überprüfung dieser Hypothese sind Experimente nötig, die die Effekte der Distinktivität von Gesichtern bei Personen mit hoher und niedriger Kontakthäufigkeit untersuchen.

Eine Hypothese war weiterhin, dass die BGS-Beamten insgesamt bessere Wiedererkennensleistungen zeigen als die Studierenden. Dem widerspricht zum einen der fehlende Haupteffekt des Faktors 'Berufsgruppe' und zum anderen die begrenzte Aufnahmefähigkeit des menschlichen Gehirns. Beide Berufsgruppen haben beim Leistungsindex \underline{A}' einen Wert von $\underline{A}' = .663$ bzw. von $\underline{A}' = .667$ im Wiedererkennentest

und von $\underline{A}' = .971$ und $\underline{A}' = .962$ in der Delayed-Matching Aufgabe. Dies bedeutet, dass bei beiden Gruppen die Mechanismen der Aufnahme und Speicherung von Informationen in gleichem Ausmaß vorhanden sind und sich die Wahrnehmung der BGS-Beamten von Fremd-Gruppen Gesichtern verschoben hat. Das beruflich induzierte 'Training' der BGS-Beamten führt nicht zu einer allgemein besseren Leistung, sondern zu einer 'Verschiebung des Fokus' in der Gesichterwahrnehmung und Speicherung, erkennbar an schlechteren Leistungen bei den Eigen-Gruppen Gesichtern und besseren Leistungen bei den Fremd-Gruppen Gesichtern. Oder wie Clifford & Bull (1978) beschrieben:

..., policemen are first and foremost humans ! They are characterized by processing and capacity limitations like any other human. Processing limitations can be overcome by both inferential and visual techniques which can be learned, but unfortunately, in the case of the police, the inferential techniques they learn are based on existing knowledge which include values, beliefs, prejudices and 'working hypotheses' about the nature of reality. (Clifford & Bull, 1978, p. 191)

Die Annahme des In-Group/Out-Group Modells von Sporer (2001a), nach der eine merkmalsbasierte Enkodierung von Fremd-Gruppen Gesichtern und eine holistische Enkodierung von Eigen-Gruppen Gesichtern stattfindet, konnte mithilfe der Delayed-Matching Aufgabe bestätigt werden. Da mit dem Faktor 'Ansicht' die Darbietung der Gesichter variiert wurde, lassen die Unterschiede in der Wiedererkennensleistung zwischen den Ansichten Aufschlüsse über die Verarbeitung von Eigen-Gruppen und Fremd-Gruppen Gesichtern zu. Eine holistische Verarbeitung postuliert gleichzeitig eine bessere Wiedererkennensleistung bei einem Posenwechsel des enkodierten Gesichts. Da das Gesicht als 'Ganzes' enkodiert wird, ist es einfacher bei einem Wechsel der Pose zu bestimmen, ob das Gesicht mit dem vorher gesehenen übereinstimmt oder nicht. Wenn nur einzelne Merkmale enkodiert werden, so ist ein Urteil über die Gleichheit zweier Gesichter schwieriger, weil diese Merkmale von mehreren Gesichtern geteilt werden können, bzw. einzelne Merkmale bei einem Posenwechsel nicht mehr sichtbar sein können. Die Teilnehmer hatten bei den deutschen (Eigen-Gruppen) Gesichtern keine Schwierigkeiten, Gesichter im Halbprofil als identisch bzw. nicht-identisch einzuordnen. Diese Fähigkeit sank mit zunehmender Distanz der Stimulusethnie von der Eigen-Gruppe. Nun ist die Delayed-Matching Aufgabe kein Wiedererkennensexperiment im engeren Sinne, da lediglich ein zeitverzögerter Abgleich von zwei Gesichtern von den Teilnehmern verlangt

wird. Aber der Posenwechsel lässt trotzdem die erwähnten Rückschlüsse auf die Verarbeitung von Eigen-Gruppen und Fremd-Gruppen Gesichtern zu, weil trotz einer sehr kurzen Zeit zwischen Enkodierung und Abruf der Information sehr große Differenzen zwischen den Gesichtsansichten und den Stimulusethnien auftreten.

Im Gegensatz zum Experiment zum Einfluss des Behaltensintervalls und des Gesichtsausschnittes in Abschnitt 4 wurde ein signifikanter Effekt bei der Reaktionsneigung zum Faktor 'Stimulusethnie' bei der Delayed-Matching Aufgabe gefunden. Beim Wiedererkennenstest liegt die Reaktionsneigung bei allen Gesichtern nahe bei null, was anzeigt, dass die Teilnehmer weder besonders konservativ noch unachtsam geantwortet haben. Warum dies so ist, lässt sich nur vermutend beantworten. Eine mögliche Erklärung ist eine besonders sorgfältige Abwägung der Teilnehmer bei der Frage, ob die Gesichter bereits gesehen wurden oder nicht. Zumindest die BGS-Beamten sollten sehr motiviert gewesen sein, gut bei dem Test abzuschneiden, weil sie selbst davon ausgingen, besonders gut bei der Personenidentifizierung zu sein. Es wurde ihnen zwar nicht vermittelt, dass es bei der Untersuchung darum ging, die Wiedererkennensleistung bei mehreren Ethnien zu testen, dennoch ist es zumindest denkbar, dass der Charakter der Untersuchung von einigen Teilnehmern in diese Richtung interpretiert worden ist. Weshalb hingegen bei der Delayed-Matching Aufgabe im Halbprofil bei den afro-amerikanischen ($B'' = .356$) und latein-amerikanischen Gesichtern ($B'' = -.401$) in der Reaktionsneigung festzustellen sind, bleibt noch zu klären.

Bezüglich der schlechteren Werte vor allem bei den afro-amerikanischen Gesichtern und in etwas schwächeren Ausmaß bei den latein-amerikanischen Gesichtern muss noch erwähnt werden, dass die Auflösung und Farbtiefe der afro-amerikanischen und latein-amerikanischen Stimuluspersonen geringer war als bei den deutschen und türkischen Stimuluspersonen. Die Bilder der deutschen und türkischen Stimuluspersonen wurden mit einer hohen Auflösung und hohen Farbtiefe aufgenommen. Dadurch ergibt sich ein größeres Bild bei afro-amerikanischen und latein-amerikanischen Gesichtern was dazu führen kann, dass feinere Nuancen eines Gesichtes, die unter Umständen bei der Enkodierung des Gesichtes hilfreich sind, verschwinden. Die Ergebnisse aber alleine der Fotoqualität zusprechen zu wollen, widersprechen die Wiedererkennensleistungen bei türkischen Gesichtern verglichen mit den deutschen Gesichtern. Trotz gleicher Qualität wurden die türkischen Gesichter schlechter wiedererkannt als die deutschen Gesichter. Es gab ebenfalls keiner der Teilnehmer an, von unterschiedlichen Bildqualitäten irritiert worden zu sein.

7 Zusammenfassende Diskussion und Ausblick

Ziel dieser Arbeit war, die Annahmen des In-Group/Out-Group Modells von Sporer (2001a) bei nicht-menschlichen Stimuli und Gesichtern zu untersuchen und diese Annahmen in den Kontext der Forschung zum Cross-Race Bias zu stellen. Zwei Faktoren beeinflussen hierbei grundlegend die Wiedererkennung von Gesichtern. Zum einen die Speicherung von Gesichtern als Exemplare in einem multidimensionalen Raum, der durch die Dimensionen aufgespannt wird, anhand derer sich Gesichter unterscheiden. Jedes Exemplar besetzt dabei einen Punkt in diesem Raum. Exemplare, die sich von den 'normalerweise' gespeicherten Exemplaren in ihrer Ausprägung auf einer oder mehreren Dimensionen unterscheiden (Fremd-Gruppen Gesichter), werden aufgrund ihrer daraus resultierenden Lage schlechter wiedererkannt als die 'normalerweise' enkodierten Gesichter. Zum anderen bestehen die Annahmen des In-Group/Out-Group Modells aus einer Kombination von sozialpsychologischen und allgemeinspsychologischen Faktoren. Die Kategorisierung anderer Menschen aufgrund eines Merkmals und die daraus folgende Zuschreibung bestimmter Charakteristika und Persönlichkeitseigenschaften ist ein sozialpsychologisches Phänomen und geschieht selbstverständlich auch oder gerade bei der Betrachtung von Gesichtern. Daraus resultierende abweichende Verarbeitungswege der Information, die Gesichter liefern, führen zu besserer (Eigen-Gruppe) bzw. schlechterer (Fremd-Gruppe) Leistung bei Wiedererkennentests. Im folgenden werden die Ergebnisse der einzelnen Untersuchungen noch einmal kurz dargestellt und ihr Bezug zum In-Group/Out-Group Modell zusammenfassend erläutert.

7.1 Experimente zur differentiell schlechteren Wiedererkennensleistung bei Fremd-Gruppen Objekten (Cross-Currency Effekt).

In den drei Experimenten zum Cross-Currency Effekt ($N = 42$) wurden die Annahmen des In-Group/Out-Group Modells auf nicht-menschliche Objekte ausgeweitet. Es wurde untersucht, ob bei Objekten die durch ihre Merkmale eine Kategorisierung als Eigen-Gruppe bzw. Fremd-Gruppe hervorrufen können, ebenfalls eine schlechtere Differenzierungsleistung bei Fremd-Gruppen Objekten auftritt oder nicht. Dies war in allen

drei in dieser Arbeit beschriebenen Experimenten der Fall; die Geldscheine der DM, die als Eigen-Gruppen Wahrung in die Experimente aufgenommen war, wurden besser erkannt als die Fremd-Gruppen Geldscheine des Britischen Pfunds und des Euro. Entgegen der Hypothese sank die Leistung bei der DM von Zeitpunkt 1 zu Zeitpunkt 2 nicht, was fur eine starke Reprasentation der DM durch den jahrelangen Gebrauch spricht. Die Leistungen beim Euro stiegen ebenfalls nicht in dem Mae wie vorhergesagt, wahrscheinlich auch deshalb, weil das Intervall zwischen den Messzeitpunkten zu kurz war, um bessere Ergebnisse zu erzielen. Ob fur Eigen-Gruppen und Fremd-Gruppen Geldscheine andere Verarbeitungswege benutzt wurden, konnte durch die Experimente nicht geklart werden. Die Annahme einer Kategorisierung der Geldscheine durch die Teilnehmer bleibt eine Annahme, die zumindest in Teilen durch die bessere Leistung bei der DM gestutzt werden kann. Einen interessanten Aspekt vermitteln die verwendeten Falschungen, zwischen denen z.T. erhebliche Unterschiede in der Leistung bestanden. So wurden die Farbfalschungen der DM und des Euro besser als Falschungen erkannt als die dazugehorigen Spiegelungsfalschungen. Die Farbe eines Objektes spielt bei der Enkodierung eine wesentliche Rolle (Gegenfurtner & Rieger, 2000). Dies war in diesem Experiment zumindest solange der Fall, wie die Farbe eines Geldscheinens eindeutig zuordnungsbar war und einen Geldschein als echt oder falsch erkennen lie. Die Spiegelungsfalschungen wurden auch deshalb schlechter erkannt, weil diese Falschungen u.U. zu subtil waren, als dass sie eindeutig als Falschungen zu erkennen gewesen waren. Die Teilnehmer hatten eine eindeutige und vor allem genauere mentale Reprasentation der Geldscheine gebraucht, um die echten Geldscheine von Spiegelungsfalschungen zu unterscheiden. Dies war offensichtlich nicht moglich, eventuell weil die Veranderungen der Spiegelungsfalschungen im Vergleich zu den normalen Geldscheinen zu unauffallig waren, bzw. wurden wirklich saliente Merkmale der normalen Geldscheine nicht eindeutig genug verandert. Zusatzlich war es Ziel dieses Experiments, die experimentellen Paradigmen auszuwahlen, die fur die folgenden Experimente verwendet werden sollten. Als adaquate Paradigmen stellten sich die Delayed-Matching Aufgabe und der Wiedererkennenstest heraus, da bei diesen Testformen die Auswertung anhand der Parameter der Signal-Entdeckungs Theorie moglich sind.

Was bedeuten diese Ergebnisse fur das In-Group/Out-Group Modell und andere Theorien uber die Wahrnehmung von Gesichtern? Zunachst bedeutet dies, dass, wie in anderen Studien zum Wiedererkennen von Objekten (z.B. Gauthier & Tarr, 1996), es moglich ist, Expertise fur diese Objekte zu entwickeln. Durch die Wahl der Geldscheine sollte eine Kategorisierung der Objekte angestoen werden, die zu einer anderen

Verarbeitung der Fremd-Gruppen Objekte führt. Welche Mechanismen im einzelnen für die unterschiedliche Leistung bei den drei Währungen verantwortlich sind, ist leider nicht eindeutig zu klären. Dass die Leistungen bei der DM insgesamt besser sind, ist auch auf die jahrelange Erfahrung mit der DM zurückzuführen. Die Darbietung von Euro und Britischen Pfund Geldscheinen kann diesen Rückstand nicht ausgleichen. Um einen wirklichen vollständigen Vergleich von Objekten und Gesichtern durchzuführen, war die Wahl der Stimulusklasse wahrscheinlich nicht ausreichend. Die Gesichter in Wiedererkennenstests gehören zwar zu einer bestimmten Gruppe, die sich je nach Person in Eigen-Gruppe bzw. Fremd-Gruppe einteilen lassen, sind aber für die Teilnehmer an solchen Experimenten unbekannt. Bei den Geldscheinen trifft die Zugehörigkeit zu einer Gruppe ebenfalls zu, jedoch sind die Geldscheine selbst den Teilnehmern bekannt, zumindest die, mit denen jahrelang gehandelt wurde. Daher sind leider bei diesem Experiment zwei Faktoren konfundiert, was die Interpretation der Daten erschwert. Es lässt sich nicht genau sagen, ob die bessere Leistung bei der DM dadurch zustande kommt, dass sie für die deutschen Teilnehmer eine Eigen-Gruppe bildet, oder ob die lange Lernerfahrung mit der DM für die bessere Leistung verantwortlich ist. Es wäre zum einen sinnvoll, wie bereits in Abschnitt 2.4.5 angesprochen wurde, das Experiment mit anderen Teilnehmern noch einmal durchzuführen, um zu testen, wie sich die Leistung nach mittlerweile zwei Jahren Erfahrung mit dem Euro verändert hat. Zum anderen wäre ein ähnliches Experiment mit einer anderen Stimulusklasse zu überlegen, bei dem ebenfalls eine eindeutige Kategorisierung der Stimuli als Eigen-Gruppe bzw. Fremd-Gruppe möglich ist, die Stimuli selbst aber den Teilnehmern unbekannt sind. Denkbar wären z.B. Häuserfronten aus verschiedenen Ländern, Automarken oder auch Weinetiketten. Dies sind lediglich Möglichkeiten, eine genauere Auswahl wäre noch vonnöten.

7.2 Pilotstudie für die Auswahl der türkischen und deutschen Gesichter

Die Pilotstudie diene zunächst der Erzeugung zweier homogener Stimulussets für die folgenden Experimente, die in den Abschnitten 4 und 5 beschrieben werden. Dadurch, dass zwei Versuchspersonen-Gruppen (Türken und Deutsche; $N = 64$) und zwei Ethnien (türkische und deutsche Gesichter) als Stimulusgesichter verwendet wurden, konnte bei den untersuchten Teilnehmergruppen ebenfalls die subjektiven Wahrnehmungen von Eigen-Gruppen und Fremd-Gruppen Gesichtern überprüft werden. Aufgrund des Out-

Group Homogenitäts Effekts sollten die Gesichter der Fremd-Gruppen als einander ähnlicher wahrgenommen werden als die Gesichter der Eigen-Gruppe. Nach dem multidimensionalen Modell der Speicherung von Gesichtern sollten ebenfalls die Eigen-Gruppen Gesichter als distinkter wahrgenommen werden als die Fremd-Gruppen Gesichter. Anhand der verwendeten Skalen konnten diese Vermutungen überprüft werden. Türkische Teilnehmer beurteilten die türkischen Gesichter als distinkter als die deutschen Gesichter und vice versa. Diese Befund spricht für die Annahmen eines multidimensionalen Modells, in dem die häufigsten Exemplare (die der Eigen-Gruppe) am besten differenzierbar sind, da die Distinktivität dieser Gruppe am höchsten ist. Die Frage, für welches Modell (Norm-basiert vs. Exemplar-basiert, siehe Tabelle 1) die Wahrnehmung der Distinktivität als Beleg gelten kann, muss noch genauer untersucht werden. Zur Beantwortung dieser Frage müßte eine genauere Analyse der Datenverteilung erfolgen, was in einer späteren Untersuchung auch geplant ist. Ebenso wurden die Gesichter der Eigen-Gruppe von beiden Versuchspersonengruppen als erinnerbarer eingestuft als die Gesichter der Fremd-Gruppe. Die Gesichter der jeweiligen Eigen-Gruppe wurden als vertrauter angesehen als die Gesichter der Fremd-Gruppe. Gleiches gilt für die eingestufte Sympathie der Gesichter.

Anhand der ermittelten Ratings konnten die Darbietungssets zusammengestellt werden, die in den Experimenten zum Einfluss des Behaltensintervalls und des Gesichtsausschnittes sowie zum Einfluss der Kontakthäufigkeit bei unterschiedlichen Berufsgruppen verwendet wurden.

7.3 Der Einfluss von Behaltensintervall und Gesichtsausschnitt auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern

In diesem Experiment wurde bei zwei Teilnehmer-Gruppen ($N = 128$) der Cross-Race Bias in Abhängigkeit vom Behaltensintervall und vom Gesichtsausschnitt untersucht. Türkische und deutsche Versuchspersonen mussten in einem Wiedererkennensexperiment bei sofortiger Testung bzw. nach drei Wochen Gesichter von Eigen-Gruppen und Fremd-Gruppen Personen wiedererkennen, bei denen entweder das vollständige Gesicht oder lediglich das Gesichtsinere zu sehen war. Die Eigen-Gruppen Gesichter für die beiden Versuchspersonen-Gruppen waren türkische und deutsche Gesichter, Fremd-Gruppen Gesichter waren von Personen afro-amerikanischer, kaukasisch-amerikanischer und, je

nach Versuchspersonen-Gruppe, türkischer bzw. deutscher Herkunft. Es zeigte sich ein Haupteffekt für den Faktor 'Zeit' mit schlechteren Wiedererkennensleistungen bei einer Testung nach drei Wochen. Kein unübliches Ergebnis in einem Wiedererkennens experiment. Allerdings konnte nach einem Zeitintervall von drei Wochen kein Cross-Race Bias gefunden werden. Türkische und deutsche Teilnehmer zeigten keine signifikanten Differenzen in der Wiedererkennensleistung nach einem längeren Zeitintervall. Die Hypothese, nach der die Eigen-Gruppen Gesichter nach einem längeren Behaltensintervall aufgrund einer holistischen Enkodierung besser wiedererkannt werden sollten als die Fremd-Gruppen Gesichter und es ebenfalls einen stärkeren Cross-Race Bias geben sollte, konnte nicht bestätigt werden. Eine mögliche Ursache für fehlende Interaktionen stellt u. a. der Bodeneffekt dar, wahrscheinlich verursacht durch die große Anzahl an Stimuli, die von den Teilnehmern enkodiert werden sollten. Bei einer Anzahl von insgesamt 160 Gesichtern, die in der Darbietungsphase gespeichert werden mussten, und bei einer Zahl von 320 Gesichtern in der Testphase, erscheint es schwierig für die Teilnehmer, adäquate Enkodierungsstrategien anzuwenden, um eine bessere Wiedererkennensleistung zu zeigen. Weiterhin zeigte sich in einer anschließenden Analyse, bei der die Leistung über die jeweiligen Fremd-Gruppen akkumuliert wurde, dass ein Cross-Race Bias nur dann gefunden wurde, wenn die Testung sofort erfolgte und die Gesichter maskiert waren. Ebenso ist insgesamt die Leistung bei maskierten Gesichtern geringer als bei vollständigen Gesichtern. Die sofortige Testung ist die einzige Bedingung, bei der die Leistung bei den maskierten Eigen-Gruppen Gesichtern mit der Leistung bei vollständigen Eigen-Gruppen Gesichtern vergleichbar ist. Bei einer Testung nach drei Wochen gleichen sich die Leistungen bei maskierten Eigen-Gruppen Gesichtern der Leistung bei maskierten Fremd-Gruppen Gesichtern an (siehe Tabelle 15). Warum wird also bei sofortiger Testung ein Cross-Race Bias gefunden und nicht bei der Testung nach drei Wochen? Als Erklärung wird wiederum zum einen die hohe Zahl an Stimuli angeführt, die nach drei Wochen zu einer schlechteren Leistung führt. Es erscheint sinnvoll, dass nach dieser Zeitspanne es sowohl für die Eigen-Gruppen als auch für die Fremd-Gruppen Gesichter einfacher war, sich auch an äußere Merkmale erinnern zu können, die ja einen zusätzlichen Hinweis auf gesehene Gesichter geben. Dies zeigen die durchweg besseren Ergebnisse bei vollständigen Gesichtern bei Testung nach drei Wochen verglichen mit den maskierten Gesichtern. Bei sofortiger Testung kommen Vergessensprozesse naturgemäß noch nicht in dem Ausmaß zum tragen. Daher ist es den Teilnehmer noch eher möglich, sich an die inneren Merkmale zu erinnern. Bezüglich des In-Group/Out-Group Modells ist dieses Ergebnis als Beleg für eine merkmalsbasierte Enkodierung von Fremd-Gruppen Gesichtern

zu werten. Eine holistische Verarbeitung der Eigen-Gruppen Gesichter, die zeitlich überdauernder ist als die Enkodierung der Fremd-Gruppen Gesichter, kann in diesem Experiment nicht gefunden werden. Um den Kritikpunkt der hohen Zahl an Stimuli, und dadurch fehlende Effekte nach längerem Zeitintervall auszuschließen, müsste ein ähnliches Experiment mit weniger Stimuli durchgeführt werden. Eine Halbierung der Anzahl scheint hier angebracht zu sein.

Hypothesenkonform zeigte sich die Leistung bei den vier untersuchten Stimulusethnien. Die afro-amerikanischen Gesichter wurden am schlechtesten wiedererkannt, gefolgt von den kaukasisch-amerikanischen, den türkischen und den deutschen Gesichtern, wobei die deutschen und türkischen Gesichter in der Reihenfolge bei türkischen Teilnehmern vertauscht sind. Diese ansteigende Wiedererkennensleistung mit zunehmender Nähe zur Eigen-Gruppe spricht für die Speicherung von Gesichtern in einem multidimensionalen Raum, in dem die ethnische Gruppe, mit der die Teilnehmer vermehrten Kontakt haben, die meisten Exemplare besitzt, und die dadurch besser differenziert werden kann. Die marginal signifikante Interaktion der Faktoren 'Ethnie der Versuchspersonen' und 'Stimulusethnie' deutet an, dass vor allem bei den deutschen Teilnehmern ein Wiedererkennensdefizit bei den afro-amerikanischen Gesichtern besteht, während bei den türkischen Teilnehmern eine 'Zweiteilung' der Stimulusethnien beobachtbar ist. Türkische und deutsche Gesichter werden gleich gut wiedererkannt und zwar besser als afro-amerikanische und kaukasisch-amerikanische Gesichter, zwischen denen bei den türkischen Teilnehmern keine Unterschiede zu vermerken sind. Ob diese Ergebnisse eher für ein Norm- bzw. Exemplar-basiertes Modell sprechen kann hier nicht geklärt werden und war auch nicht Ziel dieser Untersuchung.

Generell zeigte sich in diesem Experiment, dass die äußeren Merkmale von Gesichtern wichtige Hinweisreize für die Differenzierung von Gesichtern liefern, da die Entfernung von äußeren Gesichtsmarkmalen eine Verschlechterung der Wiedererkennensleistung zur Folge hat. Der Effekt einer schlechteren Wiedererkennensleistung bei fehlenden äußeren Merkmalen ist über alle Stimulusethnien gleich. Auffällig ist allerdings, dass beide Versuchspersonen-Gruppen die meisten Treffer bei den türkischen Gesichtern und die wenigsten falschen Alarme bei den deutschen Gesichtern hatten. Der mirror-effect sagt vorher, dass die meisten Treffer und wenigsten falschen Alarme bei den Eigen-Gruppen Gesichtern gemacht werden. Die Ergebnisse dieses Experiments sind gegensätzlich, aber analog zu der weiter oben gegebenen Erklärung zum Auftreten eines Cross-Race Bias bei maskierten Gesichtern und sofortiger Testung. Bei sofortiger Testung sind die Gesichter noch präsenter als bei Testung nach drei Wochen, eine ganzheitliche

Erfassung der Merkmale von Eigen-Gruppen Gesichtern besser erinnerbar als eine merkmalsbasierte Enkodierung. In dieser Bedingung tritt der mirror effect mit einer höheren Trefferate und weniger falschen Alarmen bei Eigen-Gruppen Gesichtern verglichen mit Fremd-Gruppen Gesichtern auf. Bei Testung nach drei Wochen ist dies nicht mehr der Fall. Die Treffer und die falschen Alarme bei maskierten Eigen-Gruppen Gesichtern unterscheiden sich nicht mehr von den maskierten Fremd-Gruppen Gesichtern. Alleine die Trefferrate bei vollständigen Eigen-Gruppen Gesichtern ist hier noch höher als bei Fremd-Gruppen Gesichtern. Unter Umständen trägt die Ähnlichkeit der kaukasisch-amerikanischen Gesichter mit den deutschen Gesichtern zu einer Konfusion der Teilnehmer bei diesen Gruppen bei. Diese Interpretation wird unterstützt durch die Ergebnisse der Studierenden im Experiment zum Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung, bei denen ein mirror-effect bei den gleichen Stimulusgesichtern wie in diesem Experiment beim Vergleich der Stimulusethnien gefunden wurde.

7.4 Pilotstudie für die Auswahl der latein-amerikanischen Gesichter

Die Pilotstudie für die Auswahl der latein-amerikanischen Gesichter ($N = 32$) diente wie die Pilotstudie der Generierung zweier Sets für das nachfolgende Experiment. Aufgrund der angesprochenen Ähnlichkeit der kaukasisch-amerikanischen und der deutschen Gesichter, die als Stimulusethnien im vorangegangenen Experiment herangezogen wurden, erschien es sinnvoll, die kaukasisch-amerikanischen Gesichter gegen Gesichter einer anderen Ethnie auszutauschen. Die Gesichter von Latein-Amerikanern wurden ausgewählt, weil sie zum einen in der Frontal- und Seitenansicht vorhanden waren und zum anderen durch ihre Ähnlichkeit mit südeuropäischen Gesichtern eine gute Ergänzung zu den bisher verwendeten Stimulusethnien darstellen.

7.5 Der Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern

Nach den Annahmen der Kontakthypothese sollte bei vermehrtem Kontakt mit anderen Ethnien die Leistungsunterschiede beim Wiedererkennen zwischen Eigen-

Gruppen und Fremd-Gruppen Gesichtern verringert werden, bzw. ganz verschwinden. In den Experimenten zum Einfluss der Kontakthäufigkeit auf die Wiedererkennensleistung von Eigen-Gruppen und Fremd-Gruppen Gesichtern ($N = 64$) wurde zwischen zwei Berufsgruppen (BGS-Beamte und Studierende) eine unterschiedliche Häufigkeit des Kontaktes mit anderen Ethnien postuliert und die Wiedererkennensleistung der beiden Berufsgruppen bei Eigen-Gruppen und Fremd-Gruppen Gesichtern untersucht. Die postulierten Unterschiede in der Kontakthäufigkeit wurden anhand selbstberichteter Daten erhoben und bestätigt. Die BGS-Beamten hatten signifikant mehr Kontakt mit Personen anderer ethnischer Gruppen als die Studierenden. Insgesamt waren die BGS-Beamten nicht, wie vermutet, besser als die Studierenden. Im Wiedererkennenstest zeigten zwar beide Gruppen, bei Betrachtung der Gesamtleistung über die Stimuluse ethnien hinweg, vergleichbare Werte im Leistungsindex A' , dieser fiel aber in Abhängigkeit von der Berufsgruppe und der Stimuluse ethnien unterschiedlich aus. Während die BGS-Beamten bei den afro-amerikanischen Gesichtern bessere Leistungen zeigten als die Studierenden, waren sie bei den deutschen Eigen-Gruppen Gesichtern schlechter als die Kontrollgruppe. Die Leistungen bei den anderen beiden Ethnien (Latein-Amerikaner und Türken) unterschieden sich zwischen den Berufsgruppen hingegen nicht. Wie in dem Experiment zum Einfluss des Behaltensintervalls und des Gesichtsausschnitts auf den Cross-Race Bias zeigten beide Berufsgruppen mit zunehmender Nähe der Gesichter zur Eigen-Gruppe bessere Wiedererkennensleistungen. Die afro-amerikanischen Gesichter wurden von beiden Berufsgruppen am schlechtesten wiedererkannt, gefolgt von den latein-amerikanischen, den türkischen und den deutschen Gesichtern. Dieser Effekt war bei den BGS-Beamten geringer als bei den Studierenden, was zunächst gegen die Hypothese eines vollständigen Verschwindens des Cross-Race Bias durch den Kontakt mit anderen Ethnien spricht. Beobachtbar ist allerdings, dass die Leistungsdifferenzen zwischen den Ethnien bei den BGS-Beamten geringer sind als bei den Studierenden. Wie können diese Ergebnisse erklärt werden? Zunächst sprechen die Ergebnisse des Wiedererkennenstests für einen Einfluss der Kontakthäufigkeit auf den Cross-Race Bias, weil, wie oben beschrieben, die Leistungsdifferenzen bei den BGS-Beamten geringer sind als bei den Studierenden. Die schlechteren Leistungen der BGS-Beamten verglichen mit den Studierenden bei den Eigen-Gruppen Gesichtern können als Ausdruck limitierter Verarbeitungskapazität gesehen werden (siehe Clifford & Bull, 1978), die selbstverständlich auch vor BGS-Beamten nicht Halt macht. Keine der beiden Berufsgruppen war benachteiligt. Zwar ist es möglich, und auch sehr wahrscheinlich, dass die BGS-Beamten sich über den Sinn und Zweck der Untersuchung austauschten und versuchten, sich ihren Vorstellungen

entsprechend zu motivieren. Sie könnten eventuell mehr Wert darauf gelegt haben, bei den Fremd-Gruppen Gesichtern gut abzuschneiden, was zu einer Vernachlässigung der Eigen-Gruppen Gesichter geführt haben könnte. Dem widersprechen aber die Aussagen der Beamten nach dem Debriefing, in denen sie überrascht waren, welchen Zweck die Untersuchung eigentlich hatte.

In der Delayed-Matching Aufgabe konnten die Ergebnisse des Wiedererkennenstests bestätigt werden. Da in diesem Teil auch die Ansicht der Gesichter variiert wurde, können Aussagen über die Enkodierung von Eigen-Gruppen und Fremd-Gruppen Gesichtern getroffen werden. Wenn beide Bilder in der Frontalansicht dargeboten wurden, war bei allen vier Ethnien die Leistung nahezu perfekt. Diese Variation stellt eine sehr leichte Aufgabe dar, weil der Teilnehmer das gesehene Gesicht nur über einen sehr kurzen Zeitraum (2000 ms) behalten muß und anschließend sein Urteil über die Gleichheit der Gesichter abgibt. Bei der Frontalansicht genügen in den meisten Fällen ein Merkmal oder wenige Merkmale, um zu einer sicheren und korrekten Entscheidung zu gelangen. Bei einem Wechsel der Ansicht wird von dem Teilnehmer mehr verlangt. Er muß zum einen das Gesicht enkodieren und eine mentale Rotation vornehmen, bei der wichtige encodierte Merkmale verschwinden können. Beschränkt sich der Teilnehmer, bewußt oder unbewußt auf die Enkodierung einzelner Merkmale, wird die Leistung abfallen, da er sich bei seiner Entscheidung im Extremfall auf Zufallsniveau begibt. Bei der Delayed-Matching Aufgabe zeigten sich in Abhängigkeit der Entfernung der Gesichter von der Eigen-Gruppe, z.T. starke Verminderungen bei der Identifikationsleistung von der Frontalansicht zur Halbprofilansicht. Die Leistung bei den (deutschen) Eigen-Gruppen Gesichter veränderte sich zwischen den beiden Darbietungsvariationen nicht. Dies bedeutet, dass Fremd-Gruppen Gesichter in weitaus höherem Maße als die Eigen-Gruppen Gesichter anhand einzelner Merkmale enkodiert werden.

7.6 Integration der Befunde

In allen drei Experimenten konnten Effekte hinsichtlich einer besseren Leistung bei der Eigen-Gruppe gefunden werden. Die Ergebnisse der drei Experimente zum Cross-Currency Effekt deuten darauf hin, dass bei Objekten ähnliche Annahmen gemacht werden können wie bei Gesichtern, allerdings müssten zunächst in einer weiteren Studie mit anderen Objekten die gleichen Annahmen überprüft werden, um eine Konfundierung von Gruppenzugehörigkeit und bekanntem Material auszuschließen. In den beiden

Experimenten in den Abschnitten 4 und 6 mit menschlichen Gesichtern finden sich ebenfalls Vorteile bei Eigen-Gruppen Gesichtern gegenüber Fremd-Gruppen Gesichtern. Beide Experimente bestätigen die Hypothese einer merkmalsbasierten Enkodierung von Fremd-Gruppen Gesichtern, sodass diese Annahme des In-Group/Out-Group Modells gute empirische Unterstützung erfährt. Da das Zeitintervall zwischen Lern- und Testphase nur im Experiment in Abschnitt 4 variiert wurde, bleibt nur dieses Experiment, um Aussagen über die Variation des Cross-Race Bias über die Zeit zu treffen. Nach einer längeren Zeitspanne werden nämlich, den Daten zufolge, Eigen-Gruppen Gesichter zu einem großen Teil über einzelne Merkmale identifiziert. Ebenso führte die Reduzierung des Umfangs der zu enkodierenden Gesichter von Experiment II zu Experiment III um die Hälfte zu besseren Leistungen, ein Fakt der die Vermutung bei der Diskussion der Ergebnisse von Experiment II stützt, dass die große Anzahl an Gesichtern zu Leistungen führten, die nahe am Zufallsniveau liegen. In beiden Experimenten mit menschlichen Gesichtern konnten die Annahmen einer Speicherung von Gesichtern in einem multidimensionalen Raum bestätigt werden. Die Wiedererkennungslleistung sank mit zunehmender Distanz der Fremd-Gruppe von der Eigen-Gruppe.

Ein weiterer Punkt, der für Modelle der Gesichtererkennung relevant sein könnte, ist der, dass alle genannten Biases (Cross-Race Bias, Own-Age Bias, Own-Sex Bias, Out-Group Homogenitäts-Effekt) natürlich eine Annahme voraussetzen, die man, in einer etwas philosophischeren Umschreibung, die Sicht der Welt aus der Ich-Perspektive nennen kann. Eine solche ego-zentrierte Weltsicht kann zum einen dazu führen, dass die Exemplare eines multidimensionalen Raums um die eigene Person verteilt sind. Diese Annahme ist jedoch mit einer rein kontaktbasierten Erfahrung und Speicherung von Exemplaren konfundiert. Üblicherweise geschieht der Kontakt mit Personen der Eigen-Gruppe, die der eigenen Person relativ ähnlich sind, bezogen auf Merkmale, die die Eigen-Gruppen von Fremd-Gruppen Mitgliedern unterscheiden. Fiske und Taylor (1991) weisen in den Erklärungsmodellen für den false consensus effect darauf hin, dass Personen selektiv die Nähe von anderen Personen suchen, die ihnen ähnlich sind und sich auch ähnlich verhalten. Eine selektive Annäherung von Personen aufgrund der eigenen Wahrnehmung des Selbst und des eigenen Verhaltens könnte dazu beitragen, die Exemplare, die gespeichert werden, nicht-intentional auszuwählen und somit Gesichter zu speichern, die dem eigenen eher ähnlich sind. Dieser Vorgang könnte nicht nur für den Cross-Race Bias von Bedeutung sein, sondern auch für den Own-Age und Own-Sex Bias in der Gesichterwiedererkennung.

7.7 Weitergehende Untersuchungen zu weiteren Modellannahmen

In dieser Arbeit wurde bezüglich der Erklärungsmodelle des Cross-Race Bias das Hauptaugenmerk auf das In-Group/Out-Group Modell von Sporer (2001a) gelegt und versucht, weitere Einflussfaktoren (Kontakthäufigkeit, Zeitintervall) in die experimentelle Planung einzubinden. Trotzdem blieben Faktoren unberücksichtigt, die im Zusammenhang mit dem Cross-Race Bias diskutiert und auch im In-Group/Out-Group Modell angesprochen werden. Es konnten Belege für eine merkmalsbasierte Enkodierung von Fremd-Gruppen Gesichtern gefunden werden. Für die Eigen-Gruppen Gesichter bedeuten die Ergebnisse von Experiment II zumindest bei kurzen Zeitintervallen zwischen Lern- und Testphase, dass sie zunächst konfigural enkodiert werden, nach einem längeren Zeitintervall die Wichtigkeit einzelner Merkmale aber zunimmt. Hier wäre, wie vorher bereits beschrieben, ein Experiment sinnvoll, das den Cross-Race Bias als Between-Subjects Design an mehreren Intervallen testet, um den Verlauf des Cross-Race Bias über die Zeit zu untersuchen. Auch die Analyse einzelner distinkte Gesichter kann darüber Aufschluß geben, da bei diesen Gesichtern eher anzunehmen ist, dass sie merkmalsbasiert enkodiert werden.

Andere Elemente des In-Group/Out-Group Modells müssen noch genauer erforscht werden. So postuliert das Modell z.B. eine automatische Verarbeitung der Eigen-Gruppen Gesichter. Eine Überprüfung dieser Annahme wäre beispielsweise möglich, indem die Reaktionszeiten analysiert werden. Für die in dieser Arbeit beschriebenen Experimente sind ebenfalls die Reaktionszeiten vorhanden, so dass anhand dieser Daten Rückschlüsse auf die Frage 'Automatische Verarbeitung von Eigen-Gruppen Gesichtern: Ja oder Nein?' gezogen werden können. Es erschien allerdings zu umfangreich, im Rahmen dieser Dissertation noch zusätzlich auf die Reaktionszeiten einzugehen.

Auf die Frage, inwiefern eine Kategorisierung der Gesichter als der Fremd-Gruppe zugehörig stattfindet, geben die Studie von MacLin und Malpass (2003) sowie die Angaben der BGS-Beamten Auskunft. MacLin und Malpass konnten zeigen, dass ein 'ethnischer Marker' ausreicht, um ein Gesicht einer bestimmten Ethnie zuzuordnen. Ebenso wurden den Gesichtern entsprechend der Stereotype über die Ethnie deren Eigenschaften auch zugeschrieben. Die BGS-Beamten gaben an, instinktiv ein Urteil über die Zugehörigkeit der Stimuli zu einer bestimmten Ethnie fällen zu wollen. Dass dieses Urteil nicht mit der eigentlichen dargebotenen Ethnie übereinstimmte, ist eine weitere

Untersuchungsoption zum Cross-Race Bias und wird hier nur der Vollständigkeit halber erwähnt. Kritisch ist hier anzumerken, dass eventuell die BGS-Beamten aufgrund ihrer Tätigkeit gelernt haben, Personen zu kategorisieren, besonders, wenn sie aufgrund ihres Aussehens zu einer Gruppe gehören, die nicht ohne weiteres in die Europäische Union einreisen kann. Leider wurden die Teilnehmer im Anschluß an die Experimente nicht nach einer Kategorisierung befragt, sodass konkrete Rückschlüsse in diesem Fall nicht möglich sind. Zieht man allerdings die bisherige Forschung zu Gruppenprozessen und sozialer Kognition heran, so ist die Kategorisierung von Personen ein alltäglicher Prozeß, der nicht alleine auf die Ethnien beschränkt ist, sondern auch Kategorien wie z. B. Geschlecht, Einkommen, Größe etc. beinhaltet (Leyens & Dardenne, 1996). Zarate und Smith (1990) konnten zeigen, dass die Geschwindigkeit, mit der ein Gesicht einer bestimmten Ethnie zugeordnet wird, einhergeht mit der Attribution von Stereotypen zu diesen Gesichtern. Damit wird eine weitere Annahme des In-Group/Out-Group Modells angesprochen, nämlich die Rolle von Stereotypen beim Wiedererkennen von Fremd-Gruppen Gesichtern. Dovidio, Evans und Tyler (1986) zeigten, dass das Priming mittels einer ethnischen Kategorie, in ihrer Studie waren es die Wörter 'Black' und 'White', dazu führten, dass die Teilnehmer bei einer Angabe von personenbezogenen Adjektiven das ethnische Stereotyp bevorzugen. Ferguson, Rhodes und Sriram (2001) fanden hingegen keinen Einfluss von Vorurteilen auf das Wiedererkennen von Fremd-Gruppen Gesichtern. Das Problem bei der Einbeziehung von Stereotypen in die Forschung ist die Erhebung adäquater Maße, da bei einem eher heiklen Thema wie Stereotypen der Einfluss der Sozialen Erwünschtheit sehr groß ist. Es ist sehr schwierig alleine anhand eines Fragebogens Unterschiede zwischen Personen bei diesen Variablen zu finden. Eichstaedt (2003, April) postuliert, dass es mittels Messung von Wortlatenzen möglich sei, implizite Einstellungen von Personen zu erfassen. Dies könnte ein interessanter Weg sein, um Einflüsse von Stereotypen auf die Wiedererkennensleistung zu erfassen.

Der Aspekt des Verbal Overshadowing wurde bereits von Fallshore und Schooler (1995) erfaßt. Die verbale Beschreibung von Eigen-Gruppen Gesichtern führte zu einer Verschlechterung der Leistung, ähnlich wie bei Fremd-Gruppen Gesichtern. Allgemein läßt sich festhalten, dass einige Einflußfaktoren des Cross-Race Bias identifiziert werden konnten (mekmalsbasierte Enkodierung, Verbal Overshadowing, Kategorisierung), der Einfluß anderer Faktoren allerdings noch genauer untersucht werden muss.

7.8 Praktische Implikationen

Die praktischen Implikationen der Forschung zum Cross-Race beziehen sich größtenteils auf die Beurteilung von Zeugenaussagen vor Gericht. Wenn der mutmaßliche Täter und der Zeuge aus unterschiedlichen ethnischen Gruppen stammen, ist es sinnvoll, die äußeren Begleitumstände, sowie die Vernehmung des Zeugen (der gebeten wurde, das Aussehen des Täters zu beschreiben, *Verbal Overshadowing*) zu erforschen. Zusätzlich wäre es sinnvoll, den Zeugen nach seiner (ggf. beruflich bedingten) Kontakthäufigkeit mit Personen anderer Ethnien zu befragen, da dies ja ebenfalls einen Einfluss auf die Wiedererkennensfähigkeit hat.

In dieser Arbeit ist durch die Gruppe der BGS-Beamten ein weiteres Berufsfeld hinzugekommen, bei dem der Cross-Race Bias Auswirkungen auf die Arbeitssituation und auch auf die Urteile, die während der Arbeit getroffen werden müssen, haben kann. Ein Teil der Arbeit von BGS-Beamten ist, wie bereits erwähnt wurde, zu entscheiden, ob eine Person, die mit einem Pass einer bestimmten Nationalität einreisen will, auch diese Nationalität besitzt. Dass der Abgleich von einem Foto und der Person, die auf dem Foto abgebildet sein soll, nicht immer eindeutig gelingt, konnte von Kemp, Towell und Pike (1997) bei Supermarktkassiererinnen gezeigt werden. Die Delayed-Matching Aufgabe, mit der die BGS-Beamten im Experiment in Abschnitt 6 konfrontiert wurden, bildet die Aufgabe des Abgleichs von Foto und Person in einer Laborsituation ab. Die BGS-Beamten zeigten hierbei lediglich bei den afro-amerikanischen Gesichtern bessere Leistungen in der Halbprofilbedingung als die Kontrollgruppe, bei den anderen untersuchten Ethnien waren sie der Kontrollgruppe ebenbürtig, mit schlechteren Leistungen bei latein-amerikanischen und türkischen Gesichtern verglichen mit der Frontalansicht. Sicherlich ist die Delayed-Matching Aufgabe nicht ganz vergleichbar mit der tatsächlichen Arbeit an der Passkontrolle. Die BGS-Beamten haben hier mehr Zeit für den Abgleich, d.h. nach Auskunft des zuständigen Beamten beim Bundesgrenzschutz für die Qualitätssicherung, ca. sieben bis zehn Sekunden für jeden einzelnen Passagier. Während dieser Zeit müssen allerdings mehr Aufgaben durchgeführt werden als bei der Delayed-Matching Aufgabe. Der BGS-Beamte muss die Person bezüglich des Herkunftslandes kategorisieren, wobei die Kategorisierung nicht immer zutreffend ist; er muß die Kategorisierung mit der Nationalität des Ausweises vergleichen und er muß entscheiden, ob die Person mit dem Foto übereinstimmt. Es ist daher gut möglich, dass die Fehler, die in der Delayed-Matching Aufgabe auch von den BGS-Beamten gemacht

wurden, auch bei der täglichen Arbeit am Flughafen Frankfurt/Main gemacht werden, und Personen illegal in die Europäische Union einreisen, weil sie aufgrund des Cross-Race Bias nicht eindeutig kategorisierbar und mit einem gefälschten Passfoto nicht vergleichbar sind. Anzuregen ist hierbei zunächst, eine ähnliche Untersuchung wie bei Kemp, Towell und Pike (1997) durchzuführen, bei der die ökologische Validität höher ist als bei der Delayed-Matching Aufgabe im Labor. Eine solche Untersuchung ist jedoch schwerer zu realisieren als ein einfaches Wiedererkennensexperiment am Computer, da das Setting so gestaltet sein muß wie bei der täglichen Arbeit, und alle Beteiligten müssen gut kooperieren, um zuverlässige Ergebnisse zu erreichen. Ausgehend von den Ergebnissen solch einer Studie müssten dann Maßnahmen geplant werden, die zu einer Verbesserung der Personenidentifikation bei der Einreise führen.

8 Literaturverzeichnis

- Allport, G. W., & Kramer, B. M. (1946). Some roots of prejudice. Journal of Psychology, 22, 9-39.
- Anderson, J. R. (1993). Rules of the mind. Hillsdale, NJ: Erlbaum.
- Anthony, T., Copper, C., & Mullen, B. (1992). Cross-racial identification: A social cognitive integration. Personality and Social Psychology Bulletin, 18, 296-301.
- Argstatter, H., Gehrke, J., Hübeler, A., Kampschulte, A., Schraub, E. M., & Sporer, S. L. (2002, März). Wiedererkennen von Eigen- und Fremdgruppen-Gesichtern. Vortrag gehalten auf der 44. Tagung experimentell arbeitender Psychologen in Chemnitz.
- Bentin, S., Allison, T., Puce, A., Perez, A., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience, 8, 551-565.
- Bentin, S., Sagiv, N., Mecklinger, A., Friederici, A., & von Cramon, Y. D. (2002). Priming visual face-processing mechanism: Electrophysiological evidence. Psychological Science, 13, 190-193.
- Blascovich, J., Wyer, N. A., Swart, L. A., & Kibler, J. L. (1997). Racism and racial categorization. Journal of Personality and Social Psychology, 72, 1364-1372.
- Bortz, J. (1999). Statistik für Sozialwissenschaftler (4. Auflage). Springer.
- Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. Personality & Social Psychology Bulletin, 15, 19-25.
- Brigham, J. C. (1990). Target person distinctiveness and attractiveness as moderator variables in the confidence-accuracy relationship in eyewitness identifications. Basic & Applied Social Psychology, 11, 101-115.
- Brigham, J. C., & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own- and other race persons. Journal of Social Issues, 41, 139-155.
- Brigham, J. C., & Wolfskeil, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identifications. Law and Human Behavior, 7, 337-349.
- Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in the superior temporal sulcus of the macaque. Journal of Neurophysiology, 46, 369-384.

-
- Bruce, V., Burton, A.M., & Dench, N. (1994). What's distinctive about a distinctive face? Quarterly Journal of Experimental Psychology, *47*, 119-141.
- Bruyer, R., & Crispeels, G. (1992). Expertise in person recognition. Bulletin of the Psychonomic Society; *30*, 501-504.
- Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. Psychological Science, *10*, 243-248.
- Caldara, R., Thut, G., Servoir, C. M., Bovet, P., & Renault, B. (2003). Face versus non-face object perception and the 'other-race' effect: a spatio-temporal event-related potential study. Clinical Neurophysiology, *114*, 515-528.
- Carey, S., & Diamond, R. (1977). From piecemeal to configural processing of faces. Science, *195*, 312-314.
- Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: factors influencing distinct processes of faces. Cognition, *83*, 1-29.
- Carter, L. F. (1948). The identification of 'racial' membership. The Journal of Abnormal and Social Psychology, *43*, 279-286.
- Chance, J. E., & Goldstein, A. G. (1987). Retention interval and face recognition: Response latency measures. Bulletin of the Psychonomic Society, *25*, 415-418.
- Chance, J. E., Goldstein, A. G., & McBride, L. (1975). Differential experience and recognition memory for faces. Journal of Social Psychology, *97*, 243-253.
- Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. The Quarterly Journal of Experimental Psychology, *48*, 879-894.
- Chung, M.-S., & Thomson, D. S. (1995). Development of face recognition. British Journal of Psychology, *86*, 55-87.
- Clifford, B. R., & Bull, R. (1978). The psychology of person identification. London: Routledge & Kegan Paul.
- Clifford, B. R., & Richards, V. J. (1977). Comparison of recall by policemen and civilians under conditions of long and short durations of exposure. Perceptual and Motor Skills, *45*, 503-512.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd Ed.). New Jersey: Lawrence Erlbaum.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. Behavior Research Methods, Instruments, and Computers, *25*, 257-271.

-
- Connors, E., Lundregan, T., Miller, N., & McEwen, T. (1996). Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial. NIH Research Report: US Department of Justice.
- Craik, F. I. M., Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11, 671-684.
- De Renzi, E., Perani, D., Carlesimo, G.A., Silveri, M.C., & Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right hemisphere – an MRI and PET study and a review of the literature. Neuropsychologia, 32, 893–902.
- Dehon, H., & Brédart, S. (2001). An 'other-race' effect in age estimation from faces. Perception, 30, 1107-1113.
- Desimone, R. (1991) Face-selective cells in the temporal cortex of monkeys. Special issue: face perception. Journal of Cognitive Neuroscience, 3, 1-8.
- Deutsche Bundesbank (n.d.). Aufgerufen im November 2001, von http://www.bundesbank.de/bargeld/bargeld_banknoten.php/.
- Devine, P. G., & Malpass, R. S. (1985). Orienting strategies in differential face recognition. Personality and Social Psychology Bulletin, 11, 33-40.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. Journal of Experimental Psychology: General, 105, 107-117.
- Dogan, E. (2003). Das Wiedererkennen von Gesichtern im Kontext des Ausländereffektes: Eine experimentelle Untersuchung. Diplomarbeit am Fachbereich Psychologie und Sportwissenschaft der Justus-Liebig-Universität Gießen.
- Donaldson, W. (1992). Measuring recognition memory. Journal of Experimental Psychology: General, 3, 275-277.
- Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: the contents of their cognitive representations. Journal of Experimental Social Psychology, 22, 22-37.
- Dunning, D., Li, J., & Malpass, R.S. (1998). Basketball fandom and cross-race identification among European-Americans: Another look at the hypothesis. Paper presented at the biennial conference of the American Psychology-Law Society in Redondo beach, CA.
- Ebbinghaus, H. (1885). Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie. Leipzig: Duncker & Humblot
- Eichstaedt, J. (2003, April). Messung differenzieller Merkmale mittels Worterkennungslatenzen. Vortrag gehalten auf der 45. Tagung experimentell arbeitender Psychologen in Kiel.

-
- Elliot, E. S., Will, E J., & Goldstein, A. G. (1973). The effects of discrimination training on the recognition of white and oriental faces. Bulletin of the Psychonomic Society, 2, 71-73.
- Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Descriptions of white and black faces by white and black subjects. International Journal of Psychology, 10, 119-123.
- Endo, M. (1986). Perception of upside-down faces: An analysis from the viewpoint of cue saliency. In H. D. Ellis, M. A. Jeeves, F. Newcombe & A. Young (Eds.), Aspects of Face Processing. Dordrecht: Martinus Nijhoff.
- Eysenck, M. W., & Keane, M. T. (1992). The use of knowledge in problem solving. In M. W. Eysenck & M. T. Keane (Eds.). Cognitive Psychology: A Student's Handbook. East Sussex: Lawrence Erlbaum.
- Faces, the ultimate Composite Picture (1998). [Computer software] InterQuest inc.
- Fallshore, M., & Schooler, J. W. (1995). Verbal vulnerability of perceptual expertise. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 1608-1623.
- Fechner G.T. (1860) . Elemente der Psychophysik. Leipzig: Breitkopf und Hartel.
- Ferguson, D. P., Rhodes, G., Lee, K., & Sriram, N. (2001). 'They all look alike to me': Prejudice and cross-race face recognition. British Journal of Psychology, 92, 567-577.
- Fiske, S. T., & Taylor, S. E. (1991). Social Cognition (2nd Ed.). Singapore: McGraw-Hill.
- Fraport AG (2002). Geschäftsbericht 2002: Fortschritt und Kontinuität. Frankfurt Airport Services Worldwide (Hrsg.). Mainz: Bödige und Partner GmbH. Aufgerufen im Juni 2003, von http://www.fraport.de/cms/investor_relations/rubrik/2/2880.downloadcenter@de.htm.
- Gauthier, I., & Tarr, M. (1996). Becoming a 'greeble' expert. Exploring mechanisms for face recognition. Vision Research, 37, 1673-1682.
- Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. Current Biology, 10, 805-808.
- Geldart, S., Mondloch, C. J., Maurer, D., de Schonen, S., & Brent, H. P. (2002). The effect of early visual deprivation on the development of face processing. Developmental Science, 5, 490-501.

-
- Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. Memory & Cognition, *13*, 8-20.
- Golby, A. J., Gabrieli, D. E., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. Nature Neuroscience, *8*, 845-850.
- Goldstein, A. G., & Chance, J. E. (1976). Measuring psychological similarity of faces. Bulletin of the Psychonomic Society, *7*, 407-408.
- Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. Journal of Psychology, *105*, 47-59.
- Goldstein, E. B. (1989). Signal Detection Theory. In: E. B. Goldstein (Ed.). Sensation and Perception (3rd Edition). Pacific Grove, CA: Brooks/Cole Publishing Company.
- Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. Psychological Bulletin, *75*, 424-429.
- Guberova, E. (2001). Die kindliche Entwicklung der Wahrnehmung für Gesichter: Eine experimentelle Überprüfung mittels einer Wiedererkennens- und einer Matching-Aufgabe. Magisterarbeit an der Universität Wien.
- Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs. ingroup face stimuli. Neuroreport, *11*, 2351-2355.
- Henning, C. (2003). Wiedererkennen von türkischen und deutschen Gesichtern: Experimentelle Untersuchung des Ausländereffekts. Diplomarbeit am Fachbereich Psychologie und Sportwissenschaft der Justus-Liebig-Universität Gießen.
- Hockley, W. E., Hemsworth, D. H., & Consoli, A. (1999). Shades of the mirror effect: Recognition of faces with and without sunglasses. Memory & Cognition, *27*, 128-138.
- Jalbert, N. L., & Getting, J. (1992). Racial and gender issues in facial recognition. In F. Lösel, D. Bender & T. Bliesener (Eds.). Psychology and Law: International Perspectives. Berlin: De Gruyter.
- Johnston, R. A., Milne, A. B., Williams, C., & Hosie, J. (1997). Do distinctive faces come from outer space? An investigation of the status of a multidimensional face-space. Visual Cognition, *4*, 59-67.
- Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of the behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), Attribution: Perceiving the causes of behavior. Morristown, NJ: General Learning Press.

-
- Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. Journal of Experimental Psychology: Learning, Memory, and Cognition, *26*, 1534-1555.
- Judd, C. M., & Park, B. (1988). Out-Group homogeneity: Judgments of variability at the individual and group levels. Journal of Personality and Social Psychology, *54*, 778-788.
- Justus-Liebig Universität Gießen (2003). Aufgerufen im November 2001, von <http://www.uni-giessen.de/uni/broschuere/JLU.html>
- Kahnemann, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, *80*, 237-251.
- Kanwisher, N., Chun, M. M., McDermott, J., & Ledden, P. (1996). Functional imaging of human visual recognition. Cognitive Brain Research, *5*, 55– 67.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in extrastriate cortex specialized for face perception. The Journal of Neuroscience, *17*, 4302-4311.
- Kemp, R, Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. Applied Cognitive Psychology, *11*, 211-222.
- Levin, D. T. (1996). Classifying faces by race: The structure of face categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, *22*, 1364-1382.
- Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. Journal of Experimental Psychology: General, *129*, 559-574.
- Leyens, J. P., & Codol, J. P. (1992). Soziale Informationsverarbeitung. In W. Stroebe, M. Hewstone, J. P. Codol & G. M. Stephenson (Hrsg.), Sozialpsychologie: Eine Einführung. Springer Verlag.
- Leyens, J. P., & Dardenne, B. (1996). Basic concepts and approaches in social cognition. In M. Hewstone, W. Stroebe, & G. M. Stephenson (Eds.), Introduction to Social Psychology: A European Perspective. Oxford: Blackwell.
- Lindsay, D. S., Jack, P. C., & Christian, M. A. (1991). Other-race face perception. Journal of Applied Psychology, *76*, 587-589.
- Lindsay, R. C. L., & Wells, G. L. (1983). What do we really know about crossrace eyewitness identification? In S. Lloyd-Bostock & B. R. Clifford (Eds.), Evaluating Witness Evidence: Recent Psychological Research and new Perspectives. Chichester: Wiley.

-
- Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492-527.
- MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. Psychology, Public Policy, and Law, 7, 98-118.
- MacLin, O. H., & Malpass, R. S. (2003). Last but not least. The ambiguous race-face illusion. Perception, 32, 249-252.
- Malpass, R. S. (1981). Training in face recognition. In: Davies, G. M.; Ellis, H. D. & Shepherd, J. W. (Eds.), Perceiving and Remembering Faces. London: Academic Press.
- Malpass, R. S. (1990). An excursion into utilitarian analysis. Behavior Science Research, 24, 1-15.
- Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other races. Journal of Personality and Social Psychology, 13, 333-334.
- Malpass, R. S., Laviguer, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. Perception & Psychophysics, 14, 285-292.
- Maurer, D., Le Grand, R., & Mondloch, C. (2002). The many faces of configural processing. Trends in Cognitive Sciences, 6, 255-260.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A metaanalytic review. Psychology, Public Policy, and Law, 7, 3-35.
- Meissner, C.A., & Brigham, J.C. (unveröffentlichtes Manuskript). Memory for own- and other-race faces: A dual-process approach. Florida International University, Miami, FL.
- Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-93.
- Molz, G., & Hopf, A. (2002). The euro - a social accentuation experiment. Economic Sociology, 3, 26-31.
- Morris, P. E., & Wickham, L. H. V. (2001). Typicality and face recognition: A critical re-evaluation of the two factor theory. The Quarterly Journal of Experimental Psychology, 54, 863-877.
- Mullen, B. (1989). Advanced basic meta-analysis. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on X-Ray image processing. Journal of Experimental Psychology: Learning, Memory and Cognition, 14, 553-557.

-
- Ng, W-J; & Lindsay, R. C. L. (1994). Cross-race facial recognition. Journal of Cross-Cultural Psychology, *25*, 217-232.
- O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other race-effect. Memory & Cognition, *22*, 208-224.
- Palmer, R. J., & Masling, J. (1969). Vocabulary for skin color in negro and white children. Developmental Psychology, *1*, 396-401.
- Pfütze, E., Sommer, W., & Schweinberger, S.R. (2002). Age-Related slowing in face and name recognition: Evidence from event-related brain potentials. Psychology and Aging, *17*, 140-160.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. Journal of Cognitive Neuroscience, *12*, 729-738.
- PsyScope (v 1.2.5). [Computer software]. Carnegie Mellon University, Department of Psychology. Pittsburgh, PA. Aufgerufen von <http://psyscope.psy.cmu.edu>.
- Publication Manual of the American Psychological Association (5th ed.). (2001). American Psychological Association. Washington, DC.
- Quattrone, G. A., & Jones, E. E. (1980). The perception of variability within ingroups and outgroups. Journal of Personality and Social Psychology, *42*, 593-607.
- Rae, G. (1976). Table of A'. Perceptual and Motor Skills, *42*, 98.
- Reder, L. M., Angstadt, P., Cary, M., Erickson, M. A., & Ayers, M. S. (2002). A reexamination of Stimulus-frequency effects in Recognition: Two mirrors for low- and high-frequency Pseudowords. Journal of Experimental Psychology: Learning, Memory, and Cognition, *28*, 138-152.
- Rhodes, G., Brake, S., & Atkinson, A. P. (1989). Expertise and configural coding in face recognition. British Journal of Psychology, *80*, 313-331.
- Rodin, M.J. (1987). Who is memorable to whom: A study of cognitive disregard. Social Cognition, *5*, 444-465.?
- Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Bream, P. (1976). Basic objects in natural categories. Cognitive Psychology, *8*, 382-439.
- Rossion, B., Gauthier, I. Tarr, M. J., Despland, P. A., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. NeuroReport, *11*, 69-74.

-
- Rossion, B., Gauthier, I., Delvenne, J.-F., Tarr, M. J., Bruyer, R., & Crommelinck, M. (1999, November). Does the N170 occipito-temporal component reflect a face-specific, structural encoding stage? Poster presented at the 7th Annual Workshop on Object Perception and Memory (OPAM) in Los Angeles, USA. Aufgerufen von http://cognitn.psych.indiana.edu/busey/Q301/pdfs/OPAMposter_99.pdf
- Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. Psychological Science, *13*, 250-257.
- Rost, J. (1995). Testmodelle. Lehrbuch Testtheorie – Testkonstruktion. Bern: Huber.
- Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. Journal of Memory and Language, *43*, 68-88.
- Schooler, J. W., & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. Cognitive Psychology, *22*, 36-71.
- Schwarzer, G. (2000). Development of face processing: The effect of face inversion. Child Development, *71*, 391-401.
- Schweinberger, S. R., Pickering, E. C., Jentsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. Cognitive Brain Research, *14*, 398-409.
- Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face memory. European Journal of Cognitive Psychology, *3*, 137-145.
- Slone, A. E., Brigham, J. C., & Meissner, C. A. (2000). Social and cognitive factors affecting the own-race bias in whites. Basic and Applied Social Psychology, *22*, 71-84.
- Sporer, S. L. (1992). Encoding strategies and the recognition of human faces. Journal of Experimental Psychology: Human Learning, Memory, and Cognition, *17*, 323-333.
- Sporer, S. L. (1992). Das Wiedererkennen von Gesichtern. Weinheim: Psychologie-Verlags-Union.
- Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D. F. Ross, J. D. Read & M. P. Toglia (Eds.) Adult Eyewitness Testimony. Cambridge University Press.
- Sporer, S. L. (1999, July). The Own-Race Bias in Germany: Testing the Contact Hypothesis with Turks and Germans. Paper presented at the 4th European Conference on Psychology and Law of the American Psychology-Law Association and the European Psychology-Law Association in Dublin, Ireland.

-
- Sporer, S. L. (2001a). Recognizing faces of other ethnic groups: An integration of theories. Psychology, Public Policy and Law, 7, 36-97.
- Sporer, S. L. (2001b). The Cross-Race: Beyond recognition of faces in the laboratory. Psychology, Public Policy, and Law, 7, 170-200.
- Sporer, S. L., Trinkl, B., & Guberova, E. (in Druck). Matching faces. Differences in processing speed of out-group faces by different ethnic groups.
- SuperLab Pro (v. 1.75). [Computer software]. Cedrus Corporation. Phoenix, Arizona.
- Swets, J. A. (1964). Signal detection and recognition by human observers. New York: Wiley.
- Tabachnik, B. G., & Fidell, L. S. (2001). Principal components and factor analysis. In B. G. Tabachnik & L. S. Fidell (Eds.), Using Multivariate Statistics. Needham Heights: Allyn & Bacon.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of social contact. In W. Austin & S. Worchel (Eds.), The Social Psychology of Intergroup Relations. Monterey: Brooks/Cole.
- Tanaka, J. W. (2001). The entry point of face recognition. Evidence for face expertise. Journal of Experimental Psychology, 130, 534-543.
- Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. Psychological Science, 12, 43-47.
- The Bank of England (n.d.). Aufgerufen im November 2001 von <http://www.bankofengland.co.uk/>
- Tickner, A. H., & Poulton, E. C. (1975). Watching for People and actions. Ergonomics, 18, 35-51.
- Tischler, L. C., Kisseler, W., & Trabert, L. (2002). Migrationsreport Hessen 2002: Bevölkerung, Ausbildung und Arbeitsmarkt. FEH-Report Nr. 637. Forschungs- und Entwicklungsgesellschaft Hessen mbH (FEH). Wiesbaden.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. Psychological Review, 95, 15-48.
- Trinkl, B. (2001). Das Wiedererkennen von Gesichtern unterschiedlicher ethnischer Gruppen: Eine experimentelle Untersuchung mit Türken und Deutschen. Magisterarbeit an der Universität Wien.
- Tulving, E. (1985). Memory and consciousness. Canadian Psychology, 26, 1-12.
- Valentine, T. (1988). Upside down faces: A review of the effect of inversion upon face recognition. British Journal of Psychology, 79, 471-491.

-
- Valentine, T. (1991a). A unified account of the effects of distinctiveness, inversion and race on face recognition. Quarterly Journal of Experimental Psychology, *44*, 671-703.
- Valentine, T. (1991b) Representation and process in face recognition. In R. Watt (Ed.), Vision and Visual Dysfunction. Vol. 14: Pattern Recognition in Man and Machine. London: Macmillan.
- Valentine, T. (1999). Face-Space models of face recognition. In M. J. Wenger und J. T. Townsend (Eds.). Computational, Geometric and Process Perspectives on Facial Recognition: Contexts and Challenges. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Valentine, T., & Bruce V. (1986). The effect of race, inversion and encoding activity upon face recognition. Acta Psychologica, *61*, 259-273.
- Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. Memory & Cognition, *20*, 291-302.
- Weimann, G., Fishman, G., & Rattner, A. (1986). Social distance and misidentification. International Journal of Comparative Sociology, *27*, 217-223.
- Wikipedia, the free encyclopedia (n.d.). Race. Aufgerufen von <http://en.wikipedia.org/wiki/Race>.
- Winer (1971). Statistical principles in experimental design (2nd ed.). New York: Wiley.
- Wright, D. B. (2003). Making friends with your data: Improving how statistics are conducted and reported. British Journal of Educational Psychology, *7*, 123-136.
- Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. Psychology, Public Policy, and Law, *7*, 119-133.
- Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. Applied Cognitive Psychology, *17*, 365-373.
- Wright, D.B., & Stroud, J. N. (2002). Age differences in lineup identification accuracy: People are better with their own age. Law and Human Behavior, *6*, 641-654.
- Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. Acta Psychologica, *114*, 101-114.
- Yin, R. K. (1969). Looking at upside down faces. Journal of Experimental Psychology, *81*, 141-145.
- Yin, R. K. (1970). Face Recognition: A dissociable ability ?. Neuropsychologica, *23*, 395-402.

-
- Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. Memory & Cognition, *25*, 747-763.
- Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source memory judgements: A formal dual-process model and an ROC analysis. Journal of Experimental Psychology: Learning, Memory, and Cognition, *25*, 1415-1434.
- Yonelinas, A. P. (2002). Components of episodic memory: The contribution of recollection and familiarity. In A. Baddeley, J. P. Aggleton, & M. A. Conway. Episodic Memory: New Definitions in Research. London: Oxford University Press.
- Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. Social Cognition, *8*, 161-185.

9 Anhang

Anhang A: Instruktionen

Anhang A1: Instruktionen für die Durchführung der Delayed-Matching-Aufgabe zum Cross-Currency Effekt. Zwischen den Instruktionen 6. und 7. wurden zwei Beispiele bearbeitet

1. Im Folgenden geht es darum, das Aussehen verschiedener Geldscheine (Euro, Deutsche Mark und Britisches Pfund) zu erlernen. Dazu werden jeweils zwei Geldscheine der gleichen Währung präsentiert.
2. Zuerst werden diese zwei Geldscheine relativ rasch hintereinander dargeboten.
3. Deine Aufgabe ist, zu entscheiden, ob die beiden Geldscheine identisch aussehen, oder nicht. Lege dazu bitte die Zeigefinger auf die Tasten 'c' und 'm'.
4. Wenn Du denkst, dass die Geldscheine identisch sind, dann drücke bitte die Taste 'c'. Wenn du denkst, dass die Geldscheine verschieden sind, dann drücke bitte die Taste 'm'.
5. Bitte drücke die entsprechende Taste erst dann, wenn der schwarze Bildschirm erscheint !
6. Jetzt kommen zwei Beispiele, damit du dich an das Experiment gewöhnen kannst.
7. Überlege nicht zu lange, da der erste Eindruck fast immer die richtige Antwort ist. Wenn du einmal gedrückt hast, kann man es nicht mehr rückgängig machen. Diese Antwort gilt. Die nächste Präsentation kommt dann automatisch. Starte das Experiment mit einer beliebigen Taste.

Anhang A2 Instruktionen für die Durchführung der AFC-Aufgabe zum Cross-Currency Effekt

1. Im Folgenden geht es darum, das Aussehen verschiedener Geldscheine (Euro, deutsche Mark und Britisches Pfund) zu erlernen. Dazu werden jeweils zwei Geldscheine der gleichen Währung präsentiert.
2. In diesem Durchgang werden zwei Geldscheine gleichzeitig dargeboten. Sie werden Dir für zwei Sekunden präsentiert.
3. Wenn dann der Bildschirm Schwarz wird, musst Du durch Tastendruck möglichst rasch entscheiden, welcher der beiden Geldscheine der 'Echte' ist. Lege dazu bitte die Zeigefinger auf die Tasten 'm' und 'c'.
4. Wenn Du denkst, dass der obere Geldschein der 'Echte' ist, dann drücke bitte die Taste 'c'. Wenn Du denkst, dass der untere Geldschein der 'Echte' ist, dann drücke bitte die Taste 'm'. Der Computer gibt Dir immer eine Rückmeldung über Deine Entscheidung. Anschließend wird Dir noch mal der richtige Geldschein gezeigt.
5. Überlege nicht zu lange, da der erste Eindruck fast immer der die richtige Antwort ist. Wenn Du einmal gedrückt hast, kann man es nicht mehr rückgängig machen. Diese Antwort gilt. Die nächste Präsentation kommt dann automatisch. Starte das Experiment mit einer beliebigen Taste.
6. In diesem Durchgang ist die Darbietungszeit der Geldscheine nicht mehr auf zwei Sekunden beschränkt. Du sollst aber wieder möglichst rasch entscheiden, welcher der beiden Geldscheine der 'Echte' ist. Wenn Du denkst, dass der obere Geldschein richtig ist, drücke die Taste 'c'. Wenn Du meinst, dass der untere Geldschein der Richtige ist, drücke die Taste 'm'.

Anhang A3 Instruktionen für die Durchführung des Wiedererkennenstests zum Cross-Currency Effekt

1. Nun werden Dir alle Geldscheine noch einmal gezeigt. Deine Aufgabe ist es, die Zeigefinger auf die Tasten 'm' und 'c' zu legen.
2. Wenn Du meinst, dass gerade ein 'echter' Geldschein präsentiert wurde, dann drücke die Taste 'c' ('m'). Wenn Du meinst, dass gerade ein 'gefälschter' Geldschein präsentiert wurde, dann drücke die Taste 'm' ('c').

Anhang A4. Instruktionen für die Durchführung der Pilotstudie zu Experiment II

Instruktionsphase

1. Bei dieser Aufgabe geht es darum, Gesichter von Männern türkischer und deutscher Herkunft nach bestimmten Eigenschaften zu beurteilen.
2. Damit Sie eine Ahnung davon bekommen, wie der Versuch abläuft, kommt zunächst ein Beispiel:
3. Zuerst wird Ihnen das Bild einer Person für zwei Sekunden präsentiert. Danach erscheint das gleiche Bild in der oberen Hälfte des Bildschirms. In der unteren Hälfte erscheinen hintereinander fünf Fragen, die Sie durch Drücken der Tasten von 1-7 beantworten sollen. Der Wert "1" bedeutet immer die niedrigste Ausprägung der Eigenschaft, der Wert "7" ist immer die höchste Ausprägung.
4. Haben Sie einmal eine Taste gedrückt, so kann dies nicht mehr rückgängig gemacht werden. Wenn Sie die fünf Beurteilungen zu einem Bild abgegeben haben, schließt sich sofort der nächste Durchgang an.
5. Wenn Sie jetzt noch Fragen haben, wenden Sie sich bitte an die Versuchsleitung.

Versuchsphase

1. Wenn Sie bereit sind, drücken Sie bitte eine Taste.

Anhang A5. Instruktionen für die Durchführung des Wiedererkennenstests in Experiment II

Darbietungsphase

1. Bei dieser Aufgabe geht es darum, wie gut sie sich Gesichter (Männern) von schwarzen und weißen Amerikanern (türkischer und deutscher Herkunft) merken können. Es werden Ihnen 80 Gesichter von schwarzen und weißen Amerikanern (Türken und Deutschen) für je 5 Sekunden in zufälliger Reihenfolge dargeboten. Ihre Aufgabe ist es, sich die Bilder anzuschauen und zu versuchen, sie sich einzuprägen.
2. (zusätzliche Instruktion für die Maskierungsbedingung) Die Gesichter, die Sie sehen werden, sind 'beschnitten', d.h. es ist nur der innere Teil der Gesichter ohne Haare und Ohren sichtbar.
3. Damit Sie sich darauf einstellen können, wie schnell die Bilder präsentiert werden, werden drei Symbole als 'Probepilder' präsentiert. Versuchen Sie bitte, nur in die Mitte des Bildschirms zu sehen, dorthin, wo die Symbole auftauchen. Starten Sie das Experiment bitte mit einer beliebigen Taste.

Testphase

1. Es werden Ihnen auf dem Computer Gesichter gezeigt. Diese Gesichter werden Ihnen im Halbprofil (45° Winkel) dargeboten. Zum Teil haben Sie sie bereits (vor drei Wochen) gesehen, zum Teil sind die Gesichter neu. Ihre Aufgabe ist es, die Zeigefinger auf die Tasten 'm' und 'c' zu legen.
2. Wenn Sie das Gesicht von der ersten Aufgabe kennen, dann drücken sie die Taste 'm' ('c'). Wenn Sie das Gesicht nicht kennen, dann drücken Sie die Taste 'c' ('m'). Überlegen Sie nicht zu lange, da der erste Eindruck fast immer die richtige Antwort ist. Wenn Sie einmal gedrückt haben, kann man es nicht mehr rückgängig machen. Diese Antwort gilt. Das nächste Bild kommt dann automatisch.

Anhang A6. Instruktionen für die Durchführung der Pilotstudie zu Experiment III

Instruktionsphase

1. Bei dieser Aufgabe geht es darum, Gesichter von Männern latein-amerikanischer Herkunft nach bestimmten Eigenschaften zu beurteilen.
2. Damit Sie eine Ahnung davon bekommen, wie der Versuch abläuft, kommt zunächst ein Beispiel:
3. Zuerst wird Ihnen das Bild einer Person für zwei Sekunden präsentiert. Danach erscheint das gleiche Bild in der oberen Hälfte des Bildschirms. In der unteren Hälfte erscheinen hintereinander fünf Fragen, die Sie durch Drücken der Tasten von 1-7 beantworten sollen. Der Wert "1" bedeutet immer die niedrigste Ausprägung der Eigenschaft, der Wert "7" ist immer die höchste Ausprägung.
4. Haben Sie einmal eine Taste gedrückt, so kann dies nicht mehr rückgängig gemacht werden. Wenn Sie die fünf Beurteilungen zu einem Bild abgegeben haben, schließt sich sofort der nächste Durchgang an.
5. Wenn Sie jetzt noch Fragen haben, wenden Sie sich bitte an die Versuchsleitung.

Versuchsphase

1. Wenn Sie bereit sind, drücken Sie bitte eine Taste.

Anhang A7. Instruktionen für die Durchführung des Wiedererkennenstests in Experiment III

Darbietungsphase

1. Bei dieser Aufgabe geht es darum, wie gut Sie sich Gesichter von Männern türkischer, deutscher, afroamerikanischer und latein-amerikanischer Herkunft merken können
2. Es werden Ihnen 80 Gesichter von Männern türkischer, deutscher, afroamerikanischer und latein-amerikanischer Herkunft für je 5 Sekunden in zufälliger Reihenfolge dargeboten. Ihre Aufgabe ist es, sich die Bilder anzuschauen und zu versuchen, sie sich einzuprägen.
3. Damit Sie sich darauf einstellen können, wie schnell die Bilder präsentiert werden, werden drei Symbole als "Probepbilder" präsentiert. Versuchen Sie bitte, nur in die Mitte des Bildschirms zu sehen, dorthin, wo die Symbole auftauchen. Starten Sie das Experiment bitte mit einer beliebigen Taste.

Testphase

1. Es werden Ihnen auf dem Computer Gesichter gezeigt. Diese Gesichter werden Ihnen im Halbprofil (45° Winkel) dargeboten. Zum Teil haben Sie sie bereits gesehen, zum Teil sind die Gesichter neu. Ihre Aufgabe ist es, die Zeigefinger auf die Tasten "m" und "c" zu legen.
2. Wenn Sie das Gesicht von der ersten Aufgabe kennen, dann drücken Sie die Taste "c" (Anm.: je nach Bedingung auch "Taste 'm'"). Wenn Sie das Gesicht nicht kennen, dann drücken Sie die Taste "m" (Anm.: s.o.) Überlegen Sie nicht zu lange, da der erste Eindruck fast immer die richtige Antwort ist. Wenn Sie einmal gedrückt haben, kann man es nicht mehr rückgängig machen. Diese Antwort gilt. Das nächste Bild kommt dann automatisch.
3. Starten Sie das Experiment bitte mit einer beliebigen Taste.

Anhang A8. Instruktionen für die Durchführung des Delayed-Matching Tests in Experiment III

1. Im Folgenden werden Ihnen wieder Gesichter gezeigt. Diese Gesichter werden Ihnen im Halbprofil (45 ° Winkel) (Anm.: dieser Satz fehlt in der Bedingung mit ausschließlich frontalansichtigen Gesichtern). Dabei werden Ihnen jeweils zwei Gesichter der gleichen ethnischen Gruppe präsentiert.
2. Zuerst werden diese zwei Gesichter relativ rasch hintereinander dargeboten. Das zweite Gesicht wird dabei immer im Halbprofil präsentiert (Anm.: bzw. "Beide Gesichter werden dabei immer in Frontalansicht präsentiert")
3. Ihre Aufgabe ist, zu entscheiden, ob die beiden Gesichter identisch aussehen oder nicht. Legen Sie dazu bitte die Zeigefinger auf die Tasten "m" und "c".
4. Wenn Sie denken, dass die Gesichter identisch sind, dann drücken Sie bitte die Taste "c" (Anm.: bzw. Taste "m"). Wenn Sie denken, dass die Gesichter verschieden sind, dann drücke Sie bitte die Taste "m" (Anm.: bzw. Taste "c")
5. Bitte drücken Sie die entsprechende Taste erst dann, wenn der schwarze Bildschirm erscheint !
6. Jetzt kommen zwei Beispiele, damit Sie sich an das Experiment gewöhnen können.

Anhang B: Fragebögen

Anhang B1: Fragebogen zum Gebrauch von Geldscheinen in den Experimenten I-III

Vp-Nr.: Geschlecht: M:___ F:___ Alter:

Key: m:___ c:___ Datum: __./__./_____

EURO - FRAGEBOGEN

Im folgenden Fragebogen geht es um den Kontakt mit der neuen Währung Euro. Dazu gibt es keine "richtigen" oder "falschen" Angaben, sondern es geht lediglich um Einschätzungen Ihrer Erfahrungen. Bitte versuchen Sie, alle Antworten so gut wie möglich nach Ihrer persönlichen Erfahrung zu beantworten und lassen Sie auf keinen Fall Fragen aus!

Wissensstand zum Euro

- (1) Welchen Wert weist der Euro-Schein mit der größten Wertigkeit auf?

/_____/ Euro

- (2) Wie lange behält das DM-Bargeld nach dem 1.1.2001 noch seinen Wert?

Gar ein zwei drei sechs ein unbegrenzt

nicht Monat Monate Monate Monate Jahr

1 2 3 4 5 6 7 |____|

- (3) In welchem Land/welchen Länder der EU wird der Euro
- nicht
- zum 1. Januar 2001 eingeführt? (Bitte ankreuzen!)

Belgien |____| Irland |____|

Dänemark |____| Luxemburg |____|

Deutschland |____| Niederlande |____|

England |____| Österreich |____|

Frankreich |____| Portugal |____|

Finnland |____| Schweden |____|

Griechenland |____| Spanien |____|

Italien |____|

- (4) Sind die Geldscheine "länderspezifisch" gestaltet?

ja |____| nein |____|

(5) Mit welchem Thema sind die Geldscheine gestaltet?

Landschaften Gesichter

Baustile nationale Symbole

Bargeldgebrauch

Stellen Sie sich eine "Durchschnittswoche" vor! Wie oft kommen Sie bei den folgenden Gelegenheiten mit Bargeld (d.h. Münzen und Scheinen) in Kontakt? Bitte tragen Sie die Häufigkeit pro Woche ein!

(6) in Geschäften/Kaufhäusern

(7) an Ihrem Arbeitsplatz, an der Universität bzw. an Ihrer Ausbildungsstätte (z.B. Mensa, Kantine)

(8) bei Freizeitaktivitäten (z.B. Kneipen oder Restaurants)

(9) Wie häufig gehen Sie für den alltäglichen Bedarf (z.B. Lebensmittel) einkaufen?

(10) Wie häufig essen Sie auwärts (z.B. Mensa oder Restaurants)?

(11) Wie häufig gehen Sie abends aus und geben dabei Geld aus (z.B. Kneipe, Kino o.ä.)

(12) Besitzen Sie eine EC/Kredit – Karte? nein ja

(13) Wie häufig benutzen Sie diese . im Vergleich zu Bargeld?

Sehr viel

Sehr viel

seltener

häufiger

1 2 3 4 5 6 7

(14) Welchen der DM-Scheine benutzen Sie am häufigsten ? Ordnen Sie die Scheine bitte der Häufigkeit nach, beginnend mit dem am häufigsten gebrauchten Schein.

10 DM 20 DM 50 DM 100 DM

(15) Welchen der Euro-Scheine werden Sie Ihrer Einschätzung nach am häufigsten benutzen ? Ordnen Sie die Scheine bitte der Häufigkeit nach, beginnend mit dem am häufigsten gebrauchten Schein.

5 Euro 10 Euro 20 Euro 50 Euro

Vertrautheit mit dem Euro

(16) Wie sind Sie dem Euro gegenüber eingestellt?

Euro- skeptisch						Euro- enthusiastisch	
1	2	3	4	5	6	7	____

(17) Wenn Sie den Euro mit der D-Mark vergleichen - wie vertraut sind Sie mit den Münzen des Euro

Sehr viel schlechter						Sehr viel besser	
1	2	3	4	5	6	7	____

(18) Wenn Sie den Euro mit der D-Mark vergleichen – wie vertraut sind Sie mit den Geldscheinen des Euro?

<i>Sehr schlecht</i>						<i>Sehr gut</i>	
1	2	3	4	5	6	7	____

(19) Wenn Sie sich mit Ihren Mitstudierenden/ Arbeitskolleg(inn)en vergleichen: wie aktiv bemühen Sie sich um Informationen zum Euro im Vergleich zu den "Durchschnittsstudierenden" an Ihrer Universität/den "Durchschnittsarbeitnehmer(inn)en" an Ihrem Arbeitsplatz?

Sehr viel weniger						Sehr viel mehr	
1	2	3	4	5	6	7	____

(20) Wie gut konnten Sie sich durch Medieninformationen (Fernsehen, Plakate, Broschüren) auf die Währungsumstellung vorbereiten?

Sehr schlecht						Sehr gut	
1	2	3	4	5	6	7	____

Anhang B2. Fragebogen zum Kontakt mit Türken respektive Deutschen

FRAGEBOGEN ÜBER KONTAKTE MIT TÜRKEN

Im folgenden Fragebogen geht es um den Kontakt mit Personen türkischer Herkunft. Dazu gibt es keine "richtigen" oder "falschen" Angaben, sondern lediglich um Einschätzungen Ihrer Erfahrungen. Bitte versuchen Sie, alle Antworten so gut wie möglich nach Ihrer persönlichen Erfahrung zu beantworten und lassen Sie auf keinen Fall Fragen aus!

Gelegenheiten zum Kontakt in der Vergangenheit

- (1) Bitte schätzen Sie den Prozentsatz türkischer Kinder, die mit Ihnen die Grundschule besucht haben. |%
- (2) Bitte schätzen Sie den Prozentsatz türkischer Kinder, die mit Ihnen die Mittelstufe (Kl. 7 - 10) besucht haben. |%
- (3) Bitte schätzen Sie den Prozentsatz türkischer Personen, die mit Ihnen die Oberstufe (Kl. 11 - 13) besucht haben bzw. mit Ihnen an Ihrem Ausbildungsplatz waren. |%
- (4) Bitte schätzen Sie, wieviel Prozent der Mitmenschen in der Nachbarschaft, in der Sie aufgewachsen sind, türkischer Herkunft waren. |%

Freundschaften in der Vergangenheit

- (5) Wie viele Freundschaften hatten Sie während Ihrer Grundschulzeit mit Kindern türkischer Herkunft? (Anzahl Personen)
- (6) Wie viele Freundschaften hatten Sie während Ihrer Zeit in der Mittelstufe mit Personen türkischer Herkunft? (Anzahl Personen)
- (7) Wie viele Freundschaften unterhielten Sie während Ihrer Zeit in der Oberstufe bzw. an Ihrem Ausbildungsplatz mit Personen türkischer Herkunft? (Anzahl Personen)

Derzeitige Alltagskontakte

Stellen Sie sich eine "Durchschnittswoche" vor! Bei welchen Gelegenheiten kommen Sie mit mehr als zwei Personen türkischer Herkunft in Kontakt? (Bitte ankreuzen!)

- (8) an der Universität bzw. an Ihrer Ausbildungsstätte
- (9) bei Freizeitaktivitäten
- (10) bei der Arbeit
- (11) in Geschäften/Kaufhäusern
- (12) im Haus, in dem ich wohne

Engere Kontakte

- (13) Wie viele Türken kennen Sie mit Vor- oder Spitznamen? (Anzahl Personen)
- (14) Bitte denken Sie nun an Ihre neun engsten Freunde: wie viele von ihnen sind türkischer Herkunft? (Anzahl Personen)
- (15) Mit wie vielen türkischen Personen des anderen Geschlechts sind sie schon abends ausgegangen (Date)? (Anzahl Personen)
- (16) Wenn Sie sich mit Ihren deutschen Mitstudierenden/ Ausbildungskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit Personen türkischer Herkunft im Vergleich zum "Durchschnittsstudierenden" an Ihrer Universität/Ihrem Arbeitsplatz?

Sehr viel
weniger

Sehr viel
mehr

1 2 3 4 5 6 7

- (17) Wie viele Jahre leben Sie schon in Deutschland?

- (18) Wenn Sie sich mit Ihren deutschen Mitstudierenden/ Ausbildungskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit "Schwarzen" (Personen schwarzer Hautfarbe) im Vergleich zum "Durchschnittsperson" an Ihrer Universität/Ihrem Arbeitsplatz/im Alltag?

Sehr viel
weniger

1 2 3 4 5 6 7

Sehr viel
mehr

|____|

- (19) Wenn Sie sich mit Ihren deutschen Mitstudierenden/ Ausbildungskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit "Schwarzen" (Personen schwarzer Hautfarbe) im Vergleich zum "Durchschnittsperson" durch Medienkonsum (Kino, Fernsehen, Video)?

Sehr viel
weniger

1 2 3 4 5 6 7

Sehr viel
mehr

|____|

Im Folgenden sollen Sie die Gesichter, die Sie vorher gesehen haben, unter einer größeren Anzahl von Gesichtern wiedererkennen. Das heißt, einige Gesichter sind "alt", andere sind "neu". Die Aufgabe wird dadurch etwas schwieriger, dass die Gesichter nicht wie vorher von vorne gezeigt werden, sondern im Halbprofil (45° Winkel):

- (20) Wieviel Prozent glauben Sie, richtig wiedererkennen zu können (zwischen 50% und 100%, denn 50% hätten Sie alleine durch Raten richtig):

(20a) Gesichter von Türken

|____|

(20b) Gesichter von Deutschen

|____|

(20c) Gesichter von schwarzen Amerikanern

|____|

(20d) Gesichter von weissen Amerikanern

|____|

FRAGEBOGEN ÜBER KONTAKTE MIT DEUTSCHEN

Im folgenden Fragebogen geht es um den Kontakt mit Personen deutscher Herkunft. Dazu gibt es keine "richtigen" oder "falschen" Angaben, sondern lediglich um Einschätzungen Ihrer Erfahrungen. Bitte versuchen Sie, alle Antworten so gut wie möglich nach Ihrer persönlichen Erfahrung zu beantworten und lassen Sie auf keinen Fall Fragen aus!

Gelegenheiten zum Kontakt in der Vergangenheit

- (1) Bitte schätzen Sie den Prozentsatz deutscher Kinder, die mit Ihnen die Grundschule besucht haben. ||%
- (2) Bitte schätzen Sie den Prozentsatz deutscher Kinder, die mit Ihnen die Mittelstufe (Kl. 7 - 10) besucht haben. ||%
- (3) Bitte schätzen Sie den Prozentsatz deutscher Personen, die mit Ihnen die Oberstufe (Kl. 11 - 13) besucht haben bzw. mit Ihnen an Ihrem Ausbildungsplatz waren. ||%
- (4) Bitte schätzen Sie, wieviel Prozent der Mitmenschen in der Nachbarschaft, in der Sie aufgewachsen sind, deutscher Herkunft waren. ||%

Freundschaften in der Vergangenheit

- (5) Wie viele Freundschaften hatten Sie während Ihrer Grundschulzeit mit Kindern deutscher Herkunft? (Anzahl Personen) ||
- (6) Wie viele Freundschaften hatten Sie während Ihrer Zeit in der Mittelstufe mit Personen deutscher Herkunft? (Anzahl Personen) ||
- (7) Wie viele Freundschaften unterhielten Sie während Ihrer Zeit in der Oberstufe bzw. an Ihrem Ausbildungsplatz mit Personen deutscher Herkunft? (Anzahl Personen)||

Derzeitige Alltagskontakte

Stellen Sie sich eine "Durchschnittswoche" vor! Bei welchen Gelegenheiten kommen Sie mit mehr als zwei Personen deutscher Herkunft in Kontakt? (Bitte ankreuzen!)

- (8) an der Universität bzw. an Ihrer Ausbildungsstätte
- (9) bei Freizeitaktivitäten
- (10) bei der Arbeit
- (11) in Geschäften/Kaufhäusern
- (12) im Haus, in dem ich wohne

Engere Kontakte

- (13) Wie viele Deutsche kennen Sie mit Vor- oder Spitznamen? (Anzahl Personen)
- (14) Bitte denken Sie nun an Ihre neun engsten Freunde: wie viele von ihnen sind deutscher Herkunft? (Anzahl Personen)
- (15) Mit wie vielen deutschen Personen des anderen Geschlechts sind sie schon abends ausgegangen (Date)? (Anzahl Personen)
- (16) Wenn Sie sich mit Ihren türkischen Mitstudierenden/ Ausbildungskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit Personen deutscher Herkunft im Vergleich zum (türkischen) "Durchschnittsstudierenden" an Ihrer Universität/Ihrem Arbeitsplatz?

Sehr viel
weniger

Sehr viel
mehr

1 2 3 4 5 6 7

- (17) Wie viele Jahre leben Sie schon in Deutschland?

- (18) Wenn Sie sich mit Ihren türkischen Mitstudierenden/ Ausbildungskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit "Schwarzen" (Personen schwarzer Hautfarbe im Vergleich zum "Durchschnittsperson" an Ihrer Universität/Ihrem Arbeitsplatz/im Alltag?

Sehr viel
weniger

Sehr viel
mehr

1 2 3 4 5 6 7

|____|

- (19) Wenn Sie sich mit Ihren türkischen Mitstudierenden/ Ausbildungskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit "Schwarzen" (Personen schwarzer Hautfarbe) im Vergleich zur "Durchschnittsperson" durch Medienkonsum (Kino, Fernsehen, Video)?

Sehr viel
weniger

Sehr viel
mehr

1 2 3 4 5 6 7

|____|

Im Folgenden sollen Sie die Gesichter, die Sie vorher gesehen haben, unter einer größeren Anzahl von Gesichtern wiedererkennen. Das heißt, einige Gesichter sind "alt", andere sind "neu". Die Aufgabe wird dadurch etwas schwieriger, dass die Gesichter nicht wie vorher von vorne gezeigt werden, sondern im Halbprofil (45° Winkel):

- (20) Wieviel Prozent glauben Sie, richtig wiedererkennen zu können (zwischen 50% und 100%, denn 50% hätten Sie alleine durch Raten richtig):

(20a) Gesichter von Türken

|____|

(20b) Gesichter von Deutschen

|____|

(20c) Gesichter von schwarzen Amerikanern

|____|

(20d) Gesichter von weissen Amerikanern

|____|

Anhang B3. Fragebogen über Kontakte mit Ausländern

**FRAGEBOGEN ÜBER KONTAKTE MIT AUSLÄNDERN
(Version für BGS-Beamte)**

Im folgenden Fragebogen geht es um den Kontakt mit Personen ausländischer Herkunft. Dazu gibt es keine "richtigen" oder "falschen" Angaben, sondern lediglich um Einschätzungen Ihrer Erfahrungen. Bitte versuchen Sie, alle Antworten so gut wie möglich nach Ihrer persönlichen Erfahrung zu beantworten und lassen Sie auf keinen Fall Fragen aus!

Gelegenheiten zum Kontakt in der Vergangenheit

- (1) Bitte schätzen Sie den Prozentsatz ausländischer Kinder, die mit Ihnen die Grundschule besucht haben. |%
- (2) Bitte schätzen Sie den Prozentsatz ausländischer Kinder, die mit Ihnen die Mittelstufe (Kl. 7 - 10) besucht haben. |%
- (3) Bitte schätzen Sie den Prozentsatz ausländischer Personen, die mit Ihnen die Oberstufe (Kl. 11 - 13) besucht haben bzw. mit Ihnen an Ihrem Ausbildungsplatz waren. |%
- (4) Bitte schätzen Sie, wieviel Prozent der Mitmenschen in der Nachbarschaft, in der Sie aufgewachsen sind, ausländischer Herkunft waren. |%

Freundschaften in der Vergangenheit

- (5) Wie viele Freundschaften hatten Sie während Ihrer Grundschulzeit mit Kindern ausländischer Herkunft? (Anzahl Personen)
- (6) Wie viele Freundschaften hatten Sie während Ihrer Zeit in der Mittelstufe mit Personen ausländischer Herkunft? (Anzahl Personen)
- (7) Wie viele Freundschaften unterhielten Sie während Ihrer Zeit in der Oberstufe bzw. an Ihrem Ausbildungsplatz mit Personen ausländischer Herkunft? (Anzahl Personen)

Derzeitige Alltagskontakte

Stellen Sie sich eine "Durchschnittswoche" vor! Bei welchen Gelegenheiten kommen Sie mit mehr als zwei Personen anderer Nationalitäten in Kontakt? (Bitte ankreuzen!)

- | | | |
|------|----------------------------|--------------------------|
| (8) | an Ihrer Arbeitsstätte | <input type="checkbox"/> |
| (9) | bei Freizeitaktivitäten | <input type="checkbox"/> |
| (10) | bei der Arbeit | <input type="checkbox"/> |
| (11) | in Geschäften/Kaufhäusern | <input type="checkbox"/> |
| (12) | im Haus, in dem Sie wohnen | <input type="checkbox"/> |
-

Im Folgenden sollen Sie die Gesichter, die Sie vorher gesehen haben, unter einer größeren Anzahl von Gesichtern wiedererkennen. Das heißt, einige Gesichter sind "alt", andere "neu". Die Aufgabe wird dadurch etwas schwieriger, dass die Gesichter nicht wie vorher von vorne gezeigt werden, sondern im Halbprofil (45° Winkel):

- | | | |
|-------|--|--------------------------|
| (13) | Wieviel Prozent glauben Sie, richtig wiedererkennen zu können (zwischen 50% und 100%, denn 50% hätten Sie allein durch Raten richtig): | |
| (13a) | Gesichter von Türken | <input type="checkbox"/> |
| (13b) | Gesichter von Deutschen | <input type="checkbox"/> |
| (13c) | Gesichter von schwarzen Amerikanern | <input type="checkbox"/> |
| (13d) | Gesichter von Latein-Amerikanern | <input type="checkbox"/> |

Engere Kontakte

(14) Wie viele Ausländer kennen Sie mit Vor- oder Spitznamen? (Anzahl Personen) |____|

(15) Bitte denken Sie nun an Ihre neun engsten Freunde: wie viele von ihnen sind ausländischer Herkunft? (Anzahl Personen) |____|

(16) Mit wie vielen ausländischen Personen des anderen Geschlechts sind sie schon abends ausgegangen? (Anzahl Personen) |____|

(17) Wie viele Jahre leben Sie schon in Deutschland? |____|

(18) Wenn Sie sich mit Ihren deutschen Arbeitskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit Personen türkischer Herkunft im Vergleich zur "Durchschnittsperson" an Ihrem Arbeitsplatz?

Sehr viel

weniger

1

2

3

4

5

6

7

Sehr viel

mehr

|____|

(19) Wenn Sie sich mit Ihren deutschen Arbeitskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit Personen türkischer Herkunft im Vergleich zur "Durchschnittsperson" durch Medienkonsum (Kino, Fernsehen, Video)?

Sehr viel

weniger

1

2

3

4

5

6

7

Sehr viel

mehr

|____|

(20) Wenn Sie sich mit Ihren deutschen Arbeitskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit "Schwarzen" (Personen schwarzer Hautfarbe) im Vergleich zur "Durchschnittsperson" an Ihrem Arbeitsplatz?

Sehr viel

weniger

1

2

3

4

5

6

7

Sehr viel

mehr

|____|

- (21) Wenn Sie sich mit Ihren deutschen Arbeitskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit "Schwarzen" (Personen schwarzer Hautfarbe) im Vergleich zur "Durchschnittsperson" durch Medienkonsum (Kino, Fernsehen, Video)?

Sehr viel
weniger

1 2 3 4 5 6

Sehr viel
mehr

7

|____|

- (22) Wenn Sie sich mit Ihren deutschen Arbeitskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit Latein-Amerikanern im Vergleich zur "Durchschnittsperson" an Ihrem Arbeitsplatz?

Sehr viel
weniger

1 2 3 4 5 6

Sehr viel
mehr

7

|____|

- (23) Wenn Sie sich mit Ihren deutschen Arbeitskolleg(inn)en vergleichen: wie viele Erfahrungen haben Sie mit Latein-Amerikanern im Vergleich zur "Durchschnittsperson" durch Medienkonsum (Kino, Fernsehen, Video)?

Sehr viel
weniger

1 2 3 4 5 6

Sehr viel
mehr

7

|____|

- (24) Wenn Sie an einen typischen Arbeitstag denken, welche ethnische Abstammung haben die Personen, denen Sie begegnen (Geben Sie bitte die geschätzte Anzahl an)

Deutsche

|____|

Weißer (weißer Mittel-Europäer, weißer Amerikaner, Australier mit mitteleuropäischer Abstammung)

|____|

Nicht-deutsche weißer Europäer (Nord-, Süd-, Osteuropa)

|____|

Ferner Osten (China, Japan, Vietnam etc.)

|____|

Mittlerer Osten (Jemen, Katar, Saudi-Arabien etc.)

|____|

Naher Osten (Türkei, Syrien, Jordanien etc.)

|____|

Afrika (Schwarze Amerikaner, schwarze Afrikaner etc.)

|____|

- (25) Geben Sie bitte die Anzahl der Dienstjahre an, in denen Sie an Ihrer Arbeitsstelle Kontakt mit unterschiedlichen ethnischen Gruppen hatten.

|____|

FRAGEBOGEN ÜBER KONTAKTE MIT AUSLÄNDERN **(Version für Studierende)**

Im folgenden Fragebogen geht es um den Kontakt mit Personen ausländischer Herkunft. Dazu gibt es keine "richtigen" oder "falschen" Angaben, sondern lediglich um Einschätzungen Ihrer Erfahrungen. Bitte versuchen Sie, alle Antworten so gut wie möglich nach Ihrer persönlichen Erfahrung zu beantworten und lassen Sie auf keinen Fall Fragen aus!

Gelegenheiten zum Kontakt in der Vergangenheit

- (1) Bitte schätzen Sie den Prozentsatz ausländischer Kinder, die mit Ihnen die Grundschule besucht haben. |____|%
- (2) Bitte schätzen Sie den Prozentsatz ausländischer Kinder, die mit Ihnen die Mittelstufe (Kl. 7 - 10) besucht haben. |____|%
- (3) Bitte schätzen Sie den Prozentsatz ausländischer Personen, die mit Ihnen die Oberstufe (Kl. 11 - 13) besucht haben bzw. mit Ihnen an Ihrem Ausbildungsplatz waren. |____|%
- (4) Bitte schätzen Sie, wieviel Prozent der Mitmenschen in der Nachbarschaft, in der Sie aufgewachsen sind, ausländischer Herkunft waren. |____|%

Freundschaften in der Vergangenheit

- (5) Wie viele Freundschaften hatten Sie während Ihrer Grundschulzeit mit Kindern ausländischer Herkunft? (Anzahl Personen) |____|
- (6) Wie viele Freundschaften hatten Sie während Ihrer Zeit in der Mittelstufe mit Personen ausländischer Herkunft? (Anzahl Personen) |____|
- (7) Wie viele Freundschaften unterhielten Sie während Ihrer Zeit in der Oberstufe bzw. an Ihrem Ausbildungsplatz mit Personen ausländischer Herkunft? (Anzahl Personen) |____|

Derzeitige Alltagskontakte

Stellen Sie sich eine "Durchschnittswoche" vor! Bei welchen Gelegenheiten kommen Sie mit mehr als zwei Personen anderer Nationalitäten in Kontakt? (Bitte ankreuzen!)

- (8) an der Universität bzw. an Ihrer Ausbildungsstätte
- (9) bei Freizeitaktivitäten
- (10) bei der Arbeit
- (11) in Geschäften/Kaufhäusern
- (12) im Haus, in dem Sie wohnen
-

Im Folgenden sollen Sie die Gesichter, die Sie vorher gesehen haben, unter einer größeren Anzahl von Gesichtern wiedererkennen. Das heißt, einige Gesichter sind "alt", andere "neu". Die Aufgabe wird dadurch etwas schwieriger, dass die Gesichter nicht wie vorher von vorne gezeigt werden, sondern im Halbprofil (45° Winkel):

- (13) Wieviel Prozent glauben Sie, richtig wiedererkennen zu können (zwischen 50% und 100%, denn 50% hätten Sie allein durch Raten richtig):
- (13a) Gesichter von Türken
- (13b) Gesichter von Deutschen
- (13c) Gesichter von schwarzen Amerikanern
- (13d) Gesichter von Latein-Amerikanern

Engere Kontakte

(14) Wie viele Ausländer kennen Sie mit Vor- oder Spitznamen? (Anzahl Personen) |____|

(15) Bitte denken Sie nun an Ihre neun engsten Freunde: wie viele von ihnen sind ausländischer Herkunft? (Anzahl Personen) |____|

(16) Mit wie vielen ausländischen Personen des anderen Geschlechts sind sie schon abends ausgegangen? (Anzahl Personen) |____|

(17) Wie viele Jahre leben Sie schon in Deutschland? |____|

(18) Wenn Sie sich mit Ihren deutschen Kommiliton(inn)en vergleichen: wie viele Erfahrungen haben Sie mit Personen türkischer Herkunft im Vergleich zur "Durchschnittsperson" an Ihrer Universität?

Sehr viel

weniger

1

2

3

4

5

6

7

Sehr viel

mehr

|____|

(19) Wenn Sie sich mit Ihren deutschen Kommiliton(inn)en vergleichen: wie viele Erfahrungen haben Sie mit Personen türkischer Herkunft im Vergleich zur "Durchschnittsperson" durch Medienkonsum (Kino, Fernsehen, Video)?

Sehr viel

weniger

1

2

3

4

5

6

7

Sehr viel

mehr

|____|

(20) Wenn Sie sich mit Ihren deutschen Kommiliton(inn)en vergleichen: wie viele Erfahrungen haben Sie mit "Schwarzen" (Personen schwarzer Hautfarbe) im Vergleich zur "Durchschnittsperson" an Ihrer Universität?

Sehr viel

weniger

1

2

3

4

5

6

7

Sehr viel

mehr

|____|

- (21) Wenn Sie sich mit Ihren deutschen Kommiliton(inn)en vergleichen: wie viele Erfahrungen haben Sie mit "Schwarzen" (Personen schwarzer Hautfarbe) im Vergleich zur "Durchschnittsperson" durch Medienkonsum (Kino, Fernsehen, Video)?

Sehr viel
weniger

1 2 3 4 5 6

Sehr viel
mehr

7

|____|

- (22) Wenn Sie sich mit Ihren deutschen Kommiliton(inn)en vergleichen: wie viele Erfahrungen haben Sie mit Latein-Amerikanern im Vergleich zur "Durchschnittsperson" an Ihrer Universität?

Sehr viel
weniger

1 2 3 4 5 6

Sehr viel
mehr

7

|____|

- (23) Wenn Sie sich mit Ihren deutschen Kommiliton(inn)en vergleichen: wie viele Erfahrungen haben Sie mit Latein-Amerikanern im Vergleich zur "Durchschnittsperson" durch Medienkonsum (Kino, Fernsehen, Video)?

Sehr viel
weniger

1 2 3 4 5 6

Sehr viel
mehr

7

|____|

- (24) Wenn Sie an einen typischen Tag an der Universität denken, welche ethnische Abstammung haben die Personen, denen Sie begegnen (Geben Sie bitte, die geschätzte Anzahl an)

Deutsche

|____|

Weißer (weißer Mittel-Europäer, weißer Amerikaner, Australier mit mitteleuropäischer Abstammung)

|____|

Nicht-deutsche weißer Europäer (Nord-, Süd-, Osteuropa)

|____|

Ferner Osten (China, Japan, Vietnam etc.)

|____|

Mittlerer Osten (Jemen, Katar, Saudi-Arabien etc.)

|____|

Naher Osten (Türkei, Syrien, Jordanien etc.)

|____|

Afrika (Schwarzer Amerikaner, schwarzer Afrikaner etc.)

|____|

- (25) Geben Sie bitte die Anzahl der Jahre an, in denen Sie an Ihrer Universität Kontakt mit unterschiedlichen ethnischen Gruppen hatten.

|____|

Anhang C: Ratings der Pilotstudien I und II

Anhang C1: Ratings der Pilotstudie I; Deutsche Teilnehmer – deutsche Gesichter

7=hoch fam 7=hoch memo 7=hoch typ 7=hoch lik 7=hoch attr

	Set A					Set B					
	Fam	mem	typ	lik	att	Fam	mem	typ	lik	att	
gu01	2.19	4.59	3.41	3.38	1.97	gu03	2.16	4.34	3.22	2.47	1.41
gu02	1.5	4.09	3.5	3.06	2.06	gu06	2.19	3.94	4.25	2.84	2.09
gu04	1.97	3.47	4.44	3.06	2.31	gu07	2.25	3.34	4.63	2.28	1.66
gu05	1.81	3.5	4.44	2.13	1.56	gu08	3	4.72	3.44	4.38	4.25
gu13	2.03	3.69	4.38	3.31	2.06	gu09	2.22	4.84	2.75	2.94	2
gu14	2.75	3.44	5.16	3.25	3.13	gu10	1.88	3.56	4.38	2.75	1.97
gu15	1.63	3.03	4.59	2.25	1.69	gu11	2.19	3.84	4.69	2.13	1.97
gu16	2.31	4.16	3.84	4.19	3.22	gu12	2.28	3.22	4.69	4.19	2.63
gu17	1.91	3.97	4.16	2.09	1.84	gu19	2.44	4.13	4.16	2.47	2.84
gu18	2.22	3.31	5.19	2.31	1.88	gu20	2.53	2.97	5.25	2.59	2.16
gu27	2.47	4.56	3.75	3.38	2.75	gu26	2.31	5.84	2.25	4.28	2.59
gu29	1.59	4.31	3.59	2.5	1.88	gu28	2.31	5.41	2.53	3.88	3.09
gu30	1.88	4.31	3.88	2.94	1.63	gu31	2	3.97	4.13	1.81	1.63
gu32	2.03	4.47	2.94	2.03	1.66	gu36	2.19	3.41	4.38	3.63	2.84
gu33	2.31	4.06	4.09	2.5	1.63	gu37	2.03	4.47	3.47	3.28	2.22
gu34	2.16	4.06	4.09	2.94	2.88	gu38	2.28	3.84	4.41	3.47	3.16
gu35	2	3.03	4.97	3.13	2.09	gu43	2.5	3.91	4.13	3.88	3.34
gu39	2.53	3.72	4.75	2.91	2.13	gu44	2.03	3.69	4.56	2.16	1.84
gu40	1.91	2.88	4.78	3.38	2.22	gu45	2.06	3.22	4.88	2.66	2.06
gu41	2	3.47	4.59	2.78	2.06	gu46	2.31	4.34	3.47	3.38	2.69
gu42	1.97	3.47	4.66	2.63	2.03	gu47	2.41	4.59	4.09	3.75	3.81
gu49	2.38	4.91	2.84	2.31	1.47	gu48	1.81	4.34	3.84	4.28	2.31
gu51	1.81	4.31	3	3.47	2.22	gu50	1.94	4.47	3.38	2.53	1.88
gu53	2.13	3.44	4.25	3.5	2.13	gu52	2.31	4	4	3.16	2.22
gu55	2.53	3.75	4.22	2.06	1.69	gu54	2.78	4.19	3.31	4.5	3.38
gu56	2.5	3.66	4.16	3.13	2.38	gu57	2.25	4.63	3.22	2.84	2.03
gu59	2.41	5.03	2.88	3.28	2.22	gu58	2.28	3.28	4.44	3.41	2.44
gu63	2.38	3.47	4.53	3.72	2.69	gu60	2.69	3.69	4.38	3.16	2.28
gu64	2.28	4.19	3.63	3.72	3.41	gu61	2.44	4.25	3.94	4.06	3.72
gu66	2.38	4.91	2.72	3.41	1.91	gu62	1.97	5.06	3.31	3.44	2.06
gu68	2.03	3.75	4.25	2.19	1.69	gu65	2.56	3.25	4.47	2.91	2.06
gu71	1.78	3.78	4.03	3.63	3.22	gu67	2.19	4.03	4.03	3.56	3
gu72	2.16	3.44	4.22	3.34	3.06	gu69	2.31	3.59	4.09	3.66	3.13
gu73	2.84	3.56	4.56	3.88	3.59	gu70	2.47	3.59	4.44	3.66	2.97
gu74	2.53	4.38	4.13	4.72	3.69	gu75	2.56	4.31	3.56	4.38	4.75
gu79	2.53	4.16	4.09	3.53	2.84	gu76	2.63	3.91	3.97	2.22	1.53
gu81	2.44	4.22	4.06	4.16	3.47	gu77	2.47	4.09	4.16	3.91	2.94
gu82	2.75	3.78	4.25	3.78	2.75	gu78	1.81	3.88	4.19	3.44	2.66
gu84	1.84	4.28	3.91	2.09	1.59	gu80	2.31	3.81	4.13	2.97	2.72
gu85	2.38	4.38	3.16	2.94	1.63	gu83	1.59	3.03	4.66	3.16	2.69

Anhang C2: Ratings der Pilotstudie I; Türkische Teilnehmer – deutsche Gesichter

7=hoch fam 7=hoch memo 7=hoch typ 7=hoch lik 7=hoch attr

	Set A					Set B					
	Fam	mem	typ	lik	att	Fam	mem	typ	lik	att	
gu01	2.22	4.22	3.81	3.06	1.72	gu03	2.41	3.72	4.22	2.5	1.31
gu02	2.25	3.78	3.41	2.81	1.63	gu06	2.88	2.78	4.19	1.97	1.5
gu04	2.69	2.81	4.13	2.56	1.81	gu07	2.56	2.59	4.22	1.59	1.38
gu05	1.88	3.25	3.94	1.81	1.25	gu08	2.47	3.88	3.91	3.63	3.34
gu13	2.66	3.47	4.13	2.41	1.5	gu09	2.13	4.25	3.28	2.38	1.81
gu14	2.72	2.78	4.63	2.41	2.13	gu10	2.19	2.84	4.63	2.16	1.53
gu15	2.25	3.16	4.28	1.91	1.63	gu11	2.19	2.88	4.22	1.69	1.56
gu16	2.47	3.66	4.19	3.31	2.44	gu12	2.78	3.63	4.41	3.25	2.19
gu17	1.88	3.72	3.97	1.81	1.44	gu19	1.75	4	3.44	1.81	2.13
gu18	2.5	3	4.38	2.19	1.94	gu20	2.53	2.69	4.78	1.84	1.88
gu27	2.13	3.69	3.78	2.78	1.97	gu26	1.78	5.16	2.5	3.84	2.22
gu29	2.16	2.78	4.5	2.22	1.31	gu28	2.5	4.25	3.34	3.34	2.81
gu30	2.28	3.81	3.91	2.56	1.45	gu31	2.44	3.28	4.22	1.84	1.41
gu32	2.13	3.72	3.13	2	1.5	gu36	2.41	2.84	4.63	2.59	2.22
gu33	2.44	3.66	3.75	1.91	1.53	gu37	2.56	4.19	3.44	2.22	1.66
gu34	2.44	2.97	3.78	2.44	2.03	gu38	2.44	3.31	4.28	3.06	2.63
gu35	2.81	3	4.19	2.44	2	gu43	2.13	3.19	4.19	3.13	2.5
gu39	2.72	3.53	4.25	2.69	1.88	gu44	2.25	2.66	4.41	2.09	1.84
gu40	2.13	2.94	4.47	3.03	2.03	gu45	2.31	2.75	4.28	2.16	1.63
gu41	2.66	3.25	4.5	2.56	2.09	gu46	2.16	3.88	3.94	3.28	2.28
gu42	2.25	2.97	4.56	2.31	1.5	gu47	3.06	3.47	4.13	2.81	2.88
gu49	1.72	4.09	3.03	2.41	1.44	gu48	2.84	3.63	4.22	3.16	1.94
gu51	2.31	3.72	3.59	3.19	2.22	gu50	2.09	3.84	4	2.09	1.5
gu53	2.41	3.34	4.63	2.34	1.5	gu52	2.16	3.13	4.66	2.34	1.66
gu55	2.38	4.03	4.03	1.94	1.5	gu54	2.13	3.41	4.13	3.75	2.41
gu56	2.28	3.16	4.22	2.38	1.81	gu57	2.69	3.84	3.72	3.03	2.13
gu59	2.66	4.03	3.72	3.16	2.25	gu58	2.47	3.09	4.25	2.97	1.81
gu63	2.56	3.25	4.31	3.38	2.66	gu60	2.72	3.78	3.59	2.69	1.88
gu64	2.66	3.78	3.47	3.41	2.88	gu61	3.03	3.69	4.22	3.34	3.25
gu66	1.75	5.28	2.53	2.56	1.44	gu62	1.97	3.91	3.53	2.41	1.59
gu68	2.53	3.56	4.28	2	1.25	gu65	2.69	2.38	4.72	2.5	1.5
gu71	2.47	3.69	4.06	4.09	3.78	gu67	2.56	3.56	4.44	3.13	2.34
gu72	2.59	2.78	4.44	2.78	2.22	gu69	2.5	3	4.59	2.72	2.44
gu73	2.44	2.94	4.56	3.25	2.78	gu70	2.66	3.25	4.09	2.97	2.56
gu74	2.84	3.66	3.91	4.25	3.03	gu75	3.16	4.06	4.03	3.5	3.75
gu79	1.91	4.19	3.56	2.81	1.81	gu76	2.59	3.66	4.25	2.09	1.63
gu81	2.63	3.13	4.25	3.66	2.97	gu77	2.72	3.72	4	3.31	2.63
gu82	2.94	3.25	4.5	2.84	2.25	gu78	2.47	3.19	4.53	2.75	2.06
gu84	1.84	3.53	4.09	1.94	1.78	gu80	2.38	3.34	4.03	2.44	2
gu85	2.16	4.13	3.72	2.5	1.59	gu83	2.41	3	4.34	2.63	2.13

Anhang C3: Ratings der Pilotstudie I; Deutsche Teilnehmer – türkische Gesichter

7=hoch fam 7=hoch memo 7=hoch typ 7=hoch lik 7=hoch attr

	Set A					Set B					
	Fam	mem	typ	lik	att	Fam	mem	typ	lik	att	
tu02	2.59	4.31	3.31	4.16	3.53	tu01	1.81	3.78	4.34	2.34	1.66
tu03	2.22	3.88	4.19	3.69	3.31	tu07	1.91	3.56	4.34	2.25	2.19
tu04	2.09	3.53	4.59	3.28	3.06	tu08	1.44	4.84	3.75	2.03	1.44
tu05	2.34	3.78	4.28	3.28	2.5	tu09	1.88	3.63	4.25	3.06	2
tu06	1.84	4.88	3.38	2.38	2.47	tu10	2.13	4.03	4.03	2.56	2.06
tu12	2.06	3.47	4.47	3.25	3	tu11	2.09	2.97	4.84	3.13	2.75
tu16	2.56	3.78	4.25	2.59	2.69	tu13	2.31	3	4.78	2.56	2.28
tu17	1.75	3.47	4.78	2.56	1.81	tu14	2.31	3.72	4.13	4.19	3.28
tu18	2.06	3.44	4.5	3.28	2.41	tu15	2.13	3.34	4.66	2.72	2.38
tu26	1.69	4.59	3.56	2.81	2	tu19	2.25	3.66	4.47	4	3.03
tu27	1.72	3.81	4.16	2.84	2.06	tu20	1.84	3.25	4.41	2.16	1.72
tu28	2.06	4.09	3.81	3.63	2.38	tu29	1.69	5.41	2.84	3.66	2.97
tu30	2	4.53	3.22	2.28	2.94	tu31	1.84	3.44	4.47	2.75	2.06
tu34	2.59	4.03	3.94	3.56	3.56	tu32	2	3.53	4.75	3.16	2.53
tu36	2.22	3.13	4.88	3.16	2.25	tu33	1.66	4.47	3.34	2.5	1.75
tu39	1.59	3.16	4.75	2.31	1.81	tu35	1.63	3.88	4.22	3.5	2.06
tu40	1.88	2.81	5	2.81	2.47	tu37	2.06	2.94	5.25	3.19	2.25
tu41	2.41	3.75	4.38	4.25	3.13	tu38	1.94	4.19	3.88	4.19	3.09
tu46	2.41	3.5	4.47	3.09	2.59	tu42	1.84	2.91	5.31	2.88	2.31
tu47	2	3.97	4	2.59	2.34	tu43	2.22	3.34	4.91	3.16	2.28
tu48	2.28	3.94	3.94	2.94	1.81	tu44	2.09	3.53	4.5	2.34	3.16
tu49	2.28	3.56	4.69	2.75	2.28	tu45	2	3.84	4.06	2.13	1.78
tu50	1.75	3.81	4.22	2.56	2	tu51	1.78	3.59	3.97	2.09	1.81
tu56	1.97	2.84	4.91	2.94	2.31	tu52	2.06	3.78	4.22	3.53	2.75
tu58	2.44	3.13	4.78	2.81	2.41	tu53	2.69	4.16	4.22	3.5	3
tu59	2.06	3.75	4.5	3.47	2.56	tu54	1.97	3.78	3.88	3.56	2.75
tu60	1.91	3	4.56	3.13	2.78	tu55	1.84	3.88	4.19	2.5	1.94
tu61	1.97	3.03	4.63	3.25	2.88	tu57	2	3.13	4.38	3.06	2.91
tu62	2.53	3.72	4.28	2.69	2.41	tu63	2.06	3.34	4.44	2.97	2.38
tu64	1.81	3.91	4.19	1.94	1.84	tu66	2.06	4.56	3.72	2.75	2.88
tu65	1.97	3.88	4.22	2.81	2.63	tu67	2.09	3.25	4.44	2.06	2.06
tu68	2.03	3.59	4.53	2.56	1.75	tu70	1.91	3	4.25	3.19	2.44
tu69	1.72	3.97	3.81	2.88	1.78	tu71	2.19	3.56	4.16	3.69	2.94
tu73	2.28	3.91	3.72	3.63	3.63	tu72	2.06	3.69	4.44	3.28	3.63
tu75	1.94	3.97	3.94	2.72	3.31	tu74	2.41	4.31	3.63	2.63	2.84
tu78	2.44	3.97	4.03	3.03	2.09	tu76	2.44	3.97	4.13	4.16	4.03
tu81	2.06	3.03	4.53	3.09	2.78	tu77	2.22	4	4.31	3.41	2.56
tu83	2.06	3.22	4.63	3.88	3.22	tu79	1.69	3.22	4.56	2.72	2.78
tu84	2.16	3.72	4.59	3.94	2.41	tu80	1.84	3.13	4.59	2.66	2.38
tu85	2.16	3.09	4.84	2.69	2.25	tu82	1.75	3.59	4.06	2.28	1.91

Anhang C4: Ratings der Pilotstudie I; türkische Teilnehmer – türkische Gesichter

7=hoch fam 7=hoch memo 7=hoch typ 7=hoch lik 7=hoch attr

	Set A					Set B					
	Fam	mem	typ	lik	att	Fam	mem	typ	lik	att	
tu02	2.41	4.19	3.47	3.31	2.88	tu01	2.25	4.47	3.59	2.63	1.44
tu03	2.94	3.88	3.31	3.53	3	tu07	2.16	4.28	3.91	2.38	1.91
tu04	2.19	4.06	3.59	2.66	2.47	tu08	1.88	4.75	2.88	2.06	1.34
tu05	2.03	4.09	3.44	3	2.31	tu09	2.59	3.94	3.28	3.16	1.91
tu06	1.78	5	3.06	2.16	1.97	tu10	2.09	4.47	3.44	2.38	1.53
tu12	2	3.47	3.97	2.69	2.59	tu11	2.75	3.5	4.19	3.13	2.22
tu16	2.19	3.72	3.69	2.22	1.97	tu13	2.44	3.41	3.81	3.03	2.66
tu17	1.97	3.69	3.81	2.66	1.94	tu14	2.53	3.34	4.34	3.59	2.91
tu18	2.28	3.97	3.97	3.09	2	tu15	2.5	3.13	4.5	2.63	2.06
tu26	2.06	4.63	3.13	2.88	1.66	tu19	2.81	3.75	3.63	3.84	3.22
tu27	2	4.06	3.69	2.28	1.81	tu20	1.72	3.44	4.59	1.94	1.47
tu28	2.84	4.41	3.69	4.19	2.63	tu29	2.16	5.19	2.88	3.03	2.19
tu30	2.13	5.22	2.47	2.28	2.44	tu31	2.5	4.09	3.69	3.25	1.59
tu34	2.5	3.88	4.13	3.44	3.09	tu32	2.19	3.31	4.09	2.81	2.34
tu36	2.69	3.09	4.19	3.09	1.91	tu33	1.72	5.41	3.06	2.94	1.91
tu39	2.09	3.41	4.19	2.38	1.75	tu35	2.19	4.09	3.66	3.19	1.78
tu40	2.53	3.13	4.19	2.69	2.13	tu37	2.44	3.34	3.94	3.19	2.22
tu41	2.34	3.94	4.16	3.69	2.38	tu38	2.31	4	3.75	3.25	2
tu46	1.91	4.16	3.53	2.66	1.63	tu42	2.25	2.63	4.72	2.84	2.41
tu47	2.84	4.25	2.88	1.97	1.72	tu43	2.75	3.94	4.09	2.91	1.88
tu48	2.53	3.41	3.56	2.53	1.69	tu44	2.59	3.56	3.59	3.16	2.72
tu49	2.38	3.66	4.22	2.53	1.88	tu45	2.22	3.84	3.78	1.78	1.44
tu50	2.13	3.75	4.25	2.25	1.72	tu51	1.81	4.72	3.69	1.91	1.44
tu56	2.09	3.28	4.13	2.31	1.91	tu52	2.72	3.91	3.63	3.25	2.53
tu58	2.16	3.5	3.84	2.56	1.81	tu53	2.31	3.84	3.06	3.19	2.47
tu59	2.16	3.91	4.13	3.59	2.28	tu54	2.34	4.41	3	3.47	3.13
tu60	2.19	3.41	3.63	2.97	2.5	tu55	2.31	4.06	3.53	2.53	1.72
tu61	2.44	3.41	4.13	2.88	1.84	tu57	2.13	3.59	3.69	2.59	2.13
tu62	2.84	4.09	3.53	3.19	2.28	tu63	2.56	3.78	3.63	2.69	2
tu64	2.13	4.13	3.09	1.81	1.66	tu66	2.38	4.28	2.84	2.53	2.66
tu65	2.13	3.53	3.88	2.72	2.19	tu67	2.09	3.44	4.06	2.41	1.78
tu68	2.47	3.34	4.28	2.28	1.72	tu70	2.28	3.19	4.34	3.13	1.94
tu69	2.03	4.19	3.25	3.06	1.69	tu71	2.91	3.88	3.75	3.34	2.28
tu73	2.06	3.88	3.5	3.06	2.38	tu72	2.5	3.38	3.78	3.13	2.66
tu75	2.47	4	3.47	2.88	2.75	tu74	2.44	3.97	3.13	2.09	1.84
tu78	3.5	4.69	3.25	2.56	1.69	tu76	2.44	4.31	3.38	3.81	3.44
tu81	2.75	3.47	4.16	2.25	1.91	tu77	2.53	3.22	3.78	3.19	1.97
tu83	2.91	3.94	3.63	3.91	3.06	tu79	2.63	3.69	3.38	2.56	2.31
tu84	2.69	3.56	4.09	4.28	2.63	tu80	1.97	3.63	3.88	2.53	1.94
tu85	2.56	3.34	4.28	2.84	1.88	tu82	1.81	3.19	4.22	2.03	1.63

Anhang C5: Ratings der Pilotstudie II; Deutsche Teilnehmer – latein-amerikanische

Gesichter 7=hoch fam 7=hoch memo 7=hoch typ 7=hoch lik 7=hoch attr

	Set A					Set B					
	Fam	mem	typ	lik	att	Fam	mem	typ	lik	att	
ho03	2.78	4	4.22	2.28	2.28	ho01	1.72	2.75	3.16	2.5	2.56
ho08	2.13	3.59	3.81	3.91	2.59	ho02	2.31	4.19	4.53	2.5	2.22
ho10	1.5	3.94	3.97	2.5	1.84	ho04	2.31	4.03	4.38	3.5	2.94
ho12	1.78	3.13	3.16	2.72	2.06	ho05	1.78	3.34	3.56	3.03	1.81
ho13	1.5	3.88	3.91	2.66	1.69	ho07	2.13	4.25	4.06	3.13	2.34
ho15	1.5	3.19	3.03	3.13	1.69	ho09	2.28	4.19	3.94	4.34	3.25
ho18	1.91	3.69	3.56	3	2.41	ho11	1.53	4	4.25	2.84	1.84
ho19	2.66	3.88	3.63	4.5	2.44	ho14	1.47	3.59	3.66	3.25	2.53
ho21	1.47	3.22	3.34	2.75	2.19	ho17	2.03	3	3.03	3	2.16
ho22	1.34	5.22	4.78	1.59	1.16	ho20	2.06	4.44	4.09	4.88	3.41
ho25	2.22	2.88	2.59	3.81	2.06	ho23	3.13	3.63	3.22	3.66	3
ho26	1.66	4.97	5.25	2.19	1.38	ho24	2.06	3.09	3.22	3.28	2.09
ho27	1.81	3.41	3.44	3.16	1.88	ho28	1.63	3.88	4.34	3.19	1.78
ho30	1.63	3.66	3.56	2.97	2.75	ho29	1.84	3.78	3.5	3.31	2.84
ho31	1.47	4.28	4.25	2.31	1.78	ho34	2.19	4.38	4.16	3.09	2.47
ho32	2.19	3.53	3.53	4.03	3.41	ho36	1.97	3.69	3.84	3.56	2.63
ho33	1.5	4.5	4.13	2.59	1.56	ho43	1.66	3.91	4.31	2.81	2.13
ho35	1.53	4.97	4.81	1.88	1.53	ho44	1.44	3.78	4.09	2.41	1.81
ho38	1.78	3.56	3.84	3.16	2.22	ho45	2.5	3.47	3.53	3.69	2.19
ho39	1.78	3.5	3.56	2.03	1.47	ho48	2.75	3.94	3.81	2.56	1.5
ho40	1.47	4	4.41	1.75	1.47	ho49	1.78	3.5	3.88	3.13	2.03
ho41	1.28	4.66	4.34	2.72	1.66	ho52	1.88	2.78	3.41	3.34	2
ho42	1.84	3.44	3.13	2.13	1.75	ho53	1.31	5.19	5.03	2.44	1.53
ho46	1.88	3.53	3.53	4.53	3.72	ho54	2.41	3.19	3.25	4.13	3.13
ho47	2.69	3.38	3.19	4.41	3.25	ho57	1.41	2.97	3.56	2.34	1.69
ho50	1.66	3.5	3.75	2.34	1.66	ho59	1.59	3.69	3.72	2.69	2.5
ho51	1.91	3.5	3.41	2.97	2.19	ho62	1.91	3.31	3.5	3.63	2.09
ho55	1.88	4.09	4.56	3.19	2.56	ho64	1.59	3.16	3.16	2.53	1.94
ho58	1.69	3.94	4.03	3.53	2.44	ho65	1.88	3.66	3.34	4.13	3.56
ho61	1.41	3.66	3.88	2.28	1.88	ho66	1.34	4.03	3.88	1.81	1.28
ho68	1.66	3.59	3.44	3.53	3.06	ho67	1.34	4.94	5.47	2.06	1.47
ho70	1.78	3.19	3.41	3.06	2.47	ho69	2.06	3.44	3.78	2.72	2.28
ho72	1.88	3.97	4.31	2.59	2.03	ho71	2	3.13	3.47	3.59	2.69
ho73	1.63	4	3.81	2.88	2.28	ho75	2.31	2.91	2.91	2.56	2.28
ho74	1.38	3.38	3.69	2.84	2.09	ho76	1.56	3.56	3.97	3.09	1.94

Anhang D: Beispielgesichter aus den Experimenten II und III

Anhang D1 Beispielgesichter aus Experiment II: links oben: Afrikanischer Amerikaner, rechts oben: Kaukasischer Amerikaner, links unten: Türke, rechts unten: Deutscher. In der Pilotstudie I wurden türkische und deutsche Gesichter beurteilt.



Anhang D2 Beispielgesichter aus Experiment III: links oben: Afrikanischer Amerikaner, rechts oben: Latein-Amerikaner, links unten: Türke, rechts unten: Deutscher.



Es war ein Tor. Ich habe es genau gesehen, meine Herren. Ich habe gesehen, wie der Ball im Netz zappelte.

(Heinrich Lübke zum Wembley-Tor 1966)

Erklärung:

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Unterschrift