

Determinanten physischer Attraktivität – der Einfluss von Durchschnittlichkeit, Symmetrie und sexuellem Dimorphismus auf die Attraktivität von Gesichtern

Habilitationsschrift
zur Erlangung des akademischen Grades
Dr. phil. habil.

vorgelegt der Philosophischen Fakultät
der Universität Regensburg

von
Dr. phil. Martin Gründl

Regensburg
2011

Inhaltsverzeichnis

1 Einführung und Ziel der Arbeit	1
2 Empirische Untersuchungen zum Einfluss der Durchschnittlichkeit auf Attraktivität	3
2.1 Stand der Forschung zur Durchschnittshypothese	3
2.2 Ziel der Untersuchungen zum Einfluss der Durchschnittlichkeit und Hypothesen.....	10
2.3 Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von männlichen Frontalgesichtern (Untersuchung <i>Durch 1</i>).....	12
2.3.1 Methode	12
2.3.1.1 Voruntersuchung	12
2.3.1.2 Versuchspersonen und Durchführung der Hauptuntersuchung.....	13
2.3.1.3 Stimulusmaterial der Hauptuntersuchung	13
2.3.2 Ergebnisse.....	16
2.4 Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von männlichen Profilgesichtern mit Texturinformation (Untersuchung <i>Durch 2</i>).....	22
2.4.1 Methode	22
2.4.1.1 Voruntersuchung	22
2.4.1.2 Versuchspersonen und Durchführung der Hauptuntersuchung.....	24
2.4.1.3 Stimulusmaterial der Hauptuntersuchung	24
2.4.2 Ergebnisse.....	24
2.5 Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von männlichen Profilgesichtern ohne Texturinformation (Schattenprofile, Untersuchung <i>Durch 3</i>)	30
2.5.1 Methode	30
2.5.2 Ergebnisse.....	31
2.6 Zusammenfassende Datenanalyse der Experimente <i>Durch 1</i> bis <i>3</i>	35
2.7 Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von weiblichen Frontalgesichtern (Untersuchung <i>Durch 4</i>)	42
2.7.1 Methode	42
2.7.1.1 Voruntersuchung	42

2.7.1.2	Versuchspersonen und Durchführung der Hauptuntersuchung.....	44
2.7.1.3	Stimulusmaterial der Hauptuntersuchung	44
2.7.2	Ergebnisse.....	45
2.8	Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von weiblichen Profilgesichtern mit Texturinformation (Untersuchung <i>Durch 5</i>)	50
2.8.1	Methode.....	50
2.8.1.1	Voruntersuchung	50
2.8.1.2	Versuchspersonen und Durchführung der Hauptuntersuchung.....	51
2.8.1.3	Stimulusmaterial der Hauptuntersuchung	52
2.8.2	Ergebnisse.....	53
2.9	Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von weiblichen Profilgesichtern ohne Texturinformation (Schattenprofile, Untersuchung <i>Durch 6</i>)	59
2.9.1	Methode.....	59
2.9.2	Ergebnisse.....	60
2.10	Zusammenfassende Datenanalyse der Experimente <i>Durch 4</i> bis <i>Durch 6</i>	64
3	Zusammenfassende Diskussion der Ergebnisse aller sechs Untersuchungen zum Einfluss der Durchschnittlichkeit.....	72
3.1	Methodische Reflexion der Untersuchungen zur Durchschnittlichkeit	73
3.1.1	Effekte der Stimulusart	73
3.1.2	Effekte von Morphing-Artefakten	75
3.1.3	Effekte der Ausgangsattraktivität der Originalgesichter	81
3.2	Alternativmodell zur Durchschnittshypothese: Der „Raffael-Effekt“	83
3.3	Deskriptive Analyse zur Identifikation von „Sexy Features“	91
3.4	Vergleich der eigenen Ergebnisse mit bisherigen Untersuchungen zur Durchschnittshypothese	103
3.5	Fazit zum Effekt der Durchschnittlichkeit auf die Attraktivität von Gesichtern und Konsequenzen für künftige Studien	116
4	Empirische Untersuchungen zum Einfluss der Symmetrie auf Attraktivität ...	122
4.1	Kurzer Überblick über den Stand der Forschung zu Symmetrie und Attraktivität.....	122

4.2	Einführung in die verschiedenen Methoden zur Untersuchung des Einflusses von Symmetrie.....	124
4.3	Ziel der Untersuchungen zum Einfluss der Symmetrie und Hypothesen.....	131
4.4	Experimentelle Manipulation der Symmetrie durch Verwendung von Chimärengesichtern und Attraktivitätsmessung durch Paarvergleiche (Untersuchung <i>Sym 1</i>).....	133
4.4.1	Stand der Forschung	133
4.4.2	Methode.....	134
4.4.2.1	Versuchspersonen.....	134
4.4.2.2	Stimulusmaterial	135
4.4.2.3	Versuchsdurchführung	136
4.4.3	Ergebnisse.....	137
4.5	Experimentelle Manipulation der Symmetrie durch Verwendung von Chimärengesichtern und Attraktivitätsmessung durch Ratings (Untersuchung <i>Sym 2</i>)	141
4.5.1	Stand der Forschung	141
4.5.2	Methode.....	142
4.5.2.1	Versuchspersonen.....	142
4.5.2.2	Stimulusmaterial und Versuchsdurchführung.....	143
4.5.3	Ergebnisse.....	144
4.6	Experimentelle Manipulation der Symmetrie durch Verwendung von Morphings <i>mit</i> Farbwertmittelung („ <i>Blends</i> “) und Attraktivitätsmessung durch Paarvergleiche (Untersuchung <i>Sym 3</i>)	151
4.6.1	Stand der Forschung	151
4.6.2	Methode.....	152
4.6.2.1	Versuchspersonen.....	152
4.6.2.2	Stimulusmaterial	152
4.6.2.3	Versuchsdurchführung	153
4.6.3	Ergebnisse.....	154
4.7	Experimentelle Manipulation der Symmetrie durch Verwendung von Morphings <i>mit</i> Farbwertmittelung („ <i>Blends</i> “) und Attraktivitätsmessung durch Ratings (Untersuchung <i>Sym 4</i>)	157
4.7.1	Stand der Forschung	157

4.7.2	Methode.....	159
4.7.2.1	Versuchspersonen.....	159
4.7.2.2	Stimulusmaterial und Versuchsdurchführung.....	159
4.7.3	Ergebnisse.....	160
4.8	Experimentelle Manipulation der Symmetrie durch Verwendung von Morphings <i>ohne</i> Farbwertmittelung („ <i>Warps</i> “) und Attraktivitätsmessung durch Paarvergleiche (Untersuchung <i>Sym 5</i>)	166
4.8.1	Stand der Forschung	166
4.8.2	Methode.....	167
4.8.2.1	Versuchspersonen.....	167
4.8.2.2	Stimulusmaterial	168
4.8.2.3	Versuchsdurchführung	169
4.8.3	Ergebnisse.....	169
4.9	Experimentelle Manipulation der Symmetrie durch Verwendung von Morphings <i>ohne</i> Farbwertmittelung („ <i>Warps</i> “) und Attraktivitätsmessung durch Ratings (Untersuchung <i>Sym 6</i>)	172
4.9.1	Stand der Forschung	172
4.9.2	Methode.....	172
4.9.2.1	Versuchspersonen.....	172
4.9.2.2	Stimulusmaterial und Versuchsdurchführung.....	173
4.9.3	Ergebnisse.....	173
4.10	Korrelatives Versuchsdesign mit direktem Rating der Symmetrie durch Versuchspersonen (Untersuchung <i>Sym 7</i>)	179
4.10.1	Stand der Forschung	179
4.10.2	Methode.....	182
4.10.2.1	Versuchspersonen.....	182
4.10.2.2	Stimulusmaterial und Versuchsdurchführung.....	183
4.10.3	Ergebnisse.....	183
4.10.4	Zusatzexperiment mit Symmetrie-Rating von „ <i>Warp</i> “-Gesichtern als Kontrollbedingung (Untersuchung <i>Sym 7 Kontrolle</i>)	185
4.11	Korrelatives Versuchsdesign mit Versuchspersonen-Ratings der Ähnlichkeit von separaten Gesichtshälften als Maß für Symmetrie (Untersuchung <i>Sym 8</i>)	190

4.11.1	Stand der Forschung	190
4.11.2	Methode	190
4.11.2.1	Versuchspersonen.....	190
4.11.2.2	Stimulusmaterial und Versuchsdurchführung.....	191
4.11.3	Ergebnisse	192
4.12	Korrelatives Versuchsdesign mit Versuchspersonen-Ratings der Ähnlichkeit von Chimärengesichtern als Maß für Symmetrie (Untersuchung <i>Sym 9</i>)	197
4.12.1	Stand der Forschung	197
4.12.2	Methode	199
4.12.2.1	Versuchspersonen.....	199
4.12.2.2	Stimulusmaterial und Versuchsdurchführung.....	199
4.12.3	Ergebnisse	201
5	Zusammenfassende Diskussion der Ergebnisse aller neun Untersuchungen zum Einfluss der Symmetrie	204
5.1	Tabellarischer Überblick über die Ergebnisse aller Symmetrie- Untersuchungen	204
5.2	Methodische Reflexion der Symmetrie-Untersuchungen.....	207
5.2.1	Effekte des Stimulusmaterials: Chimärengesichter, Morphing- Artefakte und Originalgesichter	207
5.2.1.1	Effekte von Chimärengesichtern	207
5.2.1.2	Effekte von Morphing-Artefakten	213
5.2.1.3	Effekte der Ausgangsattraktivität der Originalgesichter	213
5.2.2	Effekte des Versuchsdesigns: Paarvergleich, Rating, Between-Subjects Design und korrelative Verfahren.....	216
5.2.3	Effekte der Datenauswertung: Der Umgang mit Urteilsvarianzen.....	224
5.3	Fazit zum Effekt der Symmetrie auf die Attraktivität von Gesichtern und Konsequenzen für künftige Studien	233
6	Untersuchungen zur Hypothese des sexuellen Dimorphismus.....	239
6.1	Stand der Forschung zum Einfluss des sexuellen Dimorphismus	239
6.1.1	Begriffsdefinition des sexuellen Dimorphismus	239
6.1.2	Hintergrund der Theorie des sexuellen Dimorphismus	240

6.1.3	Methoden zur Untersuchung der Theorie des sexuellen Dimorphismus.....	242
6.1.4	Empirische Befunde zum Einfluss femininer Gesichtszüge bei Frauengesichtern.....	244
6.1.5	Empirische Befunde zum Einfluss maskuliner Gesichtszüge bei Männergesichtern.....	245
6.1.6	Ziel der Untersuchungen zum Einfluss des sexuellen Dimorphismus und Hypothesen.....	248
6.2	Experimentelle Manipulation des sexuellen Dimorphismus durch stufenweise Annäherung von Frauengesichtern an das Kindchenschema (Untersuchung <i>Dimorph 1</i>).....	253
6.2.1	Methode.....	253
6.2.1.1	Versuchspersonen.....	253
6.2.1.2	Stimulusmaterial.....	253
6.2.1.3	Versuchsdurchführung.....	256
6.2.2	Ergebnisse.....	256
6.3	Paarvergleichsexperiment mit Manipulation des sexuellen Dimorphismus durch Annäherung von Frauengesichtern an das Kindchenschema (Untersuchung <i>Dimorph 2</i>).....	260
6.3.1	Methode.....	261
6.3.1.1	Versuchspersonen.....	261
6.3.1.2	Stimulusmaterial.....	261
6.3.1.3	Versuchsdurchführung.....	263
6.3.2	Ergebnisse.....	264
6.4	Paarvergleichsexperiment mit Manipulation des sexuellen Dimorphismus durch Annäherung von Männergesichtern an das Kindchenschema (Untersuchung <i>Dimorph 3</i>).....	269
6.4.1	Methode.....	269
6.4.1.1	Versuchspersonen.....	269
6.4.1.2	Stimulusmaterial und Versuchsdurchführung.....	270
6.4.2	Ergebnisse.....	271
7	Zusammenfassende Diskussion der Ergebnisse zum Einfluss des sexuellen Dimorphismus.....	275

7.1	Vergleich der Ergebnisse mit bisherigen Studien zum sexuellen Dimorphismus bei Frauengesichtern.....	275
7.2	Vergleich der Ergebnisse mit bisherigen Studien zum sexuellen Dimorphismus bei Männergesichtern.....	277
7.3	Vergleich der Ergebnisse mit den drei meist zitierten Experimenten zum sexuellen Dimorphismus	279
7.4	Methodische Reflexion zur Verwendung des Kindchenschemas	294
7.4.1	Attraktivitätserhöhende Merkmale des Kindchenschemas.....	295
7.4.2	Mögliche neutrale oder attraktivitätssenkende Merkmale des Kindchenschemas	301
7.5	Schlussfolgerungen und Ausblick: Von globalen Theorien und vom Schema-Denken zur Analyse von konkreten Einzelmerkmalen	305
8	Experimente zum direkten experimentellen Vergleich der Hypothesen „Durchschnitt“, „Symmetrie“ und „Sexueller Dimorphismus“	311
8.1	Experiment zum direkten experimentellen Hypothesen-Vergleich mit Männer- und Frauengesichtern im Ranking-Verfahren (Untersuchung <i>Vergleich 1</i>)	313
8.1.1	Methode	313
8.1.1.1	Versuchspersonen.....	313
8.1.1.2	Stimulusmaterial	313
8.1.1.3	Versuchsdurchführung	318
8.1.2	Ergebnisse.....	318
8.2	Experiment zum direkten experimentellen Hypothesen-Vergleich mit Männer- und Frauengesichtern im Rating-Verfahren und Between-Subjects Design (Untersuchung <i>Vergleich 2</i>).....	329
8.2.1	Methode.....	329
8.2.1.1	Versuchspersonen.....	329
8.2.1.2	Stimulusmaterial und Versuchsdurchführung.....	330
8.2.2	Ergebnisse.....	332
9	Zusammenfassende Diskussion aller Experimente zum direkten experimentellen Vergleich der Hypothesen „Durchschnitt“, „Symmetrie“ und „Sexueller Dimorphismus“	342
10	Fazit	351

11 Zusammenfassung	360
12 Literaturverzeichnis	363
13 Tabellen- und Abbildungsverzeichnis	376
14 Anhang	393
Anhang A	393
Anhang B	395
Anhang C	397
Anhang D	398
Anhang E	399
Anhang F	400
Anhang G	401

1 Einführung und Ziel der Arbeit

Wohl kaum ein anderes Thema der psychologischen Grundlagenforschung dürfte bei der Bevölkerung auf ähnlich großes Interesse stoßen wie die Frage, was einen Menschen schön macht. Insbesondere die Schönheit des menschlichen Gesichts wirkt dabei immer noch zahlreiche Rätsel auf. So unterschiedlich Gesichter auch aussehen, es gibt doch bestimmte Gesichter (z. B. von Top-Models), bei denen ein hoher Konsens herrscht, dass diese sehr schön sind. Doch so leicht ein hochattraktives Gesicht auch zu erkennen ist, so schwer fällt es, zu verstehen und zu begründen, was die eigentliche Ursache dieser Attraktivität ist.

Seit nunmehr gut zwei Jahrzehnten widmet sich auch die Wissenschaft dieser Thematik. Dieser Forschungszweig der Attraktivitätsforschung, der von Psychologen und mittlerweile vor allem von Evolutionsbiologen geprägt wird, hat zwar noch lange nicht alle Ursachen der Schönheit von Gesichtern verstanden, aber in dieser Zeit zumindest eine beträchtliche Anzahl von Forschungsergebnissen produziert. Ein wichtiges Ergebnis ist beispielsweise die Erkenntnis, dass – anders als der Volksmund behauptet – Schönheit eben nicht (nur) im Auge des Betrachters liegt, sondern dass es bei allen „Geschmacksunterschieden“ dennoch eine gemeinsame Basis für menschliche Attraktivität gibt, sowohl innerhalb einer Kultur als auch mit Abstrichen zwischen verschiedenen Kulturen.

Dieser Konsens wird von der Mehrzahl der Attraktivitätsforscher als Hinweis darauf interpretiert, dass es wohl eine biologische Basis für die Präferenz bestimmter Attraktivitätsmerkmale geben müsse, also eine angeborene Tendenz, bestimmte äußere Eigenschaften eines menschlichen Gesichts schön zu finden. Nach dieser evolutionspsychologischen Sichtweise ist Attraktivität daher auch viel mehr als bloß eine mehr oder weniger willkürlich von einer Gesellschaft festgelegte Norm, der ein Individuum zu entsprechen hat. Vielmehr wird Attraktivität als leicht sichtbares Signal für andere (nicht direkt beobachtbare) Eigenschaften gesehen, die für ein Lebewesen aus biologischer Sicht (!) wichtig sind, also Eigenschaften, die im Kontext von Partnerwahl und Fortpflanzung relevant sind, wie z. B. Gesundheit, Krankheitsresistenz, Fruchtbarkeit und Ähnliches.

Vor diesem theoretischen Hintergrund sind auch die drei wichtigsten Theorien zur Attraktivität von Gesichtern zu sehen, die die Attraktivitätsforschung in den letzten 20 Jahren hervorgebracht hat: Es sind (1) die „Durchschnittshypothese“, (2) die „Symmetrie-Hypothese“ und (3) die „Theorie des sexuellen Dimorphismus“. Nach der Durchschnitts-

hypothese soll ein Gesicht dann besonders attraktiv sein, wenn seine Gesichtsproportionen dem mathematischen Durchschnitt einer Population entsprechen, nach der Symmetrie-Hypothese sollen sich attraktive Gesichter durch hohe Symmetrie auszeichnen und nach der Theorie des sexuellen Dimorphismus sollen attraktive Gesichter besonders geschlechtstypisch aussehen, d. h. *Frauengesichter* sollen besonders attraktiv sein, wenn sie typisch feminin aussehen und *Männnergesichter*, wenn sie typisch maskulin aussehen.

Zu diesen drei Theorien gibt es inzwischen etliche Forschungsarbeiten (für eine Metaanalyse siehe Rhodes, 2006). Eine genauere Analyse dieser Arbeiten macht jedoch auch zahlreiche Widersprüche und methodische Schwächen deutlich. Beispielsweise sind bei vielen Untersuchungen Durchschnittlichkeit und Symmetrie konfundiert, und die Theorie des sexuellen Dimorphismus widerspricht der Durchschnittshypothese prinzipiell. Zudem genügt bereits ein flüchtiger Blick in den Anhang der Metaanalyse von Rhodes (2006), um zu sehen, dass nicht nur die Größe, sondern teilweise sogar die Richtung eines Effekts davon abhängt, mit welchen Methoden der angeblich attraktivitätserhöhende Einfluss der drei genannten Kriterien untersucht wurde – dies gilt insbesondere für Symmetrie und geschlechtstypisches Aussehen.

Das Ziel dieser Arbeit war es daher, diesen drei bedeutendsten Theorien der Attraktivitätsforschung auf den Grund zu gehen und kritisch auf den Zahn zu fühlen. Dazu wurden insgesamt 20 verschiedene Untersuchungen durchgeführt. Besonderer Wert wurde dabei auf eine präzise und saubere Methodik gelegt. Zudem wurden gezielt für die Untersuchung derselben Fragestellungen unterschiedliche Methoden verwendet, um herauszufinden, wie sich verschiedene Methoden beispielsweise im Versuchsdesign (z. B. Paarvergleich versus Rating), bei der Erzeugung von Stimuli oder bei der Datenanalyse auf Signifikanz und Effektgröße von Ergebnissen zur Gesichterattraktivität auswirken.

Die vorliegende Arbeit ist damit die umfassendste und gründlichste, in der bislang diese drei Theorien der Gesichterattraktivität gleichzeitig untersucht wurden. Sie geht einen ganzen Schritt weiter als übliche empirische Arbeiten zu diesem Thema, denn sie untersucht neben den Hypothesen selbst gleichzeitig auch den Einfluss der unterschiedlichen Methoden, mit denen bisherige Arbeiten erstellt wurden. Dass dabei auch der Stand der gegenwärtigen Attraktivitätsforschung gründlich aufgearbeitet und kritisch diskutiert wird, versteht sich von selbst.

Damit stellt die vorliegende Arbeit gewissermaßen auch eine Art „Meta-Studie“ dar, denn sie arbeitet nicht nur den gegenwärtigen Forschungsstand auf und stellt nicht nur umfassende eigene empirische Untersuchungen zum Thema an, sondern untersucht gleichzei-

tig auch das Untersuchungsinstrumentarium der Attraktivitätsforschung. Die so gewonnenen Ergebnisse werfen dadurch auch ein neues Licht auf frühere Studien und ermöglichen somit dem Leser, sich ein besseres und eigenständiges Urteil über den aktuellen Stand der Attraktivitätsforschung zu bilden.

2 Empirische Untersuchungen zum Einfluss der Durchschnittlichkeit auf Attraktivität

2.1 Stand der Forschung zur Durchschnittshypothese

In den letzten 20 Jahren hat sich unser Verständnis davon, was ein Gesicht attraktiv macht und was die Gründe für diese Präferenzen sind, dramatisch verändert. Während früher die Meinung vorherrschte, Schönheit liege allein im Auge des Betrachters, herrscht inzwischen unter Attraktivitätsforschern ein Konsens darüber, dass sowohl innerhalb wie auch zwischen verschiedenen Kulturen eine gemeinsame Basis für die Beurteilung von Gesichtsattraktivität existiert (Cunningham, Roberts, Barbee, Druen & Wu, 1995; Langlois et al., 2000; Perrett, May & Yoshikawa, 1994; Zebrowitz, 1997).

Eine der bedeutendsten Hypothesen der Attraktivität von Gesichtern ist – zumindest in Kreisen der Attraktivitätsforscher – die sogenannte „Durchschnittshypothese“. Sie besagt, dass ein Gesicht umso attraktiver ist, je durchschnittlicher es ist. Mit „durchschnittlich“ ist dabei der mathematische Durchschnitt einer Population gemeint (Langlois & Roggman, 1990). Die Operationalisierung erfolgt durch das Herstellen sogenannter „Composites“ mit Hilfe von Morphing-Software.

Der Erste, der dieses Phänomen beobachtete, war Francis Galton (Galton, 1878, 1888), der Cousin von Charles Darwin. Galton belichtete die Gesichter von mehreren Personen übereinander auf einer Fotoplatte, so dass jedes Einzelgesicht in etwa gleichmäßig zu einem „Gesamtgesicht“ beitrug. Dabei beobachtete er, dass dieses Durchschnittsgesicht attraktiver war als die Einzelgesichter, die darin enthalten waren. Lange vor der Formulierung der Durchschnittshypothese durch Attraktivitätsforscher wurde vor einem eher künstlerischen Hintergrund durch Mehrfachbelichtungen von Fotofilmen der optische Effekt der Durchschnittsbildung von Gesichtern reproduziert und diskutiert (Daucher, 1967). In seinem Standardwerk zur Humanethologie zeigt Eibl-Eibesfeld (1997, S. 902) dieses von

Dauchner in den 60er Jahren durch Mehrfachbelichtung angefertigte Durchschnittsgesicht aus 20 Münchner Studentinnen (inklusive der Originalgesichter).

Doch zu Bekanntheit gelangte dieses Phänomen erst mit einem Artikel von Langlois und Roggman (1990) mit dem Titel „Attractive faces are only average“. Die Autorinnen überlagerten am Computer die Portraits von zahlreichen Personen (getrennt für Männer und Frauen) und stellten fest, dass die dadurch entstandenen Composites umso attraktiver waren, je mehr *component faces* darin enthalten waren, je durchschnittlicher sie also waren. Diese Ergebnisse wurden auch mit etwas anderen technischen Verfahren der Durchschnittsbildung als dieser sogenannten „*pixel matrix averaging procedure*“ repliziert, wie z. B. der traditionellen Mehrfachbelichtung von Fotofilmen (Pollard, 1995) und v. a. durch die modernere Morphing-Technik (*keypoint average method*) (DeBruine, Jones, Unger, Little & Feinberg, 2007; Langlois, Roggman & Musselman, 1994; Perrett et al., 1994; Rhodes, Sumich & Byatt, 1999; Rhodes, Yoshikawa et al., 2001; Valentine, Darling & Donnelly, 2004); für eine Übersicht siehe Rhodes (2006). Doch was ist der Grund dafür, warum diese Composites so attraktiv beurteilt werden? Insgesamt lassen sich drei wichtige Erklärungsansätze identifizieren:

1. Die Durchschnittlichkeit der Gesichtsproportionen von Composites
2. Die hohe Symmetrie von Composites
3. Die makellose Haut von Composites als Nebeneffekt (Artefakt) des Morphing-Prozesses

Die am häufigsten angeführte Erklärung (1) besagt, dass Composites in ihrer *facial shape* durchschnittlicher sind, und es gerade diese Durchschnittlichkeit sei, die von den Befragten als besonders positiv bewertet wird. Diese Auffassung wurde auch von Langlois und Roggman vertreten. Diese Interpretation wurde jedoch früh angegriffen und es wurde argumentiert, dass solche Composites ja nicht nur in ihrer Gestalt durchschnittlicher, sondern auch zugleich symmetrischer (2) seien (Grammer & Thornhill, 1994; Rhodes, Roberts & Simmons, 1999). Der Effekt sei daher auf größere Attraktivität symmetrischer Gesichtszüge zurückzuführen.

Eine unter Attraktivitätsforschern (!) als Außenseitermeinung gehandelte Interpretation (3) lautete, Composites seien deswegen attraktiver, weil sie bedingt durch ihren Herstellungsprozess auch gleichzeitig unschärfer aussähen (Alley & Cunningham, 1991). Dadurch verschwänden auch Makel wie Pickel oder Falten und die Haut sehe makelloser und jugendlicher aus. Somit sei der Effekt lediglich auf Bildverarbeitungsartefakte zurück-

zuföhren. Diese drei Erklärungsansätze schließen sich selbstverständlich nicht gegenseitig aus – möglicherweise enthält jeder einen wahren Kern.

Wenn durchschnittliche Gesichtsproportionen (*average facial shape*) tatsächlich attraktiv machen, wie die erste Interpretation besagt, was könnte dann der tiefere Sinn dieser Präferenz sein? Etliche Attraktivitätsforscher nehmen an, dass eine Vorliebe für durchschnittliche Gesichtsproportionen eine Adaptation an das Problem der Partnerwahl (*mate choice*) ist. Ursache dieser Adaptation sei, dass durchschnittliche Gesichtszüge einen hohen Partnerwert (*high mate quality*) signalisierten (Gangestad & Buss, 1993; Thornhill & Gangestad, 1993).

Gemäß diesem sogenannten „*good genes view*“ signalisieren durchschnittliche Gesichtszüge Eigenschaften eines erstrebenswerten Partners wie z. B. Gesundheit, Fruchtbarkeit oder Intelligenz. Angeblich spiegelt sich in Durchschnittlichkeit (ebenso wie in Symmetrie) (1) Entwicklungsstabilität (*developmental stability*), d. h. die Fähigkeit eines Organismus, eine normale Entwicklung trotz widriger Umweltbedingungen (*environmental stress*) aufrechtzuerhalten (Thornhill & Moller, 1997). Noch etwas spekulativer ist die Annahme (2), dass Durchschnittlichkeit Heterozygotie (*heterozygosity*) signalisiere, was sich in einer erhöhten Parasitenresistenz niederschlägt (= Parasitenresistenzhypothese) (Gangestad & Buss, 1993; Thornhill & Gangestad, 1993).

Zudem (3) wird vermutet, dass unter der Annahme einer stabilisierenden Selektion (*stabilizing selection*), die auf Gesichtszüge so wirkt wie auf viele andere Merkmale auch, Individuen mit durchschnittlichen Gesichtszügen eine höhere Fitness haben müssten als andere (Koeslag, 1990; Symons, 1979). Ein Überblick über die biologischen Theorien der Gesichtsattraktivität findet sich bei Kalick, Zebrowitz, Langlois und Johnson (1998) sowie Thornhill und Gangestad (1999). Gegen diesen „*good-genes-account*“ spricht jedoch, dass sich empirisch kaum ein nennenswerter Zusammenhang zwischen Attraktivität und Gesundheit nachweisen lässt (Henderson & Anglin, 2003; Hume & Montgomerie, 2001; Kalick, Zebrowitz, Langlois & Johnson, 1998; Langlois et al., 2000; Rhodes, Zebrowitz et al., 2001; Shackelford & Larsen, 1999); für einen Überblick siehe Weeden & Sabini, 2005).

Eine Alternativerklärung der Präferenz von Durchschnittlichkeit ist, dass Composites als schöner empfunden werden, weil sie eher wie ein typisches Gesicht aussehen. Diese Prototypen-Theorie, eine kognitive Sichtweise, wurde von Langlois und Roggman (1990) eingeföhrt. Als Argument wird angeführt, dass Gesichter, die als attraktiver beurteilt werden, ebenfalls als „more typical“ and „less unusual“ bewertet werden (Light, Hollander &

Karya-Stuart, 1981). Dadurch wirkten sie auf den Betrachter vertrauter und deswegen attraktiver. Das gelte nicht nur für Gesichter, sondern auch für andere Dinge (Vögel, Fische, Autos), die als attraktiver empfunden würden, wenn sie dem typischen Vertreter ihrer Kategorie ähnelten (Halberstadt & Rhodes, 2003). Anhänger der Prototypentheorie argumentieren, dass das prototypische Aussehen zu einer Wahrnehmungserleichterung führe, die von den Menschen als positiv erlebt werde (Winkielman, Halberstadt, Fazendeiro & Catty, 2006; Winkielman, Schwarz, Reber & Fazendeiro, 2003), weshalb auch durchschnittliche Gesichter als attraktiver empfunden würden (für eine Diskussion der Prototypentheorie siehe Rhodes, 2006). Insgesamt sind die Anhänger der Prototypentheorie unter den Attraktivitätsforschern stark in der Minderheit – was jedoch nicht heißen muss, dass sie deswegen Unrecht hätten. Die wissenschaftliche Diskussion wird jedoch von den biologisch orientierten Forschern dominiert, die bei der Bevorzugung von Durchschnittlichkeit von einem evolutionsbiologischen Hintergrund ausgehen.

Diese Erklärungen sind jedoch nur dann von Belang, wenn sich auch nachweisen lässt, dass die Schönheit von Composites tatsächlich auf die durchschnittliche Gestalt zurückzuführen ist und eben nicht auf ihre gleichzeitig erhöhte Symmetrie. Denn mittlerweile überwiegt in der Attraktivitätsforschung die Auffassung, dass Symmetrie ein Faktor ist, der für Attraktivität eine Rolle spielt. In der Metaanalyse von Rhodes (2006) wird der Effekt von Symmetrie mit $R = .25$ angegeben, der Effekt von Durchschnittlichkeit mit $R = .52$. Allerdings beruhen diese Zahlen überwiegend auf Attraktivitätsstudien zur Durchschnittlichkeit, die Effekte der Durchschnittlichkeit nicht von denen der Symmetrie trennen konnten.

Ein Versuch, diese beiden Variablen zu trennen, ist die Studie von Rhodes, Sumich und Byatt (1999). Die Autoren ließen Versuchspersonen die Symmetrie und die Durchschnittlichkeit von gemorphten Gesichtern mit Ratingskalen beurteilen und wollten damit die Symmetrie und die Durchschnittlichkeit der jeweiligen Gesichter quantifizieren. Mit Hilfe von multiplen Regressionsanalysen kamen sie dann zu dem Schluss, dass beide Faktoren unabhängig voneinander zu Attraktivität beitrugen. Die Art und Weise der Quantifizierung von Durchschnittlichkeit und Symmetrie ist jedoch skeptisch zu beurteilen. Es ist fraglich, ob Versuchspersonen überhaupt in der Lage sind, die Symmetrie oder die Durchschnittlichkeit eines Gesichts zu beurteilen und ob diese Ratings nicht jeweils (aufgrund eines Halo-Effekts) mit der Attraktivität eines Gesichts konfundiert sind. In Kapitel 4.10 wird dieses Direkt-Rating für die Variable Symmetrie mit einem eigenen Experiment ebenfalls angewandt und kritisch hinterfragt (siehe dazu auch die Kapitel 4.10.4 und 5.2.2).

Weitaus geschickter als solche direkten Ratings von Symmetrie und Durchschnittlichkeit ist eine Methode, die Valentine et al. (2004) gewählt haben. Sie testeten die Durchschnittshypothese, indem sie Aufnahmen der Profilansicht von Gesichtern verwendeten. Da ein Gesicht in der Profilansicht keine Symmetrieachse besitzt, können bei Durchschnittsprofilen höhere Attraktivitätswerte logischerweise nicht durch Symmetrie verursacht werden. Eine Konfundierung von Durchschnittlichkeit und Symmetrie ist damit von vornherein ausgeschlossen.

Diese Idee war auch Grundlage für die hier durchgeführten Experimente zur Durchschnittshypothese. In Experiment 1 sollte das Experiment von Langlois und Roggman (1990) mit 32 Männern in der Frontalansicht repliziert werden, indem der Zusammenhang zwischen Durchschnittlichkeit und Attraktivität berechnet wurde. Experiment 2 ist eine Wiederholung von Experiment 1, mit dem Unterschied, dass hierfür die *Profilansichten* derselben Männer verwendet wurden. Durch den Vergleich der beiden Effektgrößen kann herausgefunden werden, wie groß der Einfluss der Symmetrie bei der besseren Beurteilung der Composites ist, da Symmetrie als Erklärung prinzipiell bei Frontalansichten, nicht jedoch bei Profilansichten in Frage kommt. Die Experimente *Durch 4* und *Durch 5* sind analog zu den Experimenten *Durch 1* und *Durch 2*, benutzen jedoch als Stimuli Frauengesichter.

Dennoch gibt es einen dritten Grund, warum Composites so attraktiv beurteilt werden. Es ist, wie bereits oben erwähnt, die Möglichkeit von Bildverarbeitungsartefakten, die bei jeder Art der Composite-Erzeugung auftreten, egal ob in Frontal- oder Profilansicht. Wie bereits Alley und Cunningham (1991) früh kritisierten, sehen Composites unschärfer und verschwommener aus, wodurch Makel wie Pickel oder Falten weniger stark in Erscheinung treten. Bei einem *32-face-composite* sind sogar sämtliche Makel durch die Mittelwertbildung der Farbwerte (bzw. Graustufen) vollständig verschwunden. Daher ist es nicht erstaunlich, dass solch ein Gesicht mit makelloser Haut besser bewertet wird als ein Originalgesicht mit Pickeln und Falten.

In früheren Experimenten zur Durchschnittshypothese wurde im Bewusstsein dieses Problems durchaus versucht, dieser Konfundierung gegenzusteuern. Beispielsweise wurden auch die *Component Faces* mit Weichzeichnungsfiltern bearbeitet, um den Kontrast zum Durchschnittsgesicht abzumildern (z. B. Langlois & Roggman, 1990; Grammer & Thornhill, 1994). Doch dies behebt das Problem nur teilweise, da die Mittelwertbildung von Farbwerten verschiedener Bilder prinzipiell etwas anderes ist als ein Filter (im Diskussionsteil Kapitel 3.1.2 wird dies näher ausgeführt und mit Bildern belegt werden).

Gewöhnlich wird auch als Argument gegen den Einwand, dass Bildverarbeitungsartefakte für die bessere Bewertung von Composites eine Rolle spielen, die Studie von Rhodes und Tremewan (1996) angeführt, die die Attraktivität von Durchschnittsgesichtern unter Verwendung von Strichzeichnungen (*line-drawings*) nachweisen wollten. Doch diese Studie zeigt lediglich, dass es möglicherweise einen Effekt der Durchschnittsbildung (bzw. Symmetrie) gibt (wenn man diese Variablen durch Versuchspersonenratings erfasst), aber nicht, dass deswegen die makellose Gesichtsoberfläche von Composites aus Fotos zu vernachlässigen wäre. (Davon abgesehen, wird gerade diese Untersuchung häufig fehlinterpretiert und im falschen Zusammenhang zitiert, denn der Versuch, mit Strichzeichnungen die Ergebnisse von Langlois und Roggman zu replizieren, schlug fehl; mehr dazu in Kapitel 3.1.2). Dennoch ist die Grundidee von Rhodes und Tremewan völlig richtig, nämlich den Störfaktor Textur auszublenden.

In der vorliegenden Studie wurde dies durch das Experiment *Durch 3* (Männerngesichter) sowie Experiment *Durch 6* (Frauengesichter) erreicht, die Replikationen der Experimente *Durch 2* und *Durch 5* sind, jedoch mit einer Modifikation: Statt der farbigen Morphings der Profilansichten wurden hier sogenannte „Schattenprofile“ als Stimulusmaterial verwendet, bei denen lediglich der Umriss eines Profils erhalten blieb, die gesamte Textur jedoch durch eine gleichmäßig schwarze Fläche ersetzt wurde (Schatten). In beiden Experimenten wurde dann der Zusammenhang zwischen Durchschnittlichkeit und Attraktivität gemessen, und durch den Vergleich der beiden Effektgrößen konnte herausgefunden werden, wie groß der Einfluss der makelloseren Textur bei der Composite-Bildung war.

In der vorliegenden Untersuchungsreihe wird durch die verwendete Methodik einem weiteren Phänomen Rechnung getragen. Obwohl einerseits bekannt ist, dass Durchschnittsgesichter attraktiv sind, gibt es andererseits auch zahlreiche Belege dafür, dass Merkmale die Attraktivität erhöhen, die gerade vom Durchschnitt abweichen (DeBruine et al., 2007). Beispiele dafür sind bei Frauengesichtern kindchenhafte Merkmale (*babyfacedness*) wie ein kleiner, zierlicher Unterkiefer, eine kleine Nase, ein weiter Augenabstand, aber auch Reifekennzeichen wie hohe Wangenknochen (Cunningham, 1986) oder *sex-hormone-markers* wie volle Lippen (Johnston, 2000; Johnston, Solomon, Gibson & Pallares-Bejarano, 2003). Ähnliches gilt für Männerngesichter, bei denen sich die Befunde allerdings widersprechen. Z. B. zeigt sich in manchen Untersuchungen, dass maskuline Merkmale wie ein markanter Unterkiefer die Attraktivität erhöhen (Alley & Cunningham, 1991; Cunningham, Barbee & Pike, 1990; Grammer & Thornhill, 1994) während in anderen eine Femininisierung attraktiver wirkt (Perrett et al., 1998). Beides ist jedoch nicht gleichbedeutend mit Durchschnittlichkeit. Die Auswirkungen des sexuellen Dimorphismus

als Attraktivitätsmerkmal werden in den Kapiteln 6.2 bis 6.4 durch drei Experimente untersucht. In der Metaanalyse von Rhodes (2006) wird für Frauengesichter der Effekt des sexuellen Dimorphismus mit $R = .64$ und für Männergesichter mit $R = -.12$ angegeben.

Die Theorie des sexuellen Dimorphismus steht prinzipiell im Widerspruch zur Durchschnittshypothese. Dies bedeutet jedoch nicht im Umkehrschluss, dass die Durchschnittshypothese richtig ist, wenn die Theorie des sexuellen Dimorphismus falsch ist. Denn letztere macht ja sehr konkrete Vorhersagen, z. B. je männlicher ein Männergesicht, desto attraktiver ist es. Auch falls dies nicht stimmen sollte (und der marginal negative Gesamteffekt von Rhodes' Metaanalyse ist ein Hinweis darauf), dann kann es trotzdem sein, dass hochattraktive Männergesichter sich dadurch auszeichnen, dass sie *weder* durchschnittlich *noch* extrem männlich aussehen.

Die entscheidende Frage im Zusammenhang mit der Durchschnittshypothese ist daher, ob durchschnittliche Gesichtsproportionen (*facial shape*) tatsächlich das Ideal sind, wie so oft postuliert wurde (Rubenstein, Langlois & Roggman, 2002), oder ob es andere Gesichtsproportionen gibt, die systematisch vom Durchschnitt abweichen und zugleich schöner sind. Insbesondere die Experimente von Perrett, May und Yoshikawa (1994) und DeBruine et al. (2007), bei denen nicht nur ein Composite aus allen Frauen einer Stichprobe, sondern zum Vergleich auch ein Composite aus den schönsten Frauen erzeugt wurde, das attraktiver bewertet wurde, drängen diesen Schluss auf.

Zur Überprüfung dieser Frage wird in der vorliegenden Untersuchungsreihe die Methodik von Langlois & Roggman (1990) leicht modifiziert. Die Erzeugung eines *32-face-composite* in mehreren Schritten durch paarweises Morphen geschieht nicht, indem die Originalgesichter in *beliebiger* Reihenfolge miteinander gekreuzt werden. Stattdessen wird zuerst in einer Voruntersuchung die Attraktivität der Originalgesichter ermittelt, eine Rangreihenfolge gebildet und dann immer ähnlich attraktive Gesichter miteinander vermorphen. Auf das Aussehen des *32-face-composites* hat dies keinen Einfluss, da es ja *alle* 32 Gesichter zu gleichen Anteilen enthält, jedoch sehr wohl auch die „Zwischenprodukte“. Es entstehen dadurch Gesichter, die z. B. aus den attraktivsten vier oder attraktivsten acht Originalgesichtern erzeugt sind. Es wird die Hypothese aufgestellt, dass diese Gesichter attraktiver als das Gesamt-Durchschnittsgesicht (*32-face-composite*) sind. Durch den Vergleich dieser Grafiken mit dem Gesamtdurchschnittsgesicht wird man gleichzeitig sehen können, welche Merkmale ein Gesicht nicht nur ziemlich attraktiv, sondern besonders attraktiv machen.

2.2 Ziel der Untersuchungen zum Einfluss der Durchschnittlichkeit und Hypothesen

Mit den sechs Experimenten zur Durchschnittshypothese *Durch 1* bis *Durch 6* soll der aus der Attraktivitätsforschung bekannte Befund, dass mit Morphing-Software hergestellte Durchschnittsgesichter (*Composites*) besonders attraktiv sind, repliziert werden. Gleichzeitig besteht jedoch der Anspruch, die kausale Ursache dieser Attraktivitätssteigerung nachzuweisen, d. h. die Effekte von

- durchschnittlichen Gesichtsproportionen (*average facial shape*),
- Symmetrie und
- Morphing-Artefakten

zu separieren. Diese Trennung wird durch die Verwendung verschiedener Stimuli in den Experimenten umgesetzt, nämlich durch

- Composites aus Frontalgesichtern mit Textur (die eine durchschnittliche *facial shape*, hohe Symmetrie und aufgrund von Morphing-Artefakten eine makellose Haut besitzen),
- Composites aus Profilgesichtern *mit* Textur (die eine durchschnittliche *facial shape*, und aufgrund von Morphing-Artefakten eine makellose Haut besitzen, jedoch *keine* Symmetrie) und
- Composites aus Profilgesichtern *ohne* Textur (= Schattenprofile, die *nur* eine durchschnittliche *facial shape* besitzen)

Durch den Vergleich der Effekte aus den jeweiligen Experimenten können Rückschlüsse auf die drei genannten möglichen Ursachen gezogen werden.

Gleichzeitig soll jedoch auch nach Alternativen für die Durchschnittshypothese gesucht werden. Sie kann dann falsifiziert werden, wenn es gelingt, mit derselben Methode des Morphens Gesichter zu kreieren, die weniger durchschnittlich, aber dennoch attraktiver als das Durchschnittsgesicht aus allen Originalgesichtern erzeugt wurde. Dazu werden in mehreren Schritten systematisch ähnlich attraktive Originalgesichter miteinander vermorph. Es wird vermutet, dass auf diese Weise Composites entstehen, die aus den attraktivsten Originalgesichtern berechnet wurden und die

- anders aussehen als das durchschnittlichste Composite und
- deutlich attraktiver bewertet werden als das durchschnittlichste Composite.

Damit wäre ein Nachweis erbracht, dass durchschnittliche Gesichtsproportionen *nicht* das Schönheitsideal darstellen.

Folgende Hypothesen werden in jedem der sechs Experimente *Durch 1* bis *Durch 6* überprüft:

- Je durchschnittlicher ein gemorphtes Gesicht ist, desto attraktiver ist es.
- Alternativhypothese: Ein gemorphtes Gesicht ist nur so attraktiv wie die Originalgesichter, die darin enthalten sind, d. h.
 - Composites aus wenigen attraktiven Originalgesichtern sind attraktiv
 - Composites aus wenigen unattraktiven Originalgesichtern sind unattraktiv
 - Composites aus *allen* Originalgesichtern einer Stichprobe sind durchschnittlich attraktiv

Wenn sich die Alternativhypothese bestätigen lässt, dann gilt auch:

- Composites aus den attraktivsten Originalgesichtern sind attraktiver als das Durchschnittsgesicht aus *allen* Gesichtern einer Stichprobe.

Durch die Experimente mit den *Profil*gesichtern (*Durch 2*, *Durch 3*, *Durch 5* und *Durch 6*) lässt sich zusätzlich der Einfluss von Symmetrie und Morphing-Artefakten auf die Attraktivitätsbewertungen von Durchschnittsgesichtern analysieren:

- Der Effekt der Verdurchschnittlichung ist bei farbigen Profilgesichtern geringer als bei Frontalgesichtern, da bei gemorphten Frontalgesichtern ein Teil ihrer Attraktivität aus einer erhöhten Symmetrie resultiert.
- Der Effekt der Verdurchschnittlichung ist bei Schattenprofilen deutlich geringer als bei farbigen Profilen, da bei farbigen gemorphten Profilen ein Teil ihrer Attraktivität aus der makellosen Haut (= Morphing-Artefakte) resultiert.

2.3 Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von männlichen Frontalgesichtern (Untersuchung *Durch 1*)

2.3.1 Methode

2.3.1.1 Voruntersuchung

Das Fotografieren der männlichen Originalgesichter

Für die Untersuchung wurden 33 deutsche Männer („kaukasischer Typ“, 32 Studenten der Universität Regensburg verschiedener Fachbereiche und ein Model) im Alter von 21 bis 30 Jahren ($M = 25.1$, $SD = 2.5$) fotografiert. Die Teilnahme war freiwillig und wurde nicht bezahlt. Von jeder Person wurden mit einer Digitalkamera (Auflösung 1280 x 1024 Pixel) zwei Fotos gemacht, eines in der Frontalansicht und eines in der Profilansicht. Bei den Aufnahmen wurde auf möglichst standardisierte Bedingungen Wert gelegt. Alle Personen wurden unter konstanten und gleichmäßigen Lichtbedingungen aus einer konstanten Entfernung vor einem weißen Hintergrund abgelichtet. Alle trugen ein weißes T-Shirt, mussten Brillen und Piercings abnehmen und die Haare nach hinten kämmen bzw. mit Haarnadeln nach hinten stecken, so dass die Stirn nicht von Haaren bedeckt war. Sie wurden instruiert, nicht zu lächeln, sondern mit einem neutralen Gesichtsausdruck zu schauen.

Attraktivitätsbewertungen der Männergesichter in der Voruntersuchung

Die Frontalansichten aller 33 fotografierten Männer wurden in einer Größe von 320 x 400 Pixeln Versuchspersonen zur Beurteilung mit einem Computerprogramm in variierender Reihenfolge präsentiert. Jede Versuchsperson beurteilte jedes Gesicht auf einer Likert-Skala von 1 (= sehr unattraktiv) bis 7 (= sehr attraktiv) hinsichtlich seiner Attraktivität.

Versuchspersonen für die Voruntersuchung

An der Voruntersuchung zur Bewertung der Männerfotos in der Frontalansicht nahmen insgesamt 42 Versuchspersonen (17 Männer, 25 Frauen) teil. Die Altersspanne reichte von 14 bis 52 Jahren ($M = 30.2$, $SD = 11.1$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, wurde die Befragung an einem Samstag in einem Einkaufszentrum durchgeführt. Die Teilnahme war freiwillig und wurde nicht bezahlt. Die

Versuchspersonen wurden gebeten, nur die Attraktivität des Gesichts zu beurteilen und dabei die Frisur der gezeigten Personen zu ignorieren.

Ergebnisse der Attraktivitätsbewertungen der Männergesichter in der Voruntersuchung

Für die Auswertung des Vortests wurde für jedes Männergesicht ein Attraktivitätsmittelwert über alle Versuchspersonen berechnet und die Gesichter wurden auf der Basis dieser Rating-Mittelwerte in eine Rangreihenfolge gebracht. Das unattraktivste Gesicht erhielt im Mittel den Wert 2.02 ($SD = 0.98$), das attraktivste den Wert 5.04 ($SD = 1.08$). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .97 (bei 42 Versuchspersonen) sehr hoch. Dies rechtfertigte es, für jedes Gesicht die Urteile verschiedener Beurteiler zu einem Gesamtwert zusammenzufassen.

2.3.1.2 Versuchspersonen und Durchführung der Hauptuntersuchung

An der Hauptuntersuchung zur Bewertung der gemorphten Männerbilder in der Frontalansicht nahmen insgesamt 53 Versuchspersonen (27 Männer, 26 Frauen) teil. Die Altersspanne reichte von 16 bis 80 Jahren ($M = 35.0$, $SD = 13.9$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, wurde die Befragung wie auch beim Vortest an einem Samstag in einem Einkaufszentrum durchgeführt. Die Teilnahme war freiwillig und wurde nicht bezahlt. Die Versuchspersonen wurden gebeten, nur die Attraktivität des Gesichts zu beurteilen, und dabei die Frisur der gezeigten Personen zu ignorieren. Die Darbietung der Bilder erfolgte in derselben Art und Weise wie beim Vortest.

2.3.1.3 Stimulusmaterial der Hauptuntersuchung

Nachdem die Attraktivität der männlichen Originalgesichter empirisch ermittelt und alle Gesichter nach ihrer Attraktivität in eine Rangreihenfolge gebracht waren, wurden mit dem Morphing-Programm *Morpher 3.0* aus den Originalgesichtern neue Gesichter berechnet. Hierzu wurden je zwei Gesichter miteinander gekreuzt und zu einem neuen, virtuellen Gesicht gemorpht. Die Produkte dieses Morphing-Prozesses wurden dann wiederum miteinander gekreuzt, bis am Ende ein Gesicht resultierte, das 32 Originalgesichter zu gleichen Anteilen enthielt. Ein Beispiel soll dies näher erläutern:

Aus Gesicht 1 und Gesicht 2 wird ein neues Gesicht (1&2) berechnet. Aus Gesicht 3 und Gesicht 4 wird ebenfalls ein neues Gesicht (3&4) berechnet. Anschließend wird aus Ge-

sicht (1&2) und Gesicht (3&4) wieder ein neues Gesicht (1-4) berechnet, in dem dann die Gesichter 1 bis 4 zu gleichen Teilen enthalten sind. Ebenso wird mit allen anderen Gesichtern verfahren. Aus zwei „Elternteilen“ entsteht ein „Kind“. „Zwei Kinder“ sind dann wiederum die „Eltern“ für ein neues Kind. Bei 32 (2^5) Originalgesichtern kommt man so über fünf Generationen mittels eines binären Baums zu einem Durchschnittsgesicht, das alle 32 Gesichter zu gleichen Teilen enthält (eines der ursprünglich 33 Originalgesichter wurde für den Morph-Prozess weggelassen).

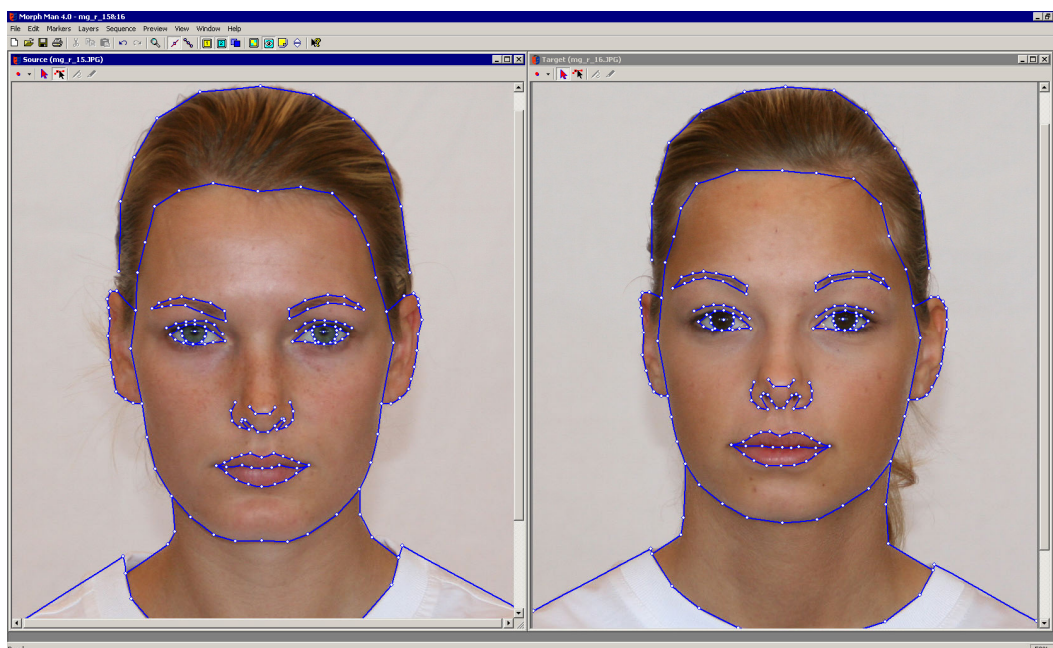


Abbildung 1: Screenshot aus einer Morphing-Software (*MorphMan 4.0*), hier mit Frauengesichtern. Die weißen Punkte sind die definierten Referenzpunkte, die blauen Linien definieren die Konturen der einzelnen Gesichtsm Merkmale. Beim Verrechnen der beiden Gesichter werden die korrespondierenden Referenzpunkte mit den entsprechenden Konturen jeweils aufeinander abgebildet.

Für das Morphen wurden die Gesichter in einer Auflösung von 1280 x 1024 Pixeln verarbeitet. Pro Gesicht wurden 250 Referenzpunkte gesetzt. Das Verrechnen zweier Gesichter erfolgte stets in einer Gewichtung von 50:50, d. h. in einem Composite sind die beiden Ausgangsgesichter zu jeweils gleichen Anteilen enthalten. Beim Morphen wurden sowohl die Koordinatenwerte als auch die Farbwerte gemittelt. Jedes Composite wurde vor einem neuen Morphing-Durchgang behutsam mit *Adobe Photoshop* retuschiert, um unvermeidliche Morphing-Artefakte (z. B. Überblendungen an den Frisuren) zu retuschieren.

Durch den hohen Arbeitsaufwand beim Morphen war es möglich, Durchschnittsgesichter zu erzeugen, die eine hohe fotorealistische Qualität aufweisen und selbst in höheren „Generationen“ kaum künstlich wirken.

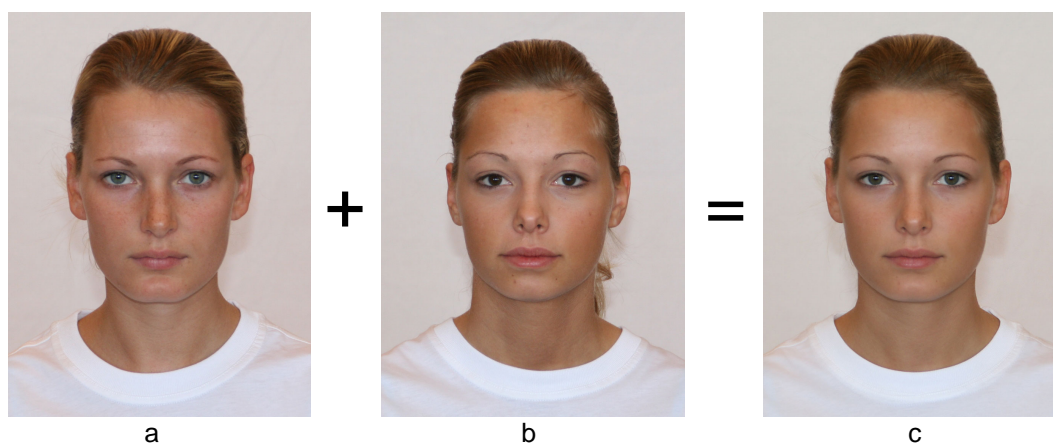


Abbildung 2: Ein Beispiel für das Morphen von Gesichtern, hier mit den Frauengesichtern aus Abbildung 1. Aus den Gesichtern a und b ergibt sich bei einer Gewichtung von 50:50 das Gesicht c, das die beiden Ausgangsgesichter zu gleichen Anteilen enthält.

Das hier gewählte Vorgehen unterscheidet sich in einem ganz wesentlichen Punkt von anderen Experimenten zur Untersuchung der Durchschnittshypothese (Grammer & Thornhill, 1994; Langlois & Roggman, 1990; Langlois et al., 1994; Little & Hancock, 2002; Valentine et al., 2004). In anderen Untersuchungen wurden die Originalgesichter in beliebiger Reihenfolge paarweise miteinander gekreuzt. Hier hingegen ging dem Morphen zuerst eine Voruntersuchung zur Ermittlung der Attraktivität der Originalgesichter voraus, das zu einem Attraktivitätsranking der Originalgesichter führte. Beim Morphen wurden dann immer ähnlich attraktive Gesichter miteinander gekreuzt, d. h., das unattraktivste mit dem zweitunattraktivsten Gesicht gekreuzt, das drittunattraktivste mit dem viertunattraktivsten usw. bis hin zum zweitattraktivsten Gesicht, das mit dem attraktivsten gekreuzt wurde.

Dieses Verfahren wurde deswegen gewählt, weil vermutet wurde, dass es (entgegen der Auffassung von Langlois & Roggman, 1990) bei der Attraktivitätsbewertung der Composites nicht nur auf deren Grad der Durchschnittlichkeit ankommt, sondern auch darauf, aus *welchen* Gesichtern sie berechnet wurden. So sollten Composites aus attraktiven Ori-

nalgesichtern besser bewertet werden als solche aus unattraktiven Originalgesichtern. Das vorliegende Stimulusmaterial erlaubt, diese Hypothese zu überprüfen. So gibt es beispielsweise in der dritten Composite-Generation ein Gesicht, „m(1-8)“, das aus den acht unattraktivsten Männergesichtern erzeugt wurde, und eines „m(25-32)“, das aus den attraktivsten acht Männergesichtern berechnet wurde. Nach Langlois & Roggman müssten diese beiden Gesichter gleich attraktiv sein, da sie gleich durchschnittlich sind. Ebenso müssten beide Gesichter weniger attraktiv sein als das Composite aus allen 32 Gesichtern „m(1-32)“, da sie beide weniger durchschnittlich sind. Wie in Kapitel 2.2 dargelegt wurde, wird hier die Auffassung vertreten, dass der Durchschnitt aus den Schönsten schöner ist als der Durchschnitt aus allen oder gar der Durchschnitt aus Unattraktiven.

Bestimmung der theoretischen Attraktivitätswerte (Nullhypothese)

Daher soll in dieser Studie nicht nur die etablierte Hypothese („je durchschnittlicher, desto schöner“) überprüft werden, sondern zusätzlich eine adäquate Alternativhypothese. Diese lautet: Ein Durchschnittsgesicht ist genau so attraktiv wie der Durchschnitt der Attraktivitätswerte der Originalgesichter, aus denen es erzeugt wurde. Diese Hypothese ließe sich auch als konkretisierte Nullhypothese bezeichnen, gegen die die Durchschnittshypothese getestet wird. Ein Beispiel: Die empirische Physische Attraktivität (PA) des Gesichts „m(25-32)“, also dem Durchschnitt aus den schönsten acht Originalgesichtern, lässt sich mit

$$PA_{m(25-32)} = \frac{\sum [PA_{m(25)} + PA_{m(26)} + \dots + PA_{m(32)}]}{8}$$

schätzen, wobei PA der Attraktivitätswert eines Gesichts ist (empirisch ermittelt aus dem Mittelwert der Rating-Urteile von Versuchspersonen) und der Index „m(25-32)“ für das Composite der Gesichter mit den Attraktivitäts-Rängen 25 bis 32 steht.

Nach diesem Prinzip wurde für jedes Composite ein theoretischer Attraktivitätswert berechnet. Der theoretische Wert ist dabei immer der Mittelwert aus den Attraktivitätswerten der jeweiligen Originalgesichter, aus denen ein Composite gemorpht wurde.

2.3.2 Ergebnisse

Analog zur Auswertung des Vortests wurde für jedes Composite ein Attraktivitätsmittelwert über alle Versuchspersonen berechnet. Das unattraktivste Composite erhielt im

Mittel den Wert 2.51 ($SD = 0.91$), das attraktivste den Wert 5.60 ($SD = 1.32$). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .98 (bei 53 Versuchspersonen) ebenso hoch wie beim Vortest.

Insgesamt wurden die gemorphten Männergesichter besser bewertet als die Original-Männergesichter. Der mittlere Attraktivitätswert der gemorphten Gesichter lag bei 4.19 ($SD = 1.13$), der der Originalgesichter nur bei 3.18 ($SD = 1.12$), d. h. das Attraktivitätsurteil lag bei den Composites im Durchschnitt um genau eine Skalenstufe höher.

Die Durchschnittshypothese von Langlois & Roggman (1990) („je durchschnittlicher, desto attraktiver“) wurde durch eine Korrelation überprüft. Dazu wurde der durchschnittliche Attraktivitätswert jedes Composites in Relation zum Grad seiner Durchschnittlichkeit gesetzt. Die Durchschnittlichkeit wurde codiert durch die „Generation“, in der das Composite entstand. So gab es insgesamt 16 Composites der Generation 1 (berechnet aus je zwei Originalgesichtern) 8 Composites der Generation 2 (berechnet aus je vier Originalgesichtern) usw. bis hin zu einem Composite der Generation 5, dem „absoluten“ Durchschnittsgesicht (berechnet aus allen 32 Originalgesichtern). Da diese Daten prinzipbedingt nicht normalverteilt sind, wurde als Test die verteilungsfreie Rangkorrelation nach Spearman ausgewählt. Die Korrelation zwischen der Durchschnittlichkeit (d. h. Generation) eines Composites und seiner Attraktivität beträgt $r_s = .67$, $N = 31$, $p < .001$. Das bedeutet, je durchschnittlicher ein Gesicht war, desto attraktiver wurde es auch bewertet.

Zusätzlich wurde jedoch auch die Alternativhypothese überprüft, die lautet: „Ein gemorphes Gesicht ist nur so attraktiv wie die Originalgesichter, die darin enthalten sind.“ Hierzu wurden die theoretischen Attraktivitätswerte (zur Berechnung siehe Kapitel 2.3.1.2) mit den empirischen korreliert. Die Korrelation beträgt hier $r_s = .68$, $N = 31$, $p < .001$. Das bedeutet, dass auch diese Alternativhypothese empirisch bestätigt werden kann. Durchschnittsgesichter aus schönen Originalgesichtern blieben schön und Durchschnittsgesichter aus hässlichen Gesichtern blieben hässlich.

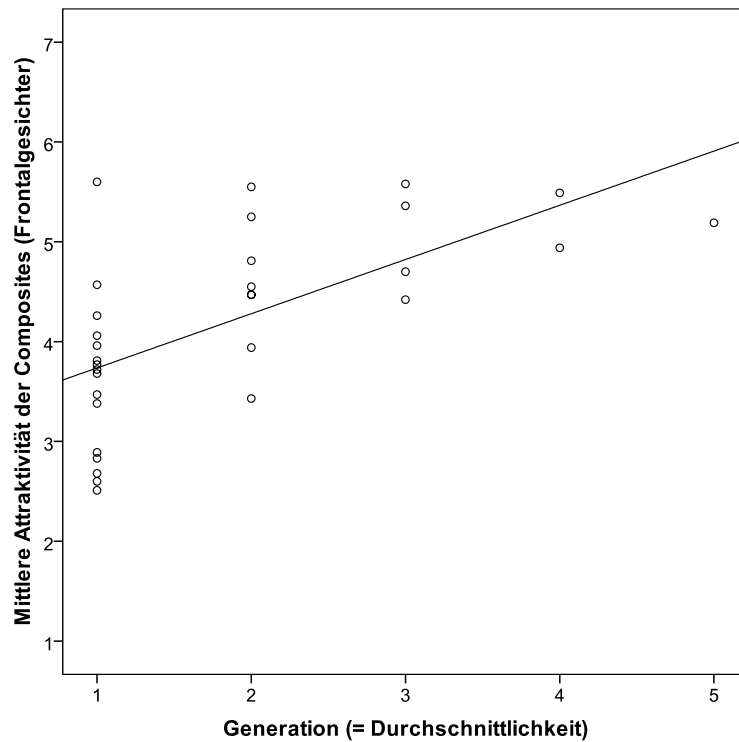


Abbildung 3: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Männergesichter in der Frontalansicht. Je durchschnittlicher ein Gesicht ist, desto attraktiver wird es tendenziell auch bewertet. Allerdings gibt es etliche Gesichter, die weniger durchschnittlich als das Gesamt-Durchschnittsgesicht aus allen 32 Gesichtern (Generation 5) sind und trotzdem attraktiver bewertet werden.

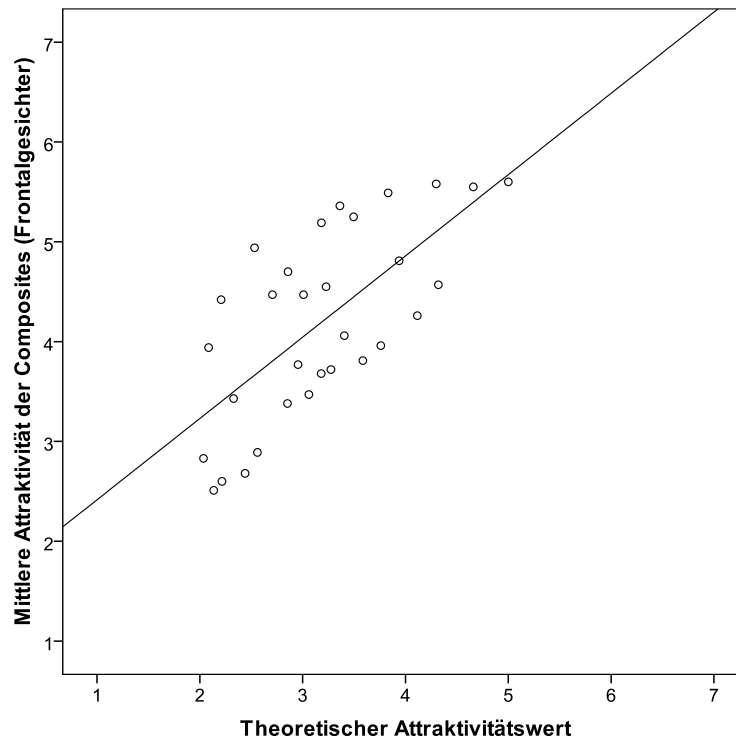


Abbildung 4: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Männergesichter in der Frontalansicht. Je attraktiver die Originalgesichter waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver wurden auch die daraus resultierenden gemorphten Gesichter beurteilt.

Daraus folgt, dass ein Composite, das aus den vier attraktivsten Originalen erzeugt wurde (m(29-32)), attraktiver sein sollte als eines aus den vier unattraktivsten Originalen (m(1-4)). Ein t -Test für gepaarte Stichproben zeigte, dass dies tatsächlich auch der Fall war ($t(52) = 9.26, p < .001$). Abbildung 5 zeigt diese beiden Beispiele.



Abbildung 5: „m(1-4)“, ein Composite aus den vier unattraktivsten männlichen Originalgesichtern (links) und „m(29-32)“, ein Composite aus den vier attraktivsten männlichen Originalgesichtern. Nach der Durchschnittshypothese müssten sie gleich attraktiv sein, da sie gleich durchschnittlich sind, tatsächlich ist jedoch der Durchschnitt aus den Schönsten weit attraktiver (mittlerer Attraktivitätswert von 5.55 ($SD = 1.01$) vs. 3.94 (1.05), $p < .001$).

Nach der Durchschnittshypothese müsste auch das absolute Durchschnittsgesicht „m(1-32)“, das alle männlichen Originalgesichter enthält, attraktiver sein als der Durchschnitt aus den acht attraktivsten Männern „m(25-32)“, da dieses ja deutlich weniger durchschnittlich ist. Das Gegenteil ist jedoch der Fall. Das Composite aus den attraktivsten acht Gesichtern ist attraktiver als das Gesamtdurchschnittsgesicht ($t(52) = 3.04$, $p = .004$).



Abbildung 6: „m(1-32)“, das Composite aus allen 32 männlichen Originalgesichtern (links) und „m(25-32)“, ein Composite aus den acht attraktivsten männlichen Originalgesichtern (rechts). Nach der Durchschnittshypothese müsste das Gesamt-Durchschnittsgesicht „m(1-32)“ attraktiver sein, da es durchschnittlicher ist als „m(25-32)“, tatsächlich ist es jedoch umgekehrt (mittlerer Attraktivitätswert von „m(1-32)“ = 5.19 ($SD = 1.24$) vs. „m(25-32)“ = 5.58 ($SD = 1.12$), $p = .004$).

Insgesamt erhielten von 31 Composites sechs einen höheren mittleren Attraktivitätswert als das Gesamt-Durchschnittsgesicht m(1-32). Bei vier davon war dieser Unterschied statistisch signifikant, darunter auch das Composite aus der „schöneren Hälfte“ aller fotografierten Personen, „m(16-32)“.

Zusammenfassend lässt sich feststellen, dass gemorphte Durchschnittsgesichter (Composites) umso attraktiver bewertet wurden, je durchschnittlicher sie waren, d. h. je mehr Originalgesichter in einem solchen Composite enthalten waren. Gleichzeitig wurden jedoch auch Belege für die Alternativhypothese gefunden, nach der ein Composite so attraktiv ist wie die Originalgesichter, aus denen es erzeugt wurde. Die Zusammenhänge waren mittelstark bis hoch und für beide Hypothesen gleich groß ($r_s = .67$ versus $r_s = .68$). Durch das systematische paarweise Morphen nach der Attraktivität der Originalgesichter konnten mehrere Composites generiert werden, die zwar aus weniger Gesichtern berechnet (also weniger durchschnittlich) waren, aber dennoch attraktiver bewertet wurden als das Durchschnittsgesicht aus *allen* 32 Männergesichtern.

2.4 Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von männlichen Profilgesichtern mit Texturinformation (Untersuchung *Durch 2*)

2.4.1 Methode

2.4.1.1 Voruntersuchung

Vorgehen bei der Voruntersuchung

Im zweiten Experiment wurde dieselbe Fragestellung wie im ersten Experiment untersucht, diesmal jedoch mit Gesichtern in der Profilansicht. Das Vorgehen war dasselbe. Zuerst erfolgte eine Voruntersuchung, bei der die Attraktivität der fotografierten Originalprofile (derselben Männergesichter wie in Experiment *Durch 1*) empirisch ermittelt wurde. Aus den Urteilen der Versuchspersonen wurde für jedes Gesicht ein Attraktivitätsmittelwert berechnet. Anschließend wurde eine Rangreihenfolge der Gesichter gebildet, beginnend vom unattraktivsten bis hin zum attraktivsten Gesicht. Das Morphen erfolgte wieder so, wie in Kapitel 2.3.1 beschrieben, indem über mehrere Generationen immer ähnlich attraktive Profile miteinander gekreuzt wurden. Für das Morphen der Profile wurde die Software *MorphMan 4.0* verwendet. Alle technischen Parameter und Arbeitsschritte beim Morphen waren genauso wie in Experiment *Durch 1*.

Beim *Vortest* gab es einen wichtigen Unterschied zu Experiment *Durch 1*. Die Beurteilung der Attraktivität der *Originalpersonen* erfolgte nicht durch die Präsentation der Profile als Farbfotos, sondern als sogenannte „Schattenprofile“. Abbildung 7 zeigt ein Beispiel.

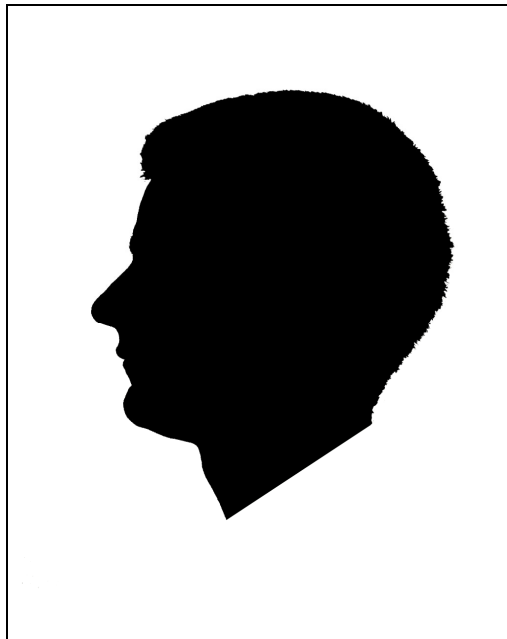


Abbildung 7: Ein Schattenprofil eines männlichen Originalgesichts, wie es beim Vortest als Stimulusmaterial verwendet wurde.

Der Grund hierfür war, dass mit Experiment *Durch 2* herausgefunden werden sollte, welche Merkmale für ein attraktives Profil charakteristisch sind. Die meiste Information steckt dabei in der Kontur der Profillinie, die auch für die Beurteilung der Ästhetik eines Gesichts für die Kieferorthopädie oder die Mund-Kiefer-Gesichtschirurgie entscheidend ist. Die Farbwerte (Textur) des Gesichts, also v. a. die Haut, sind dabei nur eine Störvariable, die durch die Präsentation der Gesichter als Schattenprofile ausgeblendet wurde.

Die Schattenprofile wurden den Versuchspersonen als Ausdrucke in einer Größe von 13 x 16 cm in variierender Reihenfolge präsentiert. Die Likert-Skala zur Attraktivitätsbewertung war dieselbe wie in Experiment *Durch 1*.

Versuchspersonen für die Voruntersuchung

An der Voruntersuchung zur Bewertung der Schattenprofile nahmen insgesamt 208 Versuchspersonen (107 Männer, 101 Frauen) teil. Die Altersspanne reichte von 14 bis 82 Jahren ($M = 32.3$, $SD = 15.5$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, wurde die Befragung an unterschiedlichen öffentlichen Orten (z. B. Einkaufszentrum, Fußgängerzone, Bahnhof) durchgeführt. Die Teilnahme war freiwillig und wurde nicht bezahlt. Die Versuchspersonen wurden gebeten, nur die Attraktivitäts-

tät des Gesichts zu beurteilen, und dabei die Frisur der gezeigten Personen zu ignorieren.

Ergebnisse der Voruntersuchung

Für die Auswertung des Vortests wurde für jedes Schattenprofil ein Attraktivitätsmittelwert über alle Versuchspersonen berechnet und die Gesichter wurden auf der Basis dieser Rating-Mittelwerte in eine Rangreihenfolge gebracht. Das unattraktivste Schattenprofil erhielt im Mittel den Wert 1.95 ($SD = 1.00$), das attraktivste den Wert 5.31 ($SD = 0.97$). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .99 (bei 208 Versuchspersonen) sehr hoch.

2.4.1.2 Versuchspersonen und Durchführung der Hauptuntersuchung

An der Hauptuntersuchung zur Bewertung der gemorphten Männerprofile mit Farbinformation nahmen insgesamt 48 Versuchspersonen (23 Männer, 25 Frauen) teil. Die Altersspanne reichte von 14 bis 69 Jahren ($M = 33.8$, $SD = 15.3$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, wurde die Befragung wieder an unterschiedlichen öffentlichen Orten (z. B. Einkaufszentrum, Fußgängerzone, Bahnhof) durchgeführt. Die Teilnahme war freiwillig und wurde nicht bezahlt. Die Versuchspersonen wurden gebeten, nur die Attraktivität des Gesichts zu beurteilen, und dabei die Frisur der gezeigten Personen zu ignorieren.

2.4.1.3 Stimulusmaterial der Hauptuntersuchung

Für die Erzeugung des Stimulusmaterials wurden die farbigen Profile (nicht die Schattenprofile!) verwendet. Das Vorgehen beim Morphing war dasselbe wie bei Experiment *Durch 1* (vgl. Kapitel 2.3.1). Dies gilt auch für die Berechnung der theoretischen Attraktivitätswerte (vgl. Kapitel 2.3.1.2). Für die Datenerhebung wurden die Composites mit einer Auflösung von 1280 x 1024 Pixeln auf Fotopapier der Größe 9 x 13 cm belichtet. Dies entspricht einer sehr hohen Auflösung von 250 dpi, wodurch selbst feine Details sichtbar blieben. Die Darbietung erfolgte in variierender Reihenfolge.

2.4.2 Ergebnisse

Analog zur Auswertung des Vortests wurde für jedes Composite ein Attraktivitätsmittelwert über alle Versuchspersonen berechnet. Das unattraktivste Composite erhielt im

Mittel den Wert 1.69 ($SD = 0.72$), das attraktivste den Wert 5.25 ($SD = 1.00$). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .98 (bei 48 Versuchspersonen) sehr hoch.

Zur Überprüfung der Durchschnittshypothese wurden dieselben Korrelationen (nach Spearman) wie in Experiment *Durch 1* bei den Gesichtern in der Frontalansicht gerechnet. Die Korrelation zwischen der Durchschnittlichkeit (d. h. Generation) eines Composites und seiner Attraktivität betrug $r_s = .60$, $N = 31$, $p < .001$. Das bedeutet, je durchschnittlicher ein Gesicht war, desto attraktiver wurde es auch bewertet.

Zusätzlich wurde jedoch auch wieder die Alternativhypothese (= Nullhypothese) überprüft, die lautet: „Ein gemorphtes Gesicht ist nur so attraktiv wie die Originalgesichter, die darin enthalten sind.“ Hierzu wurden die theoretischen Attraktivitätswerte der Schattenprofile (zur Berechnung siehe Kapitel 2.3.1.2) mit den empirischen Attraktivitätswerten der farbigen Composites korreliert. Die Korrelation betrug hier $r_s = .62$, $N = 31$, $p < .001$. Das bedeutet, dass auch diese Alternativhypothese empirisch bestätigt werden kann. Farbige Profilgesichter aus schönen Originalgesichtern (gemessen an der Attraktivität ihrer Schattenprofile) blieben schön und Durchschnittsgesichter aus hässlichen Gesichtern blieben hässlich. Damit können bei Gesichtern in der Profilansicht die Ergebnisse von Gesichtern in der Frontalansicht (Experiment *Durch 1*) bestätigt werden.

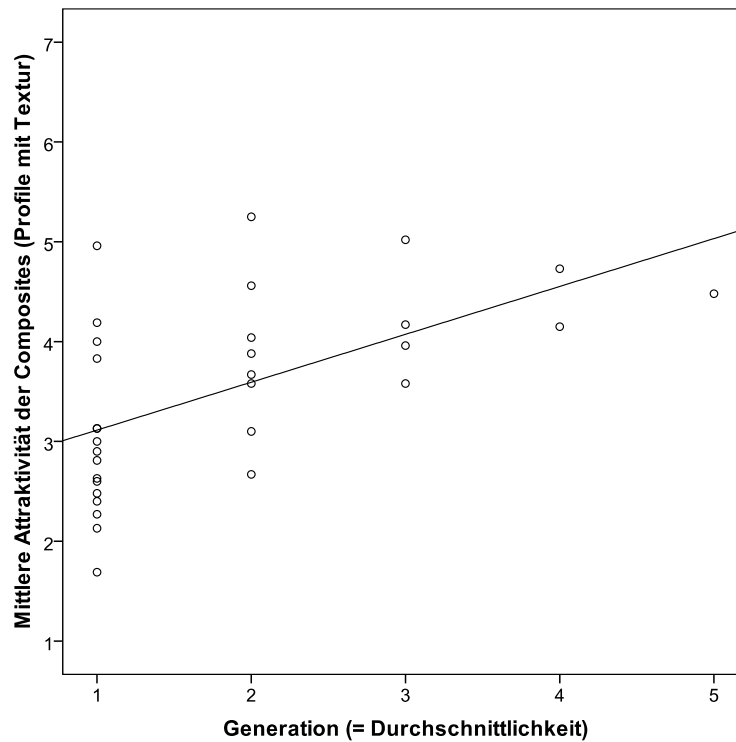


Abbildung 8: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Männerprofile mit Texturinformation. Je durchschnittlicher ein Profil ist, desto attraktiver wird es tendenziell auch bewertet. Allerdings gibt es etliche Gesichter, die weniger durchschnittlich als das Gesamt-Durchschnittsgesicht aus 32 Gesichtern (Generation 5) sind und trotzdem attraktiver bewertet werden.

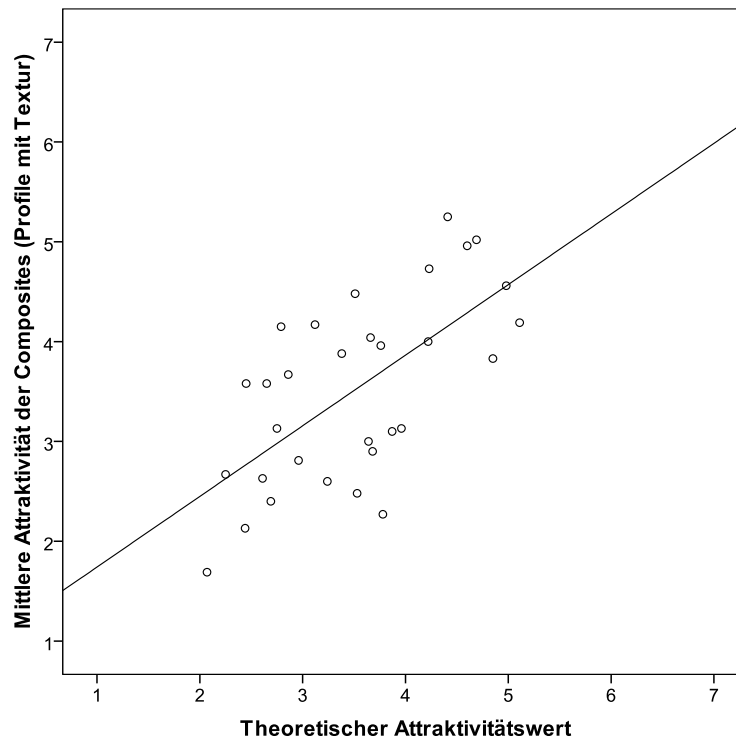


Abbildung 9: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Männerprofile mit Texturinformation. Je attraktiver die Originalgesichter waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver wurden auch die daraus resultierenden gemorphten Gesichter beurteilt.

Aus der Bestätigung der Alternativhypothese folgt, dass ein Composite, das aus den vier attraktivsten Originalen erzeugt wurde (m(29-32)) attraktiver sein sollte als eines aus den vier unattraktivsten Originalen (m(1-4)). Ein *t*-Test für gepaarte Stichproben zeigt, dass dies tatsächlich auch der Fall ist ($t(47) = 9.07, p < .001$). Abbildung 10 zeigt diese beiden Beispiele.



Abbildung 10: „m(1-4)“, ein Composite aus den vier unattraktivsten Originalprofilen (links) und „m(29-32)“, ein Composite aus den vier attraktivsten Originalprofilen. Nach der Durchschnittshypothese müssten sie gleich attraktiv sein, da sie gleich durchschnittlich sind, tatsächlich ist jedoch der Durchschnitt aus den Schönsten weit attraktiver (mittlerer Attraktivitätswert von 4.56 ($SD = 1.05$) vs. 2.67 ($SD = 1.02$), $p < .001$).

Nach der Durchschnittshypothese müsste auch das absolute Durchschnittsprofil „m(1-32)“, das alle Originalprofile enthält, attraktiver sein als der Durchschnitt aus den acht attraktivsten „m(25-32)“, da dieses ja deutlich weniger durchschnittlich ist. Das Gegenteil ist jedoch der Fall. Der Durchschnitt aus wenigen attraktiven Originalen wurde als attraktiver bewertet als das Gesamtdurchschnittsgesicht, $t(47) = 3.88$, $p < .001$.



Abbildung 11: „m(1-32)“, das Composite aus allen 32 Originalprofilen (links) und „m(25-32)“, ein Composite aus den acht attraktivsten Originalprofilen (rechts). Nach der Durchschnittshypothese müsste das absolute Durchschnittsgesicht „m(1-32)“ attraktiver sein, da es durchschnittlicher ist als „m(25-32)“, tatsächlich ist es jedoch umgekehrt (mittlerer Attraktivitätswert von „m(1-32)“ = 4.48 ($SD = 1.09$) vs. „m(25-32)“ = 5.02 ($SD = 0.89$), $p < .001$).

Zusammenfassend lässt sich feststellen, dass mit den männlichen Profilgesichtern mit farbiger Textur die Ergebnisse aus Experiment 1 bestätigt wurden. Gemorphte Durchschnittsgesichter (Composites) wurden umso attraktiver bewertet, je durchschnittlicher sie waren, d. h. je mehr Originalgesichter in einem solchen Composite enthalten waren. Gleichzeitig wurden jedoch auch Belege für die Alternativhypothese gefunden, nach der ein Composite so attraktiv ist wie die Originalgesichter, aus denen es erzeugt wurde. Die Zusammenhänge waren mittelstark bis hoch und für beide Hypothesen gleich groß ($r_s = .60$ versus $r_s = .62$). Beide Korrelationen waren jedoch nicht ganz so hoch wie bei den Frontalgesichtern (dort: $r_s = .67$ versus $r_s = .68$). Durch das systematische paarweise Morphen nach der Attraktivität der Originalgesichter konnten mehrere Composites generiert werden, die zwar aus weniger Gesichtsprofilen berechnet (also weniger durchschnittlich) waren, aber dennoch attraktiver bewertet wurden als das Durchschnittsgesicht aus *allen* 32 Männerprofilen.

2.5 Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von männlichen Profilgesichtern ohne Texturinformation (Schattenprofile, Untersuchung *Durch 3*)

2.5.1 Methode

Bei Experiment *Durch 3* handelt es sich um eine Replikation von Experiment 2, jedoch mit einem entscheidenden Unterschied. Die gemorphten Profile wurden nun nicht mehr als farbige Fotos, sondern als Schattenprofile präsentiert. Durch den Vergleich der Attraktivitätsbewertungen der farbigen Profile und der Schattenprofile sollte herausgefunden werden, wie groß der Effekt der Farbinformation (d. h. insbesondere der Farbe und Textur der Haut) der Gesichter ist. Denn durch die Präsentation als Schattenprofile wird die Variable „Textur“ komplett ausgeschaltet, da alle Gesichter dieselbe Textur aufweisen (nämlich gleichmäßig schwarz). Erhalten bleiben hingegen die Unterschiede hinsichtlich der Gesichtsproportionen eines Profils. Wenn sich mit den gemorphten Schattenprofilen die hohen Korrelationen zwischen Durchschnittlichkeit und Attraktivität aus den ersten beiden Experimenten bestätigen lassen, dann ist der Effekt eindeutig darauf zurückzuführen, dass durchschnittliche *Gesichtsproportionen* attraktiv machen. Korrelieren diese beiden Variablen jedoch nicht miteinander, dann müssen die Ergebnisse aus Experiment *Durch 2* auf die veränderte Textur der Composites zurückzuführen sein.

Versuchspersonen und Durchführung

An der Untersuchung zur Bewertung der gemorphten Schattenprofile nahmen insgesamt 198 Versuchspersonen (94 Männer, 104 Frauen) teil. Die Altersspanne reichte von 14 bis 79 Jahren ($M = 34.4$, $SD = 15.4$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, wurde die Befragung wieder an unterschiedlichen öffentlichen Orten (z. B. Einkaufszentrum, Fußgängerzone, Bahnhof) durchgeführt. Die Teilnahme war freiwillig und wurde nicht bezahlt. Die Versuchspersonen wurden gebeten, nur die Attraktivität des Gesichts zu beurteilen und dabei die Frisur der gezeigten Personen zu ignorieren.

Stimulusmaterial

Das Stimulusmaterial wurde erzeugt, indem die farbigen Composites aus Experiment 2 mit dem Programm *Adobe Photoshop* so bearbeitet wurden, dass das gesamte Profil gleichmäßig schwarz und der Rest des Bildes weiß eingefärbt wurde. Die Darbietung der Bilder war genau so wie bei der Voruntersuchung von Experiment *Durch 2*.

2.5.2 Ergebnisse

Analog zur Auswertung der beiden vorherigen Untersuchungen wurde für jedes Composite ein Attraktivitätsmittelwert über alle Versuchspersonen berechnet. Das unattraktivste Composite erhielt im Mittel den Wert 2.09 ($SD = 1.13$), das attraktivste den Wert 5.06 ($SD = 1.06$). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .99 (bei 198 Versuchspersonen) sehr hoch.

Zur Überprüfung der Durchschnittshypothese wurden wieder dieselben Korrelationen wie in den beiden vorangegangenen Experimenten gerechnet. Auch bei den Schattenprofilen ließ sich der Zusammenhang zwischen Durchschnittlichkeit und Attraktivität bestätigen, allerdings war nun der Effekt weitaus geringer als bei den farbigen Composites ($r_s = .38$, $N = 31$, $p = .037$). Der Zusammenhang zwischen „theoretischer Attraktivität“ und „empirischer Attraktivität“ war hingegen sehr stark ($r_s = .86$, $N = 31$, $p < .001$), das bedeutet, dass bei Schattenprofilen ein Composite gerade so attraktiv ist wie die Original-Schattenprofile, aus denen es berechnet wurde.

Die folgenden Abbildungen zeigen für die männlichen Schattenprofile grafisch den Zusammenhang zwischen Attraktivität und Durchschnittlichkeit (= Durchschnittshypothese) bzw. theoretischen Attraktivitätswert (= Alternativhypothese).

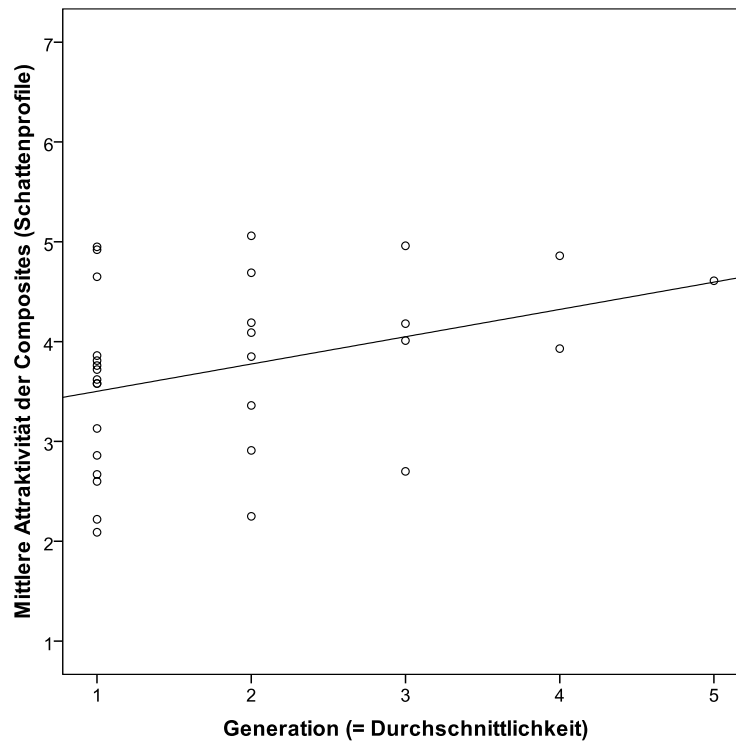


Abbildung 12: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Männerprofile ohne Texturinformation (= Schattenprofile). Je durchschnittlicher ein Profil ist, desto attraktiver wird es tendenziell auch bewertet. Allerdings gibt es etliche Schattenprofile, die weniger durchschnittlich als das Gesamt-Durchschnittsgesicht aus 32 Schattenprofilen (Generation 5) sind und trotzdem attraktiver bewertet werden. Der Zusammenhang ist jedoch eher gering ($r_s = .38$, $p = .037$).

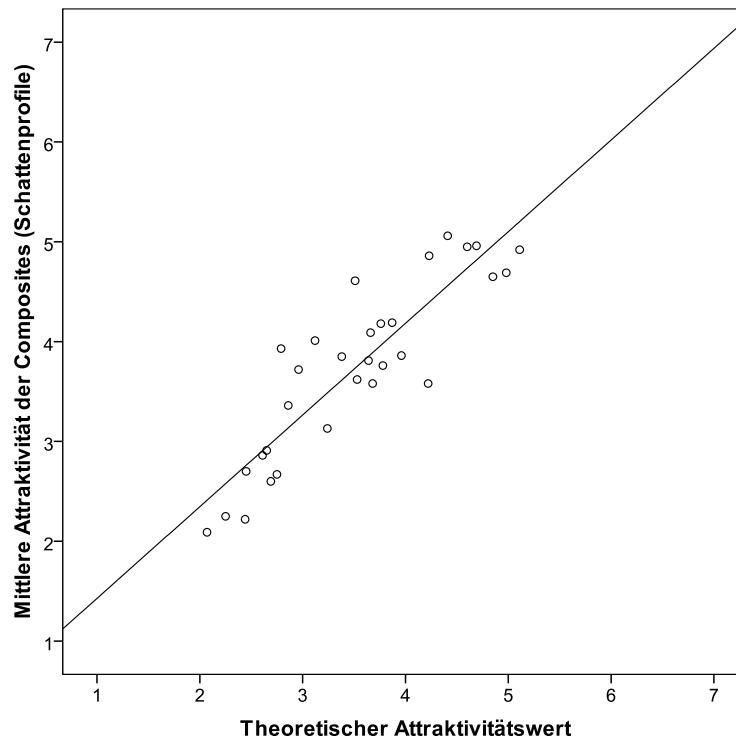


Abbildung 13: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Männerprofile ohne Texturinformation (= Schattenprofile). Je attraktiver die originalen Schattenprofile waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver wurden auch die daraus resultierenden gemorphten Schattenprofile beurteilt. Der Zusammenhang ist sehr hoch ($r_s = .86$, $p < .001$).

Insgesamt wurden sieben Schattenprofile attraktiver bewertet als das absolute Durchschnittsschattenprofil, das nach der Durchschnittshypothese das attraktivste sein müsste. Bei fünf Profilen war der Unterschied statistisch signifikant.

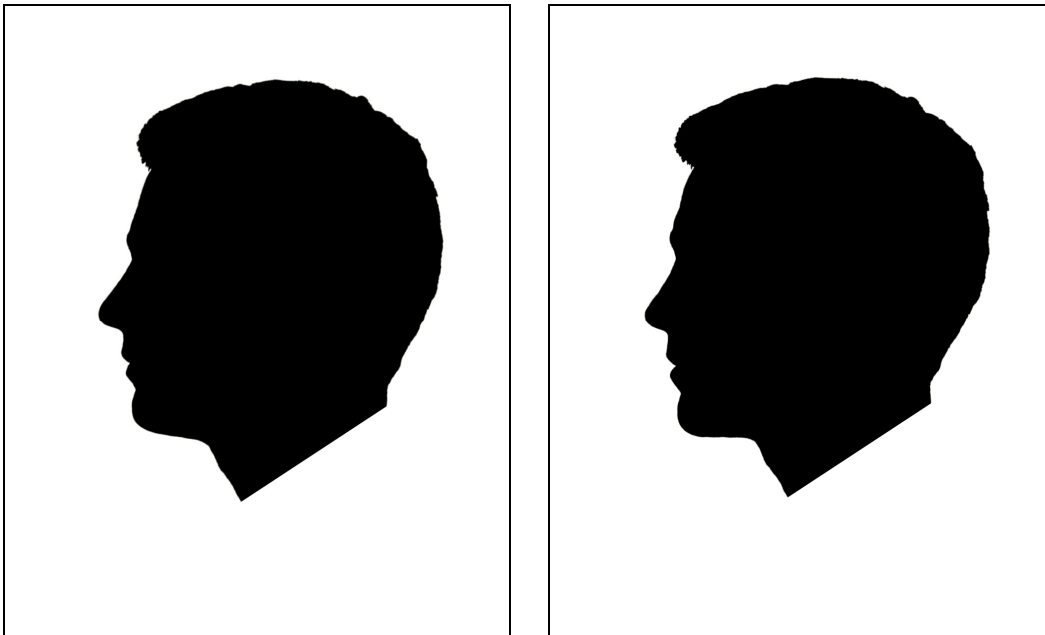


Abbildung 14: „m(1-32)“, das Composite aus allen 32 Originalschattenprofilen (links) und „m(25-32)“, ein Composite aus den acht attraktivsten Originalschattenprofilen (rechts). Nach der Durchschnittshypothese müsste das absolute Durchschnittsgesicht „m(1-32)“ attraktiver sein, da es durchschnittlicher ist als „m(25-32)“, tatsächlich ist es jedoch auch hier wieder umgekehrt (mittlerer Attraktivitätswert von „m(1-32)“ = 4.61 ($SD = 1.02$) vs. „m(25-32)“ = 4.96 ($SD = 1.11$), ($t(197) = 4.02$, $p < .001$).

Zusammenfassend lässt sich festhalten, dass das aus den ersten beiden Experimenten erzielte Ergebnis, dass ein Gesicht umso attraktiver ist, je durchschnittlicher es ist, sich mit Schattenprofilen zwar gerade noch replizieren ließ, der Zusammenhang war jedoch nur noch schwach ($r_s = .38$, $N = 31$, $p = .037$). Der Zusammenhang zwischen „theoretischer Attraktivität“ und „empirischer Attraktivität“ war hingegen sehr stark ($r_s = .86$, $p < .001$), das bedeutet, dass bei Schattenprofilen ein Composite gerade so attraktiv bewertet wurde wie die Original-Schattenprofile, aus denen es berechnet wurde. Trotz geringerer Durchschnittlichkeit wurden mehrere Composites attraktiver bewertet als das Durchschnittsprofil aus allen 32 Gesichtern. Die auf $r_s = .38$ geschrumpfte Höhe der Korrelation zwischen Durchschnittlichkeit und Attraktivität bei den Schattenprofilen lässt den Schluss zu, dass bei den gemorphten Profilen *mit* Textur (Experiment *Durch 2*, $r_s = .60$) ein erheblicher Teil des Effekts auf die makellosere Haut der Composites und nicht auf die durchschnittlichen Gesichtsproportionen zurückzuführen ist, da die Textur der einzige Unterschied zwischen den Stimuli der beiden Experimente war.




2.6 Zusammenfassende Datenanalyse der Experimente *Durch 1 bis 3*

In einem weiteren Schritt wurden die Daten der Experimente *Durch 1 bis 3* varianzanalytisch ausgewertet. Dafür wurden für alle drei Experimente drei ANOVAs durchgeführt, bei denen die Attraktivitätsratings, die jeweils die abhängige Variable waren, und der Grad der „Durchschnittlichkeit“ (für drei Generationen) sowie der „Attraktivität der Originalgesichter“ (theoretischer Attraktivitätswert) die jeweiligen unabhängigen Faktoren waren.

Für die Variable „Grad der Durchschnittlichkeit“ wurden die Stimuli nach drei Faktorstufen unterschieden (*2-face-*, *4-face-* and *8-face-composites*). Für die Variable „Attraktivität der Originalgesichter“ wurden sie in vier Kategorien eingeteilt. In der ersten Kategorie (= 1. Quartil) waren alle Composites, die aus den acht unattraktivsten Originalgesichtern erzeugt wurden (Ränge 1-8), in der zweiten Kategorie (= 2. Quartil) alle Composites, die aus den zweitunattraktivsten Originalen (Ränge 9-16) erzeugt wurden usw. Das vierte Quartil enthielt die Composites aus den acht schönsten Gesichtern. In jeder der vier Klassen gab es damit die gleiche Anzahl an Composites verschiedener Durchschnittlichkeit (jeweils vier *2-face-composites*, zwei *4-face-composites* und ein *8-face-composite*). Drei Composites konnten prinzipbedingt keiner Klasse zugeordnet werden (das Composite aus *allen* Gesichtern, sowie die beiden Composites aus der unattraktiven und der attraktiven Hälfte der Originalgesichter) und wurden bei den ANOVAs nicht berücksichtigt.

Für jedes der drei Experimente wurde eine zweifaktorielle ANOVA gerechnet mit dem Faktor „Durchschnittlichkeit“ (3 Stufen, d. h. Generation 1 bis 3) und dem Faktor „Attraktivität der Originalgesichter“ (4 Stufen, d. h. Quartil 1 bis 4). Die Ergebnisse sind in Tabelle 1 zusammengefasst.

Tabelle 1: Die Ergebnisse der drei ANOVAs mit Tests der Haupteffekte und Interaktionseffekte bei den drei Experimenten mit männlichen Durchschnittsgesichtern

	<i>Durch 1</i>	<i>Durch 2</i>	<i>Durch 3</i>
	Frontalgesichter	Profilgesichter mit Texturinformation	Schattenprofile ohne Texturinformation
			
Durchschnittlichkeit (= Generation) (<i>df</i> = 2)	$F = 26.011$ $p < .001$ $\eta^2 = .765$	$F = 17.041$ $p < .001$ $\eta^2 = .681$	$F = 2.662$ $p = .100$ $\eta^2 = .250$
Attraktivität der Originalgesichter (= Quartil) (<i>df</i> = 3)	$F = 13.236$ $p < .001$ $\eta^2 = .713$	$F = 14.962$ $p < .001$ $\eta^2 = .737$	$F = 25.139$ $p < .001$ $\eta^2 = .825$
Interaktion (Durchschnittlichkeit * Attraktivität der Originalgesichter) (<i>df</i> = 6)	$F = 0.425$ $p = .851$ $\eta^2 = .138$	$F = 0.226$ $p = .962$ $\eta^2 = .078$	$F = 0.125$ $p = .992$ $\eta^2 = .045$

Die Ergebnisse bestätigten die Ergebnisse der Rangkorrelationen nach Spearman: Benutzt man für das Experiment Frontalgesichter oder Profilgesichter mit Textur, dann war der Effekt durch die Faktoren „Durchschnittlichkeit“ (Frontalgesichter: $\eta^2 = .77$, Profilgesichter: $\eta^2 = .68$) und „Attraktivität der Originalgesichter“ (Frontalgesichter: $\eta^2 = .71$, Profilgesichter: $\eta^2 = .74$) in etwa gleich groß und in allen Fällen hoch signifikant. Beim dritten

Experiment, bei dem die Profilgesichter keine Textur mehr enthielten, wurde hingegen der kleine Haupteffekt der Durchschnittlichkeit (= Generation) ($\eta^2 = .25$) nicht mehr signifikant ($p = .100$); entscheidend für die Attraktivität eines Composites war hier vielmehr die Attraktivität der ihnen zugrunde liegenden Originalgesichter ($\eta^2 = .83$, $p < .001$). Bei allen drei Experimenten gab es keinerlei Interaktionseffekte zwischen den beiden Faktoren „Durchschnittlichkeit“ und „Attraktivität der Originalgesichter“.

Abbildung 15 illustriert die Haupteffekte der Durchschnittlichkeit und Attraktivität der Originalgesichter für Frontalgesichter, Abbildung 16 dasselbe für Profilgesichter mit Textur.

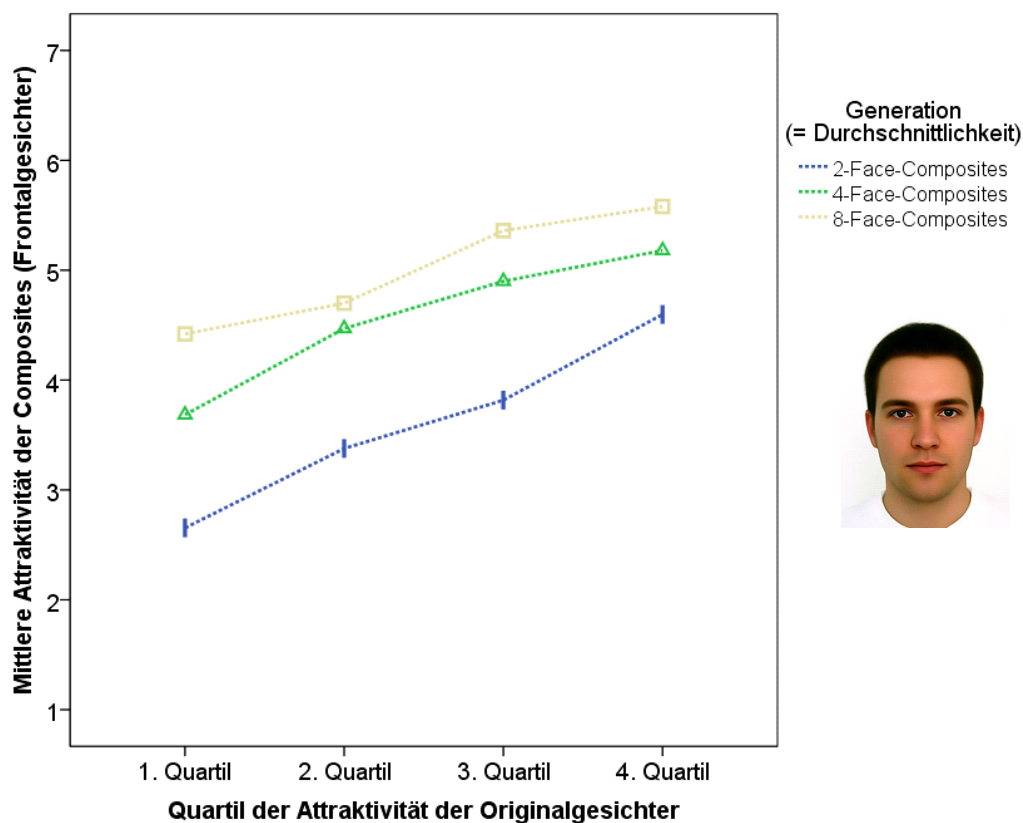


Abbildung 15: Die mittlere Attraktivitätsbewertung von gemorphten Männergesichtern in der Frontalansicht in Abhängigkeit von der ursprünglichen Attraktivität der Originalgesichter (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). Die beiden Haupteffekte sind klar erkennbar.

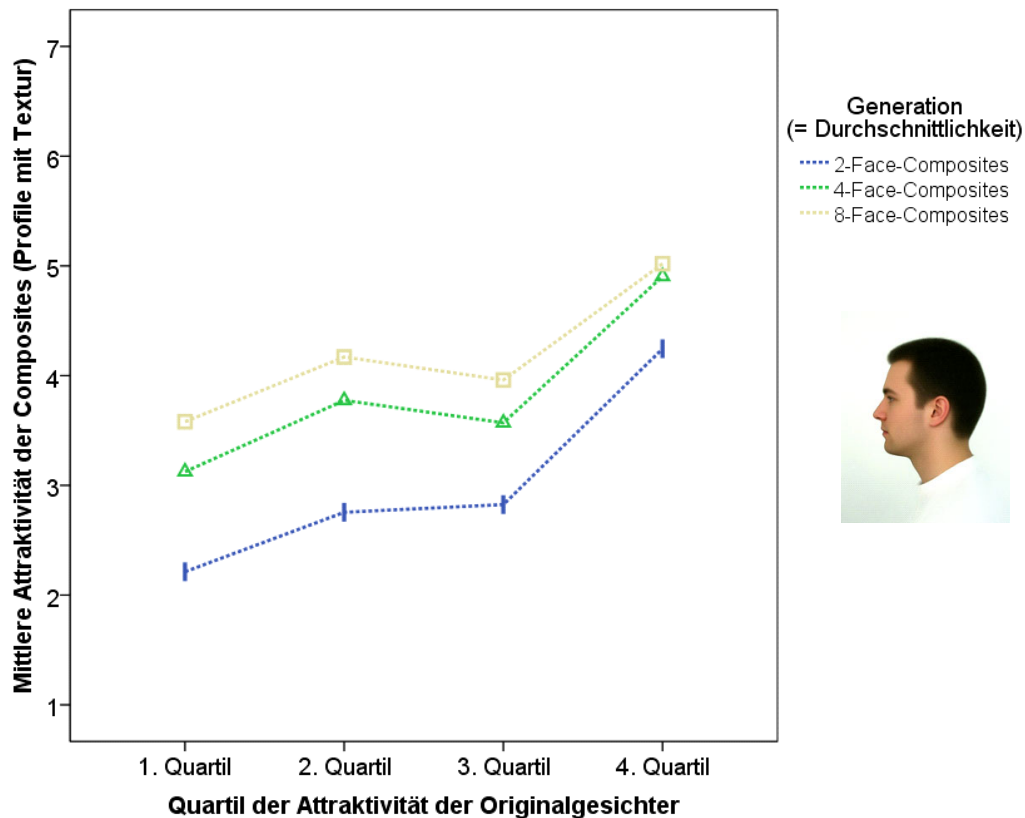


Abbildung 16: Die mittlere Attraktivitätsbewertung von gemorphten Männerprofilen mit Textur in Abhängigkeit von der ursprünglichen Attraktivität der Originalprofile (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation).

Bei Abbildung 16 fällt auf, dass *2-face-* und *4-face-composites*, die aus dem 3. Attraktivitätsquartil erstellt wurden, etwas unattraktiver bewertet werden als solche aus dem 2. Quartil. Der Grund dafür ist nicht bekannt. Der Mittelwertsunterschied zwischen Gesichtern des 2. und 3. Quartils wird in einem Post-hoc-Test jedoch auch nicht signifikant.

In Abbildung 17 erkennt man gut, dass bei den Schattenprofilen der Haupteffekt der Attraktivität der Originalgesichter genauso vorhanden ist wie bei den anderen beiden Stimulusarten, der Haupteffekt der Durchschnittlichkeit jedoch nur noch marginal (und nicht signifikant) ist.

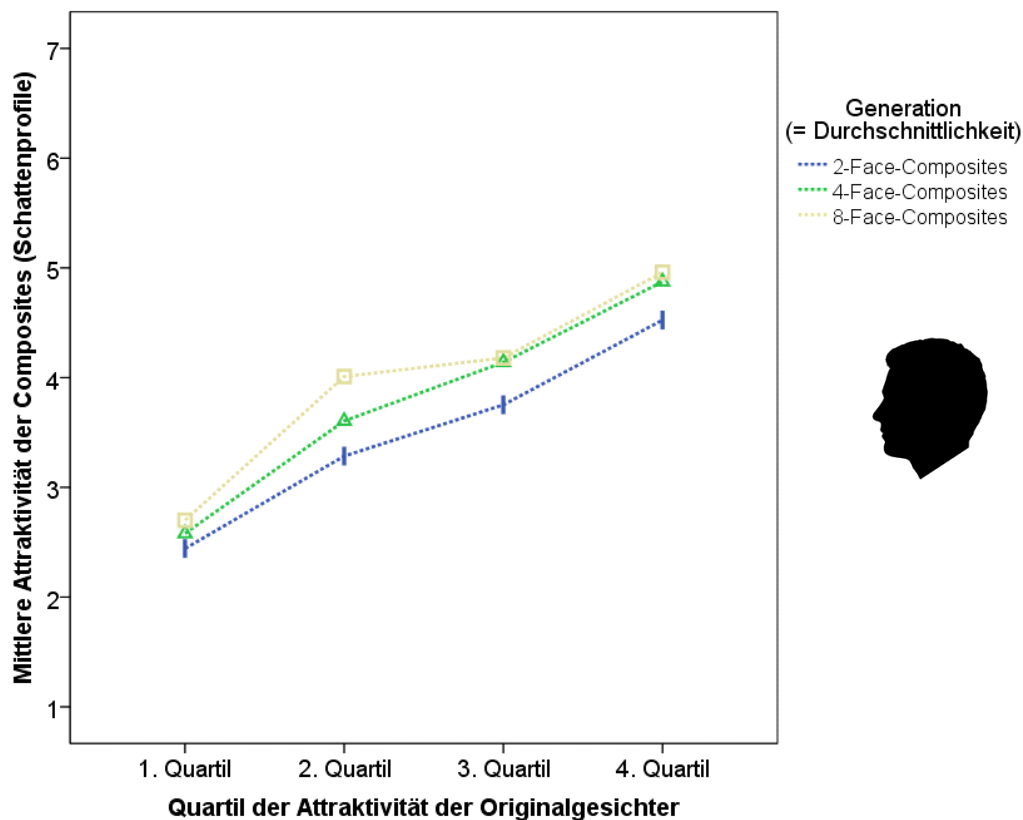


Abbildung 17: Die mittlere Attraktivitätsbewertung von gemorphten Männerprofilen ohne Textur (= Schattenprofile) in Abhängigkeit von der ursprünglichen Attraktivität der Originalprofile (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). Hier besteht nur noch ein Haupteffekt der Attraktivität der Originalgesichter. Der Effekt der Durchschnittlichkeit wird nicht mehr signifikant.

Um den Effekt der Stimulusart zu untersuchen, wurde eine dreifaktorielle ANOVA durchgeführt, bei der die Stimulusart (Frontalgesichter, Profilgesichter mit Textur und Schattenprofile ohne Textur) als dritter Faktor hinzugenommen wurde. Es zeigten sich zusätzlich zu den signifikanten Haupteffekten von „Durchschnittlichkeit“ und „Attraktivität der Originalgesichter“ ein signifikanter Haupteffekt des Stimulustyps sowie eine signifikante Interaktion zwischen Durchschnittlichkeit und Stimulusart. Diese Interaktion ist auf die abweichenden Ergebnisse der Schattenprofil-Stimuli zurückzuführen, wie Abbildung 18 (S. 41) illustriert. Das Ergebnis bestätigt damit den obigen Befund, dass die Durchschnittlichkeit zwar bei den Profilen mit Textur etwas mit Attraktivität zu tun hat, jedoch nicht, wenn man Schattenprofile verwendet. Ob also Durchschnittlichkeit einen Effekt auf Attraktivität hat, hängt davon ab, welches Stimulusmaterial man für eine Untersuchung verwendet.

Tabelle 2: Ergebnisse der dreifaktoriellen ANOVA. Einfluss von Durchschnittlichkeit, Attraktivität der Originalgesichter und Stimulusart auf die Attraktivität eines männlichen Composites

	<i>df</i>	<i>F</i>	<i>p</i>	η^2
Attraktivität der Originalgesichter (= Quartil)	3	49.602	<.001	.756
Durchschnittlichkeit (= Generation)	2	39.320	<.001	.621
Stimulusart	2	17.943	<.001	.428
Quartil * Generation (Interaktion)	6	0.342	.911	.041
Quartil * Stimulusart (Interaktion)	6	1.789	.121	.183
Generation * Stimulusart (Interaktion)	4	3.164	.022	.209
Quartil * Generation * Stimulusart (Interaktion)	12	.214	.997	.051

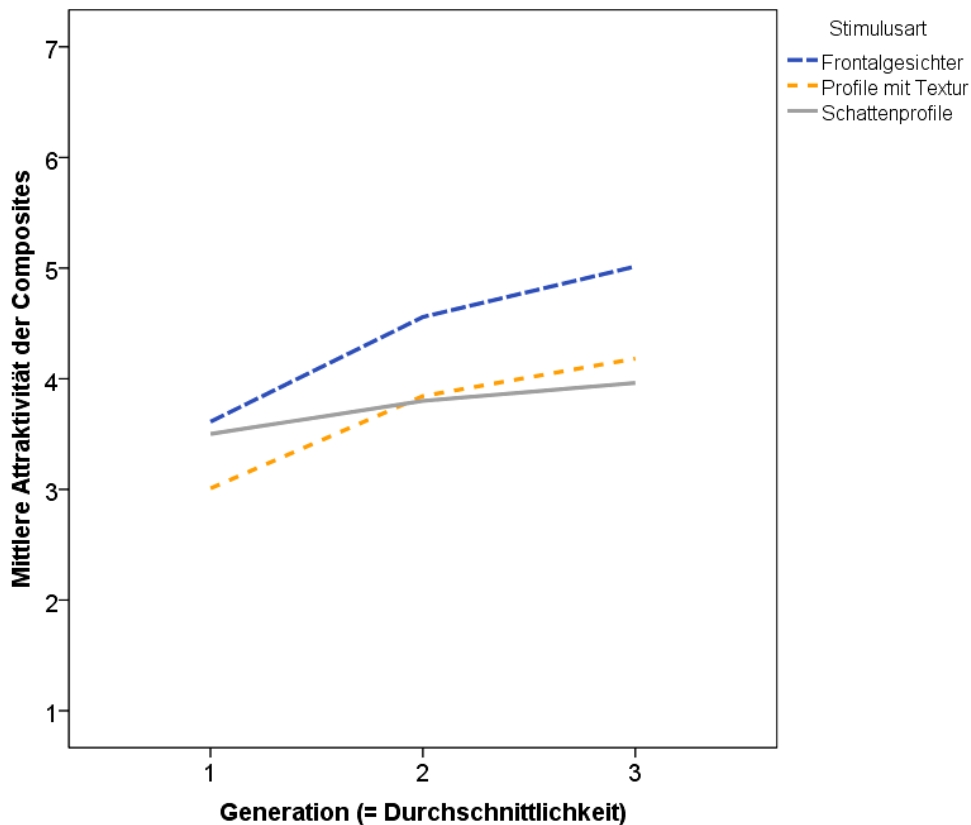


Abbildung 18: Die mittlere Attraktivität der gemorphten Männergesichter in Abhängigkeit von ihrer Durchschnittlichkeit und der Art der Stimuli. Der Interaktionseffekt zwischen Durchschnittlichkeit und Stimulusart wird durch die Schattenprofile ohne Textur verursacht, da bei ihnen die Attraktivität fast unabhängig von der Durchschnittlichkeit ist.

Daraus folgt, dass die Attraktivitätszugewinne bei den gemorphten Composites nicht nur auf die höhere Durchschnittlichkeit – genauer gesagt, die durchschnittlichen Gesichtsproportionen – zurückzuführen sind, denn sonst müsste bei den Schattenprofilen als Stimulusmaterial ein ähnliches Ergebnis herauskommen. Da die Textur der einzige Unterschied zwischen Schattenprofilen und Profilgesichtern mit Textur ist, muss ein wesentlicher Grund für die Attraktivitätszunahme gemorphter Composites in den Farbwerten der Textur liegen. Die naheliegende Erklärung ist, dass die bekannten und unvermeidlichen Morphing-Artefakte in Form der glatteren und makelloseren Haut die Ursache dieses Ergebnisses sind. Diese Konfundierung zwischen makelloser Haut und Durchschnittlichkeit ist bei den beiden Stimulusarten mit Textur gegeben, bei den Schattenprofilen jedoch

nicht; aus diesem Grund fällt dort der Effekt der Durchschnittlichkeit so gering aus und wird auch nicht mehr signifikant.

Zusammenfassend kann festgestellt werden, dass die varianzanalytische Auswertung der Daten aus den drei Experimenten *Durch 1* bis *Durch 3* den klaren Nachweis erbracht hat, dass die Attraktivitätssteigerung von Composites *mit* Textur (frontal oder im Profil) zu einem Großteil auf Morphing-Artefakte der makelloseren Haut zurückzuführen ist. Dies belegt der signifikante Interaktionseffekt ($p = .022$) zwischen „Durchschnittlichkeit“ und „Stimulusart“ (vgl. Tabelle 2). Waren bei den Stimuli keine Texturunterschiede mehr vorhanden (Schattenprofile in Experiment *Durch 3*), dann wurde der Effekt der Durchschnittlichkeit statistisch nicht mehr signifikant ($p = .100$), da der Effekt von durchschnittlichen Gesichtsproportionen allein zu schwach war.

2.7 Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von weiblichen Frontalgesichtern (Untersuchung *Durch 4*)

2.7.1 Methode

2.7.1.1 Voruntersuchung

Das Fotografieren der weiblichen Originalgesichter

Das methodische Vorgehen für die Untersuchung des Effekts der Durchschnittlichkeit auf die Attraktivität von Frauengesichtern war analog zum Vorgehen bei Männergesichtern. Der wichtigste Unterschied war, dass mehr als doppelt so viele Frauengesichter fotografiert wurden. Die Berechnung des weiblichen frontalen Durchschnittsgesichts erfolgte somit aus 64 Originalgesichtern, anstatt wie bei den Männern nur aus 32. Dies verdoppelte den ohnehin schon sehr hohen Zeitaufwand für das Morphen der Gesichter, doch dafür repräsentiert das so gewonnene Durchschnittsgesicht noch besser die durchschnittlichen Gesichtsproportionen von jungen Frauen.

Für die Untersuchung wurden 78 deutsche Frauen („kaukasischer Typ“, 71 Studentinnen der Universität Regensburg verschiedener Fachbereiche und 7 Models) im Alter von 17 bis 30 Jahren ($M = 23.3$, $SD = 2.3$) fotografiert. Die Teilnahme war freiwillig und wurde nicht bezahlt. Von jeder Person wurden mit einer Digitalkamera (Auflösung 1280 x 1024

Pixel) zwei Fotos gemacht, eines in der Frontalansicht und eines in der Profilansicht. Bei den Aufnahmen wurde auf möglichst standardisierte Bedingungen Wert gelegt. Alle Personen wurden unter konstanten und gleichmäßigen Lichtbedingungen aus einer konstanten Entfernung vor einem weißen Hintergrund abgelichtet. Alle trugen ein weißes T-Shirt, mussten Brillen und Piercings abnehmen und die Haare nach hinten kämmen bzw. mit Haarnadeln nach hinten stecken, so dass die Stirn nicht von Haaren bedeckt war. Sie wurden instruiert, nicht zu lächeln, sondern mit einem neutralen Gesichtsausdruck zu schauen.

Attraktivitätsbewertungen der Frauengesichter in der Voruntersuchung

Die Frontalansichten aller 78 fotografierten Frauen wurden in einer Größe von 320 x 400 Pixeln Versuchspersonen zur Beurteilung mit einem Computerprogramm in variierender Reihenfolge präsentiert. Jede Versuchsperson beurteilte jedes Gesicht auf einer Likert-Skala von 1 (= sehr unattraktiv) bis 7 (= sehr attraktiv) hinsichtlich seiner Attraktivität.

Versuchspersonen für die Voruntersuchung

An der Voruntersuchung zur Bewertung der Fotos in der Frontalansicht nahmen insgesamt 39 Versuchspersonen (24 Männer, 15 Frauen) teil. Die Altersspanne reichte von 13 bis 62 Jahren ($M = 33.7$, $SD = 12.9$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, wurde die Befragung an einem Samstag in einem Einkaufszentrum durchgeführt. Die Teilnahme war freiwillig und wurde nicht bezahlt. Die Versuchspersonen wurden gebeten, nur die Attraktivität des Gesichts zu beurteilen, und dabei die Frisur der gezeigten Personen zu ignorieren.

Ergebnisse der Attraktivitätsbewertungen der Frauengesichter in der Voruntersuchung

Für die Auswertung des Vortests wurde für jedes Gesicht ein Attraktivitätsmittelwert über alle Versuchspersonen berechnet und die Gesichter wurden auf der Basis dieser Rating-Mittelwerte in eine Rangreihenfolge gebracht. Das unattraktivste Gesicht erhielt im Mittel den Wert 1.85 ($SD = 0.84$), das attraktivste den Wert 5.38 ($SD = 1.27$). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .96 (bei 39 Versuchspersonen) sehr hoch. Dies rechtfertigte es, für jedes Gesicht die Urteile verschiedener Beurteiler zu einem Gesamtwert zusammenzufassen.

2.7.1.2 Versuchspersonen und Durchführung der Hauptuntersuchung

An der Hauptuntersuchung zur Bewertung der gemorphten Frauenbilder in der Frontalan-sicht nahmen insgesamt 63 Versuchspersonen (37 Männer, 26 Frauen) teil. Die Alters-spanne reichte von 16 bis 80 Jahren ($M = 32.0$, $SD = 11.5$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, wurde die Befragung wie auch beim Vortest an einem Samstag in einem Einkaufszentrum durchgeführt. Die Teilnahme war freiwillig und wurde nicht bezahlt. Die Versuchspersonen wurden gebeten, nur die Attrak-tivität des Gesichts zu beurteilen, und dabei die Frisur der gezeigten Personen zu ignorie-ren. Die Darbietung der Bilder erfolgte in derselben Art und Weise wie beim Vortest.

2.7.1.3 Stimulusmaterial der Hauptuntersuchung

Nachdem die Attraktivität der weiblichen Originalgesichter empirisch ermittelt und alle Gesichter nach ihrer Attraktivität in eine Rangreihenfolge gebracht waren, wurden mit dem Morphing-Programm *Morpher 3.0* aus den Originalgesichtern neue Gesichter be-rechnet. Hierzu wurden je zwei Gesichter miteinander gekreuzt und zu einem neuen, virtuellen Gesicht gemorpht. Die Produkte dieses Morphing-Prozesses wurden dann wiederum miteinander gekreuzt, bis am Ende ein Gesicht resultierte, das 64 Originalge-sichter zu gleichen Anteilen enthielt. Von den ursprünglich 78 Originalgesichtern wurden 14 nicht verwendet. Die Entscheidung, welche Gesichter nicht verwendet wurden, erfolg-ten dabei nach zwei Kriterien: (1) Gesichter, bei denen durch den Morphprozess starke Bildverarbeitungsartefakte zu erwarten waren (z. B. aufgrund von Haaren, die Teile der Stirn verdeckten, Locken oder sehr voluminösen Frisuren, die trotz nach hinten gebunde-ner Haare teilweise sichtbar waren) wurden aussortiert. (2) Bei Gesichtern mit sehr ähnli-chen oder gleichen Attraktivitätswerten (v. a. im unteren Mittelfeld) wurden ebenfalls eini-ge aussortiert.

Das gesamte Vorgehen war grundsätzlich identisch mit dem bei den Männergesichtern (vgl. Kapitel 2.3). Der wesentliche Unterschied bestand darin, dass für die Erzeugung des weiblichen Durchschnittsgesichts doppelt so viele Originalgesichter verwendet wurden als für die Erzeugung des männlichen Durchschnittsgesichts (64 statt 32). Dadurch war der Arbeitsaufwand doppelt so hoch. Die Bestimmung der theoretischen Attraktivitätswerte, d. h. der jeweilige Mittelwert aus den Attraktivitätswerten der jeweiligen Originalgesichter, aus denen ein Composite gemorpht wurde, wurde auf dieselbe Weise berechnet wie bei den Männergesichtern (vgl. Kapitel 2.3.1.2).

2.7.2 Ergebnisse

Analog zur Auswertung des Vortests wurde für jedes Composite ein Attraktivitätsmittelwert über alle Versuchspersonen berechnet. Das unattraktivste Composite erhielt im Mittel den Wert 1.87 ($SD = 0.85$), das attraktivste den Wert 6.33 ($SD = 0.72$). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .99 (bei 63 Versuchspersonen) ebenso hoch wie beim Vortest. Insgesamt wurden die gemorphten Frauengesichter besser bewertet als die Original-Frauengesichter. Der mittlere Attraktivitätswert der gemorphten Gesichter lag bei 4.29 ($SD = 1.07$), der der Originalgesichter nur bei 3.58 ($SD = 1.26$), d. h. das Attraktivitätsurteil lag bei den Composites im Durchschnitt um 0.7 Skalenstufen höher.

Die Durchschnittshypothese von Langlois & Roggman (1990) („je durchschnittlicher, desto attraktiver“) wurde analog zur Auswertung bei den Männergesichtern durch eine Korrelation überprüft. Dazu wurde der durchschnittliche Attraktivitätswert jedes Composites in Relation zum Grad seiner Durchschnittlichkeit gesetzt. Die Durchschnittlichkeit wurde codiert mit der „Generation“, in der das Composite entstand. So gab es insgesamt 32 Composites der Generation 1 (berechnet aus je zwei Originalgesichtern) 16 Composites der Generation 2 (berechnet aus je vier Originalgesichtern) usw. bis hin zu einem Composite der Generation 6, dem „absoluten“ Durchschnittsgesicht (berechnet aus allen 64 Original-Frauengesichtern). Da diese Daten prinzipbedingt nicht normalverteilt sind, wurde als Test die verteilungsfreie Rangkorrelation nach Spearman ausgewählt. Die Korrelation zwischen der Durchschnittlichkeit (d. h. Generation) eines Composites und seiner Attraktivität betrug $r_s = .61$, $N = 63$, $p < .001$. Das bedeutet, je durchschnittlicher ein Gesicht war, desto attraktiver wurde es auch bewertet.

Zusätzlich wurde jedoch auch die Alternativhypothese überprüft, die lautet: „Ein gemorphes Gesicht ist nur so attraktiv wie die Originalgesichter, die darin enthalten sind.“ Hierzu wurden die theoretischen Attraktivitätswerte (zur Berechnung siehe Kapitel 2.3.1.2) mit den empirischen korreliert. Die Korrelation betrug hier $r_s = .74$, $N = 63$, $p < .001$. Das bedeutet, dass auch diese Alternativhypothese empirisch bestätigt werden kann. Durchschnittsgesichter aus schönen Originalgesichtern blieben schön und Durchschnittsgesichter aus hässlichen Gesichtern blieben hässlich.

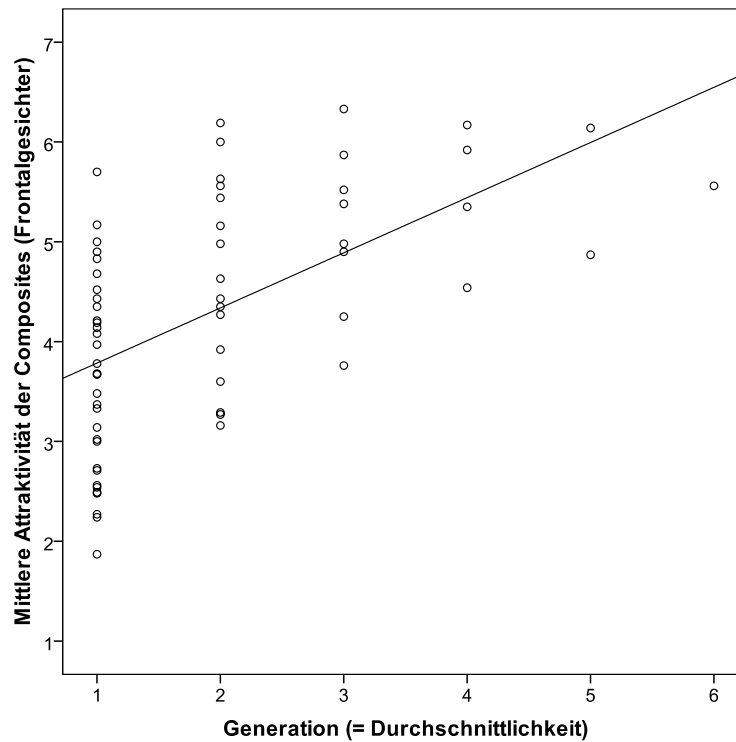


Abbildung 19: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Frauengesichter in der Frontalansicht. Je durchschnittlicher ein Gesicht ist, desto attraktiver wird es tendenziell auch bewertet. Allerdings gibt es etliche Gesichter, die weniger durchschnittlich als das Gesamt-Durchschnittsgesicht aus allen 64 Gesichtern (Generation 6) sind und trotzdem attraktiver bewertet werden.

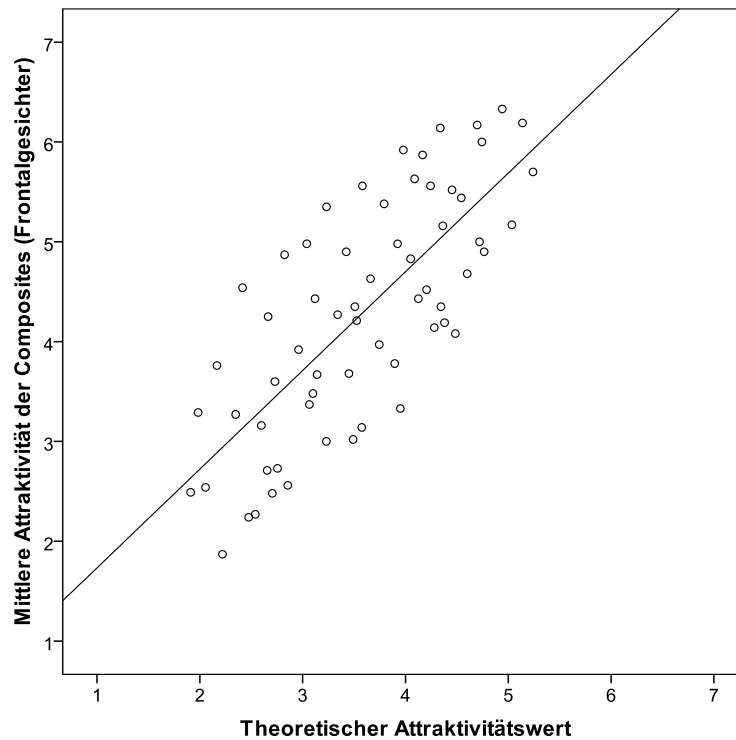


Abbildung 20: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Frauengesichter in der Frontalansicht. Je attraktiver die Originalgesichter waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver werden auch die daraus resultierenden gemorphten Gesichter beurteilt.

Daraus folgt, dass ein Composite, das aus den vier attraktivsten Originalen erzeugt wurde (w(61-64)), attraktiver sein sollte als eines aus den vier unattraktivsten Originalen (w(1-4)). Ein *t*-Test für gepaarte Stichproben zeigte, dass dies tatsächlich auch der Fall war ($t(62) = 21.17, p < .001$). Abbildung 21 zeigt diese beiden Beispiele.



Abbildung 21: „w(1-4)“, ein Composite aus den vier unattraktivsten weiblichen Originalgesichtern (links) und „w(61-64)“, ein Composite aus den vier attraktivsten weiblichen Originalgesichtern. Nach der Durchschnittshypothese müssten sie gleich attraktiv sein, da sie gleich durchschnittlich sind, tatsächlich ist jedoch der Durchschnitt aus den Schönsten weit attraktiver (mittlerer Attraktivitätswert von 6.19 ($SD = 0.80$) vs. 3.29 ($SD = 0.96$), $p < .001$).

Nach der Durchschnittshypothese müsste auch das absolute Durchschnittsgesicht „w(1-64)“, das alle Originalgesichter enthält, attraktiver sein als der Durchschnitt aus den acht attraktivsten „w(57-64)“, da dieses ja deutlich weniger durchschnittlich ist. Doch trotz der höheren Durchschnittlichkeit (das Gesamtdurchschnittsgesicht wurde aus achtmal mehr Originalgesichtern berechnet) wurde das Composite, das aus den schönsten Originalgesichtern erzeugt wurde, klar attraktiver bewertet ($t(62) = 5.62$, $p < .001$). Es erhielt sogar von allen Originalgesichtern und allen gemorphten Gesichtern den höchsten Attraktivitätswert.



Abbildung 22: „w(1-64)“, das Composite aus allen 64 weiblichen Originalgesichtern (links) und „w(57-64)“, ein Composite aus den acht attraktivsten Originalgesichtern (rechts). Nach der Durchschnittshypothese müsste das Gesamtdurchschnittsgesicht „w(1-64)“ attraktiver sein, da es durchschnittlicher ist als „w(57-64)“, tatsächlich ist es jedoch umgekehrt (mittlerer Attraktivitätswert von 6.33 ($SD = 0.72$) vs. 5.56 ($SD = 1.10$), $p < .001$).

Insgesamt erhielten von 63 Composites 10 einen höheren mittleren Attraktivitätswert als das Gesamtdurchschnittsgesicht w(1-64). Bei 7 davon war dieser Unterschied statistisch signifikant, darunter auch das Composite aus der „schöneren Hälfte“ aller fotografierten Frauengesichter, „w(33-64)“, sowie die Composites aus dem schönsten und dem zweit schönsten Quartil der Originalgesichter, „w(33-48)“ und „w(49-64)“.

Zusammenfassend lässt sich feststellen, dass mit gemorphten Durchschnittsgesichtern von Frauen in der Frontalansicht die Ergebnisse mit Männergesichtern (Experiment *Durch 1*) repliziert wurden. Die Composites wurden umso attraktiver bewertet, je durchschnittlicher sie waren, d. h. je mehr Originalgesichter in einem solchen Composite enthalten waren. Gleichzeitig wurden jedoch auch hier wieder Belege für die Alternativhypothese gefunden, nach der ein Composite so attraktiv ist wie die Originalgesichter aus denen es erzeugt wurde. Die Zusammenhänge waren mittelstark bis hoch und für die Alternativhypothese etwas größer als für die Durchschnittshypothese ($r_s = .61$ versus $r_s = .74$). Durch das systematische paarweise Morphen nach der Attraktivität der Originalgesichter konnten mehrere Composites generiert werden, die zwar aus weniger Gesichtern

berechnet (also weniger durchschnittlich) waren, aber dennoch attraktiver bewertet wurden als das Durchschnittsgesicht aus *allen* 64 Frauengesichtern.

2.8 Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von weiblichen Profilgesichtern mit Texturinformation (Untersuchung *Durch 5*)

2.8.1 Methode

2.8.1.1 Voruntersuchung

Vorgehen bei der Voruntersuchung

Im Experiment *Durch 5* wurde dieselbe Fragestellung wie im Experiment *Durch 4* untersucht, diesmal jedoch mit Frauengesichtern in der Profilansicht. Das Vorgehen war dasselbe. Gleichzeitig sollte das Experiment *Durch 5* die Ergebnisse des Experiments *Durch 2* replizieren, jedoch mit dem Unterschied, dass nun Frauenprofile statt Männerprofile verwendet wurden.

Zuerst erfolgte eine Voruntersuchung, bei der die Attraktivität der fotografierten Originalprofile (derselben Frauen wie in Experiment *Durch 4*) empirisch ermittelt wurde. Aus den Urteilen der Versuchspersonen wurde für jedes Gesicht ein Attraktivitätsmittelwert berechnet. Anschließend wurde eine Rangreihenfolge der Gesichtsprofile gebildet, beginnend vom unattraktivsten bis hin zum attraktivsten Gesichtsprofil. Das Morphen erfolgte wieder so, wie in Kapitel 2.3.1 beschrieben, indem über mehrere Generationen immer ähnlich attraktive Profile miteinander gekreuzt wurden. Für das Morphen der Profile wurde die Software *MorphMan 4.0* verwendet. Alle technischen Parameter und Arbeitsschritte beim Morphen waren genau so wie in Experiment *Durch 1*.

Wie auch im Experiment *Durch 2* bei den Männerprofilen erfolgte die Beurteilung der Attraktivität der *Original*personen nicht durch die Präsentation der Profile als Farbfotos, sondern als sogenannte „Schattenprofile“. Der Grund war derselbe: Mit Experiment *Durch 5* sollte herausgefunden werden, welche Merkmale für ein attraktives Profil charakteristisch sind. Die meiste Information steckt dabei in der Kontur der Profillinie. Die Farbwerte (Textur) des Gesichts, also v. a. die Haut, sind dabei nur eine Störvariable, die durch die Präsentation der Gesichter als Schattenprofile ausgeblendet wurde.

Die Schattenprofile wurden den Versuchspersonen als Ausdrucke in einer Größe von 13 x 16 cm in variierender Reihenfolge präsentiert. Die siebenstufige Likert-Skala zur Attraktivitätsbewertung war dieselbe wie in allen Experimenten zur Durchschnittshypothese.

Versuchspersonen für die Voruntersuchung

An der Voruntersuchung zur Bewertung der Schattenprofile nahmen insgesamt 80 Versuchspersonen (43 Männer, 37 Frauen) teil. Die Altersspanne reichte von 14 bis 57 Jahren ($M = 37.7$, $SD = 11.5$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, wurde die Befragung an unterschiedlichen öffentlichen Orten (z. B. Einkaufszentrum, Fußgängerzone, Bahnhof) durchgeführt. Die Teilnahme war freiwillig und wurde nicht bezahlt. Die Versuchspersonen wurden gebeten, nur die Attraktivität des Gesichts zu beurteilen, und dabei die Frisur der gezeigten Personen zu ignorieren.

Ergebnisse der Voruntersuchung

Für die Auswertung des Vortests wurde für jedes Schattenprofil ein Attraktivitätsmittelwert über alle Versuchspersonen berechnet und die Gesichter wurden auf der Basis dieser Rating-Mittelwerte in eine Rangreihenfolge gebracht. Das unattraktivste Schattenprofil erhielt im Mittel den Wert 1.39 ($SD = 0.58$), das attraktivste den Wert 5.61 ($SD = 0.82$). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .98 (bei 80 Versuchspersonen) sehr hoch.

2.8.1.2 Versuchspersonen und Durchführung der Hauptuntersuchung

An der Hauptuntersuchung zur Bewertung der gemorphten Frauenprofile mit Farbinformation nahmen insgesamt 108 Versuchspersonen (57 Männer, 51 Frauen) teil. Die Altersspanne reichte von 15 bis 83 Jahren ($M = 32.4$, $SD = 14.9$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, wurde die Befragung wieder an unterschiedlichen öffentlichen Orten (z. B. Einkaufszentrum, Fußgängerzone, Bahnhof) durchgeführt. Die Teilnahme war freiwillig und wurde nicht bezahlt. Die Versuchspersonen wurden gebeten, nur die Attraktivität des Gesichts zu beurteilen und dabei die Frisur der gezeigten Personen zu ignorieren.

2.8.1.3 Stimulusmaterial der Hauptuntersuchung

Die Erzeugung des Stimulusmaterials erfolgte analog zur Erzeugung der gemorphten Männerprofile (Experiment *Durch 2*). Es wurden die farbigen Profile (nicht die Schattenprofile!) verwendet. Das Vorgehen beim Morphing war dasselbe wie bei Experiment *Durch 1* und *4* (vgl. Kapitel 2.3.1). Dies gilt auch für die Berechnung der theoretischen Attraktivitätswerte (vgl. Kapitel 2.3.1.2).

Es gibt jedoch einen wesentlichen Unterschied zum Experiment *Durch 4* mit den gemorphten Frauenbildern in der Frontalansicht. Bei diesem Experiment wurden nicht 64 Originalgesichter für die Erzeugung des Durchschnittsgesichts verwendet, sondern nur eine Teilmenge von 32 Gesichtern. Dafür gab es zwei Gründe: Zum einen sollte der immens hohe Zeitaufwand reduziert werden, den die Erzeugung eines Durchschnittsgesichts aus 64 Originalen mit sich bringt. Zum anderen erwiesen sich zahlreiche Gesichtsprofile aufgrund ihrer Frisuren als sehr problematisch zu morphen. Während ein Vermorphen bei Männergesichtern aufgrund der Kurzhaarfrisuren unproblematisch war, bedeutete die hohe Variabilität der Frauenfrisuren durch ihr Volumen einen großen Störfaktor.

Ein Beispiel: Während bei der einen Frau nach hinten gebundene, voluminöse Haare das Bild dominieren, Teile des Nackens und der Schulter verdecken und die Beurteilung der Schädelform erschweren, sind bei einer anderen Frau mit einer Kurzhaarfrisur die Haare unproblematisch und verdecken nichts. Würde man allerdings diese beiden Bilder miteinander vermorphen, würden starke Bildverarbeitungsartefakte resultieren, da Haare des einen Bilds auf Bildpartien abgebildet würden, an denen sich gar keine Haare befinden. Aus diesem Grund wurden solche Frauenprofile für das Morphen ausgewählt, bei denen die Frisur beim Morphen weniger Schwierigkeiten machte. Um die Varianz der Gesichterattraktivität nicht zu reduzieren, wurde als weiteres Kriterium versucht, Originalprofile hohen bzw. niedrigen Attraktivitätsbewertungen eher zu behalten, während von den zahlreichen mittelattraktiven Profilen (mit ähnlichen Attraktivitätsmittelwerten) eher weniger verwendet wurden.

Da bei den gemorphten Bildern großer Wert auf ein natürliches Aussehen gelegt wurde, wurden nach jedem Morphvorgang die resultierenden Bilder mit hohem Aufwand nachretuschiert, um Bildverarbeitungsartefakte auf ein Minimum zu reduzieren. Da (trotz der Vorauswahl) aufgrund der hohen Variabilität der Frisuren sich Überblendungen bei Haaren (z. B. durch Pferdeschwänze, die auf unterschiedlicher Höhe zusammengebunden

waren), sich nicht vermeiden ließen, wurden alle diese Artefakte wieder wegretuschiert. Das Endergebnis war ein Stimulusmaterial von hoher Natürlichkeit, bei dem (zumindest auf den ersten Blick) nicht erkennbar ist, dass es sich bei den präsentierten Personen um Computergrafiken anstatt Fotos realer Personen handelt.

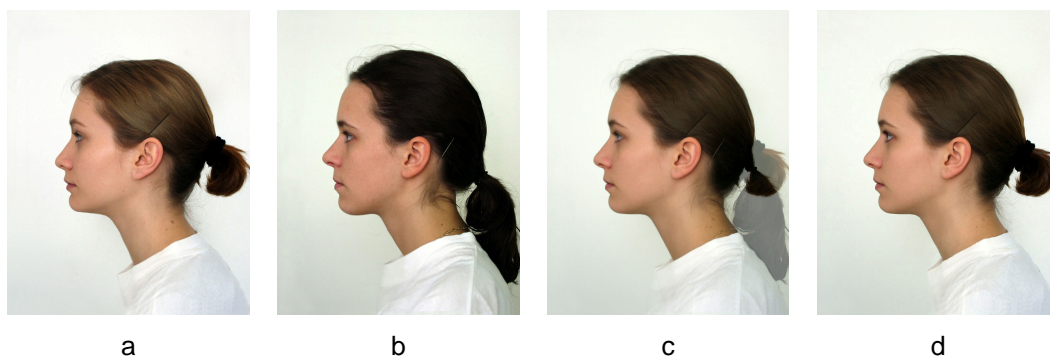


Abbildung 23: Beispiel für Bildretusche. Zwei Frauenprofile (a und b) werden miteinander vermorpht und ergeben ein Composite (c), das an den Haaren starke Artefakte enthält. Diese Bildstörungen werden wegretuschiert und ergeben ein natürlich aussehendes Frauenprofil (d), dem man die Künstlichkeit praktisch nicht ansieht. Bereits die Bilder a und b sind gemorphte Profile (ebenfalls mit nachträglichen Bildretuschen an der Frisur), die aus je zwei Originalfotos erstellt wurden.

Für die Datenerhebung wurden die Composites mit einer Auflösung von 1280 x 1024 Pixeln auf Fotopapier der Größe 9 x 13 cm belichtet. Dies entspricht einer sehr hohen Auflösung von 250 dpi, wodurch selbst feine Details sichtbar blieben. Die Darbietung erfolgte in variierender Reihenfolge.

2.8.2 Ergebnisse

Analog zur Auswertung des Vortests wurde für jedes weibliche Composite ein Attraktivitätsmittelwert über alle Versuchspersonen berechnet. Das unattraktivste gemorphte Frauenprofil erhielt im Mittel den Wert 1.69 ($SD = 1.02$), das attraktivste den Wert 5.76 ($SD = 0.99$). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .995 (bei 108 Versuchspersonen) sehr hoch.

Zur Überprüfung der Durchschnittshypothese wurden dieselben Korrelationen (nach Spearman) wie in Experiment *Durch 4* bei den Frauengesichtern in der Frontalansicht gerechnet. Die Korrelation zwischen der Durchschnittlichkeit (d. h. Generation) eines

Composites und seiner Attraktivität betrug $r_s = .42$, $N = 31$, $p = .020$. Das bedeutet, je durchschnittlicher ein Frauenprofil war, desto attraktiver wurde es auch bewertet.

Zusätzlich wurde jedoch auch wieder die Alternativhypothese (= Nullhypothese) überprüft, die lautet: „Ein gemorphtes Gesicht ist nur so attraktiv wie die Originalgesichter, die darin enthalten sind.“ Hierzu wurden die theoretischen Attraktivitätswerte der Schattenprofile (zur Berechnung siehe Kapitel 2.3.1.2) mit den empirischen Attraktivitätswerten der farbigen Composite-Profile korreliert. Die Korrelation betrug hier $r_s = .83$, $N = 31$, $p < .001$. Das bedeutet, dass auch diese Alternativhypothese empirisch bestätigt werden kann. Farbige Profilgesichter aus schönen Originalgesichtern (gemessen an der Attraktivität ihrer Schattenprofile) blieben schön und Durchschnittsprofile aus hässlichen Gesichtern blieben hässlich. Damit können bei Gesichtern in der Profilansicht die Ergebnisse von Gesichtern in der Frontalansicht (Experiment 4) bestätigt werden. Zudem bestätigen die Ergebnisse der Frauenprofile die Ergebnisse, die mit Männerprofilen gewonnen wurden.

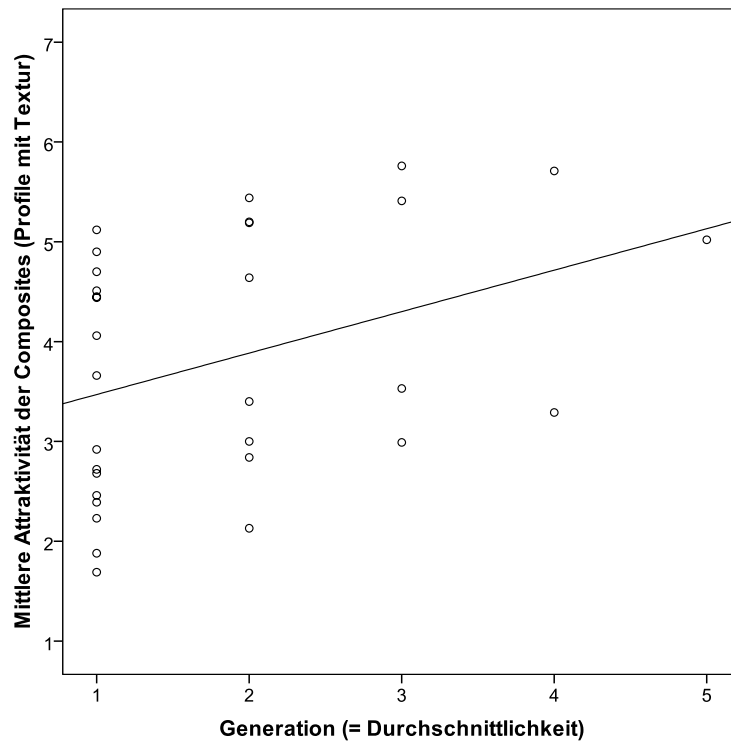


Abbildung 24: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Frauenprofile mit Texturinformation. Je durchschnittlicher ein Profil ist, desto attraktiver wird es tendenziell auch bewertet. Allerdings gibt es etliche Gesichter, die weniger durchschnittlich als das Gesamt-Durchschnittsgesicht aus 32 Gesichtern (Generation 5) sind und trotzdem attraktiver bewertet werden.

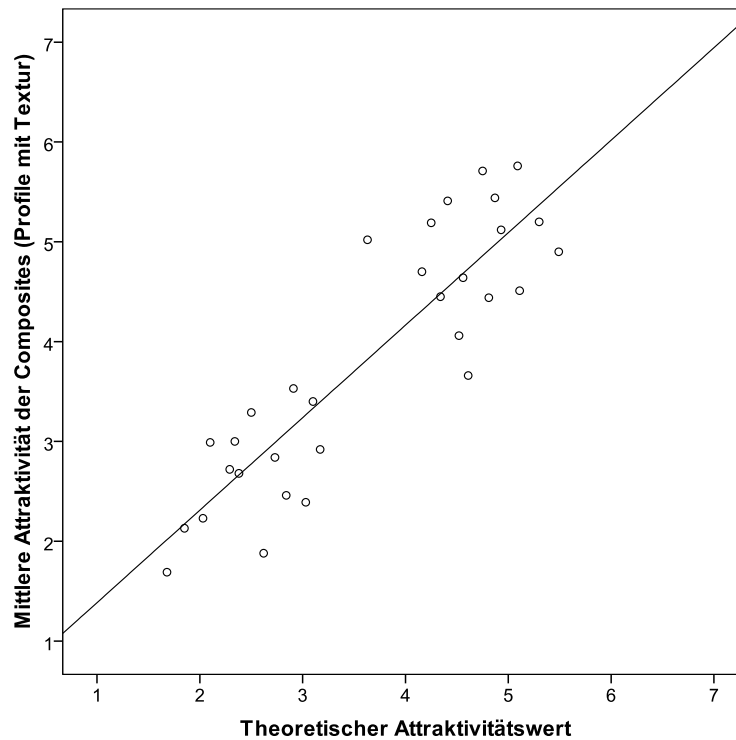


Abbildung 25: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Frauenprofile mit Texturinformation. Je attraktiver die Originalgesichter waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver wurden auch die daraus resultierenden gemorphten Gesichter beurteilt. Die Lücke in der Punkteverteilung resultiert aus dem Umstand, dass bei der Auswahl der 32 zu vermorphenden Gesichter (aus 78 Originalen) weniger Gesichter mit mittlerer Attraktivität verwendet wurden.

Aus der Bestätigung der Alternativhypothese folgt, dass ein Composite, das aus den vier attraktivsten Originalen erzeugt wurde (w(29-32)) attraktiver sein sollte als eines aus den vier unattraktivsten Originalen (w(1-4)). Ein *t*-Test für gepaarte Stichproben zeigte, dass dies tatsächlich auch der Fall war ($t(107) = 23.76, p < .001$). Abbildung 26 zeigt diese beiden Beispiele.

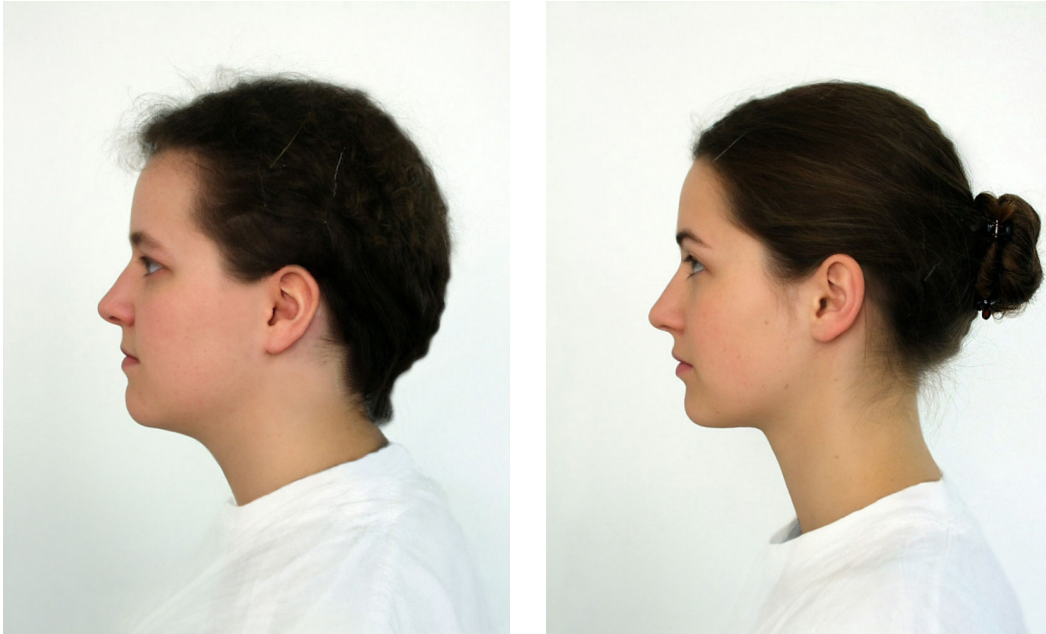


Abbildung 26: „w(1-4)“, ein Composite aus den vier unattraktivsten weiblichen Originalprofilen (links) und „w(29-32)“, ein Composite aus den vier attraktivsten weiblichen Originalprofilen. Nach der Durchschnittshypothese müssten sie gleich attraktiv sein, da sie gleich durchschnittlich sind, tatsächlich ist jedoch der Durchschnitt aus den Schönsten weit attraktiver (mittlerer Attraktivitätswert von 5.20 ($SD = 1.11$) vs. 2.13 ($SD = 1.06$), $p < .001$).

Nach der Durchschnittshypothese müsste auch das Gesamt-Durchschnittsprofil „w(1-32)“, das alle 32 weiblichen Originalprofile enthält, attraktiver sein als der Durchschnitt aus den acht attraktivsten „w(25-32)“, da dieses ja deutlich weniger durchschnittlich ist. Das Gegenteil ist jedoch der Fall. Der Durchschnitt aus wenigen attraktiven Originalen wird als attraktiver bewertet als das Gesamtdurchschnittsgesicht, $t(107) = 6.86$, $p < .001$.



Abbildung 27: „w(1-32)“, das Composite aus 32 Originalprofilen (links) und „w(25-32)“, ein Composite aus den acht attraktivsten Originalprofilen (rechts). Nach der Durchschnittshypothese müsste das Gesamt-Durchschnittsgesicht „w(1-32)“ attraktiver sein, da es durchschnittlicher ist als „w(25-32)“, tatsächlich ist jedoch umgekehrt (mittlerer Attraktivitätswert von „w(1-32)“ = 5.02 ($SD = 1.18$) vs. „w(25-32)“ = 5.76 ($SD = 0.99$), $p < .001$).

Zusammenfassend lässt sich feststellen, dass mit den weiblichen Profilgesichtern mit Textur die Ergebnisse der weiblichen Frontalgesichter (Experiment *Durch 4*) und der Männerprofile (Experiment *Durch 2*) bestätigt wurden. Gemorphte Durchschnittsgesichter (Composites) wurden umso attraktiver bewertet, je durchschnittlicher sie waren, d. h. je mehr Originalgesichter in einem solchen Composite enthalten waren. Gleichzeitig wurden jedoch auch Belege für die Alternativhypothese gefunden, nach der ein Composite so attraktiv ist wie die Originalgesichter, aus denen es erzeugt wurde. Der Zusammenhang war jedoch für die Alternativhypothese weit größer ($r_s = .83$) als für die Durchschnittshypothese ($r_s = .42$). Die Indizien für die Gültigkeit der Durchschnittshypothese waren damit bei weiblichen Profilgesichtern auch deutlich schwächer als bei weiblichen Frontalgesichtern ($r_s = .61$). Durch das systematische paarweise Morphen nach der Attraktivität der Originalgesichter konnten – wie schon bei den vier Experimenten zuvor – etliche Composites generiert werden, die zwar aus weniger Gesichtsprofilen berechnet (also weniger durchschnittlich) waren, aber dennoch attraktiver bewertet wurden als das Durchschnittsgesicht aus *allen* 32 Frauenprofilen.

2.9 Experimentelle Manipulation der Durchschnittlichkeit durch Verwendung von weiblichen Profilgesichtern ohne Texturinformation (Schattenprofile, Untersuchung *Durch 6*)

2.9.1 Methode

Bei Experiment *Durch 6* handelt es sich um eine Replikation von Experiment *Durch 5*, jedoch mit dem Unterschied, dass die gemorphten Profile nun nicht mehr als farbige Fotos, sondern als Schattenprofile präsentiert wurden. Gleichzeitig ist das Experiment der Versuch, die Ergebnisse aus Experiment *Durch 3* zu replizieren – statt Schattenprofilen von Männern wurden hier jedoch Schattenprofile von Frauen verwendet. Durch den Vergleich der Attraktivitätsbewertungen der farbigen Profile und der Schattenprofile sollte herausgefunden werden, wie groß der Effekt der Farbinformation (d. h. insbesondere der Farbe und Textur der Haut) der Gesichter ist. Denn durch die Präsentation als Schattenprofile wird die Variable „Textur“ komplett ausgeschaltet, da alle Gesichter dieselbe Textur aufweisen (nämlich gleichmäßig schwarz). Erhalten bleiben hingegen die Unterschiede hinsichtlich der Gesichtsproportionen eines Profils. Wenn sich mit den gemorphten Schattenprofilen die Korrelationen zwischen Durchschnittlichkeit und Attraktivität aus den beiden Experimenten *Durch 4* und *Durch 5* bestätigen lassen, dann ist der Effekt eindeutig darauf zurückzuführen, dass durchschnittliche *Gesichtsproportionen* attraktiv machen. Korrelieren diese beiden Variablen jedoch nicht miteinander, dann müssen die Ergebnisse aus Experiment *Durch 5* auf die veränderte Textur der Composites zurückzuführen sein.

Versuchspersonen und Durchführung

An der Untersuchung zur Bewertung der gemorphten Schattenprofile nahmen insgesamt 40 Versuchspersonen (25 Männer, 15 Frauen) teil. Die Altersspanne reichte von 19 bis 61 Jahren ($M = 29.1$, $SD = 11.7$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, wurde die Befragung wieder an unterschiedlichen öffentlichen Orten (z. B. Einkaufszentrum, Fußgängerzone, Bahnhof) durchgeführt. Die Teilnahme war freiwillig und wurde nicht bezahlt. Die Versuchspersonen wurden gebeten, nur die Attraktivität des Gesichts zu beurteilen und dabei die Frisur der gezeigten Personen zu ignorieren.

Stimulusmaterial

Das Stimulusmaterial wurde erzeugt, indem die farbigen Composites aus Experiment *Durch 5* mit dem Programm *Adobe Photoshop* so bearbeitet wurden, dass das gesamte Profil gleichmäßig schwarz und der Rest des Bildes weiß eingefärbt wurde. Die Darbietung der Bilder war genau so wie bei der Voruntersuchung von Experiment *Durch 5*.

2.9.2 Ergebnisse

Analog zur Auswertung der vorherigen Untersuchungen wurde für jedes Composite ein Attraktivitätsmittelwert über alle Versuchspersonen berechnet. Das unattraktivste Composite erhielt im Mittel den Wert 2.05 ($SD = 1.26$), das attraktivste den Wert 5.63 ($SD = 0.90$). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .98 (bei 40 Versuchspersonen) sehr hoch.

Zur Überprüfung der Durchschnittshypothese wurden wieder dieselben Korrelationen wie in den beiden vorangegangenen Experimenten gerechnet. Bei den Schattenprofilen ließ sich der Zusammenhang zwischen Durchschnittlichkeit und Attraktivität – im Gegensatz zu den Ergebnissen mit den farbigen Composites – *nicht* bestätigen; der schwach positive Effekt war nicht signifikant ($r_s = .30$, $N = 31$, $p = .101$). Der Zusammenhang zwischen „theoretischer Attraktivität“ und „empirischer Attraktivität“ war hingegen sehr stark ($r_s = .87$, $N = 31$, $p < .001$), das bedeutet, dass bei Schattenprofilen ein Composite gerade so attraktiv ist wie die Original-Schattenprofile, aus denen es berechnet wurde.

Die folgenden Abbildungen zeigen für die weiblichen Schattenprofile grafisch den Zusammenhang zwischen Attraktivität und Durchschnittlichkeit (= Durchschnittshypothese, Abbildung 28, S. 61) bzw. theoretischem Attraktivitätswert (= Alternativhypothese, Abbildung 29, S. 62).

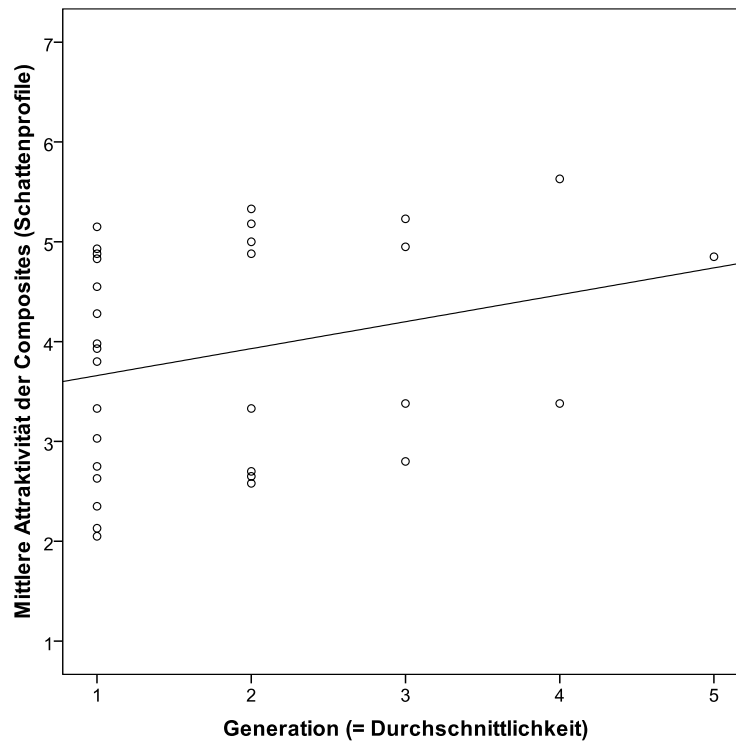


Abbildung 28: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Frauenprofile ohne Texturinformation (= Schattenprofile). Je durchschnittlicher ein Profil ist, desto attraktiver wird es tendenziell auch bewertet. Der Zusammenhang ist jedoch nur sehr schwach ($r_s = .30$) und nicht signifikant. Zudem gibt es etliche Schattenprofile, die weniger durchschnittlich als das Durchschnittsgesicht aus 32 Schattenprofilen (Generation 5) sind und trotzdem attraktiver bewertet werden.

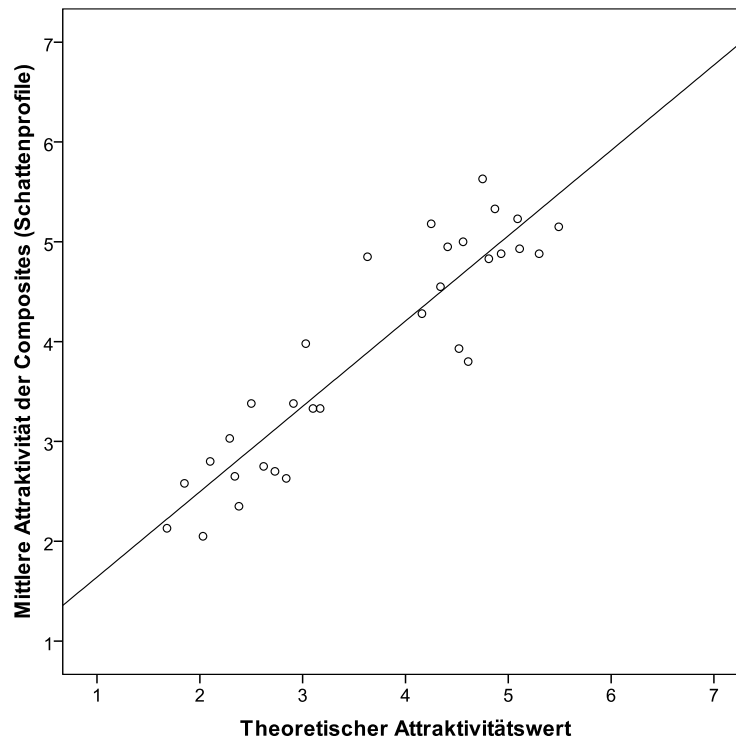


Abbildung 29: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Frauenprofile ohne Texturinformation (= Schattenprofile). Je attraktiver die originalen Schattenprofile waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver werden auch die daraus resultierenden gemorphten Schattenprofile beurteilt. Der Zusammenhang ist sehr stark ($r_s = .87$, $p < .001$).

Insgesamt wurden zehn Schattenprofile attraktiver bewertet als das absolute Durchschnittsschattenprofil, das nach der Durchschnittshypothese das attraktivste sein müsste. Der Unterschied war jedoch aufgrund der geringen Probandenzahl ($N = 40$) nur in einem Fall statistisch signifikant (in einem weiteren Fall, genau bei $p = .05$).

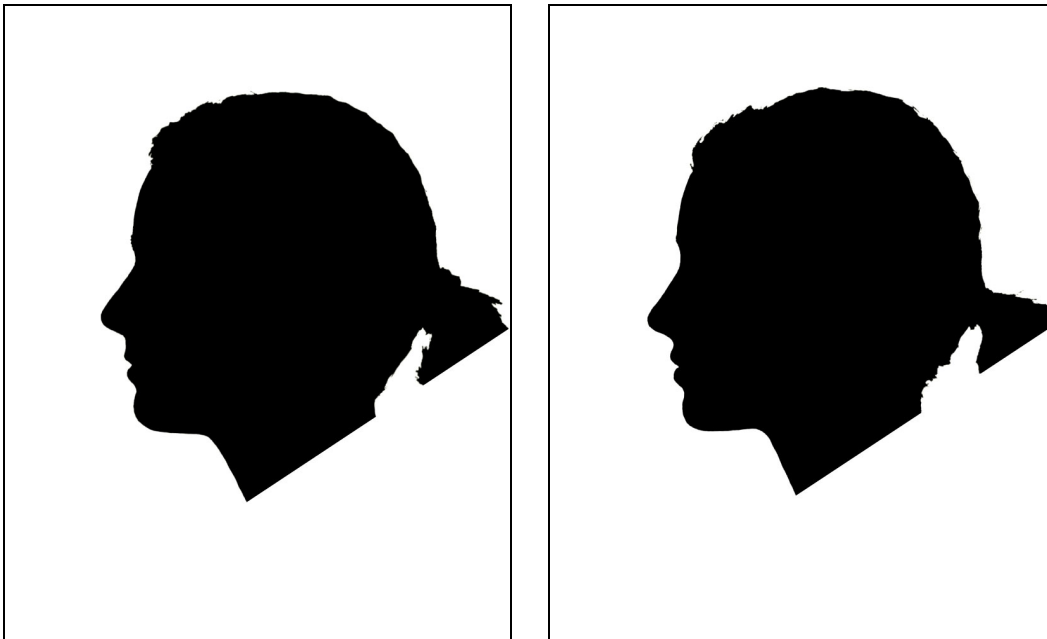


Abbildung 30: „w(1-32)“, das Composite aus 32 Originalschattenprofilen (links) und „w(25-32)“, ein Composite aus den acht attraktivsten Originalschattenprofilen (rechts). Nach der Durchschnittshypothese müsste das absolute Durchschnittsgesicht „w(1-32)“ attraktiver sein, da es durchschnittlicher ist als „w(25-32)“, tatsächlich ist es jedoch auch hier wieder umgekehrt (mittlerer Attraktivitätswert von „w(1-32)“ = 4.85 ($SD = 1.00$) vs. „w(25-32)“ = 5.23 ($SD = 1.10$), ($t(39) = 2.02$, $p = .050$).

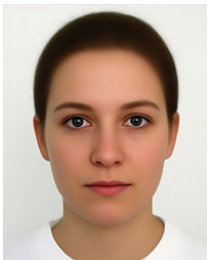


Zusammenfassend lässt sich festhalten, dass sich die Durchschnittshypothese mit Schattenprofilen *nicht* bestätigen ließ. Der Zusammenhang war jedoch nur schwach ($r_s = .30$, $p = .101$) und nicht signifikant. Der Zusammenhang zwischen „theoretischer Attraktivität“ und „empirischer Attraktivität“ (= Alternativhypothese) war hingegen sehr stark ($r_s = .87$, $p < .001$). Das bedeutet, dass bei weiblichen Schattenprofilen ein Composite gerade so attraktiv bewertet wurde wie die Original-Schattenprofile, aus denen es berechnet worden war. Die Ergebnisse mit den weiblichen Schattenprofilen bestätigen damit die Ergebnisse der männlichen Schattenprofile, auch wenn dort der Effekt der Durchschnittlichkeit noch knapp signifikant war ($r_s = .38$, $p = .037$). Da der Effekt der Durchschnittlichkeit bei den weiblichen Schattenprofilen nicht signifikant war, lässt dies den Schluss zu, dass die Attraktivitätssteigerung gemorphter weiblicher Composite-Profile *mit* Textur (*Durch 5*) auf Morphing-Artefakte der makelloseren Haut zurückzuführen ist und nicht auf die durchschnittlichen Gesichtsproportionen, da die Textur der einzige Unterschied zwischen den Stimuli der beiden Experimente war.

2.10 Zusammenfassende Datenanalyse der Experimente *Durch 4 bis Durch 6*

Analog zum Vorgehen bei den Männergesichtern wurden auch bei den Frauengesichtern die Daten der Experimente *Durch 1 bis Durch 3* varianzanalytisch ausgewertet. Dafür wurden drei ANOVAs durchgeführt, bei denen die Attraktivitätsratings jeweils die abhängige Variable waren und der „Grad der Durchschnittlichkeit“ (für drei Generationen) sowie der „Attraktivität der Originalgesichter“ (theoretische Attraktivitätswert) die jeweiligen unabhängigen Faktoren waren. Die Zuordnung der jeweiligen Composites zu den beiden Faktoren „Grad der Durchschnittlichkeit“ und „Attraktivität der Originalgesichter“ erfolgte wie bei den Männergesichtern (vgl. Kapitel 2.6). Der einzige Unterschied war, dass beim Experiment *Durch 4* (Frauengesichter in der Frontalansicht) die Anzahl der Originalgesichter 64 statt 32 betrug. Deswegen wurden hier zusätzlich die vier *16-face-composites* (= 4. Composite-Generation) in die Auswertung mit einbezogen, die aus je einem Attraktivitätsquartil erzeugt worden waren.

Für jedes der drei Experimente wurde eine zweifaktorielle ANOVA gerechnet mit dem Faktor „Durchschnittlichkeit“ (4 Stufen bei Frontalgesichtern, d. h. Generation 1 bis 4; 3 Stufen bei Profilgesichtern) und dem Faktor „Attraktivität der Originalgesichter“ (4 Stufen, d. h. Quartil 1 bis 4). Die Ergebnisse sind in Tabelle 3 zusammengefasst.

Tabelle 3: Die Ergebnisse der drei ANOVAs mit Tests der Haupteffekte und Interaktionseffekte bei den drei Experimenten mit weiblichen Durchschnittsgesichtern

	<i>Durch 1</i>	<i>Durch 2</i>	<i>Durch 3</i>
	Frontalgesichter	Profilgesichter mit Texturinformation	Schattenprofile ohne Texturinformation
			
Durchschnittlichkeit (= Generation) ($df = 2$, bei Frontalgesichtern $df = 3$)	$F = 42.452$ $p < .001$ $\eta^2 = .743$	$F = 10.801$ $p = .001$ $\eta^2 = .574$	$F = 2.697$ $p = .098$ $\eta^2 = .252$
Attraktivität der Originalgesichter (= Quartil) ($df = 3$)	$F = 33.634$ $p < .001$ $\eta^2 = .696$	$F = 49.079$ $p < .001$ $\eta^2 = .902$	$F = 47.763$ $p < .001$ $\eta^2 = .900$
Interaktion (Durchschnittlichkeit * Attraktivität der Originalgesichter) ($df = 6$)	$F = 0.325$ $p = .962$ $\eta^2 = .062$	$F = 0.244$ $p = .955$ $\eta^2 = .084$	$F = 0.985$ $p = .467$ $\eta^2 = .270$

Die Ergebnisse bestätigen die Ergebnisse der Rangkorrelationen nach Spearman und bestätigen grundsätzlich die Befunde der Männergesichter: Benutzt man für das Experiment Frontalgesichter, dann ist der Effekt der Durchschnittlichkeit ($\eta^2 = .74$) in etwa gleich groß wie der Effekt der Attraktivität der Originalgesichter ($\eta^2 = .70$) und jeweils hoch signifikant. Bei Profilgesichtern mit Textur ist der Effekt der Durchschnittlichkeit bereits deutlich

kleiner ($\eta^2 = .57$) und der Effekt der Attraktivität der Originalgesichter deutlich größer ($\eta^2 = .90$, p jeweils $<.001$). Beim dritten Experiment, bei dem die Profilgesichter keine Textur mehr enthielten, wurde hingegen der kleine Haupteffekt der Durchschnittlichkeit (= Generation) ($\eta^2 = .25$) nicht mehr signifikant ($p = .098$); entscheidend für die Attraktivität eines Composites war hier vielmehr die Attraktivität der ihnen zugrunde liegenden Originalgesichter ($\eta^2 = .90$, $p < .001$). Bei allen drei Experimenten gab es keinerlei Interaktionseffekte zwischen den beiden Faktoren „Durchschnittlichkeit“ und „Attraktivität der Originalgesichter“.

Damit replizieren die Daten der drei Experimente mit Frauengesichtern ziemlich präzise die Ergebnisse der Experimente mit Männergesichtern. Die zusammenbrechende Effektstärke der Durchschnittlichkeit bei den Schattenprofilen ist bei männlichen und weiblichen Stimuli sogar exakt gleich ($\eta^2 = .25$, $p = .10$). Der Unterschied ist, dass bereits bei Verwendung von Profilen mit Textur bei Frauengesichtern ein deutlicher Rückgang des Effekts der Durchschnittlichkeit im Vergleich zur Verwendung von Frontalgesichtern zu verzeichnen ist, während er bei Männergesichtern nicht ganz so stark ausfällt (vgl. Kapitel 2.6).

Abbildung 31 (S. 67) illustriert die Haupteffekte der Durchschnittlichkeit und Attraktivität der Originalgesichter für weibliche Frontalgesichter, Abbildung 32 (S. 68) dasselbe für Profilgesichter mit Textur.

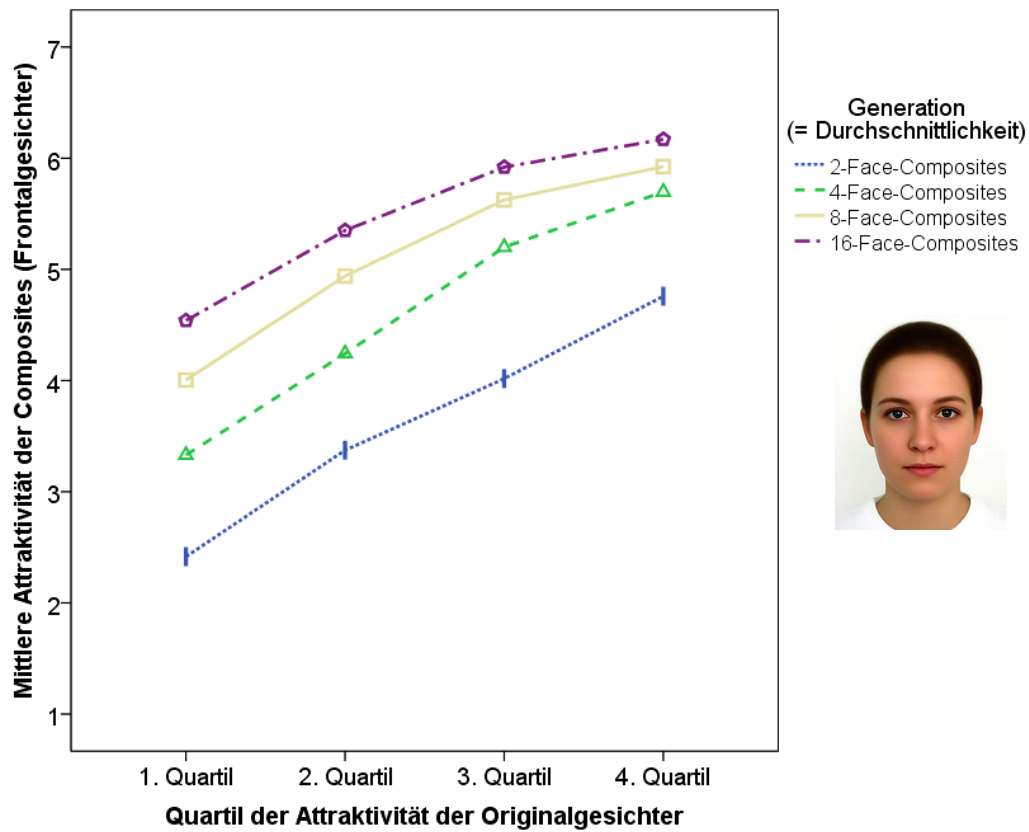


Abbildung 31: Die mittlere Attraktivitätsbewertung von gemorphten Frauengesichtern in der Frontalansicht in Abhängigkeit von der ursprünglichen Attraktivität der Originalgesichter (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). Die beiden Haupteffekte sind klar erkennbar.

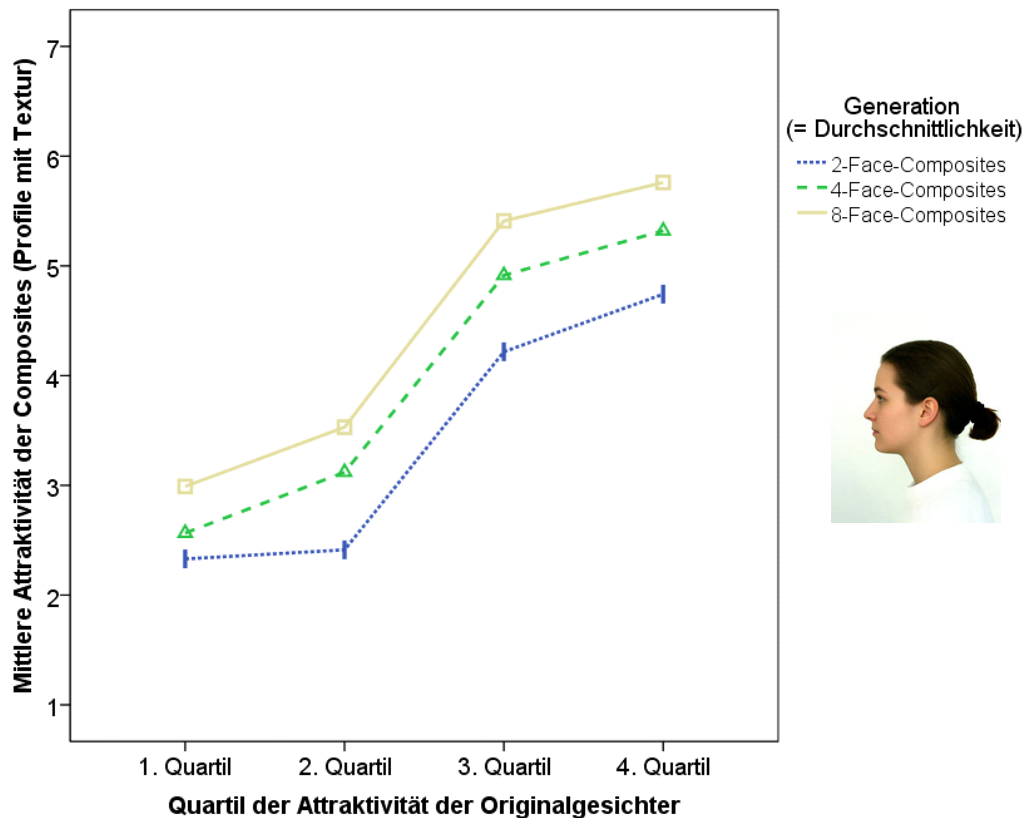


Abbildung 32: Die mittlere Attraktivitätsbewertung von gemorphten Frauenprofilen mit Textur in Abhängigkeit von der ursprünglichen Attraktivität der Originalprofile (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). Der „Sprung“ zwischen dem 2. und 3. Quartil resultiert aus dem Umstand, dass in der Auswahl der 32 (von den ursprünglich 78 Originalgesichtern) zu vermorphenden Profile solche Gesichter mit mittlerer Attraktivität unterrepräsentiert waren.

Bei Abbildung 32 fällt der „Attraktivitätssprung“ zwischen dem 2. und 3. Quartil auf, der bereits im Streudiagramm Abbildung 28 (S. 61) zu beobachten war und daher kommt, dass bei der Auswahl der 32 zu vermorphenden Gesichter (aus 78 Originalen) weniger Gesichter mit mittlerer Attraktivität verwendet wurden.

In Abbildung 33 (S. 69) erkennt man gut, dass bei den Schattenprofilen der Haupteffekt der Attraktivität der Originalgesichter genauso vorhanden ist wie bei den anderen beiden Stimulusarten, der Haupteffekt der Durchschnittlichkeit jedoch nur noch marginal (und nicht signifikant) ist.

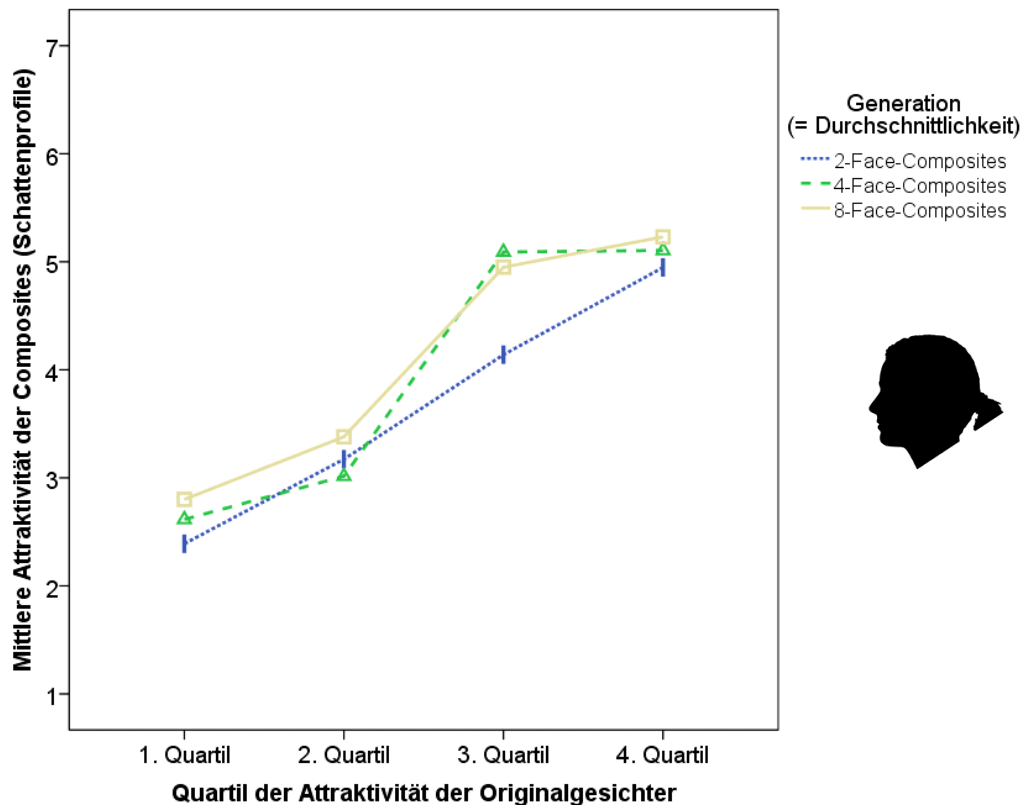


Abbildung 33: Die mittlere Attraktivitätsbewertung von gemorphten Frauenprofilen ohne Textur (= Schattenprofile) in Abhängigkeit von der ursprünglichen Attraktivität der Originalprofile (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). Hier besteht nur noch ein Haupteffekt der Attraktivität der Originalgesichter. Der Effekt der Durchschnittlichkeit wird nicht mehr signifikant.

Um den Effekt der Stimulusart zu untersuchen, wurde – analog zum Vorgehen bei den Männergesichtern – eine dreifaktorielle ANOVA durchgeführt, bei der die Stimulusart (Frontalgesichter, Profilgesichter mit Textur und Schattenprofile ohne Textur) als dritter Faktor hinzugenommen wurde. Es zeigte sich zusätzlich zu den signifikanten Haupteffekten von „Durchschnittlichkeit“ und „Attraktivität der Originalgesichter“ ein signifikanter Haupteffekt der „Stimulusart“ sowie eine signifikante Interaktion zwischen „Durchschnittlichkeit“ und „Stimulusart“. Wie schon bei den Männergesichtern ist diese Interaktion auf die abweichenden Ergebnisse der Schattenprofil-Stimuli zurückzuführen, wie Abbildung 34 (S. 71) illustriert. Das Ergebnis bestätigt damit den obigen Befund, dass die Durchschnittlichkeit zwar bei den Profilen *mit* Textur etwas mit Attraktivität zu tun hatte, jedoch nicht, wenn man Schattenprofile verwendete. Ob also Durchschnittlichkeit einen Effekt

auf Attraktivität hatte, hing davon ab, welches Stimulusmaterial man für eine Untersuchung verwendete.

Tabelle 4: Ergebnisse der dreifaktoriellen ANOVA. Einfluss von Durchschnittlichkeit, Attraktivität der Originalgesichter und Stimulusart auf die Attraktivität eines weiblichen Composites

	<i>df</i>	<i>F</i>	<i>p</i>	η^2
Attraktivität der Originalgesichter (= Quartil)	3	88.050	<.001	.777
Durchschnittlichkeit (= Generation)	3	33.856	<.001	.572
Stimulusart	2	15.725	<.001	.293
Quartil * Generation (Interaktion)	9	0.603	.791	.067
Quartil * Stimulusart (Interaktion)	6	2.481	.030	.164
Generation * Stimulusart (Interaktion)	4	4.452	.003	.190
Quartil * Generation * Stimulusart (Interaktion)	12	0.316	.985	.047

Zusätzlich zeigte sich ein Interaktionseffekt zwischen der Attraktivität der Originalgesichter und der Stimulusart. Dieser Effekt resultiert aus dem „Knick“ zwischen dem 2. und 3. Attraktivitätsquartil bei den Attraktivitätsbewertungen von gemorphten Profilgesichtern. Dieses Phänomen wird durch Abbildung 32 (S. 68) und Abbildung 33 (S. 69) illustriert. Es resultiert aus dem Umstand, dass mittelattraktive Originalgesichter als Ausgangsmaterial für die gemorphten Gesichter zahlenmäßig unterrepräsentiert waren, als eine Auswahl für 32 aus 78 Profilgesichtern getroffen wurde, die für das Morphen verwendet wurden. Diese Interaktion hat jedoch weder Bedeutung für die aufgestellten Hypothesen noch für die Theorie.

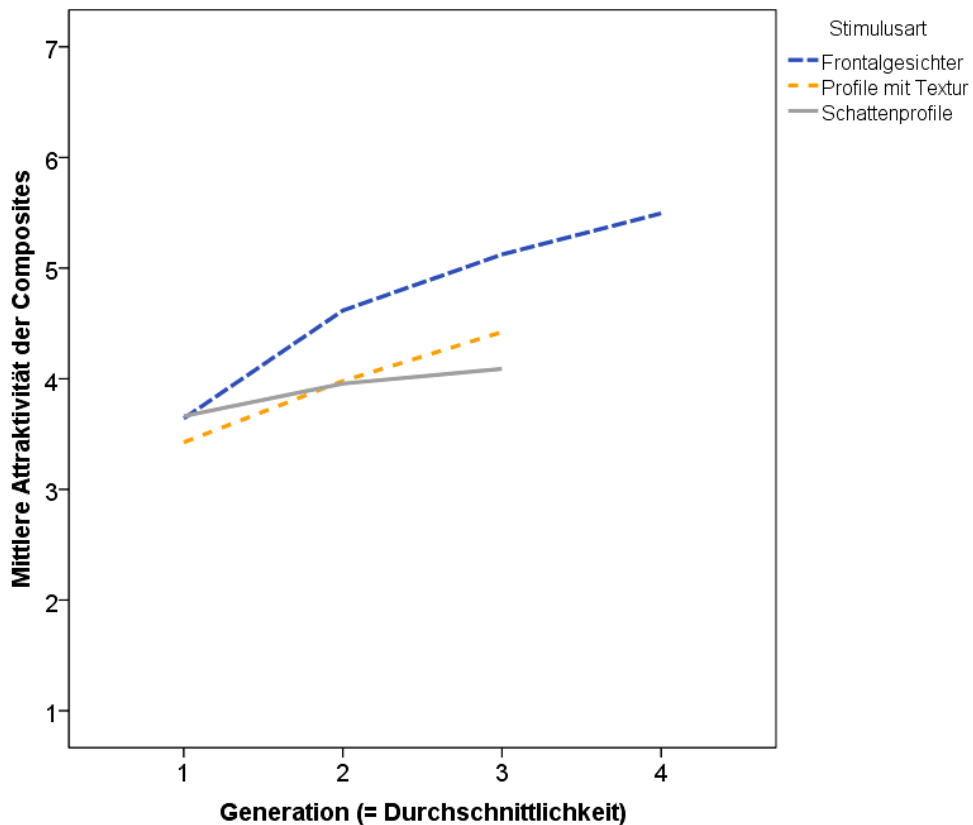


Abbildung 34: Die mittlere Attraktivität der gemorphten Frauengesichter in Abhängigkeit von ihrer Durchschnittlichkeit und der Art der Stimuli. Der Interaktionseffekt zwischen Durchschnittlichkeit und Stimulusart wird durch die Schattenprofile ohne Textur verursacht, da bei ihnen die Attraktivität fast unabhängig von der Durchschnittlichkeit ist.

Aus den Ergebnissen der dreifaktoriellen ANOVA folgt, dass die Attraktivitätszugewinne bei den gemorphten weiblichen Composites nicht nur auf die höhere Durchschnittlichkeit – genauer gesagt, die durchschnittlichen Gesichtsproportionen – zurückzuführen sind, denn sonst müsste bei den Schattenprofilen als Stimulusmaterial ein ähnliches Ergebnis herauskommen. Da die Textur der einzige Unterschied zwischen Schattenprofilen und Profilgesichtern mit Textur ist, muss ein wesentlicher Grund für die Attraktivitätszunahme gemorphter Composites in den Farbwerten der Textur liegen. Die naheliegende Erklärung ist, dass die bekannten und unvermeidlichen Morphing-Artefakte in Form der glatteren und makelloseren Haut die Ursache dieses Ergebnisses sind. Diese Konfundierung zwischen makelloser Haut und Durchschnittlichkeit ist bei den beiden Stimulusarten mit Tex-

tur gegeben, bei den Schattenprofilen jedoch nicht; aus diesem Grund fällt dort der Effekt der Durchschnittlichkeit so gering aus und wird auch nicht mehr signifikant.

Zusammenfassend kann festgehalten werden, dass die Ergebnisse der Frauengesichter sehr genau die Befunde der Experimente mit den Männergesichtern bestätigten. Die varianzanalytische Auswertung der Daten aus den drei Experimenten *Durch 4* bis *Durch 6* erbrachte den klaren Nachweis, dass die Attraktivitätssteigerung von weiblichen Composites *mit* Textur (frontal oder im Profil) zu einem Großteil auf Morphing-Artefakte der makelloseren Haut zurückzuführen ist. Dies belegt der signifikante Interaktionseffekt ($p = .003$) zwischen „Durchschnittlichkeit“ und „Stimulusart“ (vgl. Tabelle 4). Waren bei den weiblichen Stimuli keine Texturunterschiede mehr vorhanden (Schattenprofile in Experiment *Durch 6*), dann wurde der Effekt der Durchschnittlichkeit statistisch nicht mehr signifikant ($p = .098$), da der Effekt von durchschnittlichen Gesichtsproportionen allein zu schwach war.

3 Zusammenfassende Diskussion der Ergebnisse aller sechs Untersuchungen zum Einfluss der Durchschnittlichkeit

Die sechs durchgeführten Experimente zur Durchschnittshypothese ergeben ein sehr schlüssiges und widerspruchsfreies Gesamtbild. Die Ergebnisse von Frauengesichtern und Männergesichtern stimmen gut überein, und die Daten der Profilgesichter mit Textur bestätigen die Befunde, die mit Frontalgesichtern gewonnen wurden. Bei Verwendung von Schattenprofilen bricht bei Männer- und Frauengesichtern der Effekt der Durchschnittlichkeit zusammen, obwohl die Stimuli dieselben sind wie bei der Verwendung der farbigen Profilgesichter, nur mit dem Unterschied, dass die Schattengesichter keine Texturinformation mehr enthalten. Im Folgenden wird daher die Frage diskutiert, welche Schlussfolgerung aus diesem Befund für die Gültigkeit der Durchschnittshypothese zu ziehen ist. Diskutiert werden folgende Fragen:

- Welche Rolle spielt die Art der Stimuli (Frontalgesichter, Profile mit Textur, Schattenprofile) für den Einfluss der Durchschnittlichkeit (Kapitel 3.1.1)?
- Inwieweit sind die Attraktivität erhöhenden Effekte der Durchschnittlichkeit auf Morphing-Artefakte wie eine makellose Haut zurückzuführen (Kapitel 3.1.2)?

- Welche Rolle spielt die Attraktivität der Originalgesichter, aus denen Composites erstellt werden (Kapitel 3.1.3)?

Anschließend wird ein Alternativmodell zur Durchschnittshypothese aufgestellt. Es trägt dem Umstand Rechnung, dass die attraktivsten Gesichter sich eben nicht durch maximale Durchschnittlichkeit auszeichnen, sondern vielmehr Composites sind, die aus den attraktivsten Originalgesichtern erzeugt wurden. Diese Gesichter verfügen offensichtlich über das „gewisse Extra“, durch das sich hochattraktive Gesichter von Gesichtern mit durchschnittlichen Gesichtszügen unterscheiden.

In einer zusätzlichen Untersuchung wird dazu aus hochattraktiven Gesichtern (die acht attraktivsten Gesichter von Endrundenteilnehmerinnen der Miss-Germany-Wahl 2007) ein Composite gemorpht, dessen Attraktivität in einer zusätzlichen Datenerhebung evaluiert wird (Kapitel 3.2).

In einer qualitativen Untersuchung werden durch den Vergleich von attraktiven und unattraktiven Prototypen attraktivitätsrelevante Merkmale („Sexy Features“) für Gesichter in der Frontalansicht und der Profilansicht herausgearbeitet (Kapitel 3.3).

Abschließend werden die Ergebnisse vor dem Hintergrund anderer Untersuchungen zur Durchschnittshypothese diskutiert (Kapitel 3.4) und Konsequenzen für künftige Studien aufgezeigt (Kapitel 3.5).

3.1 Methodische Reflexion der Untersuchungen zur Durchschnittlichkeit

3.1.1 Effekte der Stimulusart

Für die sechs Experimente zur Durchschnittshypothese wurden sowohl Männergesichter (*Durch 1 bis 3*) als auch Frauengesichter (*Durch 4 bis 6*) verwendet. Die Ergebnisse zeigen, dass das Geschlecht der Gesichterstimuli für die Durchschnittshypothese keine Rolle spielt. Die Befunde von Männer- und Frauengesichtern zeigen eine hohe Übereinstimmung. Dies deckt sich auch mit der Metaanalyse von Rhodes (2006), in der beim Geschlecht der Stimuluspersonen ebenfalls kein Unterschied hinsichtlich des Einflusses der Durchschnittlichkeit auf die Attraktivität festgestellt wurde.

Bei der Verwendung von Frontalgesichtern und Profilgesichtern mit Textur bzw. Schattenprofilen als Stimuli hat jedoch die Stimulusart teilweise einen Einfluss auf den Effekt der Durchschnittlichkeit. Bei farbigen Gesichtern mit Textur (frontal oder Profilgesichter) sind die Unterschiede bei den Ergebnissen nur gering (mit etwas kleineren Effekten bei Profilgesichtern) und bestätigen den Einfluss der Durchschnittlichkeit, bei Schattengesichtern hingegen bricht der Effekt der Durchschnittlichkeit zusammen und wird nicht mehr signifikant. Die Gründe für den Zusammenbruch dieses Effekts werden im folgenden Kapitel 3.1.2 diskutiert.

Der etwas schwächere Effekt der Durchschnittlichkeit bei den Gesichtsprofilen mit Textur im Vergleich zu den Frontalgesichtern ist jedoch auch aus theoretischer Sicht von Interesse. Denn seit Entstehung der Durchschnittshypothese mit dem Artikel von Langlois und Roggman (1990) gab es immer wieder die Kritik, die beobachteten Effekte der höheren Attraktivität von Composites seien nicht darauf zurückzuführen, dass sie so durchschnittlich seien, sondern vielmehr darauf, dass Composites zugleich auch eine hohe Symmetrie aufwiesen (Grammer & Thornhill, 1994; Rhodes, Roberts et al., 1999). Es sei also in Wirklichkeit ein Effekt der höheren Symmetrie. Bei Composites in der Profilansicht fällt jedoch das Argument der Symmetrie von vornherein weg, weil es bei Profilen keine Symmetrieachse gibt.

Das bedeutet im Umkehrschluss: Wenn es bei gemorphten Frontalgesichtern tatsächlich einen Effekt der Symmetrie geben sollte, dann kann dieser maximal (!) so groß sein wie die Differenz der Effekte von Frontalgesichtern und Profilgesichtern mit Textur für den Faktor Generation (= Durchschnittlichkeit), denn die Symmetrie eines Composites kovariert mit seiner Durchschnittlichkeit (da in höheren „Morph-Generationen“ sowohl die Durchschnittlichkeit als auch die Symmetrie der Composites ansteigt). Für Männergesichter beträgt diese Differenz $\eta^2 = 0.084$ ($0.765 - 0.681$, vgl. Tabelle 2, S. 36, in Kapitel 2.6), für Frauengesichter $\eta^2 = 0.169$ ($0.743 - 0.574$, vgl. Tabelle 3, S. 65, in Kapitel 2.10). Die genaue Höhe des Effekts der Symmetrie lässt sich für Frontalgesichter jedoch nicht beziffern, er könnte auch weit kleiner als die genannten Differenzen der Effektgrößen sein bis hin zu gar keinem Einfluss. Sicher lässt sich aus logischen Gründen nur sagen, dass bei Profilgesichtern der Einfluss der Symmetrie Null ist.

Kritiker könnten einwenden, dass ein Rückschluss auf die Bedeutung – auch nicht auf die maximale Bedeutung – der Symmetrie nicht zulässig sei, da es prinzipiell auch denkbar sei, dass diese beiden Stimulusarten qualitativ so verschieden seien (obwohl sie ja von denselben fotografierten Personen stammen), dass die Ergebnisse sich deswegen nicht

vergleichen ließen. Theoretisch könnten nämlich für die Beurteilung der Frontalansicht eines Gesichts andere ästhetische Regeln gelten als für die Beurteilung der Profilansicht. Angenommen, dieser Einwand wäre berechtigt, dann hätten allerdings die Vertreter der Durchschnittshypothese erst recht ein Problem: Denn evolutionspsychologisch orientierte Forscher könnten dann nicht mehr erklären, warum allgemeine evolutionäre Mechanismen (*developmental stability*, *heterozygosity* und *stabilizing selection*, vgl. Kapitel 2.1), die Durchschnittlichkeit angeblich begünstigen, sich nur auf das menschliche Gesicht in der Frontalansicht, nicht jedoch auch in der Profilansicht auswirken sollten. Und auch die (zahlenmäßig wenigen) Anhänger der kognitiven Sichtweise („Prototypentheorie“) kämen in Erklärungsnot, denn sie dürften schwerlich plausibel erklären können, warum eine allgemeine Präferenz des Menschen für das typische, durchschnittliche, normale Aussehen selbst auf Dinge wie Vögel, Fische oder Autos anwendbar sein sollte (Halberstadt & Rhodes, 2003), gleichzeitig jedoch *nur* auf die Frontalansicht eines Gesichts und *nicht* auf seine Profilansicht.

Zur Klärung des Einflusses von Symmetrie wurden jedoch in dieser Arbeit zusätzlich neun Untersuchungen durchgeführt, mit denen diese Frage besser zu klären ist (Kapitel 4.4 bis 4.12) als durch den Vergleich der Effektgrößen zwischen gemorphten Frontalgesichtern und Profilgesichtern. Und um herauszufinden, ob Durchschnittlichkeit oder Symmetrie einen größeren Einfluss auf die Attraktivität hat, wurden weitere zwei Experimente durchgeführt, bei denen diese beiden Hypothesen direkt gegeneinander getestet wurden (Kapitel 8.1 bis 8.2). Daher sollen an dieser Stelle die Konsequenzen der Ergebnisse aus den Experimenten zur Durchschnittshypothese nicht weiter vertieft werden, sondern es wird auf die zusätzlichen 11 Untersuchungen dieser Arbeit verwiesen, die die Bedeutung der Symmetrie für die Attraktivität eines Gesichts eindeutig klären werden.

3.1.2 Effekte von Morphing-Artefakten

Die Durchschnittshypothese ist seit ihrer Entstehung untrennbar mit einer bestimmten Methodik verknüpft, nämlich mit der Methode des Morphens. Die Frage, ob Durchschnittlichkeit attraktiv macht, ist in der Attraktivitätsforschung praktisch identisch mit der Frage, ob gemorphte Composites attraktiv sind. Genauso alt wie die Beobachtung, dass Composites attraktiv sind, ist jedoch auch die Kritik an der Interpretation der Durchschnittlichkeit. Bereits Alley und Cunningham (1991) argumentierten, Composites seien deswegen attraktiver, weil bedingt durch ihren Herstellungsprozess sie auch gleichzeitig unschärfer aussähen (Alley & Cunningham, 1991). Dadurch verschwänden auch Makel wie Pickel oder Falten und die Haut sehe makelloser und jugendlicher aus. Somit sei der Effekt

lediglich auf Bildverarbeitungsartefakte zurückzuführen. Dennoch wurde diese Technik in der Folge häufig verwendet und zur Grundlage für zahlreiche Attraktivitätsstudien.

Da sich bei den hier durchgeführten Experimenten klar herausstellte, dass sich der Attraktivität erhöhende Effekt der Durchschnittlichkeit dann – und nur dann – replizieren ließ, wenn eine Textur vorhanden war (farbige Frontalgesichter und farbige Gesichtsprofile), nicht jedoch, wenn die Textur fehlte (Schattenprofile), ist damit offensichtlich, dass es eben primär *nicht* die durchschnittlichen Gesichtsproportionen sind, die den Effekt verursachen, sondern offensichtlich Effekte, die beim Vermorphen von Texturen entstehen. Aus diesem Grund wird im Folgenden herausgearbeitet, welche Effekte bzw. Bildverarbeitungsartefakte durch das Morphen entstehen und wie sie die Attraktivität eines Gesichts beeinflussen, da diese Grundlagen der Bildverarbeitung offensichtlich zahlreichen experimentell arbeitenden Psychologen auch über 20 Jahre nach Erfindung der Morphing-Technik immer noch unbekannt sind.

Gemorphte Composites besitzen zwei Besonderheiten: Die erste ist, dass ihre Gesichtsproportionen (*facial shape*) verdurschnittlicht wurden, und zwar durch die Mittelung der Koordinatenwerte von *Facial Landmarks* der Originalgesichter (*component faces*). Die zweite Besonderheit ist jedoch, dass die Textur der Composites eine völlig andere ist als bei Originalgesichtern. Dies zeigt sich v. a. in einer sehr glatten, makellosen Haut. Diese perfekte Haut entsteht zwar (technisch gesehen) durch die Mittelwertbildung der Farbwerte der einzelnen Bildpunkte am Computer, aber ihre Erscheinung ist alles andere als durchschnittlich (Benson & Perrett, 1992; Little & Hancock, 2002).

Im Gegenteil, eine so perfekte Haut, wie sie ein *32-face-composite* (bzw. *64-face-composite*) hat, gibt es in der Realität nicht (vgl. Abbildung 6, S. 21 und Abbildung 27, S. 58). Allenfalls bei professioneller Anwendung von Make-up kann man ihr nahekommen. Das Besondere der hier durchgeführten Experimente ist nun, dass sie quantifizieren können, wie groß der Einfluss der Verdurschnittlichung der Gesichtsproportionen (*facial shape*) einerseits und der perfekten, makellosen Haut andererseits ist. Man sieht dies durch den Vergleich der Effektstärken des Faktors Durchschnittlichkeit (= Generation) in den Experimenten *Durch 2* und *Durch 3* bzw. *Durch 5* und *Durch 6*, also jeweils den Experimenten mit den Profilgesichtern, jedoch einmal mit und einmal ohne Textur (vgl. Kapitel 2.6 und 2.10). Bei den farbigen Männerprofilen (bei denen auch die Pixelwerte für die Haut gemittelt wurden und die Composites mit zunehmender Durchschnittlichkeit eine immer perfektere Haut bekommen) beträgt die Effektgröße $\eta^2 = .68$. Bei den Schattenprofilen jedoch, bei denen der Faktor „Textur“ durch gleichmäßiges Schwärzen des gesam-

ten Profils ausgeschaltet war, beträgt die Effektgröße nur $\eta^2 = .25$. Dieser geringere Effekt ist also der Effekt, der allein durch die Verdurchschnittlichung der Gesichtsproportionen entsteht, die Differenz auf die Effektgröße $\eta^2 = .68$ ist hingegen der Effekt, der auf den Einfluss der glatten, makellosen Gesichtsoberfläche (v. a. Haut) zurückzuführen ist (Differenz: $\eta^2 = .43$). Bei Frauenprofilen ist der Befund ähnlich: Bei den farbigen Frauenprofilen mit Textur beträgt die Effektgröße $\eta^2 = .57$, bei den Schattenprofilen jedoch, bei denen der Faktor Textur durch gleichmäßiges Schwärzen des gesamten Profils ausgeschaltet war, beträgt die Effektgröße ebenfalls nur $\eta^2 = .25$ (Differenz: $\eta^2 = .32$). In beiden Fällen sind die Effektgrößen der Schattengesichter nicht mehr signifikant (vgl. Kapitel 2.6 und 2.10).

Diese Zahlen bedeuten, dass für Männer- wie für Frauenprofile jeweils deutlich mehr als die Hälfte des attraktivitätserhöhenden Effekts des Faktors Generation in Wirklichkeit *nicht* auf die Durchschnittlichkeit der Gesichtsproportionen zurückzuführen ist, sondern auf den Effekt der Texturveränderungen, die durch den Morphprozess entstehen. In allen bisherigen Untersuchungen konnten diese beiden Einflüsse nicht voneinander getrennt werden und deshalb wurde fälschlicherweise der *gesamte* Effekt der Attraktivitätssteigerung der Durchschnittlichkeit zugeschrieben. Die vorliegenden Daten der vier Experimente mit Profilgesichtern zeigen jedoch klar, dass der Effekt der morphingtypischen Texturveränderungen größer ist als der Effekt der durchschnittlichen Gesichtsproportionen.

Dies gilt zumindest für Gesichter in der Profilansicht. Dass dies auch auf Gesichter in der Frontalansicht übertragbar ist, ist damit zwar nicht bewiesen, aber der Schluss erscheint doch naheliegend. Empirisch überprüfbar wäre dies, indem man ein Experiment mit Frontalgesichtern macht, bei denen die Farbwerte weitestgehend ausgeschaltet sind, z. B. wie beim Experiment von Rhodes und Tremewan (1996) durch die Umwandlung der Fotos in Strichzeichnungen. Dort hatten die Autorinnen (in ihrem Experiment 2) versucht, das Experiment von Langlois und Roggman (1990) mit Strichzeichnungen zu replizieren und scheiterten damit: „However, the attractiveness of the composites did not increase with the number of component faces. [...] This result contrasts with the linear increase in attractiveness with the number of component faces found by Langlois und Roggman (1990)“ (p. 108).

Befürworter der Durchschnittshypothese dürften nun möglicherweise einwenden, dass in den einschlägigen Studien die Durchschnittsgesichter doch gar keine makelloser Haut als die Originalgesichter gehabt hätten, weil vor dem Experiment alle Originalgesichter (und weniger durchschnittliche Gesichter, d. h. *2-face-*, *4-face-composites* etc.) mit

Weichzeichnungsfiltern eines Bildverarbeitungsprogramms stark bearbeitet wurden. So schreiben Langlois et al. (1994): „[...] the individual faces were smoothed and blurred to make the averaged and individual images visually similar in degree of blur“ (p. 218). Deshalb könnten die Unterschiede in der Attraktivität nicht auf Bildverarbeitungsartefakte des Morphings zurückzuführen sein.

Doch dieses Argument stimmt nicht. Der Effekt des Morphings ist nicht vergleichbar mit der Anwendung eines Weichzeichnungsfilters. Die folgenden Bilder sollen dies deutlich machen. Abbildung 35 zeigt Bildausschnitte von vier Originalgesichtern A bis D, die in dieser Studie zu einem Gesicht verrechnet wurden.



Abbildung 35: Die Mäulern von vier Originalgesichtern (A, B, C and D, von links nach rechts), die in der Studie zu einem Durchschnittsgesicht gemorphet wurden.

Betrachtet man einmal den Mund von Person A, so fällt auf, dass er eine relativ schlechte Hautqualität mit vielen Pickeln hat. Kreuzt man nun Person A mit Person B erhält man das Gesicht A&B, das bereits eine deutlich gesünder aussehende Haut aufweist (Abbildung 36, linkes Bild). Nimmt man auch noch die Gesichter B und C hinzu, entsteht Gesicht A&B&C&D (Abbildung 36, zweites Bild). Zum Vergleich daneben das Originalgesicht A, das mit einem Gaußschen Weichzeichnungsfilter (Radius vier Pixel, der gesamte Bildausschnitt hat eine Größe von 340 x 170 Pixeln) behandelt wurde. Wie man sieht, werden die Pickel im weichgezeichneten Originalbild lediglich unscharf, aber sie verschwinden nicht. Die Haut von A&B&C&D (Abbildung 36, zweites Bild) sieht gesünder und schöner aus als die des Originals A (vgl. vorherige Abbildung 35, linkes Bild). Der Unterschied bleibt auch dann erhalten, wenn man nachträglich A&B&C&D ebenfalls gleich stark weichzeichnet wie Gesicht A (Abbildung 36, drittes und viertes Bild).

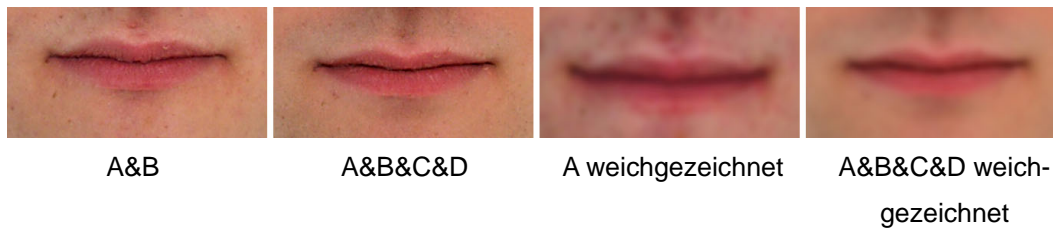


Abbildung 36: Bildausschnitte (240 x 170 Pixel) einzelner Gesichter im Vergleich: Bild ganz links: Das Composite aus den Gesichtern A und B; zweites Bild von links: Das Composite aus den Gesichtern A, B, C und D; drittes Bild von links: Das Gesicht A weichgezeichnet (Gaußscher Weichzeichnungsfilter mit 4 Pixeln Radius); Bild ganz rechts: Das Composite aus den Gesichtern A, B, C und D weichgezeichnet (Gaußscher Weichzeichnungsfilter mit 4 Pixeln Radius).

Der Grund dafür ist, dass beim weichgezeichneten Original A die Konturen von Pickeln und Muttermalen lediglich verwischt werden. Bei Gesicht A&B&C&D dagegen scheint jeder Pickel nur noch zu 25 % durch, weil er von gesunden Hautpartien aus den anderen drei Gesichtern überlagert wird. Der Effekt ist dadurch vergleichbar mit einem Make-up, das nicht vollständig abdeckt. Beim *32-face-composite* ist sogar jeder Pickel von Gesicht A (und auch von allen anderen 31 Originalgesichtern) nur zu $1/32$ im Bild vorhanden und damit überhaupt nicht mehr sichtbar.

Ähnlich ist der Effekt mit Hautstellen, die fettig glänzen. Bei einem Composite gibt es keinen Fettglanz, sobald darin wenigstens ein Originalgesicht enthalten ist, das an der entsprechenden Hautstelle *nicht* fettig glänzt. Gemorphte Gesichter haben daher in der Regel einen schönen, seidig schimmernden Teint. Durch einen Weichzeichnungsfilter kann man hingegen Fettglanz prinzipiell nicht zum Verschwinden bringen, weil der Pixelradius des Weichzeichnungsalgorithmus weitaus kleiner ist als fettig glänzende Hautpartien. Abbildung 37 demonstriert dies am Beispiel einer großen, leicht geröteten Hautpartie (was vom Prinzip her keinen Unterschied zu einer glänzenden Hautpartie macht).



Abbildung 37: Drei Ausschnitte (200 x 430 Pixel) aus einem Originalgesicht (die ersten drei Bilder von links) im Vergleich mit einem gemorphten Gesicht (Bild ganz rechts). Eine gerötete Hautstelle bei einem Originalgesicht unterhalb des Ohres ohne Filter (linkes Bild), derselbe Bildausschnitt mit einem Gaußschen Weichzeichnungsfilter von 4 Pixeln Radius (zweites Bild von links) und mit 8 Pixeln Radius (drittes Bild von links). Egal wie stark man weichzeichnet, die Rötung bleibt immer erkennbar. Zum Vergleich (ganz rechts) das *2-face-composite*, das dieses Original zu 50 % enthält und zu 50 % ein anderes Originalgesicht, das an dieser Stelle eine gesunde Haut hatte. Obwohl es sich *nur* um ein *2-face-composite* handelt, ist die Rötung nur noch schwach zu erkennen.

Das Prinzip der Verschönerung der Haut beim Morphen ist, dass auf unschöne, ungesund aussehende Hautstellen (Pickel, Falten, Flecken) gesunde, makellose Hautpartien aus anderen Gesichtern abgebildet und von diesen zu einem bestimmten Prozentsatz (50 %, 75 % usw.) überdeckt werden. Daher ist es auch kein Gegenargument, wenn Langlois et al. (1994) ihre Effekte von 1990 replizieren konnten, indem sie das *32-face-composite* mit solchen Composites verglichen, die jeweils aus 32 Fotos derselben Person erzeugt waren. Denn wenn man eine Person mit sich selbst morpht, werden (vorausgesetzt, die Fotos sind präzise aufgenommen) Makel immer auf dieselbe Stelle im Bild abgebildet, also ein Pickel immer auf sich selbst und nicht auf eine gesunde Hautpartie, eine fettglänzende Stirn immer auf ihren eigenen Fettglanz und nicht auf eine nicht glänzende Stirn aus einem anderen Gesicht. Dadurch blieben bei den von Langlois et al. mit sich selbst gemorphten Composites Makel erhalten (wurden bestenfalls unschärfer, ähn-

lich wie bei der Verwendung eines Weichzeichnungsfilters), womit auch die Attraktivitätsunterschiede zwischen Originalen und dem *32-face-composite* erhalten blieben.

Dies macht deutlich, dass die Haut gemorphter Gesichter etwas Besonderes hat, das sich nicht durch Weichzeichnungseffekte (*blurring* und *smoothing*) von Originalgesichtern erreichen lässt. Gemorphte Gesichter haben prinzipbedingt grundsätzlich eine makelloser Haut als Originalgesichter, selbst dann, wenn versucht wird, durch Weichzeichnungsfiler diese Unterschiede zu reduzieren. Das bedeutet, dass die Effekte von Experimenten zur Durchschnittlichkeit, bei denen Composites mit Originalgesichtern verglichen wurden, aufgrund der Konfundierung der Durchschnittlichkeit der Gesichtsproportionen mit einer makellosen Haut teilweise oder gar größtenteils auf Bildverarbeitungsartefakten beruhen. Dies zeigt auch das Experiment von Little & Hancock (2002), die *facial shape* und *texture* durch Morphing unabhängig voneinander manipulierten und damit zeigen konnten, dass beim Morphen sowohl die Verdurchschnittlichung der Gesichtsproportionen als auch die der Textur (mit seinen Smoothing-Effekten) zur Attraktivitätssteigerung beitragen.

Man sollte die Sache jedoch nicht so negativ sehen, indem man darauf verweist, dass die Schönheit von Durchschnittsgesichtern zum Großteil auf Bildverarbeitungsartefakten basiert. Technisch betrachtet handelt es sich zwar um Bildverarbeitungsartefakte, aber auf der Wahrnehmungsebene liegt dies an einer glatten, faltenlosen, makellosen Haut. Die Aussage, dass eine makellose Haut attraktiv macht, ist allerdings wenig überraschend und bei weitem nicht so spektakulär und provokativ wie die Überschrift des viel zitierten Klassikers der Attraktivitätsforschung von Langlois und Roggman: „Attractive faces are only average.“

3.1.3 Effekte der Ausgangsattraktivität der Originalgesichter

Die vorliegende Untersuchungsreihe weist über bisherige Experimente zur Durchschnittshypothese hinaus. Denn neben der Replikation bekannter Befunde zur Durchschnittlichkeit konnte ebenfalls gezeigt werden, dass die „absoluten“ Durchschnittsgesichter (also der Durchschnitt aus *allen* Gesichtern einer Stichprobe) *nicht* am attraktivsten sind, denn es konnten mit derselben Methodik Gesichter generiert werden, die attraktiver als alle Originalgesichter und attraktiver als der Durchschnitt aus allen Gesichtern sind. Dies sind die Composites, die aus den attraktivsten Einzelgesichtern berechnet wurden, z. B. die Composites aus den acht schönsten einer Gesichterstichprobe. In allen sechs Experimenten erhielten sie signifikant höhere Bewertungen als die „absoluten“ Durchschnittsgesichter. Ein Blick auf die Abbildungen (vgl. S. 21 und S. 49) macht sofort deut-

lich, dass diese Gesichter Merkmale aufweisen, die klar vom Durchschnitt abweichen. Dies bestätigt die Ergebnisse von Perrett et al. (1994) und DeBruine et al. (2007), die dies bereits mit Frauengesichtern zeigen konnten.

Wie die Varianzanalysen gezeigt haben, ist der Effekt der Attraktivität der Originalgesichter in den vier Experimenten mit Gesichterstimuli mit farbiger Textur in etwa genauso groß wie der Effekt der Durchschnittlichkeit. Und in den beiden Experimenten mit Schattenprofilen ist der Effekt sogar weit größer als der Effekt der Durchschnittlichkeit (bei Männer-Schattenprofilen: $\eta^2 = .83$ versus $\eta^2 = .25$; bei Frauen-Schattenprofilen: $\eta^2 = .90$ versus $\eta^2 = .25$; vgl. Kapitel 2.6 und 2.10).

Möglich wurde der Nachweis des unabhängigen Effekts der Attraktivität der Originalgesichter, weil das ursprüngliche Versuchsdesign von Langlois & Roggman (1990) leicht modifiziert wurde. Statt die Originalgesichter in beliebiger Reihenfolge paarweise miteinander zu kreuzen, wurde hier zuerst eine Voruntersuchung durchgeführt, die die Attraktivität der Originalgesichter ermittelte. Anschließend wurden immer ähnlich attraktive Originalgesichter miteinander gekreuzt. Das Endprodukt, nämlich das *32-face-composite* (bzw. *64-face-composite* bei den weiblichen Frontalgesichtern), bleibt dabei dasselbe wie beim zufälligen Kreuzen, aber die Zwischenprodukte auf dem Weg dorthin unterscheiden sich stark. Denn beim systematischen Morphen nach Attraktivität gibt es innerhalb jeder Morph-Generation eine große Varianz hinsichtlich der Attraktivität (insbesondere in den frühen Generationen), d. h. Gesichter, die sehr attraktiv sind und solche die ziemlich unattraktiv sind. (Beim unsystematischen Morphen wird die Varianz hingegen reduziert, weil auch attraktive mit unattraktiven Gesichtern gekreuzt werden.) Dadurch steigt die Wahrscheinlichkeit, dass sich unter den Schönen auch solche befinden, die schöner sind als das „absolute“ Durchschnittsgesicht aus allen Originalgesichtern.

Bisherige Untersuchungen zur Durchschnittshypothese verzichteten auf eine Voruntersuchung zur Ermittlung der Attraktivität der Originalgesichter und das systematische paarweise Morphen nach Attraktivität (z. B. Langlois & Roggman, 1990; Langlois, Roggman & Musselman, 1994; Grammer & Thornhill, 1994; Pollard, 1995; Valentine, Rhodes, Sumich & Byatt, 1999; Rhodes, Yoshikawa et al., 2001; Little & Hancock, 2002; Darling, & Donnelly, 2004). Dadurch verspielten sie die Chance, auch solche Gesichter zu erzeugen, die deutlich attraktiver als das absolute Durchschnittsgesicht sind. Dadurch ist es natürlich auch wesentlich leichter, die eigene Hypothese („average faces are attractive“) zu bestätigen. Das gilt insbesondere dann, wenn in einem einzigen Arbeitsschritt (ohne Zwischen-

schritte) das Gesamtdurchschnittsgesicht gemorpht wird, das dann nur noch mit den Originalgesichtern verglichen wird.

Da die Composites aus den attraktivsten Originalgesichtern attraktiver sind als alle Originalgesichter und als die „absoluten“ Durchschnittsgesichter, scheint genau in diesen Composites der Schlüssel zum Verständnis von Attraktivität zu liegen. Denn dafür ist es ja gerade entscheidend, was hochattraktive Gesichter von bloß einigermaßen attraktiven unterscheidet. Hierfür lohnt auch noch mal ein Blick zurück auf die absoluten Attraktivitätsbewertungen der einzelnen Composites die aus den Streudiagrammen gut ersichtlich sind. Ein Attraktivitätswert von 5 bedeutet auf der verwendeten Attraktivitätsskala „eher attraktiv“, 6 bedeutet „ziemlich attraktiv“ und 7, der Maximalwert (der bei einem Durchschnittsurteil natürlich nie erreicht wird) bedeutet „sehr attraktiv“. Die „absoluten“ Durchschnittsgesichter kommen bestenfalls gerade über den Wert 5 (= „eher attraktiv“) hinaus. Auch bei früheren Studien (Grammer & Thornhill, 1994; Langlois & Roggman, 1990; Rhodes, Sumich et al., 1999; Rhodes & Tremewan, 1996; Rhodes, Yoshikawa et al., 2001; Rubenstein, Kalakanis & Langlois, 1999) fällt auf, dass Durchschnittsgesichter zwar relativ gesehen attraktiver als andere (nicht systematisch nach Attraktivität gemorphte) Composites oder Originalgesichter waren, jedoch hinsichtlich ihrer absoluten Attraktivitätswerte keinesfalls überzeugen können, da sie gerade einmal Werte knapp oberhalb der Mitte der verwendeten Attraktivitätsskala erzielen. So gesehen ist es verwunderlich, dass diese Daten von den Autoren dieser Studien aber so interpretiert werden, dass Durchschnittlichkeit ein Kriterium für Schönheit ist – würden sie nämlich die Aussagen ihrer Versuchspersonen und die Bedeutung der Stufen ihrer verwendeten Attraktivitätsskalen ernst nehmen, dann müssten sie zum Schluss kommen, dass Durchschnittlichkeit ein Kriterium für ein passables, mittelmäßiges Aussehen ist, keineswegs jedoch für Schönheit.

3.2 Alternativmodell zur Durchschnittshypothese: Der „Raffael-Effekt“

Da in allen sechs Untersuchungen zur Durchschnittshypothese nachgewiesen wurde, dass nicht die „absoluten“ Durchschnittsgesichter aus allen Originalgesichtern einer Stichprobe, sondern vielmehr Composites aus den attraktivsten Einzelgesichtern am attraktivsten bewertet wurden, ist dies eine überzeugende Grundlage, um eine Alternativhypothese zur Durchschnittshypothese zu formulieren. Sie lautet:

„Ein Gesicht ist dann attraktiv, wenn es die Merkmale aufweist, die für attraktive Gesichter typisch und für unattraktive oder mittelmäßig attraktive Gesichter untypisch sind.“

Im ersten Moment mag diese Hypothese wenig hilfreich erscheinen, weil sie erstens wie ein Zirkelschluss klingt und zweitens nur eine Floskel bleibt, so lange es nicht gelingt, klar zu benennen, was konkret diese typisch attraktiven Merkmale sind. Dennoch ist diese Hypothese nützlich, da mit der Technik des Morphens ein Instrumentarium zur Verfügung steht, genau dieses Charakteristische zu visualisieren. Indem man also ein Composite aus wenigen sehr attraktiven Gesichtern herstellt und dieses mit dem Composite aus allen Originalgesichtern einer Stichprobe (d. h. dem Durchschnitt) vergleicht, kann man durch den direkten visuellen Vergleich die einzelnen Attraktivitätsmerkmale identifizieren.

Um diese Alternativhypothese zu überprüfen und die Nützlichkeit der Morphing-Technik für diese Fragestellung zu demonstrieren, wurde eine zusätzliche empirische Untersuchung durchgeführt, in der als Stimulusmaterial Fotos der Endrundenteilnehmerinnen der Miss-Germany-Wahl 2007 verwendet wurden.

Methode

Als Stimulusmaterial dienten Frontalfotos von den 22 Endrundenteilnehmerinnen der Miss-Germany-Wahl 2007 (genauere Informationen zur Aufnahme dieser Fotos finden sich im Kapitel 4.4.2.2). In einer Voruntersuchung wurden diese 22 Gesichter auf Fotopapier belichtet (Bildabmessungen: 10 x 13.3 cm bei 1600 x 2134 Pixeln; Auflösung 406 dpi) und 56 Versuchspersonen (24 Männer und 32 Frauen) im Alter von 19 bis 47 Jahren ($M = 25.0$, $SD = 5.5$) zur Attraktivitätsbeurteilung auf derselben siebenstufigen Skala wie bei den Experimenten zuvor vorgelegt. Die Beurteilerübereinstimmung war mit einem Cronbach's α von .95 sehr hoch.

Anschließend wurde für jedes Gesicht ein mittlerer Attraktivitätswert berechnet. Das unattraktivste Gesicht wurde mit $M = 2.95$ ($SD = 1.02$) und das attraktivste Gesicht mit $M = 5.11$ ($SD = 1.17$) bewertet (der Durchschnitt über alle Gesichter und alle Versuchspersonen betrug $M = 4.05$). Die acht attraktivsten Miss-Germany-Finalistinnen lagen im Bereich zwischen $M = 4.21$ ($SD = 1.04$) und $M = 5.11$ ($SD = 1.17$).

Anschließend wurde aus den acht attraktivsten Miss-Germany-Finalistinnen nach dem in Kapitel 2.3.1.2 beschriebenen Vorgehen durch paarweises Morphen (*MorphMan 4.0*) ein *8-face-composite* hergestellt.



Abbildung 38: Die „Virtuelle Miss Germany“, ein aus den acht attraktivsten Finalistinnen der Miss-Germany-Wahl 2007 berechnetes Composite (links). Zum Vergleich daneben das Durchschnittsgesicht „w(1-64)“ aus 64 „normalen“ Frauengesichtern (rechts). Abgesehen von den üblichen Morphing-Artefakten wie der makellosen Haut und dem Strukturverlust an den Haaren visualisiert das linke Bild, was den schönsten Frauen gemeinsam ist, während das rechte Bild lediglich zeigt, was für junge Frauen im Alter zwischen 17 und 30 Jahren normal ist.

In einer weiteren Datenerhebung wurde dieses *8-face-composite* zusammen mit den „Zwischenprodukten“ zu diesem Gesicht (vier *2-face-composites* und zwei *4-face-composites*) sowie acht zufällig ausgewählten Original-Missen (darunter auch die tatsächliche Gewinnerin der Miss-Germany-Wahl 2007) einer neuen Stichprobe von Versuchspersonen zur Attraktivitätsbewertung vorgelegt. An dieser Evaluation nahmen insgesamt 52 Versuchspersonen (9 Männer und 43 Frauen) im Alter von 20 bis 50 Jahren ($M = 25.1$, $SD = 6.0$) teil. Die Beurteilerübereinstimmung war mit einem Cronbach's α von .99 sehr hoch. Einige Fotos von Original-Missen wurden deshalb noch einmal zusätzlich zu den gemorphten Gesichtern präsentiert, da schon vom bloßen Augenschein bereits vor der Datenerhebung offensichtlich war, dass die gemorphten Gesichter deutlich attraktiver aussahen. Es bestand daher die Gefahr, dass die Versuchspersonen bei alleiniger Präsentation von gemorphten Gesichtern an das höhere Attraktivitätsniveau adaptieren und insgesamt ein strengeres Attraktivitätskriterium anlegen könnten (Haubensak, 1985; Henss, 1993; Higgins & Stangor, 1988; Kenrick & Gutierres, 1980; Kenrick, Gutierres &

Goldberg, 1989). Durch die zusätzliche Präsentation einiger Original-Missen wurden vergleichbare Bedingungen wie in der ersten Datenerhebung hergestellt und damit Kontexteffekten vorgebeugt (zur Problematik von Kontexteffekten vgl. auch die sehr gut geschriebene Übersicht von Henss, 1993).

Ergebnisse

Die statistische Analyse der Attraktivitätsbewertungen ergab eindeutige Ergebnisse. Jedes der gemorphten 7 Miss-Gesichter wurde signifikant besser bewertet als jedes der präsentierten Original-Miss-Gesichter (*t*-Tests *within groups*, jeweils $p < .01$). Am attraktivsten wurde das *8-face-composite* aus den 8 attraktivsten Original-Missen bewertet ($M = 6.04$, $SD = 0.56$). Zum Vergleich: Das Gesicht der Gewinnerin des Schönheitswettbewerbs wurde nur mit $M = 4.06$ ($SD = 1.04$) bewertet, also genau zwei Skalenstufen schlechter. Das Gesicht dieser „Virtuellen Miss Germany“ war zudem auch attraktiver als die Gesichter aller 22 Original-Missen, die in der Voruntersuchung vollständig präsentiert worden waren (*t*-Tests *between groups*, jeweils $p < .001$).

Das Composite aus den 8 attraktivsten Miss-Germany-Finalistinnen ist damit auch deutlich attraktiver als das Durchschnittsgesicht aus 64 normalen Frauen, das im Experiment *Durch 4* lediglich mit $M = 5.56$ ($SD = 1.10$) bewertet wurde ($p < .01$). Auffallend ist auch die sehr niedrige Standardabweichung ($SD = 0.56$) bei den Attraktivitätsbewertungen der „Virtuellen Miss Germany“, die zeigt, dass unter den Versuchspersonen eine hohe Einigkeit darin herrschte, dass dieses Gesicht sehr attraktiv ist. Eine derart geringe Standardabweichung gab es bei keiner einzigen Attraktivitätsbewertung aller in dieser gesamten Arbeit durchgeführten Untersuchungen (das schlechteste Einzel-Urteil für dieses *8-face-composite* war der Wert 5 = „eher attraktiv“, und 69 % der Befragten vergaben den Wert 6 = „ziemlich attraktiv“).

Schlussfolgerung

Die Ergebnisse dieser Evaluation zeigen eindeutig, dass das *8-face-composite* aus Abbildung 38 tatsächlich etwas sichtbar macht, was dem gegenwärtigen Schönheitsideal für ein Frauengesicht ziemlich nahe kommt. Der direkte Vergleich dieses Composites mit dem weiblichen Durchschnittsgesicht aus 64 „normalen“ Frauengesichtern macht unmittelbar klar, dass die Schönheit eines Gesichts *nicht* gleichzusetzen ist mit dessen Durchschnittlichkeit. Es gilt also das Gegenteil der von den Erfinderinnen der Durchschnittshypothese aufgestellten Behauptung „attractive faces are only average“ (= Titel des berühmten Aufsatzes von Langlois und Roggman, 1990), auch wenn diese bis heute von

universitären Attraktivitätsforschern als gesicherte Tatsache verkauft wird. Das Charman- te daran ist, dass für diese Erkenntnis noch nicht einmal Expertenwissen über Psycholo- gie, Morphing, Versuchsplanung oder Statistik notwendig ist – es reicht aus, sehen zu können. Die Gesichter in Abbildung 38 sprechen für sich. Manchmal sagt ein Bild nicht nur mehr als tausend Worte, sondern auch mehr als tausend statistisch signifikante Test- ergebnisse.

Die Idee, das Gemeinsame von hoch attraktiven Einzelgesichtern in einem einzigen Bild zu verschmelzen und damit ein Schönheitsideal sichtbar zu machen, ist an sich nicht neu. Bereits vor 500 Jahren ging Raffael so vor: In seinem Gemälde „Triumph der Galatea“ stellte er eine Frauenfigur dar, die offenbar genau dem Schönheitsideal seiner Zeit entsprach, denn die Schönheit seiner dargestellten Seegöttin begeisterte Raffaels Zeitge- nossen.



Abbildung 39: „Der Triumph der Galatea“ von Raffael (Deckenfresko der Villa Farnesina in Rom, um 1512). Die dargestellte Seegöttin Galatea (Bildmitte) entsprach offenbar dem Schönheitsideal des 16. Jahrhunderts, zumindest wurde die Frauenfigur wegen ihrer außerordentlichen Schönheit bewundert. Raffael selbst erklärte in einem Brief auf die Frage, wie es ihm gelungen sei, eine derart schöne Frau zu malen, er habe in dieser Figur die Essenz und die Idee von vielen einzelnen schönen Frauen vereint.

Interessant ist Raffaels Erklärung, wie es ihm gelungen sei, eine solch schöne Frau zu malen. In einem Brief (Raffaello & Camesasca, 1994) schrieb er an seinen Gönner, den Grafen von Castiglione:

Aber [...] ich muss Euch sagen, daß ich, um eine Schöne zu malen, deren mehrere sehen müßte, und zwar unter der Bedingung, daß Eure Herrlichkeit sich bei mir befände, um eine Auswahl der Allerschönsten zu treffen. Da nun aber immer Mangel an richtigem Urteil wie an schönen Frauen ist, bediene ich mich einer gewissen Idee, die in meinem Geist entsteht. Ob diese nun künstlerische Vortrefflichkeit in sich trägt, weiß ich nicht; wohl aber bemühe ich mich, sie zu erreichen. (Raffaello & Camesasca, 1994, S. 166; deutsche Übersetzung nach Krems, 2005, S. 46)

Nach eigener Aussage vereinte Raffael also die Essenz, das Gemeinsame, von zahlreichen schönen einzelnen Frauen in seinem Geiste und versuchte dann, diese „gewisse Idee“ von Schönheit in der Person der Galatea wiederzugeben (Krems, 2005).

Als Rezept, um eine schöne Frau darzustellen, nannte er genau das, was in dem hier durchgeführten Experiment tatsächlich gemacht wurde, nämlich (1) viele schöne Modelle als „Rohmaterial“ und (2) Attraktivitätsurteile. Als schöne Modelle dienten hier Miss-Germany-Endrundenteilnehmerinnen und als Attraktivitätsurteile zur „Auswahl der Allerschönsten“ allerdings nicht die Urteile des Grafen von Castiglione, sondern Urteile von Versuchspersonen einer für die aktuelle Bevölkerung möglichst repräsentativen Stichprobe (das angebliche Angewiesensein auf die Expertise des Grafen in Sachen Schönheit ist wohl auch eher als Schmeichelei gegenüber seinem Mäzen zu verstehen). Denn nach Auffassung der Attraktivitätsforschung ist Schönheit auch nichts, was vom Urteil eines oder weniger Experten abhängt, sondern per Definition das, was durch das Urteil der Bevölkerung bestimmt wird.

Für den letzten Schritt (3), der Visualisierung dieses Schönheitsideals benötigte Raffael einerseits Abstraktionsfähigkeit und Phantasie, um aus vielen schönen einzelnen Gesichtern und Figuren das Gemeinsame zu erkennen und zu verbinden und andererseits handwerkliches Können, um seine „gewisse Idee“ des Schönen dann auch zu malen. Im Fall der „Virtuellen Miss Germany“ ging dies viel einfacher und zugleich objektiver: Das Herausarbeiten des Gemeinsamen aller Originalgesichter und das Visualisieren erledigte die Morphing-Software durch Mittelwertbildungen von Bildkoordinatenwerten und Farbwerten automatisch. Damit hatte Raffael bereits vor einem halben Jahrtausend die Idee formuliert, mit der in dieser Untersuchung ein abstraktes Schönheitsideal visualisiert wurde: Ein Gesicht ist dann schön, wenn es das Wesentliche enthält, das allen Gesichtern einer Stichprobe, die als schön beurteilt wurden, gemeinsam ist. Aus diesem Grund soll

die hohe Attraktivität von Composites aus den attraktivsten Einzelgesichtern einer Stichprobe als „Raffael-Effekt“ bezeichnet werden.

Hätte Raffael vor 500 Jahren bereits Morphing-Software zur Verfügung gehabt, hätte er sie bestimmt mit Begeisterung eingesetzt. Hätte er allerdings damals schon von den Theorien der Attraktivitätsforschung zu Beginn des 21. Jahrhunderts gewusst und hätte kritiklos an sie geglaubt, dann hätte sein Rezept für die Erzeugung von Schönheit nicht gelautet: „Suche viele schöne Frauen als Modelle, beurteile ihre Schönheit, selektiere die allerschönsten und male dann die ‚Idee‘ dieser wenigen Einzelschönheiten“. Stattdessen hätte er empfohlen: „Wähle per Zufall irgendwelche Frauen aus (unabhängig von deren Aussehen) und male die ‚Idee‘ einer Frau“. Mit einer solchen normalen, durchschnittlichen Frau hätte er allerdings mit seinem Bild „Triumph der Galatea“ bei seinen Zeitgenossen wohl kaum den gleichen Erfolg gehabt.

Dass der „Raffael-Effekt“ funktioniert, hat letztlich einen statistischen Grund: Es wäre durchaus möglich, dass eine reale Person mit idealen Gesichtszügen tatsächlich existiert, doch die Wahrscheinlichkeit, eine solche zu finden, ist äußerst gering (bereits Raffael beklagte ja den „Mangel an schönen Frauen“ s. o.). Der Grund dafür ist, dass die Attraktivität eines Gesichts durch sehr viele Merkmale (d. h. Einzelmerkmale und Proportionen) beeinflusst wird. Erst wenn alles passt, dann entspricht das Gesicht dem Ideal. Je mehr Parameter jedoch die Attraktivität eines Gesichts beeinflussen, desto unwahrscheinlicher ist es, ein solches Gesicht auch zu finden. Die Wahrscheinlichkeit fällt sogar exponentiell mit der Anzahl der Parameter.

Ein Gedankenexperiment macht dies deutlich: Angenommen, es gäbe (1) nur 10 Parameter, die für die Attraktivität eines Gesichts notwendig wären, und (2) jede zehnte Person würde das jeweilige Kriterium dieses Merkmals erfüllen (d. h. 10 % der Personen hätten eine perfekte Nase, 10 % ein perfektes Kinn usw.); angenommen weiterhin, (3) die Ausprägung dieser Merkmale wäre voneinander statistisch unabhängig, dann betrüge in diesem Fall die Wahrscheinlichkeit, dass eine Person alle Kriterien gleichzeitig erfüllt:

$$\left(\frac{1}{10}\right)^{10} = \frac{1}{10000000000}$$

Auf 10 Milliarden Menschen (mehr als es auf der Welt gibt) käme statistisch eine einzige, die sämtlichen Kriterien gleichzeitig entspricht. In der Realität ist es allerdings so, dass die Ausprägung einzelner Gesichtsm Merkmale *nicht* statistisch unabhängig voneinander

ist, denn es gibt Einflüsse z. B. durch Hormone oder durch das Alter, die mehrere Merkmale gleichzeitig beeinflussen (insofern wäre die tatsächliche Wahrscheinlichkeit höher). Andererseits gibt es ganz sicher weit mehr als nur 10 relevante Parameter, und die Annahme, dass 10 % aller Personen diesen Kriterien *exakt* entsprechen, ist eher auch zu optimistisch. Realistischer ist, dass ein idealer Wert mehr oder weniger knapp verfehlt wird, eine Nase z. B. etwas zu groß oder etwas zu klein ist.

An dieser Stelle setzt das Prinzip des Morphens mit seiner Berechnung eines Durchschnittsgesichts an: Durch die Bildung von Mittelwerten heben sich die „Fehler“ gegenseitig auf, so dass das Gesamtergebnis dem „wahren Wert“, dem Ideal, näher kommt als jedes individuelle einzelne Gesicht. Konkret für den „Raffael-Effekt“ der Virtuellen Miss Germany heißt dies: Auch die Originalgesichter der einzelnen Missen sind zwar attraktiv, aber trotz der Tatsache, dass sie im Finale der Miss-Germany-Wahl teilnahmen, bei weitem nicht perfekt. Doch durch die Durchschnittsbildung beim Morphen wird bei allen acht Missen das Untypische, die individuelle Abweichung vom Ideal, der „Fehler“, herausgemittelt. Dadurch wird sichtbar, was das Gemeinsame aller Missen ist und was das Charakteristische ist, durch das sie sich vom Durchschnitt normaler Frauen unterscheiden (vgl. Abbildung 38).

Der „Raffael-Effekt“ zeigt also, dass es möglich ist, ein hochattraktives Gesicht herzustellen, indem man mehrere attraktive Gesichter zu einem Gesamt-Gesicht vereint. Doch nur, weil man in der Lage ist, Schönheit herzustellen, bedeutet dies noch nicht, dass man auch verstanden hat, was sie ausmacht. Hierzu fehlt noch eine entsprechende Analyse. Die formulierte Alternativhypothese weist auf den entscheidenden Punkt hin: „Ein Gesicht ist dann attraktiv, wenn es die Merkmale aufweist, die für attraktive Gesichter typisch und für unattraktive oder mittelmäßig attraktive Gesichter untypisch sind.“

Was also noch fehlt, ist, dieses Typische oder Charakteristische herauszuarbeiten. Ein simples Beispiel: Die Virtuelle Miss Germany hat zwei Augen. Daraus folgt jedoch noch nicht, dass das Merkmal „zwei Augen“ bereits ein Schönheitskriterium ist. Denn das Durchschnittsgesicht hat schließlich auch zwei Augen. Entscheidend ist also, was die Virtuelle Miss Germany hat, was das Durchschnittsgesicht nicht hat.

Dass zu Schönheit das gewisse Extra gehört, das sich vom bloßen Durchschnitt abhebt, schrieb bereits Kant vor über 200 Jahren (Kant, 1796): „Das Mittelmaß scheint das Grundmaß und die Basis der Schönheit, aber noch lange nicht die Schönheit selbst zu sein, weil zu dieser etwas Charakteristisches erfordert wird“ (zitiert nach Grammer, 2000,

S. 174). Mit der Analyse und Identifizierung dieses „Charakteristischen“ befasst sich das folgende Kapitel.

3.3 Deskriptive Analyse zur Identifikation von „Sexy Features“

Was ist aber nun das Besondere, das die Virtuelle Miss Germany so attraktiv macht und von einem bloß durchschnittlichen Aussehen unterscheidet? Was hat dieses Gesicht, was das weibliche Durchschnittsgesicht aus 64 normalen jungen Frauengesichtern nicht hat?

Um dies herauszufinden, wurden die beiden Gesichter aus Abbildung 38 einer qualitativen Analyse unterzogen. Diese Analyse erfolgte ohne Befragung von Versuchspersonen, da es in diesem Fall nicht um subjektive Empfindungen (wie ein ästhetisches Urteil), sondern um ein Beschreiben objektiv vorhandener Tatsachen ging. Alle gefundenen Unterschiede ließen sich auf folgende drei Merkmalskategorien zurückführen:

- Qualitative Unterschiede (Farben)
- Quantitative Unterschiede von Einzelmerkmalen (Größe, Länge, Dicke)
- Unterschiede in Gesichtsproportionen

Die Analyse wurde nicht nur mit der Virtuellen Miss Germany und dem weiblichen frontalen Durchschnittsgesicht durchgeführt, sondern zwischen

- allen *8-face-composites* aus den acht attraktivsten Originalgesichtern und dem jeweiligen Durchschnittsgesicht und
- allen *4-face-composites* aus den vier attraktivsten Originalgesichtern und den jeweiligen *4-face-composites* aus den vier *unattraktivsten* Originalgesichtern.

Insgesamt wurden also folgende Composites miteinander verglichen:

Tabelle 5: Übersicht der Composites, die in der qualitativen Analyse zur Identifikation von Attraktivitätsmerkmalen miteinander verglichen wurden

	Frauengesichter	Männnergesichter
Frontalansicht	<p>„w(1-64)“ versus „Virtuelle Miss Germany“ → Abbildung 38, S. 85</p> <p>„w(1-64)“ versus „w(57-64)“ → Abbildung 22, S. 49</p> <p>„w(1-4)“ versus „w(61-64)“ → Abbildung 21, S. 48</p>	<p>–</p> <p>„m(1-32)“ versus „m(25-32)“ → Abbildung 6, S. 21</p> <p>„m(1-4)“ versus „m(29-32)“ → Abbildung 5, S. 20</p>
Profilansicht	<p>„w(1-32)“ versus „w(25-32)“ → Abbildung 27, S. 58</p> <p>„w(1-4)“ versus „w(29-32)“ → Abbildung 26, S. 57</p>	<p>„m(1-32)“ versus „m(25-32)“ → Abbildung 11, S. 29</p> <p>„m(1-4)“ versus „m(29-32)“ → Abbildung 10, S. 28</p>

Zwar ist aus theoretischer Sicht der Vergleich zwischen einem attraktiven 8-face-composite und einem Durchschnittsgesicht aussagekräftiger, weil als Kontrollbedingung dabei das normale, durchschnittliche Aussehen dient. Aus praktischer Sicht ist der Vergleich jedoch teilweise schwierig durchzuführen, weil die optischen Unterschiede zwischen diesen Gesichtern nur klein sind. Dies gilt v. a. für jene Composites, bei denen das Gesamtdurchschnittsgesicht nur aus 32 Originalgesichtern berechnet wurde, denn die acht attraktivsten Gesichter sind ja eine Teilmenge der 32 Originalgesichter. Daher besteht das *32-face-composite* auch zu 25 % aus dem *8-face-composite*, sieht diesem also zwangsläufig relativ ähnlich. Anders ist dies bei dem *64-face-composite* aus weiblichen Frontalgesichtern, das das attraktive *8-face-composite* nur zu 12.5 % enthält. Noch einfacher ist es beim Vergleich des weiblichen *64-face-composite* mit der „Virtuellen Miss Germany“, da diese ja aus einer anderen Stichprobe an Gesichtern berechnet wurde und daher in dem Durchschnittsgesicht aus „normalen“ Frauen gar nicht enthalten ist. Aus diesem Grund wurden zusätzlich auch die *4-face-composites* aus den jeweils attraktivsten bzw. unattraktivsten Originalgesichtern miteinander verglichen. Denn hier sind Unterschiede besonders leicht erkennbar.

Tabelle 6: Ergebnisse der deskriptiven Analyse von Composites in der *Frontalansicht*. Unterschiede zwischen Composites aus hochattraktiven Gesichtern und durchschnittlichen Composites bzw. Composites aus unattraktiven Gesichtern, getrennt nach Frauengesichtern und Männergesichtern. Aufgelistet sind die Merkmale der attraktiven Composites – für die unattraktiven gelten entgegengesetzte Merkmale

	Frauengesichter	Männergesichter
Qualitative Unterschiede (Farben)	braunere Haut dunklere Wimpern dunklere Augenbrauen hellere Sklera (Augenweiß) keine dunklen Augenringe	braunere Haut – dunklere Augenbrauen hellere Sklera (Augenweiß) –
Quantitative Unterschiede von Einzelmerkmalen (Größe, Länge, Dicke)	schmalere Augenbrauen nach lateral stärker ansteigende Augenbrauen nach lateral sich stärker verjüngende Augenbrauen dichtere Wimpern längere Wimpern kleinerer Abstand zwischen Augenlid und Lidfalte	– – – dichtere Wimpern längere Wimpern –

Fortsetzung

Fortsetzung von Tabelle 6:

	Frauengesichter	Männergesichter
Quantitative Unterschiede von Einzelmerkmalen (Größe, Länge, Dicke)	stärker ausgeprägte Wangenknochen	stärker ausgeprägte Wangenknochen
	konkavere Wangen	konkavere Wangen
	schmalerer Nasenrücken	breiterer Nasenrücken
	kleinerer Abstand der Nasenflügel	kleinerer Abstand der Nasenflügel
	–	schwächere Nasolabialfalten
	kürzeres Philtrum (Abstand zwischen Nase und Oberlippe, <i>sn–ls</i>)	–
	vollere Lippen	vollere Lippen
	–	symmetrische Oberlippe (<i>crista philtri</i> auf gleicher Höhe)
	schmalerer Hals	schmalerer Hals
	längerer Hals	–
	–	keine Geheimratsecken
	markantere Kieferwinkel	markantere Kieferwinkel
	–	markanteres Kinn mit größerem Abstand zwischen Mund und Kinnunterseite (<i>sto–gn</i>)

Fortsetzung

Fortsetzung von Tabelle 6:

	Frauengesichter	Männergesichter
Unterschiede in Gesichtsproportionen	<p>schmaleres Gesicht (größeres Verhältnis Länge : Breite)</p> <p>größeres Verhältnis der vertikalen Distanz zwischen Scheitel und Haaransatz zu Stirnhöhe ($v-tr : tr-g$)</p> <p>kürzeres Mittel- und Untergesicht im Verhältnis zum gesamten Schädel ($n-gn : v-gn$)</p>	<p>schmaleres Gesicht (größeres Verhältnis Länge : Breite)</p> <p>größeres Verhältnis der vertikalen Distanz zwischen Scheitel und Haaransatz zu Stirnhöhe ($v-tr : tr-g$)</p> <p>längerer Unterkiefer im Verhältnis zum Obergesicht ($sn-gn : tr-n$)</p>

Anmerkung: Die aufgelisteten Merkmale basieren bei Frauengesichtern auf den Vergleichen von „w(1-64)“ versus „Virtuelle Miss Germany“ (→ Abbildung 38, S. 85), „w(1-64)“ versus „w(57-64)“ (→ Abbildung 22, S. 49) und „w(1-4)“ versus „w(61-64)“ (→ Abbildung 21, S. 48) sowie bei Männergesichtern auf den Vergleichen von „m(1-32)“ versus „m(25-32)“ (→ Abbildung 6, S. 21) und „m(1-4)“ versus „m(29-32)“ (→ Abbildung 5, S. 20). Die Bedeutungen der Abkürzungen für einzelne *Facial Landmarks* sowie eine Skizze finden sich in Anhang A (S. 393).

Tabelle 7: Ergebnisse der deskriptiven Analyse von Composites in der *Profilansicht*. Unterschiede zwischen Composites aus hochattraktiven Gesichtern und durchschnittlichen Composites bzw. Composites aus unattraktiven Gesichtern, getrennt nach Frauengesichtern und Männergesichtern. Aufgelistet sind die Merkmale der attraktiven Composites – für die unattraktiven gelten entgegengesetzte Merkmale

	Frauengesichter	Männergesichter
Qualitative Unterschiede (Farben)	keine dunklen Augenringe	–
Quantitative Unterschiede von Einzelmerkmalen (Größe, Länge, Dicke)	keine Geheimratsecken	keine Geheimratsecken
	–	weniger fliehende Stirn
	–	weniger protrusive Orbitalwulst
	–	weniger protrusive Nasenwurzel
	–	konkav (statt konvex) gekrümmter Nasenrücken
	–	weniger protrusive Nasenspitze
	schmalere Augenbrauen	–
	nach lateral stärker ansteigende Augenbrauen	–
	nach lateral sich stärker verjüngende Augenbrauen	–
	dichtere Wimpern	–
	längere Wimpern	–

Fortsetzung

Fortsetzung von Tabelle 7:

	Frauengesichter	Männergesichter
	schwächere Nasolabialfalten	–
	vollere Oberlippe	–
	vollere Unterlippe	–
	–	markanteres Kinn mit größerem Abstand zwischen Mund und Kinnunterseite (<i>sto–gn</i>)
	konturierter (kleinerer) Kiefer-Hals-Winkel (<i>gn–cp–h</i>)	konturierter (kleinerer) Kiefer-Hals-Winkel (<i>gn–cp–h</i>)
	schmalerer Hals	–
	längerer Hals	–
Unterschiede in Gesichtsproportionen	Oberkiefer im Verhältnis zum Unterkiefer nicht zurückliegend	–
	–	in der Vertikalen längerer Unterkiefer im Verhältnis zum Obergesicht (<i>sn–gn : tr–n</i>)

Anmerkung: Die aufgelisteten Merkmale basieren bei Frauengesichtern auf den Vergleichen von „w(1-32)“ versus „w(25-32)“ (→ Abbildung 27, S. 58) und „w(1-4)“ versus „w(29-32)“ (→ Abbildung 26, S. 57) sowie bei Männergesichtern auf den Vergleichen von „m(1-32)“ versus „m(25-32)“ (→ Abbildung 11, S. 29) und „m(1-4)“ versus „m(29-32)“ (→ Abbildung 10, S. 28). Die Bedeutungen der Abkürzungen für einzelne *Facial Landmarks* sowie eine Skizze finden sich in Anhang A (S. 393).

Ein Blick auf Tabelle 6 mit den Unterschieden der Frontalgesichter und Tabelle 7 mit den Unterschieden der Profilgesichter sowie ein aufmerksames Vergleichen dieser Kriterien

mit den Abbildungen der jeweiligen Gesichter machen deutlich, dass es eine Fülle von Einzelmerkmalen gibt, durch die sich attraktive Gesichter von bloß durchschnittlichen (bzw. unattraktiven) unterscheiden. Zwar sind die Unterschiede oft nur dezent, doch die Summe der Details macht letztlich den Unterschied aus.

Diese „Sexy Features“ lassen sich größtenteils auf zwei übergeordnete Attraktivitätsprinzipien zurückführen, nämlich auf

1. Jugendlichkeit und
2. Sexuellen Dimorphismus bzw. typisch feminine Merkmale.

In die Kategorie „Jugendlichkeit“ (1) fallen folgende Merkmale:

- Keine Geheimratsecken (vgl. nicht nur „m(1-4)“, Abbildung 5, S. 20, sondern auch „w(1-4)“, Abbildung 26, S. 57), da mit zunehmendem Alter – auch bei Frauen – das Haar an Dichte verliert und sich der Haaransatz nach oben verschiebt
- Eine hellere Sklera, da bei älteren Personen das Augenweiß eintrübt, also an Helligkeit verliert und durch Hineinwachsen von Äderchen in Richtung Iris der Farbton weniger farbneutral wird (Denniston & Murray, 2009; Gründl, Knoll, Eisenmann-Klein & Prantl, 2011)
- *Keine* dunklen Augenringe (vgl. „w(1-4)“, Abbildung 21, S. 48), da im Alter mit der Abnahme und dem Absacken des Fettgewebes unterhalb des Auges Blutgefäße stärker durchscheinen (Coleman & Grover, 2006; Gosain, Klein, Sudhakar & Prost, 2005; Hoenig, Shorr & Shorr, 1997)
- Dichtere und längere Wimpern, da im Alter die Anzahl und Länge der Wimpern abnimmt
- Ein kleinerer Abstand zwischen Augenlid und Lidfalte (vgl. „w(1-4)“, Abbildung 21, S. 48), da es im Alter zu einem Abbau des suborbitalen Fettgewebes kommt, was v. a. bei schlanken Frauen zu einem nach oben Wandern der Augenlidfalte führt (Fagien, 2002; Heinz & Kikkawa, 1997; Jelks & Jelks, 1993; Rohrich, Coberly, Fagien & Stuzin, 2004) (– bei übergewichtigen Personen und bei Männern kommt es hingegen durch das erschlaffte überschüssige Hautgewebe eher zu Schlupflidern)
- Eine größere Steigung der Augenbrauen in Richtung lateral, da es im Alter mit dem Erschlaffen des Bindegewebes und der Haut zu einem Absacken des lateralen Teils der Augenbraue kommt (Alex, 2004; Cartwright, Kurumety, Nelson, Frueh & Musch, 1992, 1994; Friedman, 2005; Westmore, 1974; Yaremchuk, 2004)

- Ein leicht konkaver statt konvexer Nasenrücken mit leicht angehobener Nasenspitze (vgl. Abbildung 10, S. 28), da im Alter durch den Abbau von Knochensubstanz am Nasenbein und das Absinken der Nase die Tendenz zu einem sog. „Altershöcker“ besteht und die Nasenspitze ebenfalls nach unten sinkt (Coleman & Grover, 2006; Guyuron, 1997)
- Schwächer ausgeprägte Nasolabialfalten (vgl. „m(1-4)“, Abbildung 5, S. 20), da Falten generell ein Alterskriterium sind
- Ein kürzeres Philtrum (Abstand zwischen Nase und Oberlippe, *sn-ls*), da es im Alter zu einem Abbau des perioralen Fettgewebes kommt, wodurch die gesamte (anatomische) Oberlippe (nicht nur das sichtbare Lippenrot) schlaffer nach unten hängt, die oberen Schneidezähne stärker verdeckt und dadurch die Distanz zwischen Subnasale (*sn*) und dem Beginn des Lippenrots an der Oberlippe (*ls*) vergrößert (Coleman & Grover, 2006)
- Vollere Lippen, da im Alter durch den Abbau des perioralen Fettgewebes das Lippenvolumen und das sichtbare Lippenrot abnehmen (Donofrio, 2000; Zimble, Kokoska & Thomas, 2001)
- *Keine* Rücklage des Oberkiefers im Verhältnis zum Unterkiefer, da es im hohen Alter nach dem Verlust der eigenen Zähne zu einem Abbau der Knochensubstanz an den Kiefern kommt, in stärkerem Ausmaß jedoch am Oberkiefer (Maxilla) (Bartlett, Grossman & Whitaker, 1992)
- Ein konturierter (kleinerer) Kiefer-Hals-Winkel (*gn-cp-h*), da es im Alter durch das Erschlaffen des Gewebes an den Wangen und im Bereich des Unterkiefers zu einem schlaffen Herabhängen der Hautpartien unterhalb des Unterkiefers kommt (Coleman & Grover, 2006; Friedman, 2005; Hamra, 1994)

In die Kategorie „Sexueller Dimorphismus“ (2), also geschlechtstypisches Aussehen, fallen bei *Frauengesichtern* folgende Merkmale:

- Schmalere Augenbrauen, die nach lateral stärker ansteigen und sich stärker verjüngen, da Frauen von Natur aus weniger buschige Augenbrauen haben als Männer und zusätzlich einem kulturell bedingten Ideal folgen und sich die Augenbrauen in die entsprechende Form zupfen (Tipple, Atkinson & Young, 2002; Volpe & Ramirez, 2005)
- Ein schmalere Nasenrücken und ein kleinerer Abstand der Nasenflügel, da bei Frauen im Unterschied zu Männern im Durchschnitt die Nase nicht nur weniger protrusiv, sondern auch insgesamt kleiner und schmaler ist (Enlow, 1989; Farkas, 1994)

- Vollere Lippen, da diese ein Merkmal sind, das durch einen hohen Östrogenspiegel während der Pubertät ausgebildet wird (Johnston & Franklin, 1993; Symons, 1995; Thornhill & Moller, 1997)
- Ein schmalere und längerer Hals, da bei Frauen die Hals-, Nacken- und Schultermuskulatur im Durchschnitt schwächer ausgeprägt ist als bei Männern und weil Frauen einen kleineren Kehlkopf haben, wodurch zusätzlich in der Profilansicht der Hals optisch schmaler wirkt
- Ein größeres Verhältnis der vertikalen Distanz zwischen Scheitel und Haaransatz zu Stirnhöhe ($v-tr : tr-g$), da dies den optischen Eindruck eines großen Hirnschädels verstärkt, der als Merkmal des Kindchenschemas für eine weibliche Schädelform charakteristischer ist als für eine männliche (Eibl-Eibesfeldt, 1997; Jones, 1995)
- Ein kürzeres Mittel- und Untergesicht im Verhältnis zum gesamten Schädel ($n-gn : v-gn$), da dies ein Merkmal des Kindchenschemas ist, das sich bedingt durch den Einfluss von Testosteron während der Pubertät bei Männern in Richtung eines längeren Mittel- und Untergesichts verschiebt (Eibl-Eibesfeldt, 1997; Enlow, 1989; Farkas, 1994)
- Ein konturierter (kleinerer) Kiefer-Hals-Winkel ($gn-cp-h$), weil bei Frauen der Hals schlanker und der Kehlkopf kleiner ist, was zu einem kleineren Kiefer-Hals-Winkel beiträgt (vgl. dazu auch die Profilansichten des weiblichen Durchschnittsgesichts, Abbildung 27, S. 58, und des männlichen Durchschnittsgesichts Abbildung 11, S. 29)

Für *Frauengesichter* lassen sich also bei dieser deskriptiven Analyse etliche Merkmale identifizieren, die mit der Theorie des sexuellen Dimorphismus in Einklang stehen. Anders ist dies hingegen für *Männergesichter*. Denn wie der Vergleich der jeweiligen Spalten von Tabelle 6 und Tabelle 7 zeigt, treffen etliche der oben aufgelisteten Merkmale auch auf männliche Gesichter zu, *obwohl* es eigentlich typisch feminine Merkmale sind. Auch die attraktiven männlichen Composites unterscheiden sich von den unattraktiven männlichen Composites darin, dass sie vollere Lippen, einen schmaleren Hals und einen klarer konturierten (kleineren) Kiefer-Hals-Winkel ($gn-cp-h$) haben. Besonders deutlich ist dies bei der Nase: Eine zierlichere Nase (schmalere Nasenrücken, geringerer Abstand der Nasenflügel, weniger protrusive Nasenwurzel und Nasenspitze) ist offensichtlich auch ein Attraktivitätsmerkmal für Männer, obwohl eine zierliche Nase typisch für Frauengesichter ist. Umgekehrt sind Geheimratsecken, die fliehende Stirn und die ausgeprägte Orbitalwulst typisch männlich und trotzdem eine Eigenschaft der unattraktiven Männergesichter.

Insgesamt finden sich für Männergesichter weit mehr Indizien *gegen* die attraktivitätssteigernde Wirkung typisch maskuliner Merkmale als dafür. Insofern wäre es angemessener, nicht von sexuell dimorphen Merkmalen als „Sexy Features“ zu sprechen, sondern stattdessen von femininen Merkmalen.

Das einzige Merkmal, das eine klare Ausnahme darstellt, ist ein markanter, kräftiger Unterkiefer mit einem größeren vertikalen Abstand zwischen Mund und Kinn (*sto-grn*). Hier passen bei Männergesichtern die beobachteten Unterschiede zu den Vorhersagen der Theorie des sexuellen Dimorphismus. Diese wichtige Theorie der Attraktivitätsforschung wird jedoch in Kapitel 6 mit drei eigenen Experimenten untersucht.

Das angewandte Verfahren, Composites aus attraktiven Gesichtern mit Gesamtdurchschnittsgesichtern oder Composites aus unattraktiven Gesichtern als Kontrollbedingung zu vergleichen und deskriptiv herauszuarbeiten, hat jedoch auch Nachteile und Grenzen. Diese sind:

- Es gibt bestimmte Merkmale, die mit dieser Methode nicht gefunden werden können, obwohl sie möglicherweise sehr bedeutsam sind. Dies sind solche Merkmale, in denen sich die Composites aus technischen Gründen des Morphings nicht unterscheiden *können*, wie z. B.
 - im Attraktivitätskriterium der glatten und makellosen Haut (deren extrem wichtiger Einfluss ja bereits klar belegt wurde, vgl. Kapitel 2.6 und Kapitel 2.10), denn *alle* Composites – auch die aus nur vier unattraktiven Originalen – haben durch die unvermeidlichen Morphing-Artefakte eine glatte, makellose Haut
 - bei Haarfarbe und Augenfarbe, denn beim Vermorphen von verschiedenfarbigen Haarfarben und Augenfarben ist das Ergebnis immer dunkel, wodurch es keine blonden oder blauäugigen Composites gibt
- Effekte von Merkmalen, die in der Stimuluspersonen-Stichprobe nicht vertreten oder unterrepräsentiert waren – egal ob attraktive oder unattraktive Merkmale – (z. B. Falten, alle deutlichen Alterskennzeichen, Merkmale sehr maskuliner Männer) können sich auch nicht in den Composites wiederfinden. Die Zusammensetzung des Stimulusmaterials und dessen Konsequenzen auf die Ergebnisse ist zwar ein generelles Problem der Attraktivitätsforschung. Dennoch sollte im Hinterkopf behalten werden: Auch wenn ein bestimmtes Merkmal in einem hochattraktiven Composite nicht vorhanden ist, könnte es trotzdem attraktiv machen. Denn es könnte bloß deswegen nicht im Composite sichtbar sein, weil in der Stichprobe der fotografierten Personen

eben keine Person vertreten war, die dieses Merkmal besaß. Beispiel: Es könnte sein, dass sehr markante, wuchtige Unterkiefer ein Männergesicht attraktiv machen. Da dieses Merkmal jedoch in der Bevölkerung selten ist, war es in der fotografierten Stichprobe ($N = 33$) nicht vertreten.

- Da die Composites nur aus acht bzw. vier Originalgesichtern berechnet sind, das N damit klein ist, spielen Zufallseinflüsse für das Aussehen eines Composites eine relativ große Rolle. Für künftige Untersuchungen wären daher größere Stichproben von fotografierten Gesichtern oder die Verwendung bereits vorselektierter Stichproben (wie das Finale einer Miss-Germany-Wahl, dem mehrere Wettbewerbe auf Landes- und Regionalebene vorangehen) eine Abhilfe.
- Die Vergleiche der jeweiligen Prototypen können zwar Unterschiede beschreiben, sie können jedoch keine Aussage über die Wichtigkeit eines Merkmals machen.

Aus diesen Gründen sind die aufgelisteten „Sexy Features“ unter einem gewissen Vorbehalt zu sehen. Um sichergehen zu können, dass diese Merkmale tatsächlich auch kausal Attraktivität bedingen, sind zusätzliche Untersuchungen notwendig – am besten experimentelle Untersuchungen, bei denen gezielt ein oder mehrere Merkmale (z. B. mit Morphing-Software) manipuliert und die Auswirkungen dieser Veränderungen auf die Attraktivitätsbewertung evaluiert werden.

Daher ist die durchgeführte deskriptive Analyse auch weniger als Erkenntnismethode, sondern eher zur Generierung von Hypothesen geeignet. Hierfür jedoch erstaunlich gut – dafür spricht die große Anzahl an gefundenen Unterschieden sowie deren Plausibilität, da sie sich inhaltlich auf Merkmale der Jugendlichkeit und geschlechtstypische Merkmale (bzw. Femininität) zurückführen lassen. Als Fundgrube für neue Ideen künftiger Attraktivitätsstudien sind die in Tabelle 6 und Tabelle 7 aufgelisteten Unterschiede sehr wertvoll.

Dies ist umso bemerkenswerter, da die analysierten prototypischen Composites gar nicht das Ziel der durchgeführten Untersuchungen waren. Sie waren lediglich „Zwischenprodukte“ auf dem Weg, jeweils ein Gesamt-Durchschnittsgesicht zu morphen, um damit in den Untersuchungen *Durch 1* bis *Durch 6* die Durchschnittshypothese zu testen. Möglich wurden diese „Zwischenprodukte“ erst, weil bei den hier durchgeführten Experimenten das paarweise Morphen systematisch nach der Attraktivität der Originalgesichter erfolgte und nicht, wie bei bisherigen Studien, per Zufall. Erst dadurch entstanden durch den „Rafael-Effekt“ die hochattraktiven Composites, die sich empirisch als attraktiver als die jeweiligen Gesamt-Durchschnittsgesichter erwiesen. Durch die anschließende deskriptive Analyse konnten damit zahlreiche Merkmale identifiziert werden, bei denen zumindest ein

starker Verdacht besteht, dass sie tatsächlich Attraktivität kausal bedingen – Merkmale, die bisher allen Attraktivitätsforschern bei ihren Versuchen, die Gültigkeit der Durchschnittshypothese zu beweisen, verborgen geblieben sind.

Wie die Ergebnisse aus Kapitel 2.6 und Kapitel 2.10 zeigten, ist die Durchschnittshypothese nicht länger haltbar, da der Großteil des attraktivitätssteigernden Effekts auf Bildverarbeitungsartefakte beim Morphen zurückzuführen ist und durchschnittliche Gesichtsproportionen an sich nur einen sehr geringen Effekt ausmachen. Dieselben Composites, die die recht pauschale Durchschnittshypothese widerlegten, werfen jedoch gleichzeitig eine Fülle von neuen Fragen auf. Aus ihnen lassen sich unzählige neue, viel konkretere, plausiblere und anwendungsnähere Hypothesen formulieren, mit denen in zukünftigen Studien der Frage nachgegangen werden wird, was ein attraktives Gesicht ausmacht.

3.4 Vergleich der eigenen Ergebnisse mit bisherigen Untersuchungen zur Durchschnittshypothese

In den vorangegangenen Kapiteln wurde ausführlich dargelegt, dass der in dieser Untersuchungsreihe erzielte, scheinbar große attraktivitätserhöhende Effekt der Durchschnittlichkeit in Wahrheit größtenteils auf Artefakte des Morphens zurückzuführen ist (vgl. Kapitel 2.6 und 2.10). Durch den Vergleich verschiedener Stimulusarten (Frontalgesichter und Profilgesichter mit und ohne Textur), konnte klar nachgewiesen werden, dass die Ausgangsattraktivität der Originalgesichter hauptverantwortlich für die Attraktivität eines Composites ist (vgl. Kapitel 3.1.3) und dass die allgemeine Attraktivitätssteigerung bei Composites durch Bildverarbeitungsartefakte entsteht, weil dadurch die Haut von Composites glatt und makellos wird (vgl. Kapitel 3.1.2).

In diesem Kapitel wird dargelegt, warum bisherige Untersuchungen trotz teils sehr ähnlicher Methoden zu so ganz anderen Schlussfolgerungen gelangten. Die Probleme vieler anderer Studien lassen sich auf folgende Punkte zurückführen:

1. Einseitige Interpretation der Ergebnisse
2. Fehlende plausible Alternativhypothese
3. Fokussierung auf relative statt absolute Effekte
4. Konfundierung mit Bildverarbeitungsartefakten
5. Invalide Methoden zur Manipulation der Durchschnittlichkeit

6. Invalide Methoden zur Messung von Durchschnittlichkeit

(1) Einseitige Interpretation der Ergebnisse: Dass es attraktivere Gesichter als ein Durchschnittsgesicht gibt, ist kein Phänomen, das sich erst in dieser Untersuchungsreihe zeigte. Bereits im berühmten Artikel von Langlois und Roggman (1990), der die Durchschnittshypothese begründete, waren mehrere Originalgesichter signifikant attraktiver bewertet worden als die jeweiligen Gesamt-Durchschnittsgesichter. Diese Information steht nicht nur im Ergebnisteil des Artikels, sondern sogar in der Zusammenfassung. Diese interessante Beobachtung wurde jedoch als statistisches Rauschen abgetan, ebenso wie bei ihrem neueren Experiment vier Jahre später (Langlois, Roggman, Musselman, 1994, S. 218). Stattdessen fokussieren die Autorinnen auf Durchschnittswerte und unterstreichen, dass in ihrer Arbeit *32-face-composites* im Durchschnitt (!) besser bewertet wurden als Originalgesichter im Durchschnitt (!). Aus den dort berichteten Minima und Maxima der Originalgesichter lässt sich sogar ersehen, dass diese Attraktivitätsunterschiede noch nicht einmal gering waren – detailliertere Angaben fehlen bei ihnen jedoch (In den sechs eigenen Untersuchungen der vorliegenden Arbeit wurden hingegen extra Streudiagramme abgedruckt, damit der Leser auf den ersten Blick erkennen kann, wie viele andere Gesichter attraktiver als das Gesamt-Durchschnittsgesicht waren).

Anstatt jedoch der Frage nachzugehen, was diese attraktiven Einzelgesichter haben, was die Durchschnittsgesichter nicht haben (wie hier im Kapitel 3.3, Deskriptive Analyse zur Identifikation von „Sexy Features“) ignorierten Langlois und Roggman ihren eigenen Befund und interpretierten lediglich Durchschnittswerte. Dabei ist der Umstand, dass es sich nur um einige wenige Gesichter handelte, ein schlechtes Argument. Denn hohe Attraktivität ist nun mal ein Merkmal, das in der Bevölkerung selten vorkommt.

Dass es Attraktivitätsforscher gibt, die offensichtlich keinen Blick oder kein echtes Interesse für Attraktivität haben und die interessantesten Aspekte in ihren eigenen Daten übersehen, ist eigentlich schon bedauerlich genug, doch neun Jahre später überbot Langlois dies in einer weiteren Studie zur Durchschnittshypothese. Dort argumentierten Rubinstein, Kalakanis and Langlois (1999) dreist, dass es doch gar kein Widerspruch zur Theorie sei, dass einige wenige Gesichter attraktiver als das absolute Durchschnittsgesicht seien. Diese Gesichter seien eben deshalb so attraktiv, weil sie zufälligerweise schon sehr durchschnittlich aussähen. Auf die Tatsache, dass ein Originalgesicht ihrer Stichprobe gleich attraktiv wie ihr Composite aus 32 Gesichtern war, entgegneten sie:

We would not expect a significant difference in attractiveness ratings between the most attractive individual faces and the 32-face averaged face be-

cause, by definition, the most attractive individual faces should be the closest to the mathematical average of the population in configuration. (Rubinstein, Kalakanis & Langlois, 1999, p. 849)

Das Argument der Autorinnen mit anderen Worten: Composites sind schön, weil sie durchschnittlich sind. Wenn wider Erwarten ein Nicht-Composite trotzdem schön ist, dann kann es nur daran liegen, dass es zufälligerweise auch durchschnittlich ist. Diese Argumentation klingt recht bizarr, weil sie einen Zirkelschluss beinhaltet nach dem Motto: Ein Gesicht ist schön, weil es durchschnittlich ist, und ein Gesicht ist durchschnittlich, weil es schön ist. Die Autorinnen setzen also etwas voraus, was sie erst beweisen wollen. Leider enthält der Artikel von Rubinstein et al. keine Abbildungen, sonst hätte sich jeder Leser selbst ein Bild davon machen können, ob das schönste Originalgesicht tatsächlich durchschnittlich aussieht. An den Abbildungen der hier durchgeführten Experimente (vgl. Abbildung 22, S. 49, Abbildung 27, S. 58, Abbildung 6, S. 21 und Abbildung 11, S. 29) kann jedenfalls jeder sehen, dass die Composites aus den Schönsten anders aussehen als die Composites aus allen und es wurde statistisch nachgewiesen, dass sie höhere Attraktivitätsbewertungen erhalten.

(2) Fehlende plausible Alternativhypothese: In allen bisherigen Untersuchungen, in denen die Durchschnittlichkeit eines Gesichts experimentell manipuliert wurde (DeBruine et al., 2007; Langlois & Roggman, 1990; Langlois et al., 1994; Perrett et al., 1994; Pollard, 1995; Rhodes, Sumich et al., 1999; Rhodes, Yoshikawa et al., 2001; Valentine et al., 2004), fehlte es an einer plausiblen Alternativhypothese. Meist wurden nur die Gesamtdurchschnittsgesichter gegen die Originalgesichter getestet. Der beobachtete (durchschnittliche!) Attraktivitätsgewinn aufgrund von Morphing-Artefakten der makelloseren Haut bzw. Experimenten mit „Warping-Verfahren“ (d. h. nur durch Verzerrung der Gesichtsproportionen bei konstanter Textur) aufgrund der Harmonisierung einzelner stark von der Norm abweichender Gesichtsproportionen wurde dann dahingehend interpretiert, dass durchschnittliche Gesichtsproportionen das Schönheitsideal schlechthin seien.

In der vorliegenden Untersuchung hingegen wurde eine extra Alternativhypothese formuliert, die separat zur Durchschnittshypothese getestet wurde, nämlich: „Ein gemorphtes Gesicht ist nur so attraktiv wie die Originalgesichter, die darin enthalten sind“. Durch das systematische paarweise Morphen nach Attraktivität der Originalgesichter (anstatt planlos per Zufall wie bei allen anderen Untersuchungen) konnte nicht nur (wie bei anderen Untersuchungen) ein Gesamt-Durchschnittsgesicht berechnet werden; zusätzlich resultierten auch zahlreiche „Zwischenergebnisse“, die aus unterschiedlich attraktiven Originalgesichtern berechnet waren. Auch konnte nachgewiesen werden, dass Composites aus

attraktiven Originalen attraktiver als das Gesamt-Durchschnittsgesicht waren und Composites aus unattraktiven Originalen entsprechend unattraktiver.

Andere Untersuchungen gaben jedoch dieser plausiblen Alternativhypothese („Durchschnittsgesichter sind durchschnittlich attraktiv“) von vornherein gar keine Chance für eine Bestätigung. Wenn man eine bestimmte Frage gar nicht erst stellt, erhält man natürlich auch keine Antwort darauf. Aus diesem Grund kamen andere Untersuchungen zu der Schlussfolgerung, dass Durchschnittlichkeit an sich bereits ein Attraktivitätskriterium ist, weil ihnen der signifikante durchschnittliche Attraktivitätszuwachs gemorphter Gesichter als Beleg bereits hinreichend erschien.

(3) Fokussierung auf relative statt absolute Effekte: Dass heutzutage unter Attraktivitätsforschern der Konsens vorherrscht, Durchschnittlichkeit mache attraktiv, mag noch einen weiteren sehr einfachen Grund haben. In den entsprechenden Untersuchungen ist auf Seiten der Autoren offenkundig eine geringe Motivation erkennbar, bei der Interpretation ihrer Daten sich in Erinnerung zu rufen, was ihre selbst gemessenen Attraktivitätsurteile eigentlich inhaltlich bedeuten. Dies beginnt bereits mit dem „Klassiker“ von Langlois und Roggman (1990). Für die Attraktivitätsmessung wurde dort eine Skala von 1 (= sehr unattraktiv) bis 5 (= sehr attraktiv) verwendet. Die neutrale Mitte der Skala lag also bei der Stufe 3. Im Ergebnis zeigte sich, dass die insgesamt sechs *32-face-composites* im Durchschnitt nur den Wert 3.27 (Männerngesichter) bzw. 3.25 (Frauengesichter) erreichten, also Bewertungen bekamen, die nur im Mittelfeld der Bewertungsskala lagen.

Jeder unvoreingenommene Leser, der „nur“ mit seinem gesunden Menschenverstand urteilt, würde daraus folgern, dass Durchschnittsgesichter eben nur mittelmäßig attraktiv sind. Doch die Autorinnen interpretieren das Ergebnis ganz anders und rücken den Umstand in den Mittelpunkt, dass diese Composites eben *auch* statistisch signifikant attraktiver bewertet wurden als der Durchschnitt der Originalgesichter. Anstatt die Urteile ihrer Versuchspersonen ernst zu nehmen, ignorierten sie die *Bedeutung* ihrer gemessenen Zahlenwerte und diskutierten statt des gemessenen *absoluten* Effekts lediglich den *relativen* Effekt der Verdurchschnittlichung von Gesichtern. So kamen sie schließlich zum spektakulär klingenden Titel ihres Artikels „Attractive faces are only average“, obwohl ihre Durchschnittsgesichter lediglich als mittelmäßig attraktiv bewertet worden waren.

Insofern ist es erstaunlich, dass dieses Experiment trotz der nur mäßigen Ergebnisse zu so vielen Folgeuntersuchungen geführt hat und die Durchschnittshypothese mittlerweile als die am besten bestätigte Theorie der Attraktivitätsforschung gilt (vgl. Metaanalyse von Rhodes, 2006). Aber auch bei späteren Untersuchungen war diese Art der Dateninterpre-

tation der Normalfall und nicht etwa eine Ausnahme – dafür genügt ein Blick in die jeweiligen Ergebnisteile der Artikel.

Ein paar Beispiele: Rubenstein, Kalakanis und Langlois (1999) benutzten für Attraktivitätsratings eine 5er-Skala (5 = sehr attraktiv) und ihr *32-face-composite* wurde lediglich mit dem Wert 3.38 bewertet (S. 849, Angaben zu Minimum und Maximum bei den Originalgesichtern fehlen leider, berichtet wird nur der Mittelwert aller Originalgesichter). In der Untersuchung von Rhodes, Sumich und Byatt (1999) lag die Bewertung für verdurchschnittlichte Gesichter sogar unterhalb der Mitte der 7er-Skala (S. 56). In einer weiteren Studie von Rhodes und Tremewan (1996) lagen verdurchschnittlichte Gesichter im Experiment 1 nur knapp oberhalb der Mitte der 7er-Attraktivitätsskala (S. 107); in ihrem Experiment 2, das als Versuch gedacht war, mit Strichzeichnungen die Ergebnisse von Langlois und Roggman (1990) zu replizieren, erreichten die drei männlichen *32-face-composites* im Mittel nur den Wert 4.0 auf einer 7er-Skala, die drei weiblichen *32-face-composites* erreichten sogar nur den Wert 3.7 – die Ergebnisse waren im Übrigen auch nicht statistisch signifikant und der Replikationsversuch scheiterte (S. 108 f). Und in der Untersuchung von Rhodes, Yoshikawa et al. (2001) erhielten verdurchschnittlichte Gesichter (*high average faces*) auf einer 10er-Skala (10 = sehr attraktiv) im Mittel nur Bewertungen um den Wert 4 (die Mitte der Skala lag bei 5.5, S. 616), also Bewertungen deutlich im unattraktiven Bereich der Skala. Im Experiment von Grammer und Thornhill (1994) wurden alle Durchschnittsgesichter als eher unattraktiv bewertet (S. 238); auf einer 7er-Skala (7 = sehr attraktiv) bekamen weibliche Composites im Mittel nur den Wert 3.26 und männliche sogar nur 2.88 (Durchschnittsgesichter von Männern waren sogar signifikant noch unattraktiver als männliche Originalgesichter).

Selbstverständlich ist es nicht falsch, auf relative Unterschiede zu fokussieren, wenn man sich eben dafür interessiert, erst recht nicht, wenn diese Unterschiede durch experimentelle Manipulation der unabhängigen Variablen hervorgerufen wurden. Andererseits ist diese Art der Dateninterpretation aber wenig hilfreich für jemanden, der ernsthaft verstehen möchte, was ein hochattraktives Gesicht ausmacht.

(4) Konfundierung mit Bildverarbeitungsartefakten: In den hier durchgeführten Experimenten wurden die Ergebnisse von Profilgesichtern mit und ohne Textur miteinander verglichen, und es wurde nachgewiesen, dass ohne die morphingtypischen Bildverarbeitungsartefakte der glatten, makellosen Haut der Effekt der Durchschnittlichkeit zusammenbricht (vgl. Kapitel 2.6 und 2.10). In Kapitel 3.1.2 wurde erklärt und mit Abbildungen

illustriert, *warum* dies so ist. Diese Argumente sollen an dieser Stelle nicht noch einmal wiederholt werden. Es sei daher nur auf die entsprechenden Kapitel verwiesen.

Die Kritik an der Gleichsetzung gemorphter Composites mit Durchschnittlichkeit ist so alt wie die Durchschnittshypothese selbst (Alley & Cunningham, 1991; Pittenger, 1991). Die Bemühungen, diese Einwände zu entkräften (wie es z. B. Langlois et al., 1994, in ihrem Artikel versuchen), sind insbesondere bei den typischen Hautartefakten wenig überzeugend. So wiederholten Langlois und Roggman im Standardwerk über Attraktivitätsforschung von Rhodes und Zebrowitz (2002) im Kapitel über die Durchschnittshypothese ihre schwachen Argumente:

In our facial averaging studies, however, individual faces are always "smoothed and blurred" to match any smoothing and blurring that occurs in averaged faces as a result of the computer manipulation (Langlois & Roggman, 1990; Langlois et al., 1994; Rubenstein et al., 1999). (Rubenstein, Langlois & Roggman, 2002, S. 12)

Mit dieser Behauptung zeigen sie damit nur, dass es ihnen an einem tieferen Verständnis für Bildverarbeitung mangelt und sie auch zwölf Jahre später immer noch nicht den Unterschied zwischen Morphing und einem Gaußschen Weichzeichnungsfilter begriffen haben. Ein Weichzeichnungsfilter ahmt keineswegs das nach, was beim Morphen mit der Textur von Gesichtern passiert (vgl. dazu Kapitel 3.1.2 sowie insbesondere Abbildung 35 bis Abbildung 37).

(5) Invalide Methoden zur Manipulation der Durchschnittlichkeit: Ein weiterer Grund für Ergebnisse anderer Studien, die Durchschnittlichkeit als Attraktivitätskriterium zu bestätigen scheinen, ist die Verwendung von ungeeigneten Verfahren zur Manipulation von Durchschnittlichkeit. Es handelt sich um solche Studien, die die *caricature technique* benutzen, um Gesichter zu manipulieren (Deffenbacher, Vetter, Johanson & O'Toole, 1998; Rhodes, Sumich et al., 1999; Rhodes & Tremewan, 1996; Rhodes, Yoshikawa et al., 2001; Rhodes, Zebrowitz et al., 2001). Bei der Karikatur-Methode werden Gesichter (meist Originalgesichter) in ihren Proportionen von den Proportionen des Durchschnittsgesichts (desselben Geschlechts) weg verzerrt (meist zu einem Grad von 50 %), während die Textur der Gesichter unverändert bleibt. Der Grundgedanke ist zwar gut, nämlich Gesichtsproportionen zu verändern, ohne sich dabei die typischen Morphing-Artefakte der glatten Haut einzuhandeln. Aber die Gefahr ist sehr groß, mit dieser Karikatur-Technik Gesichter zu produzieren, die tatsächlich aussehen wie Karikaturen. Dieses Verfahren wurde (und wird weiterhin) auch zur Verstärkung von Asymmetrien und zum Herstellen von sexuell dimorphen Gesichtsproportionen eingesetzt (vgl. Kapitel 6.1.3. und 7.3). Ins-

besondere in letzterem Zusammenhang wurde das Verfahren häufig wegen fehlender Validität kritisiert (Enlow, 1966, 1979; Grumbach, 2000; Johnston, Hagel, Franklin, Fink & Grammer, 2001; Meyer & Quong, 1999; Tanner, 1978).

Mit Morphing-Software ist es auch möglich, Originalgesichter an die Proportionen eines Durchschnittsgesichts anzunähern, anstatt sie davon zu entfernen (in Kapitel 8 wurde dies ebenfalls durchgeführt). Dieses Verfahren (*anti-caricature*) ist weitaus geeigneter, da hierbei keine unrealistischen und seltsam aussehenden Verzerrungen auftreten. Denn die Proportionen eines Originalgesichts an ein Schema anzunähern, ist etwas anderes, als sie von einem Schema zu entfernen, da bei der Entfernung bestehende Unterschiede zwischen zwei Gesichtern nicht verkleinert, sondern vergrößert werden. Bei der Vergrößerung ist der Software-Algorithmus jedoch auf Interpolationen angewiesen, die erstens ungenau sind und zweitens bei Gesichtern zu unrealistischen Resultaten führen, weil die Software lediglich Abstände zwischen definierten Punkten in einer Grafik verändert ohne dabei zu berücksichtigen, was anatomisch in einem Gesicht überhaupt möglich ist und was nicht. Hat beispielsweise eine Person größere Augen als das Durchschnittsgesicht, werden bei der Karikatur die Augen noch mehr vergrößert (inklusive Iris und Pupille), selbst dann, wenn es solch große Augen in einem Gesicht in der Realität nicht geben kann.

In den oben genannten Studien werden beide Techniken gleichzeitig eingesetzt, *anti-caricature-faces*, unveränderte Originalgesichter und *caricature faces*. Durch das Karikieren der Originalgesichter wird zuverlässig das Ergebnis produziert, dass die karikierten Gesichter unattraktiver bewertet werden. Da das Karikieren dadurch erfolgte, dass die Originalgesichter in ihren Proportionen vom Durchschnittsgesicht weg verzerrt wurden, also *weniger* durchschnittlich gemacht wurden, wird daraus typischerweise die Schlussfolgerung gezogen, dass Durchschnittlichkeit attraktiv mache, da weniger durchschnittliche Proportionen hässlicher bewertet wurden.

Dass alle verwendeten Gesichter dieser Studien insgesamt bestenfalls mittelmäßig bewertet wurden und kein einziges als sehr attraktiv, sei nur nebenbei erwähnt (siehe Punkt 3 oben) und kümmerte die Autoren jedoch wenig, denn ihre produzierten Unterschiede waren ja statistisch signifikant. In Wahrheit sind ihre Ergebnisse jedoch wenig aussagekräftig, da sie mit ihren Manipulationen der Verdurchschnittlichung nicht hohe Attraktivität erreicht haben – vielmehr haben sie durch die Entstellung der Gesichter zu Karikaturen Hässlichkeit produziert. Attraktivitätsforscher, die zur Herstellung ihrer Stimuli die Karikatur-Methode benutzen, missbrauchen eine Funktion eines Bildverarbeitungsprogramms,

die eigentlich vom Hersteller als *Fun-Tool* gedacht ist, und tun so, als betrieben sie seriöse Forschung. Der Name „Karikatur-Technik“ sollte eigentlich schon Warnung genug sein: Bereits ein leicht gedreht fotografiertes Gesicht, das nicht exakt frontal abgebildet ist, wird bei einer 50 % Karikatur unnatürlich verzerrt. Besitzt dieses Gesicht zusätzlich einige sichtbare Asymmetrien oder auffällige Merkmale (z. B. eine große Nase), wird in der Karikatur aus der leichten, normalen Asymmetrie in Kombination mit der nicht-frontalen Aufnahme eine pathologisch wirkende Asymmetrie und aus der großen Nase eine entstehende Missbildung. Mit Karikaturen als Stimulusmaterial kann man nicht Attraktivität erforschen. Wer es dennoch versucht, produziert damit keine Erkenntnisse, sondern karikiert die Idee des experimentellen Arbeitens in der empirischen Sozialforschung.

(6) Invalide Methoden zur Messung von Durchschnittlichkeit: Der letzte Grund, warum die Durchschnittshypothese innerhalb der Szene der Attraktivitätsforscher so hoch gehandelt wird, liegt an der Art und Weise, wie Durchschnittlichkeit gemessen wurde. Dieses Problem betrifft die Mehrheit aller Untersuchungen zur Durchschnittshypothese. In der eigenen Untersuchungsreihe wurde in allen sechs Experimenten (*Durch 1 bis Durch 6*) Durchschnittlichkeit über die Anzahl der in einem Composite enthaltenen Originalgesichter operationalisiert (= Variable „Generation“). In Composites der ersten Generation waren zwei Originalgesichter enthalten, in Composites der zweiten Generation vier Originalgesichter, in der dritten Generation acht usw. Die Bestimmung der Durchschnittlichkeit als unabhängige Variable ist damit eindeutig, objektiv und nachvollziehbar.

In den meisten anderen Untersuchungen wurde hingegen versucht, Durchschnittlichkeit über Urteile von Versuchspersonen zu erfassen. Diese Durchschnittlichkeit wurde zudem noch nicht einmal durch ein direktes Fragen nach Durchschnittlichkeit (z. B. „Wie durchschnittlich ist dieses Gesicht?“) erfasst, da die Autoren zu Recht befürchteten, dass Versuchspersonen mathematisch durchschnittliche Gesichtsproportionen mit mittelmäßiger Attraktivität verwechseln könnten. Daher lautet die standardmäßige Frage an die Probanden: „How distinctive is this face?“ Die Idee dahinter ist, dass der Begriff *distinctiveness* das Gegenteil von Durchschnittlichkeit erfasst. Der schwer ins Deutsche zu übersetzende Begriff bedeutet so etwas wie „Unterscheidbarkeit“. In den Studien von Rhodes, die häufig mit solchen *distinctiveness*-Ratings arbeitete und immer noch arbeitet (ebenso wie mit Ratings von Symmetrie, Maskulinität und Femininität), wird der Begriff so operationalisiert: „[...] *distinctiveness* (1 = *very difficult to pick out of a crowd at a busy railway station*, 7 = *easy to pick out of a crowd at a busy railway station*)“ (Rhodes & Tremewan, 1996, S. 106). Diese Operationalisierung stammt aus einer Studie (Valentine, 1991), die sich je-

doch weder mit der Durchschnittshypothese noch mit Attraktivität befasste, sondern mit dem Wiedererkennen von Gesichtern (*face recognition*).

Der typische Befund zahlreicher Studien ist, dass Ratings von *distinctiveness* negativ mit Ratings von Attraktivität korrelieren. Da *distinctiveness* von Anhängern der Durchschnittshypothese als das Gegenteil von Durchschnittlichkeit definiert (!) wird, ziehen sie daraus den Schluss, dass deswegen Durchschnittlichkeit attraktiv mache (Rhodes, 2006). Diese Interpretation ist jedoch sehr problematisch. Denn selbstverständlich ist ein Gesicht leicht wiederzuerkennen, also *distinctive*, (z. B. in einem belebten Bahnhof, um die Formulierung von Rhodes aufzugreifen), wenn es Merkmale aufweist, die stark von der Norm abweichen, z. B. durch eine viel zu große Nase, Kieferfehlstellungen, Narben, Missbildungen usw. Je entstellender das Merkmal, desto größer der Wiedererkennungswert. Auch alte Gesichter mit tiefen, charakteristischen Falten führen zu einer größeren *distinctiveness* als junge Gesichter (Deffenbacher et al., 1998; Wickham & Morris, 2003). Insofern ist es trivial, dass Gesichter mit hoher *distinctiveness* unattraktiver bewertet werden. Das bedeutet jedoch im Umkehrschluss noch lange nicht, dass ein Gesicht schon deswegen attraktiv ist, nur weil es keine auffallend hässlichen Merkmale besitzt. Schönheit ist mehr als nur das Fehlen von Hässlichkeit.

Zudem mangelt es der Operationalisierung von Durchschnittlichkeit durch das Gegenteil von *distinctiveness* an Konstruktvalidität. Denn ein hochattraktives Gesicht eines Models (dessen Gesichtsproportionen klar vom Durchschnitt abweichen) würde in einem belebten Bahnhof ebenfalls leicht wiedererkannt. Nicht nur Hässlichkeit ist also *distinctive*, sondern ebenfalls herausragende Schönheit. Dass sich dieser Umstand jedoch nicht in den Ergebnissen dieser Studien niederschlägt, hat einen einfachen statistischen Grund: Hässlichkeit ist viel häufiger als Schönheit; dass stark von der Norm abweichende Gesichtsmerkmale ein Gesicht unattraktiver machen, ist viel wahrscheinlicher als dass sie es attraktiver machen. Aus diesem Grund würden hochattraktive Gesichter in einer tatsächlich für die Bevölkerung repräsentativen Stichprobe keine große Rolle spielen, weil sie zahlenmäßig im Vergleich zu unattraktiven Menschen stark unterrepräsentiert sind (inwieweit im Stimulusmaterial von Attraktivitätsforschern hochattraktive Gesichter überhaupt vertreten sind, ist noch mal ein eigenes Problem). Die widersprüchliche Beziehung zwischen *distinctiveness* und Attraktivität belegt auch die Studie von Wickham und Morris (2003).

Hinzu kommt bei der Gleichsetzung von Durchschnittlichkeit mit dem Gegenteil von *distinctiveness* noch ein weiteres Problem für die Konstruktvalidität: Ein Gesicht mit nied-

riger *distinctiveness* (also hoher postulierter Durchschnittlichkeit) mag unauffällig für ein Gesicht an sich sein, hierbei wird jedoch nicht zwischen Männer- und Frauengesichtern unterschieden. Die Operationalisierung von Durchschnittlichkeit ist damit nicht sauber abgrenzbar von einer anderen wichtigen Theorie der Attraktivitätsforschung, nämlich der Theorie des sexuellen Dimorphismus. Beispiel: Eine Frau mit typisch männlichen Gesichtsproportionen ist als Gesicht an sich überhaupt nicht auffällig (denn viele Männer sehen ja auch so aus), sie wäre also wenig *distinctive* und angeblich durchschnittlich. Andererseits jedoch hat eine Frau mit sehr männlichen Gesichtszügen einen sehr hohen Wiedererkennungswert, weil sie völlig untypisch für eine Frau aussieht (sehr *distinctive*, d. h. wenig durchschnittlich). In diesem Bereich bleibt das Konzept der *distinctiveness* unscharf und widersprüchlich. Hinzu kommt natürlich noch die Unschärfe der Messung selbst, da die Bestimmung der (angeblichen) Durchschnittlichkeit ja über Befragungen einer Stichprobe von Versuchspersonen umgesetzt wird.

Die Autoren, die mit *distinctiveness*-Ratings arbeiten, bleiben den Nachweis schuldig, dass ihre Operationalisierung tatsächlich das misst, was sie zu messen vorgibt. Die Beweislast liegt dabei selbstverständlich bei den Autoren, die dieses Verfahren verwenden. Eine Möglichkeit wäre beispielsweise, Gesichter nach *Facial Landmarks* zu vermessen, daraus ein Maß (Index) für die Durchschnittlichkeit von Gesichtsproportionen abzuleiten und diese „Durchschnittlichkeitsindizes“ dann mit den mittleren *distinctiveness*-Ratings zu korrelieren. Dieser Nachweis fehlt bislang, denn wer sich schon die Mühe macht, Gesichter anthropometrisch zu vermessen (Baudouin & Tiberghien, 2004; Grammer & Thornhill, 1994; Jones & Hill, 1993; Pollard, Shepherd & Shepherd, 1999), ist natürlich auf ein schwammiges *distinctiveness*-Rating durch Versuchspersonen als Maß für Durchschnittlichkeit nicht mehr angewiesen. Der Grund dafür, warum *distinctiveness*-Ratings so verbreitet sind, ist wohl eher, weil sie erstens so leicht und bequem durchzuführen sind und zweitens die gewünschten Ergebnisse dabei herauskommen.

In Studien, bei denen die Durchschnittlichkeit von Gesichtern (durch Morphing, einfache Bildüberlagerungen, *caricature-technique* oder *anti-caricature-technique*) experimentell manipuliert wurde, gibt es eigentlich gar keinen Grund, über *distinctiveness*-Ratings die Durchschnittlichkeit noch einmal zu messen, denn man hat ja bereits von jedem Gesicht mindestens zwei Varianten und man kennt auch die Stärke des Treatments, das man angewendet hat. Trotzdem gibt es Studien, die in einem solchen Fall *distinctiveness* noch einmal extra von Versuchspersonen bewerten lassen (Deffenbacher et al., 1998; Little & Hancock, 2002; Rhodes, Harwood, Yoshikawa, Nishitani & McLean, 2002; Rhodes, Rob-

erts et al., 1999; Rhodes, Sumich et al., 1999; Rhodes & Tremewan, 1996; Rhodes, Yoshikawa et al., 2001).

Mitunter kommt es sogar zu recht fragwürdigen Argumentationen wie z. B. in der Studie von Rhodes und Tremewan (1996), in der sie die beurteilte *distinctiveness* gegen die objektiv feststellbare Durchschnittlichkeit ausspielen! In ihrem Experiment 2 wollten die Autorinnen mit Strichzeichnungen von Gesichtern das Experiment von Langlois und Roggman (1990) replizieren und zogen ihr Experiment genau gleich auf. Da bei Strichzeichnungen jedoch keine Texturinformationen vorhanden sind, es also bei diesen gemorphten Gesichtern auch keine Haut gibt, die dadurch glatter und makelloser wird, scheiterten sie mit ihrem Versuch auf ganzer Linie. Attraktivität stand in keinerlei Zusammenhang mit Durchschnittlichkeit, d. h. mit der Anzahl der Gesichter, die in einem Composite enthalten waren. Alle Composites wurden im Durchschnitt ungefähr gleich mittelmäßig attraktiv bewertet (auf einer 7er-Skala alle um den Wert 4; bei Frauengesichtern bekamen *2-face-composites* sogar deutlich bessere Werte ($M = 4.4$) als *32-face-composites* ($M = 3.7$)).

However, the attractiveness of the composites, with sex of face and number of component faces as repeated measures factors, showed no significant effects. This result contrasts with the linear increase in attractiveness with the number of component faces found by Langlois and Roggman (1990). (Rhodes & Tremewan, 1996, S. 108)

Doch statt es einfach bei dieser Feststellung zu belassen, legten die Autorinnen im nächsten Satz eine geradezu abenteuerliche Kehrtwende hin und stellten den Sinn ihrer eigenen Daten auf den Kopf:

However, it does not contradict the claim that attractiveness is attractive, because the composites did not get less distinctive (more average) as more faces were included either. A two-way ANOVA on the mean distinctiveness ratings for the composites, with sex of face and number of component faces as repeated measures factors, showed only a marginal effect of number of faces, $F(4,16) = 2.88$, $p < .06$. Apparently the composites were all about equally average and equally attractive. No other effects were significant. (Rhodes & Tremewan, 1996, S. 108)

Mit ihrer Varianzanalyse wiesen sie selbst nach, dass die *distinctiveness*-Ratings ihrer Versuchspersonen überhaupt nichts mit der objektiven Durchschnittlichkeit (operationalisiert durch die Anzahl der in einem Composite enthaltenen Originalgesichter) zu tun hatten. Anstatt nun aber die logische Konsequenz daraus zu ziehen und einzusehen, dass *distinctiveness*-Ratings eine völlig invalide Methode zur Bestimmung der Durchschnittlichkeit sind und daher überhaupt nicht verwendbar, stützten sie ihre Argumentation ge-

nau auf dieses unsinnige Maß. Sie behaupteten ernsthaft, dass ihre Composites aus vielen Originalgesichtern (z. B. 32) gar nicht durchschnittlicher seien als Composites aus nur wenigen Originalgesichtern (z. B. zwei), bloß deswegen, weil es bei den *distinctiveness*-Ratings durch die Versuchspersonen keine Unterschiede gab. Und deswegen seien alle ihre Composites gleich durchschnittlich und gleich attraktiv („equally average and equally attractive“, S. 108) gewesen. Mit anderen Worten: Die Autorinnen hatten die Hypothese, dass durchschnittliche Gesichter attraktiv sind. Sie manipulierten experimentell die Durchschnittlichkeit von Gesichtern durch Verwendung von Morphing-Software. Ihre eigenen Daten belegten, dass die Hypothese falsch war. Daraufhin behaupteten sie, ihre Durchschnittsgesichter wären ja gar nicht durchschnittlich und benutzten als Argument ihre *distinctiveness*-Ratings, deren mangelnde Validität sie mit ihren eigenen Daten nachgewiesen hatten. Bemerkenswert ist, dass diese absurde Argumentation nicht von irgendwem stammt, sondern von Gillian Rhodes, die zu den weltweit renommiertesten Attraktivitätsforschern zählt.

Da bewertete *distinctiveness* als Operationalisierung für Durchschnittlichkeit im Sinne der Durchschnittshypothese ungeeignet ist, sind damit auch automatisch all diejenigen Untersuchungen für die Überprüfung der Durchschnittshypothese wertlos, bei denen Durchschnittlichkeit ausschließlich über *distinctiveness*-Ratings erfasst wurde. Dies sind alle Studien, bei denen Durchschnittlichkeit nicht experimentell manipuliert wurde, sondern lediglich die *distinctiveness* von normalen, unveränderten Gesichtern von Versuchspersonen beurteilt wurde (Light et al., 1981; Morris & Wickham, 2001; O'Toole, Deffenbacher, Valentin & Abdi, 1994; Peskin & Newell, 2004; Rhodes, Simmons & Peters, 2005; Sarno & Alley, 1997; Vokey & Read, 1992).

Gerade diese Studien wurden jedoch in der Metaanalyse von Rhodes (2006) ebenfalls zu den Studien gezählt, die angeblich belegen, dass Durchschnittlichkeit Gesichter attraktiv macht. Fast alle diese Untersuchungen hatten jedoch gar nicht zum Ziel, die Durchschnittshypothese der Attraktivitätsforschung zu überprüfen, sondern befassten sich mit dem Wiedererkennen von Gesichtern (*face recognition*). In diesem Kontext mag es richtig und sinnvoll sein, Versuchspersonen die *distinctiveness* von Gesichtern bewerten zu lassen, indem man sie fragt, wie leicht jemand innerhalb einer Menschenmenge wiederzuerkennen ist („easy to pick out of a crowd at a busy railway station“, Rhodes & Tremenwan, 1996, S. 106), als Maß für die Durchschnittlichkeit von Gesichtsproportionen ist es jedoch aus oben genannten Gründen nicht geeignet.

Warum in ihrer Metaanalyse Rhodes (2006) dennoch Untersuchungen auflistete und auswertete, die angeblich etwas über den Einfluss von Durchschnittlichkeit auf Attraktivität aussagen, obwohl diese ein ganz anderes Ziel hatten und Attraktivität von Gesichtern nur als eine von mehreren Variablen (die im Kontext der Gesichtserkennung relevant sein könnten) erfassten, wird im Anhang ihrer Arbeit deutlich: In der Anmerkung 1 in Anhang 1 zu ihrer Metaanalyse nennt Rhodes die Suchbegriffe, mit denen sie nach Studien zur Durchschnittshypothese recherchierte: „Databases MEDLINE, PSYCINFO, BIOL ABSTRACTS, ERIC were searched using keywords, “face” + “attractive” (or “attractiveness”) + “average” (or “distinctive” or “distinctiveness”)” (Rhodes, 2006). Sie suchte also gezielt auch nach Studien, die *distinctiveness* zum Thema hatten, und schloss damit auch Studien zur Gesichtserkennung in ihre Metaanalyse mit ein.

Von den insgesamt 20 Studien zur Durchschnittshypothese, die in Rhodes' Metaanalyse ausgewertet wurden, arbeiteten 14 mit *distinctiveness*-Ratings und basierten damit auf einer Operationalisierung von Durchschnittlichkeit, die als sehr problematisch anzusehen ist. Die Durchschnittshypothese beruht damit auf einer empirischen Grundlage, die bei weitem nicht so solide ist, wie unter Attraktivitätsforschern gern der Anschein erweckt wird. Hinzu kommen die weiteren Kritikpunkte (Punkte 1 bis 5), die oben ausgeführt wurden. Auf die meisten Studien treffen gleich mehrere Schwachpunkte gleichzeitig zu.

Zusammenfassend bleibt festzuhalten, dass das aus Experimenten *Durch 1* bis *Durch 6* gezogene Fazit, dass durchschnittliche Gesichtsproportionen eben nicht attraktiv, sondern nur mittelmäßig attraktiv machen, ein deutlicher Widerspruch zum Stand der Forschung ist, wie es zunächst scheint. Diese Diskrepanz lässt sich jedoch erklären, indem man bisherige Arbeiten zu diesem Thema einmal genau liest (inklusive Methoden- und Ergebnisteile) und kritisch unter die Lupe nimmt. Dann findet man nämlich reihenweise Beispiele dafür, wie man Forschung *nicht* betreiben sollte: Ergebnisse wurden (1) einseitig und oberflächlich (z. B. nur nach Mittelwertsunterschieden) analysiert und interpretiert, (2) naheliegende Alternativhypothesen wurden nicht geprüft und (3) es wurde stillschweigend übergangen, dass Durchschnittsgesichter lediglich mittelmäßige und keine hohen Attraktivitätsbewertungen bekamen (d. h. absolute Effekte wurden zugunsten relativer Effekte ignoriert). Zudem waren zahlreiche Ergebnisse (4) durch Morphing-Artefakte der glatten, makellosen Haut von Composites konfundiert, (5) Durchschnittlichkeit wurde mit nicht validen Verfahren experimentell manipuliert, indem Karikaturen von Gesichtern verwendet wurden und (6) das Ausmaß von Durchschnittlichkeit wurde methodisch fehlerhaft gemessen, indem *distinctiveness*-Ratings von Versuchspersonen verwendet wur-

den, die das Gegenteil von Durchschnittlichkeit messen sollten, in Wirklichkeit jedoch kein valides Maß für Durchschnittlichkeit sind.

3.5 Fazit zum Effekt der Durchschnittlichkeit auf die Attraktivität von Gesichtern und Konsequenzen für künftige Studien

Wie im vorigen Kapitel 3.4 ausführlich dargelegt, ist der größte Teil aller Untersuchungen zur Durchschnittshypothese mit erheblichen Mängeln behaftet – und meist nicht nur mit einem einzigen, sondern mehreren gleichzeitig. Umso erstaunlicher ist es, dass die Durchschnittshypothese es zur einflussreichsten Theorie innerhalb der Szene der Attraktivitätsforscher schaffen konnte. Bereits der „Startschuss“ für diese Forschung, der Artikel von Langlois und Roggman (1990), ist bei sorgfältiger und kritischer Lektüre alles andere als überzeugend: Deren Durchschnittsgesichter waren (bedingt durch den damaligen Stand der Technik mit einfachen Überblendungen anstatt Morphing) (1) von optisch schlechter, verwaschener Qualität, (2) es gab etliche Originalgesichter, die attraktiver bewertet wurden als die Durchschnittsgesichter, (3) alle Durchschnittsgesichter bekamen nur mäßige Bewertungen im mittleren Bereich der Attraktivitätsskala, und (4) die Durchschnittlichkeit der Composites war mit deren glatter, makelloser Haut konfundiert. Dennoch setzte dieser Artikel einen Boom ähnlicher Forschung in Gang.

Bei den folgenden Studien verbesserte sich zwar einerseits mit dem Aufkommen von Morphing-Programmen die Bildqualität der Composites (auch wenn diese mangels Sorgfalt und Genauigkeit beim Morphen bis heute weit hinter den technischen Möglichkeiten zurückbleiben), doch wurden andererseits den bisherigen methodischen Schwächen noch weitere hinzugefügt, indem auch mit Karikaturen von Gesichtern experimentiert wurde und Durchschnittlichkeit mit nicht validen *distinctiveness*-Ratings gemessen wurde. Auf der anderen Seite wurde aber bis heute der gewaltige Störfaktor der glatten, makellosen Haut von Composites nicht angemessen gewürdigt (bzw. verstanden) und Ergebnisse, die dessen Einfluss belegen, wie das Experiment 2 von Rhodes und Tremewan (1996) nicht zur Kenntnis genommen bzw. fehlinterpretiert (vgl. Kapitel 3.4).

Die Fülle an methodisch fehlerhaften Studien mit ihren gleichzeitig oft einseitigen und interessengeleiteten Ergebnisinterpretationen haben zu der Situation geführt, dass gegenwärtig unter Attraktivitätsforschern die Meinung herrscht, dass die Durchschnittshypothese empirisch bestens fundiert sei und obendrein hervorragend erklären könne, was ein Gesicht attraktiv macht. So schreiben beispielsweise Rubenstein, Langlois und

Roggman in dem einflussreichen Standardwerk „Facial Attractiveness“ von Rhodes und Zebrowitz (2002) im Kapitel über die Durchschnittshypothese als Fazit:

Averageness ist the only characteristic discovered to date that is both necessary and sufficient to ensure facial attractiveness – without a facial configuration close to the average of the population, a face will not be attractive no matter how smooth, youthful, or symmetrical. Averageness is *fundamental*. (Rubenstein, Langlois & Roggman, 2002, S. 21)

Diese Behauptung ist falsch. Durchschnittlichkeit ist für die Attraktivität eines Gesichts nicht notwendig und schon gar nicht hinreichend. Bereits ein kritisches Durcharbeiten der publizierten Studien zu diesem Thema würde klar machen, dass die empirische Basis dieser Behauptung sehr dürftig ist. Die sechs Experimente dieser Arbeit belegen zudem eindeutig, dass die Postulierung von Durchschnittlichkeit als Schönheitsideal auf einem Irrtum beruht.

Es konnte eindeutig nachgewiesen werden, dass der in dieser Untersuchungsreihe erzielte, scheinbar große attraktivitätserhöhende Effekt der Durchschnittlichkeit in Wahrheit größtenteils auf Artefakte des Morphens zurückzuführen ist (vgl. Kapitel 2.6 und 2.10) und dass die allgemeine Attraktivitätssteigerung bei Composites durch die Bildverarbeitungsartefakte der glatteren und makelloseren Haut entsteht (vgl. Kapitel 3.1.2). Durch den Vergleich verschiedener Stimulusarten (Frontalgesichter und Profilgesichter mit und ohne Textur), konnte belegt werden, dass die Ausgangsattraktivität der Originalgesichter hauptverantwortlich für die Attraktivität eines Composites ist (vgl. Kapitel 3.1.3).

Die Ergebnisse der sechs durchgeführten Experimente lassen nur den Schluss zu, dass durchschnittliche Gesichtsproportionen (also ohne die Morphing-Artefakte der makellosen Haut) zwar die Attraktivität vieler (aber nicht aller) Gesichter geringfügig steigern können, dass jedoch Durchschnittsproportionen bei weitem nicht das Attraktivitätsideal darstellen (vgl. dazu auch die beiden Experimente *Vergleich 1* und *Vergleich 2* in Kapitel 8). Die plausibelste Erklärung für diesen kleinen Effekt ist, dass bei durchschnittlichen Gesichtsproportionen keine unattraktiv machenden Merkmale mehr vorhanden sind, wie beispielsweise einzelne zu große oder zu kleine Gesichtsmerkmale oder unharmonisch wirkende Gesichtsproportionen. Ein durchschnittliches, normales Gesicht ist also in erster Linie ein guter Schutz vor hässlichem Aussehen – aber diese Erkenntnis ist trivial.

Die Behauptung, dass die durchschnittlichsten Gesichter auch die schönsten seien, ist jedoch nicht haltbar. Denn es wurden in den sechs Experimenten zahlreiche Gesichter produziert, die deutlich attraktiver als die jeweiligen durchschnittlichsten Gesichter bewertet wurden. Dass diese Gesichter keine Originalgesichter, sondern auch gemorpht waren,

tut überhaupt nichts zur Sache. Denn bei der Frage, was das Schönheitsideal ist und wie ein solches ideales Gesicht aussieht, spielt es gar keine Rolle, wie häufig ein solch schönes Gesicht ist und ob es in der sehr begrenzten untersuchten Stichprobe an fotografierten Stimuluspersonen zufälligerweise vorhanden ist oder nicht. Genau diesen Fehler machen alle Forscher, die die Attraktivität eines gemorphten Durchschnittsgesichts nur mit den jeweiligen Originalgesichtern vergleichen, aus denen dieses Composite berechnet wurde.

Dem Schönheitsideal der Bevölkerung deutlich näher kommen Composites aus attraktiven Originalgesichtern. Diese Composites sind zugleich auch deutlich attraktiver als die jeweiligen Originalgesichter, aus denen sie erzeugt wurden. Der Mechanismus dieses „Raffael-Effekts“ ist dabei derselbe, der auch für den Attraktivitätszuwachs gemorphter Gesichter generell (auch des „absoluten“ Durchschnittsgesichts) verantwortlich ist: Vereinzelte Abweichungen von einem idealen Schema (von denen auch attraktive Gesichter nicht frei sind) werden herausgemittelt und die Haut erscheint glatt und makellos. Gerade weil der Mechanismus derselbe ist und diese hochattraktiven Composites auf dieselbe Weise erzeugt wurden wie die „absoluten“ Durchschnittsgesichter, ist dies eine so überzeugende Widerlegung der Durchschnittshypothese. Und dass dieser Nachweis nicht nur mit statistischen Mitteln gelang, sondern über die hergestellten Gesichter selbst sinnlich erfahrbar ist (vgl. Abbildung 38, S. 85, Abbildung 22, S. 49, Abbildung 27, S. 58, Abbildung 6, S. 21 und Abbildung 11, S. 29), macht den Irrtum der Befürworter der Durchschnittshypothese selbst für jeden Laien unmittelbar deutlich.

Die Prototypen für attraktive Gesichter widerlegen nicht nur die Durchschnittshypothese, sondern liefern eine Vielzahl von Indizien, welche Merkmale tatsächlich für hohe Attraktivität verantwortlich sind. Dies zeigte deutlich die deskriptive Analyse, bei der diese Prototypen mit den jeweiligen „absoluten“ Durchschnittsgesichtern verglichen wurden (Kapitel 3.3). Es wurde ebenfalls gezeigt, dass sich diese Einzelmerkmale auf die Faktoren „Jugendlichkeit“ und „sexueller Dimorphismus“ bzw. „Femininität“ zurückführen lassen.

Diese beiden Faktoren, Jugendlichkeit und sexueller Dimorphismus, werfen ein bezeichnendes Licht auf die „akademische“ Attraktivitätsforschung und weisen auf ein tief liegendes Problem dieses Forschungsgebiets hin. Attraktivitätsforscher behandeln diese beiden Bereiche völlig unterschiedlich. Während es eine Menge Forschungsarbeiten zum sexuellen Dimorphismus gibt, existiert Jugendlichkeit als Erklärung für Attraktivität innerhalb der Attraktivitätsforscher-Szene überhaupt nicht und wird ignoriert. Der Grund dafür ist einfach: Hinter dem sexuellen Dimorphismus steckt die Idee, dass Attraktivität mehr ist als

nur ästhetisches Aussehen, sondern dass sie ein Signal ist für Merkmale, die aus biologischer Sicht für Partnerwahl von Bedeutung sind, z. B. für Fruchtbarkeit, reproduktiven Wert, Parasitenresistenz oder allgemein „gute Gene“. Hinter der Theorie des sexuellen Dimorphismus steckt die evolutionäre Psychologie, die seit Beginn der 90er-Jahre (Buss, 1999) mit dem Boom in den Biowissenschaften einen ungeheuren Aufschwung erlebt hat. Von dieser Strömung hat die Attraktivitätsforschung, die etwa zur gleichen Zeit aufkam, stark profitiert und sie sichert ihr nach wie vor eine Beachtung und durch die Nähe zu den Biowissenschaften ein positives naturwissenschaftliches Image, das sie ansonsten nicht hätte.

Anders hingegen Jugendlichkeit als Attraktivitätsfaktor. Dass jugendliche Gesichter attraktiver sind als alte Gesichter, gilt als so trivial, dass es nicht weiter erforschenswert erscheint. Und doch stecken hinter dieser Einstellung zwei Missverständnisse: Wenn man sich ernsthaft für die Ursachen von Attraktivität interessiert, dann sollte man dies (1) ohne Vorbehalte tun und nicht bestimmte sehr plausible und umfassende Erklärungsansätze von vornherein ignorieren, nur weil sie aus theoretischer Sicht nicht besonders spannend sind und man stattdessen lieber ein bisschen innerhalb der evolutionären Psychologie mitmischen möchte.

Zudem (2) ist der Faktor Jugendlichkeit alles andere als trivial, denn aus wissenschaftlicher Sicht geht es hier um die Details: Welches Gesichtsmerkmal altert auf welche Weise und wie stark, wie hängt dies mit dem Elastizitätsverlust der Haut, mit dem Abbau des Unterhautfettgewebes und sogar mit der Veränderung der Knochensubstanz zusammen (Bartlett et al., 1992; Coleman & Grover, 2006; Fagien, 2002; Friedman, 2005; Gosain et al., 2005; Guyuron, 1997; Hamra, 1994; Heinz & Kikkawa, 1997; Hoenig et al., 1997; Jelks & Jelks, 1993; Rohrich et al., 2004)? Ein altes Gesicht unterscheidet sich von einem jugendlichen Gesicht in weitaus mehr Merkmalen als einfach nur grauen Haaren und vielen Falten (wie die meisten medizinischen Laien pauschal denken). Zudem erklären typische Alterskriterien ja nicht nur fehlende Attraktivität bei alten Gesichtern, sondern auch bei jungen, nämlich dann, wenn bei einer jungen Person aufgrund ihrer genetischen Veranlagung Gesichtsmerkmale nur zufälligerweise so aussehen wie Merkmale, die typisch sind für alte Gesichter (z. B. eine Höckernase mit abwärts gerichteter Nasenspitze und gleichzeitiger Oberkieferrücklage, vgl. Kapitel 3.3). All dies wird intensiv erforscht – aber eben nicht von Attraktivitätsforschern, sondern von Medizinern.

Dort, wo es wirklich darum geht, Schönheit zu verstehen (mit dem Ziel, sie auch herzustellen oder wiederherzustellen), findet die Forschung ohne Attraktivitätsforscher statt.

Diese haben sich abgesondert, untersuchen ihre eigenen Hypothesen und zitieren die Arbeiten anderer Attraktivitätsforscher. Ob sich die Ergebnisse ihrer Forschung für irgendetwas anwenden lassen, ist ihnen offensichtlich gleichgültig. Während die ästhetische Medizin boomt und sich immer mehr Menschen medizinischen Eingriffen unterziehen, um attraktiver auszusehen, herrscht zwischen den beiden Disziplinen Attraktivitätsforschung und Medizin gegenseitiges Desinteresse. Akademische Attraktivitätsforscher befassen sich lieber mit abstrakten Theorien über die (biologische) Funktion von Attraktivität anstatt mit Detailfragen des Aussehens, während ästhetisch-plastische Chirurgen und Mund-Kiefer-Gesichtschirurgen mit den Ergebnissen der akademischen Attraktivitätsforschung ohnehin wenig anfangen können.

Die Durchschnittshypothese – unter Attraktivitätsforschern die renommierteste Attraktivitätstheorie – steht dabei symptomatisch für die Situation der Mainstream-Attraktivitätsforschung: Isoliert in einer akademischen Nische und ohne Aussicht auf eine Anwendung. Und darüber können Attraktivitätsforscher sogar noch froh sein. Denn man stelle sich einmal vor, wenn „akademische“ Attraktivitätsforscher mit ihrer Behauptung, durchschnittliche Gesichtsproportionen machten attraktiv, tatsächlich ernst genommen würden. Wenn auf einmal ohnehin schon sehr attraktive Menschen in ihrem Ehrgeiz, noch attraktiver auszusehen, den Wunsch nach einem durchschnittlicheren Aussehen äußern würden und Chirurgen dem tatsächlich nachkommen würden. Die Patienten sähen unattraktiver aus als vorher, die Chirurgen würden als unfähige Operateure dastehen und die akademischen Attraktivitätsforscher mit ihrer Durchschnittshypothese wären durch die Praxis widerlegt.

So gesehen überrascht es auch nicht, dass „akademische“ Attraktivitätsforscher an einer anwendungsorientierten Forschung gar kein Interesse haben. Solange die Ergebnisse der eigenen Forschung für nichts anwendbar sind, können sie von der Realität auch nicht widerlegt werden. Doch langfristig führt diese Einstellung in eine Sackgasse. Wer dauerhaft und kategorisch praktische Anwendungsmöglichkeiten seiner Forschung ignoriert und auf den Sinn und Zweck seiner Arbeit nur mit „Grundlagenforschung“ zu antworten weiß, manövriert sich früher oder später ins akademische Abseits. Im Fall der Durchschnittshypothese befindet sich die Attraktivitätsforschung schon von Anfang an dort, denn in der ästhetischen Medizin hatte und hat diese Theorie keinerlei Bedeutung.

Die Hypothese, Durchschnittlichkeit mache ein Gesicht attraktiv, ist nicht mehr länger zu halten. Für die Attraktivitätsforschung ist es Zeit, von dieser Behauptung abzulassen und sich neuen Fragen zuzuwenden. Eine interdisziplinäre Zusammenarbeit mit der ästheti-

schen Medizin könnte dabei für beide Seiten sehr produktiv sein. Die Durchschnittshypothese ist nicht nur nicht wahr, sondern der selbstzufriedene Glaube an ihre Gültigkeit behindert sogar einen weiteren Erkenntnisfortschritt. Denn wie die deskriptive Analyse in Kapitel 3.3 gezeigt hat, liegt gerade in den Abweichungen von einem bloß durchschnittlichen Aussehen der Schlüssel zum Verständnis hoher Attraktivität. Das optimale Aussehen dieser Details zu erforschen, bietet reichlich Stoff für viele künftige Untersuchungen zur Attraktivität von Gesichtern.

4 Empirische Untersuchungen zum Einfluss der Symmetrie auf Attraktivität

4.1 Kurzer Überblick über den Stand der Forschung zu Symmetrie und Attraktivität

In den letzten eineinhalb Jahrzehnten entstanden zahlreiche Studien zur Symmetrie des Gesichts (aber auch des Körpers), die von der Idee geprägt sind, dass eine hohe Symmetrie ein Signal für einen hohen Partnerwert (*mate value*) ist (Gangestad & Thornhill, 1997; Gangestad, Thornhill & Yeo, 1994; Palmer & Strobeck, 1986; Thornhill & Gangestad, 1999; Thornhill & Moller, 1997). Inzwischen liegen zahlreiche Belege dafür vor, dass hohe Symmetrie auf Entwicklungsstabilität (*developmental stability*) während der Ontogenese zurückschließen lässt. Dies kann entweder an einem Fehlen von negativen Umwelteinflüssen liegen, so dass das Wachstum des Individuums sich ungestört entwickeln konnte, oder daran, dass zwar Umweltstressfaktoren vorhanden waren, das Individuum jedoch aufgrund seiner guten genetischen Ausstattung (*genetic quality*, „good genes“) gut damit zu Recht kam, oder eine Kombination beider Faktoren.

Wenn hohe Symmetrie tatsächlich ein Signal (Fitness-Indikator) für eine gute genetische Ausstattung ist, dann sollten (nach der *sexual selection theory*) hoch symmetrische Individuen als Fortpflanzungspartner bevorzugt werden (Anderson, 1994). Für Attraktivitätsforschung ergibt sich daraus die Vorhersage, dass hoch symmetrische Gesichter als attraktiver bewertet werden müssten als Gesichter mit Asymmetrien.

Studien zum Zusammenhang zwischen Symmetrie und Attraktivität gibt es seit nunmehr 15 Jahren. Eine gute Übersicht bietet eine aktuelle Metaanalyse (Rhodes, 2006). Je nach Studie und je nach verwendeter Methodik fallen jedoch die Ergebnisse völlig unterschiedlich aus: Während die Mehrzahl der Untersuchungen einen schwach positiven Effekt von Symmetrie fand, fanden andere (Little & Jones, 2003; Mealey, Bridgstock & Townsend, 1999; Rhodes, Roberts et al., 1999; Rhodes, Sumich et al., 1999; Rikowski & Grammer, 1999) mittelstarke bis hohe Effekte (bis 0.83), andere (Langlois et al., 1994) hingegen einen negativen Effekt („je symmetrischer, desto unattraktiver“, Effektgrößen bis -0.77) und wieder andere überhaupt keinen Effekt (Hönekopp, Bartholome & Jansen, 2004; Jones & Hill, 1993; Shackelford & Larsen, 1997).

Ein Grund für die widersprüchlichen Ergebnisse ist, dass Symmetrie bzw. Asymmetrie nicht einheitlich definiert ist, bzw. verschieden operationalisiert wird. Während v. a. in früheren Studien unter Asymmetrie jede Abweichung von der perfekten Spiegelbildlichkeit gesehen wurde, wird mittlerweile zwischen direktonaler (*directional asymmetry, DA*) und fluktuierender Asymmetrie (*fluctuating asymmetry, FA*) differenziert. Fluktuierende Asymmetrien sind die zufälligen, nicht-direktionalen Abweichungen von der perfekten Symmetrie bei bilateralen, gepaarten Merkmalen (Rhodes, 2006, S. 205). Der Erwartungswert aller Abweichungen pro Gesicht ist dabei Null. Als direkte Asymmetrie werden die systematischen Abweichungen verstanden, z. B. dann, wenn eine Gesichtshälfte gleichmäßig etwas größer ist als die andere (was tatsächlich der Fall zu sein scheint, da offenbar rechte Gesichtshälften beim Menschen durchschnittlich größer sind als linke (Farkas & Cheung, 1981; Peck, Peck & Kataja, 1991; Sackeim, 1985; Simmons, Rhodes, Peters & Koehler, 2004)). Der Erwartungswert ist hierbei von Null verschieden. Aktuelle Forschungen v. a. aus dem Bereich der Biologie über den Zusammenhang Symmetrie und Entwicklungsstabilität, Gesundheit, Krankheitsresistenz u. a. beziehen sich in der Regel auf die Fluktuierende Asymmetrie (FA). Direktionale Asymmetrie gilt hier als das „Rauschen“, das das eigentliche „Signal“, nämlich die FA, maskiert und ihre Untersuchung erschwert, da es bei der Messung dieser Asymmetrien teilweise Konfundierungen zwischen beiden Größen gibt (Graham, Emlen, Freeman, Leamy & Kieser, 1998; Palmer & Strobeck, 1986).

Erschwerend kommt hinzu, dass FA nicht einheitlich operationalisiert wird, da hiermit meistens zufällige Abweichungen von der perfekten Bilateralsymmetrie gemeint sind, manche Autoren (Palmer & Strobeck, 1986) jedoch als Kriterium, ob eine Abweichung vorliegt oder nicht, den Populationsdurchschnitt zugrunde legen. Beispiel: Wenn bei einem Individuum der rechte äußere Augenwinkel zwei Millimeter oberhalb des linken äußeren Augenwinkels liegt, dann handelt es sich *nicht* um eine Asymmetrie, wenn im Durchschnitt bei allen Individuen einer Population dieses Merkmal so ausgeprägt ist. Erst bei einer vertikalen Abweichung von mehr als zwei Millimetern oder weniger als zwei Millimetern liegt nach dieser Definition eine fluktuierende Asymmetrie vor.

4.2 Einführung in die verschiedenen Methoden zur Untersuchung des Einflusses von Symmetrie

Diese oben genannten Operationalisierungen sind jedoch bereits gebunden an eine bestimmte Methode zur Erfassung von Symmetrie, nämlich das (physikalische) Vermessen der Distanzen zwischen verschiedenen, definierten markanten Punkten im Gesicht, den *Facial Landmarks*. Dies ist eine sehr präzise Methode, die zwar innerhalb der Biologie (bei Körpern von Tieren) häufig, jedoch innerhalb der Attraktivitätsforschung nur von einem Teil der Forscher verwendet wird. Viele andere verwenden ungenauere, „weichere“ oder indirekte Verfahren. Die folgende Grafik zeigt einen systematischen Überblick über die unterschiedlichen Verfahren.

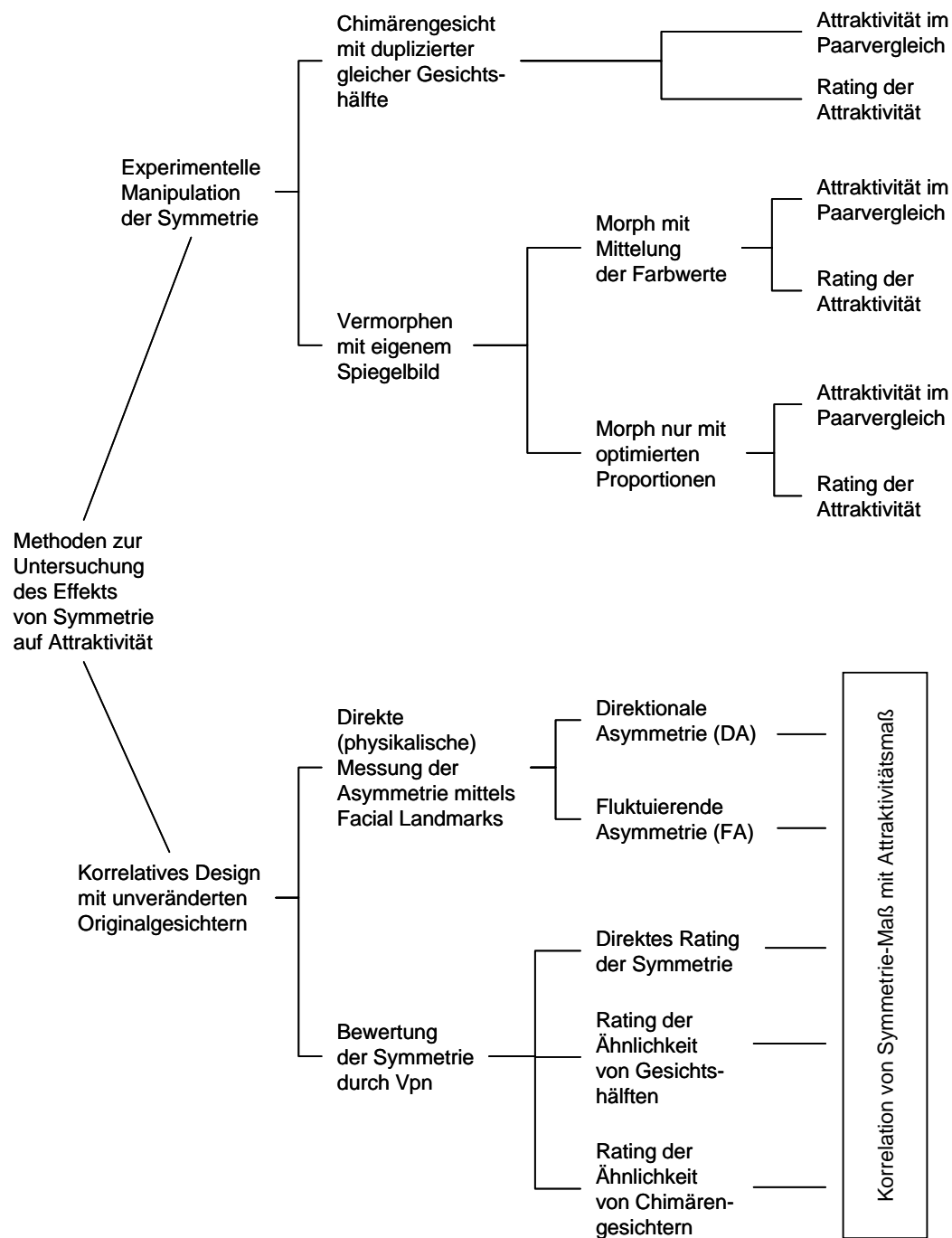


Abbildung 40: Übersicht über die verschiedenen Methoden zur Untersuchung des Einflusses der Symmetrie auf die Attraktivität.

Zunächst lassen sich zwei grundsätzliche Methoden unterscheiden:

- Die Experimentelle Manipulation der unabhängigen Variablen Symmetrie, bei der normale Gesichter mit natürlichen Asymmetrien symmetrisch optimiert werden, während Versuchspersonen die Attraktivität der jeweiligen Gesichter-Varianten beurteilen.
- Ein korrelatives Design, bei der die Asymmetrie natürlicher Gesichter ermittelt wird und dann mit den jeweiligen Attraktivitätsbewertungen dieser Gesichter korreliert wird.

Die simpelste Form eines experimentellen Designs ist das Produzieren von Chimärengesichtern, bei dem mit einem Bildverarbeitungsprogramm ein frontal fotografiertes Gesicht entlang der Symmetrieachse in seine zwei Gesichtshälften geteilt und durch horizontales Spiegeln ein neues Gesicht kreiert wird, das entweder aus zwei linken oder aus zwei rechten Hälften konstruiert wird. Diese Art der Symmetrisierung wurde erstmalig zu Beginn des 20. Jahrhunderts eingesetzt (Hallevorden, 1902, 1929), wurde schon damals für Experimente verwendet (Wolff, 1933) und ist unter dem Namen „Wolff's split-face technique“ bekannt (zit. nach Kowner, 1996b).



Abbildung 41: Chimärengesicht aus zwei rechten Gesichtshälften (links), Originalgesicht (Mitte) und Chimärengesicht aus zwei linken Gesichtshälften (rechts).

Die Attraktivität eines Chimärengesichts wird dann mit der Attraktivität des asymmetrischen Originalgesichts durch Paarvergleich oder Rating verglichen. Die Methode gilt als sehr schlecht, dennoch gibt es einige Studien, in denen sie eingesetzt wurde (Kowner, 1996a; Langlois et al., 1994; Samuels, Butterworth, Roberts, Graupner & Hole, 1994). Hauptkritikpunkt ist, dass die dadurch produzierten symmetrischen Gesichter meist unnatürlich starr oder seltsam aussehen. Dies gilt insbesondere dann, wenn der Experimenta-

tor handwerklich nicht sauber arbeitet und Personen auf den Fotos nicht ganz frontal abgelichtet sind, sondern mit leicht gedrehtem Kopf, wenn die Ausleuchtung des Gesichts auf dem Foto unregelmäßig ist oder die Symmetrieachse vor dem Duplizieren einer Gesichtshälfte nicht exakt durch die Gesichtsmitte verläuft. Das typische Ergebnis solcher Studien ist, dass symmetrische Gesichter schlechter abschneiden als asymmetrische Originale.

Ein anderes experimentelles Verfahren ist der Einsatz von Morphing-Software zur Herstellung perfekt symmetrischer Gesichter. Hierbei wird jedes Gesicht mit seinem eigenen Spiegelbild gekreuzt („*blends*“). Bei der Neuberechnung der Proportionen werden beide Bilder mit 50 % gewichtet, die Farbwerte jedes Gesichts gehen ebenfalls zu 50 % in das neue Bild ein (Rhodes, Proffitt, Grady & Sumich, 1998; Rhodes, Roberts et al., 1999; Rhodes, Sumich et al., 1999; Swaddle & Cuthill, 1995). Der Vorteil gegenüber der Chimären-gesichter-Methode ist, dass die symmetrisierten Morphs sehr natürlich aussehen. Der Nachteil ist, dass durch die Neuberechnung der Farbwerte zwar auch die Farbverteilung zwischen den beiden Gesichtshälften exakt gleich ist, jedoch gleichzeitig die Haut weicher und makelloser erscheint. Je unreiner die Haut eines Individuums ist, desto größer ist der positive Effekt, wenn man das Gesicht mit dieser Methode symmetrisch optimiert. Derart veränderte Gesichter wirken meist attraktiver als die entsprechenden Originale, jedoch ist nicht mehr zu unterscheiden, ob der Effekt auf die Symmetrie oder die makelloosere Haut zurückzuführen ist. Experimente mit dieser Methode erzielten meist kleine bis mittelgroße Effektstärken.

Eine methodisch korrektere Variante, die Morphing-Methode zu verwenden, ist die Studie von Perrett et al., die lediglich die Gesichtsproportionen symmetrisch optimierten, nicht aber die Farbwerte neu berechneten (Perrett et al., 1999). Dadurch behält das Gesicht dieselbe Hautstruktur wie beim Original. Die Unterschiede zum Original sind hierbei allerdings nicht mehr so deutlich wie bei der Methode *mit* Neuberechnung der Farbwerte (vgl. Abbildung 42).



Abbildung 42: Das asymmetrische Originalgesicht (links), ein symmetrisch optimiertes Gesicht nach dem herkömmlichen Morphing-Verfahren mit Neuberechnung der Farbwerte, wodurch die Haut glatter und makelloser erscheint (Mitte), sowie ein symmetrisch optimiertes Gesicht (rechts), bei dem nur die Proportionen mit dem eigenen Spiegelbild vermorpht wurden, die Original-Farbwerte jedoch beibehalten wurden.

Bei allen experimentellen Verfahren, bei denen die UV Symmetrie manipuliert wird, gibt es verschiedene Möglichkeiten, die Attraktivität der beiden Gesichter-Varianten zu ermitteln. Häufig werden Paarvergleiche verwendet (mit *forced choice*) – dadurch sind auch geringe Attraktivitätsunterschiede noch statistisch nachweisbar. Seltener ist das Rating der Varianten, das jedoch meist dann verwendet wird, wenn mehrere Variablen gleichzeitig (varianzanalytisch) untersucht werden (wie z. B. Symmetrie, Durchschnitt, Ausdruck und verschiedene Versuchspersonen-Variablen, vgl. beispielsweise Rhodes, Sumich & Byatt, 1999, Rhodes, Yoshikawa et al., 2001, Rhodes, Zebrowitz et al., 2001). Dieses hat den Vorteil, dass nicht nur der Unterschied zwischen den beiden Varianten erfasst wird, sondern zugleich auch, wie attraktiv ein Gesicht generell ist. Sinnvoll ist hier ein Between-Subject-Design, bei dem jede Vp entweder nur die originale oder nur die symmetrisch optimierte Version eines Gesichts zu sehen bekommt. Nachteil ist, dass aufgrund der geringen Mittelwertsdifferenzen große Stichproben notwendig sind, um signifikante Effekte nachweisen zu können.

Bei den korrelativen Versuchsdesigns gibt es wiederum zwei völlig konträre Herangehensweisen. Während die einen Forscher anthropometrische Messungen vornehmen und Symmetrie mit Hilfe von *Facial Landmarks* physikalisch vermessen, lassen andere wiederum die Symmetrie von Gesichtern einfach von Versuchspersonen beurteilen, in der Regel mit Hilfe eines Ratingverfahrens (z. B. Frage: „Wie symmetrisch auf einer Skala

von 1 bis 7 ist dieses Gesicht?“). Obwohl Symmetrie-Ratings relativ häufig eingesetzt werden (z. B. Langlois et al., 1994; Rhodes et al., 1998; Rhodes, Roberts et al., 1999; Rhodes, Sumich et al., 1999; Rhodes, Zebrowitz et al., 2001; Simmons et al., 2004; Zebrowitz, Voinescu, & Collins, 1996) – selbst von renommierten Attraktivitätsforschern wie Rhodes – bestehen jedoch aus methodischer Sicht erhebliche Zweifel, ob Versuchspersonen überhaupt in der Lage sind, Symmetrie isoliert von anderen Eigenschaften des Gesichts zu bewerten. Gerade wenn man sich die zahlreichen Befunde zum Attraktivitätsstereotyp (Eagly, Ashmore, Makhijani & Longo, 1991; Feingold, 1992; Hosoda, Stone-Romero & Coats, 2003; Jackson, Hunter & Hodge, 1995; Langlois et al., 2000; Mazzella & Feingold, 1994) vor Augen hält, so drängt sich der Verdacht geradezu auf, dass die kleinen bis mittleren Effektstärken, die bei diesen Symmetrie-Untersuchungen bisher erzielt wurden, zum Großteil auf einen Halo-Effekt der Attraktivität des Gesichts zurückzuführen sind, dass also attraktive Gesichter eher als symmetrisch *bezeichnet* werden, obwohl sie es in Wirklichkeit gar nicht sind.

Ein alternatives Verfahren zur Bewertung der Symmetrie durch Vpn ist ein Ähnlichkeits-Rating von jeweils zwei Chimärengesichtern, die entweder aus zwei rechten oder zwei linken Hälften eines Gesichts hergestellt wurden. Je größer die Ähnlichkeit der Chimärengesichter, desto symmetrischer ist das Original (Mealey et al., 1999; Penton-Voak, Jones et al., 2001). Alternativ dazu lassen sich die beiden Hälften des Originalgesichts unverbunden nebeneinander präsentieren (Hönekopp et al., 2004; Scheib, Gangestad & Thornhill, 1999) und somit ein Ähnlichkeitsurteil erheben (eine Hälfte wird dabei horizontal gespiegelt). Das Erheben eines Ähnlichkeitsurteils für beide Gesichtshälften ist aufgrund der geringeren Konfundierung des Urteils mit der Attraktivität des Gesichts methodisch geschickter als die direkte Frage nach Symmetrie, wurde aber nur selten eingesetzt.

Aus theoretischer Sicht ist zudem bei allen Symmetrie-Bewertungen durch Vpn problematisch, dass völlig offen bleibt, welche Art von Symmetrie die Vpn eigentlich bewerten, die direktionale oder die fluktuierende.

Die präziseste Messung von Symmetrie ist das bereits oben erwähnte Vermessen von *Facial Landmarks* mit dem Berechnen der Abweichungen dieser Referenzpunkte von der idealen Symmetrie. Für jedes Gesicht wird dann basierend auf der Summe der Abweichungen ein Gesamtwert für Asymmetrie berechnet (*Composite Index of Fluctuating Asymmetry* = *CFA*), der mit dem Attraktivitätsurteil korreliert werden kann (Leung, Forbes & Houle, 2000). Attraktivitätsstudien, bei denen Symmetrie durch Landmarks gemessen

wurde, existieren etliche (Baudouin & Tiberghien, 2004; Grammer & Thornhill, 1994; Hönekopp et al., 2004; Hume & Montgomerie, 2001; Jones et al., 2001; Jones & Hill, 1993; Kowner, 1996a; Penton-Voak & Perrett, 2001; Rhodes, Zebrowitz et al., 2001; Rikowski & Grammer, 1999; Shackelford & Larsen, 1997; Simmons et al., 2004), allerdings gibt es zum Teil große Unterschiede in den verwendeten Landmarks, den zwischen Landmarks gemessenen Strecken (Simmons et al., 2004) und der Formel, mit der ein Gesamt-Index für die Asymmetrie eines Gesichts berechnet wird. In einem Review von 1986 zählten Palmer und Strobeck bereits 22 verschiedene Berechnungsverfahren. Unterschiede sind beispielsweise,

- ob mit den Beträgen oder den Varianzen von Abweichungen gerechnet wird,
- ob berücksichtigt wird, welche Größe ein bestimmtes Merkmal hat (*allometry*), und wenn ja, wie eine Standardisierung erfolgt (am Individuum oder an der Population) (Leung, 1998)
- ob berücksichtigt wird, dass bei bestimmten Gesichtsmerkmalen Asymmetrien natürlicherweise ausgeprägter sind als bei anderen (Palmer & Strobeck, 1986).

Bei der anthropometrischen Messung von Symmetrie erzielten bisherige Studien manchmal kleine (Baudouin & Tiberghien, 2004; Hume & Montgomerie, 2001; Penton-Voak, Jones et al., 2001; Rhodes, Zebrowitz et al., 2001), seltener mittlere Effektgrößen (Grammer & Thornhill, 1994; Jones et al., 2001; Rikowski & Grammer, 1999; Scheib et al., 1999), am häufigsten aber Nulleffekte (Hönekopp et al., 2004; Jones & Hill, 1993; Kowner, 1996a; Rhodes, Zebrowitz et al., 2001; Rikowski & Grammer, 1999; Shackelford & Larsen, 1997; Simmons et al., 2004).

Bei einer solchen Vielfalt unterschiedlicher Methoden – experimentell mit primitiven Chimärengesichtern oder raffinierten Morphings, korrelativ mit simplen Symmetrie-Ratings oder komplizierten Landmark-Vermessungen – verwundert es nicht, wenn dies zu völlig verschiedenen und teilweise einander widersprechenden Ergebnissen führt. Zusammenfassend muss festgestellt werden, dass trotz zahlreicher Untersuchungen zum Zusammenhang von Symmetrie und Attraktivität völlig unklar ist, wie groß der Einfluss von Symmetrie auf die Attraktivitätsbewertung eines Gesichts ist.

4.3 Ziel der Untersuchungen zum Einfluss der Symmetrie und Hypothesen

Gegenwärtig herrscht die Situation vor, dass die Ergebnisse der verschiedenen Studien zum Zusammenhang von Symmetrie und Attraktivität untereinander gar nicht vergleichbar sind, da sie mit ganz unterschiedlichen Methoden unterschiedlicher Qualität gewonnen wurden. Entsprechend findet jeder unter der Fülle publizierter Untersuchungen Belege, die seine eigene Meinung zu bestätigen scheinen. Manche Arbeiten fanden einen großen Effekt der Symmetrie, manche gar keinen und manche sogar einen negativen Effekt.

Ziel der geplanten eigenen Untersuchungen ist es daher, an ein und demselben Stimulusmaterial alle bekannten Methoden zur Untersuchung der Symmetrie-Hypothese anzuwenden und die Effektgröße der erzielten Ergebnisse zu vergleichen.

Auf eine Auflistung aller Methoden wird an dieser Stelle verzichtet, da bereits die schematische Übersicht (Abbildung 7) einen guten Überblick über die verschiedenen Verfahren bietet.

Ergänzend dazu werden weitere Experimente durchgeführt, um methodische Streitfragen zu klären. Ein Beispiel:

So werden (bei der experimentellen Manipulation von Symmetrie) Morphings mit symmetrisch optimierten Proportionen *und* gemittelten Farbwerten mit ihren asymmetrischen Pendants verglichen. Parallel dazu werden Morphings mit *nur* symmetrisch optimierten Proportionen *ohne* gemittelte Farbwerte mit ihren asymmetrischen Pendants verglichen. Zu erwarten ist, dass der Effekt beim Experiment, bei dem zusätzlich auch Farbwerte gemittelt wurden, größer ausfällt, da hierbei (als Nebeneffekt) auch die Hautqualität verbessert wird.

Ein weiteres Zusatzexperiment wird dem Nachweis dienen, dass es methodisch unsinnig ist, Versuchspersonen direkt die Symmetrie eines Gesichts auf einer Ratingskala beurteilen zu lassen, wie dies jedoch häufig praktiziert wird. Hierzu werden ausschließlich in Gesichtsproportionen symmetrisch optimierte Morphings präsentiert und die Vpn um eine Bewertung der Symmetrie gebeten. Da alle präsentierten Gesichter perfekt symmetrisch sind, sollten alle gleich hohe Symmetrie-Ratings bekommen. Die Vorhersage ist jedoch, dass das Symmetrie-Rating von der Attraktivität des präsentierten Gesichts abhängt. Es

wird zudem vorhergesagt, dass die Höhe der Korrelation genauso groß ist wie bei der Präsentation von unveränderten Originalgesichtern. Das wäre der Beweis dafür, dass Versuchspersonen Symmetrie nicht isoliert von Attraktivität bewerten können (= Halo-Effekt der Attraktivität) und damit bei allen Studien, die mit dieser Methode gearbeitet haben, eine Konfundierung zwischen UV und AV vorliegt.

Das Stimulusmaterial, das für alle Untersuchungen verwendet wird, besteht aus 15 „normalen“ jungen frontal fotografierten Frauengesichtern sowie 15 frontal fotografierten Gesichtern der Endrundenteilnehmerinnen der Miss-Germany-Wahl 2007. Durch die Verwendung von Endrundenteilnehmerinnen eines Schönheitswettbewerbs wird sichergestellt, dass tatsächlich auch attraktive Individuen im verwendeten Stimulusmaterial vorhanden sind, da ein häufiges Problem hinsichtlich der externen Validität der Umstand ist, dass Untersuchungen zur Attraktivität gemacht werden, ohne dazu wirklich attraktives Stimulusmaterial zu verwenden.

Zusammenfassung der Fragestellungen und Hypothesen zu Untersuchungen zur Symmetrie:

- Die Größe und Richtung eines Effekts von Symmetrie auf die Attraktivitätsbewertung hängt stark von der verwendeten Untersuchungsmethode ab.
- Je exakter und valider die Untersuchungsmethode, desto stärker tendieren erzielte Effekte gegen Null.
- Die Methode der Chimärengesichter führt zu einem negativen Effekt der Symmetrie auf die Attraktivität.
- Experimente mit Morphings, die in den Farbwerten gemittelt wurden, erzielen größere Effekte als solche ohne Farbwertmittelung.
- Experimente mit direktem Paarvergleich (symmetrisch vs. asymmetrisch) erzielen deutlichere Ergebnisse als Experimente mit Rating-Verfahren und *Between-Subjects Design*.
- Bei allen Experimenten im *Between-Subjects Design* sind die Effektgrößen marginal.
- Direktionale Asymmetrie ist für Vpn leichter wahrnehmbar als fluktuierende.
- Vertikale Abweichungen (y-Wert-Differenzen) haben einen größeren Effekt auf die Attraktivität als horizontale Abweichungen (x-Wert-Differenzen zur Symmetrieachse).
- Alle Methoden der anthropometrischen Messung von *Facial Landmarks* führen nur zu geringen Effektgrößen.

- Die Methode des direkten Ratings der Symmetrie von Gesichtern führt zu mittleren bis großen Effekten.
- Es kann nachgewiesen werden, dass bei der Methode des direkten Ratings der Symmetrie von Gesichtern UV und AV konfundiert sind.
- Es gibt bei den Versuchspersonen ein ausgeprägtes „Symmetrie-Stereotyp“ nach dem Motto: „Was schön ist, ist auch symmetrisch.“

4.4 Experimentelle Manipulation der Symmetrie durch Verwendung von Chimärengesichtern und Attraktivitätsmessung durch Paarvergleiche (Untersuchung *Sym 1*)

4.4.1 Stand der Forschung

Die einzige Studie, bei der Chimärengesichter im direkten Paarvergleich mit Originalgesichtern präsentiert wurden, ist die von Kowner (1996). Er verwendete drei Altersgruppen von Gesichtern als Stimuli, nämlich Kinder, junge Erwachsene und alte Menschen. Kowner gab vor, die Nullhypothese beweisen zu wollen, dass also die kleinen Asymmetrien, die in normalen, gesunden Gesichtern vorhanden sind, von Versuchspersonen gar nicht bemerkt werden und sich deshalb auch nicht auf die Attraktivitätsbewertung auswirken. Bei Kindergesichtern und Gesichtern von jungen Erwachsenen wurde in der Untersuchung tatsächlich kein Attraktivitätsunterschied zwischen symmetrisierten Chimärengesichtern und Originalen gefunden. Gesichter von alten Menschen wurden hingegen besser bewertet, wenn sie symmetrisiert waren. Mit Daten aus einem zusätzlichen Experiment kam Kowner zu der Schlussfolgerung, dass Symmetrie als Signal für Jugendlichkeit diene, da in hohem Alter die Asymmetrien in Gesichtern zunehmen.

Zusätzlich erwähnenswert an dieser Untersuchung ist, dass Kowner bei der Erzeugung seiner Chimärengesichter recht sorgfältig arbeitete. So wurden beispielsweise die typischen Texturverdopplungsartefakte wie doppelte Muttermale auf beiden Seiten nachträglich wieder weggerechnet. Zudem wurde die Frisur des Originalgesichts nachträglich auf die jeweiligen Chimärengesichter kopiert, so dass es dort keine Artefakte wie doppelte Scheitel oder symmetrische Locken gibt. Grundsätzlich ist dies natürlich vernünftig, andererseits sind durch diesen Eingriff die verwendeten Gesichter keine „reinen“ Chimärengesichter mehr, bei denen auch die Textur vollständig symmetrisch ist. Dieses Stimulusma-

terial hat vielmehr bereits Ähnlichkeit mit der Methode des Warpens (vgl. Kapitel 4.8 und 4.9, *Sym 5* und *6*).

Kritisch anzumerken ist, dass Kowner bei der statistischen Datenanalyse wie etliche andere Autoren auch (z. B. Perrett et al., 1999, Jones et al., 2007, Little & Jones, 2003, sowie Zaidel et al., 1995) zu einem statistischen „Trick“ greift, um die winzigen gemessenen Effekte signifikant zu bekommen. Bei den binomialverteilten Daten der Paarvergleiche wurden keine Binomialtests gerechnet, sondern stattdessen „the mean selection ratio of the symmetrical portraits of each participant“ (Kowner, 1996, S. 667), so dass mit einem Ein-Stichproben-*t*-Test geprüft wurde, ob die mittlere Selektionsrate der symmetrischen Varianten signifikant von der A-priori-Wahrscheinlichkeit von 0.50 abwich. Nur durch diesen Kunstgriff der Urteilsvarianzreduktion wurde die Präferenz für die symmetrischen Varianten bei den alten Gesichtern, die im Durchschnitt (!) nur knapp über der Ratewahrscheinlichkeit liegt (55.1 %), mit nur 64 Urteilern signifikant.

4.4.2 Methode

4.4.2.1 Versuchspersonen

An der Untersuchung *Sym 1* nahmen insgesamt 128 Versuchspersonen teil. 59 Personen waren männlich (46.1 %), 69 weiblich (53.9 %). 26 Personen (20.3 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 14 Jahre, die älteste 83. Der Altersdurchschnitt betrug 37.9 Jahre (*SD* = 17.3). Die 128 Versuchspersonen verteilten sich zu gleichen Teilen auf zwei Versuchsbedingungen. Hinsichtlich der erhobenen demografischen Variablen unterscheiden sich die Teilstichproben nicht (Geschlecht: $\chi^2(1, N = 128) = 0.283, p = .595$; Alter: $t(191) = -0.698, p = .486$; Studierenden-Status: $\chi^2(1, N = 128) = 0.000, p = 1.000$).

Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, fanden die Datenerhebungen im öffentlichen Raum statt, z. B. an Bahnhöfen, Fußgängerzonen und Einkaufszentren. Die Teilnahme am Experiment erfolgte freiwillig und wurde nicht entlohnt.

4.4.2.2 Stimulusmaterial

Als Stimulusmaterial dienten Fotos von 30 jungen Frauen. 15 Frauen waren Endrundenteilnehmerinnen der Miss-Germany-Wahl 2007, 15 Frauen waren „normale“ junge Frauen. Alle Stimuluspersonen wurden von professionellen Fotografen in einem Fotostudio aufgenommen. Aus organisatorischen Gründen wurden Miss-Germany-Finalistinnen und „normale“ Frauen in zwei unterschiedlichen Lokalitäten von zwei verschiedenen Fotografen aufgenommen. Aufgrund der professionellen Aufnahmebedingungen und der genauen Vorgaben für das Fotografieren der Gesichter sind jedoch die Fotos beider Fotoserien sehr gut vergleichbar und weisen in Belichtung, Farbgebung etc. praktisch keine Unterschiede auf.



Abbildung 43: Beispiel für das verwendete Stimulusmaterial. Das mittlere Bild zeigt ein unverändertes Originalgesicht. Das linke Bild zeigt ein Chimärengesicht, das aus der rechten Gesichtshälfte des Originalgesichts erzeugt wurde. Das rechte Bild zeigt ein Chimärengesicht aus der linken Gesichtshälfte.

Mit einem Bildverarbeitungsprogramm (*Adobe Photoshop*) wurden aus jedem Originalgesicht zwei perfekt symmetrische Chimärengesichter erzeugt. Dazu wurde jedes Gesicht exakt in der Mitte geteilt und aus einer Gesichtshälfte durch Verdopplung und horizontale Spiegelung um eine vertikale Achse ein neues Gesicht zusammengesetzt, das aus zwei gleichen Gesichtshälften bestand. Dasselbe wurde mit der anderen Gesichtshälfte gemacht. Da es bei diesem Verfahren leicht zu Artefakten kommen kann, wurde gegebenenfalls bei Gesichtern, die eine leicht geneigte Kopfhaltung hatten, vor dem Setzen der Bisymmetrieachse durch die Gesichtsmittte das Gesicht rotiert, damit das Gesicht exakt senkrecht im Bildausschnitt stand. Die Beurteilung, wann ein Gesicht senkrecht stand, erfolgte nach dem Augenmaß, das in diesem Fall ein weitaus valideres Kriterium ist als beispielsweise eine Ausrichtung an der Augachse. Denn erstens führt eine Justierung des Gesichts anhand einer horizontalen Augachse zu unterschiedlichen Ergebnissen je nach-

dem, welche *Facial Landmarks* man heranzieht (Pupillenmitte, reflektierender Glanzpunkte durch das Blitzlicht, innere Augenwinkel oder äußere Augenwinkel), und zweitens eignet sich die Augachse prinzipiell nicht, da es nicht selten vorkommt, dass bei einem Gesicht ein Auge tiefer liegt als das andere. Dies würde bei einigen Gesichtern zu sehr starken Drehungen des Gesichts im Bildausschnitt und damit aufgrund der nicht vertikal durch das Gesicht verlaufenden Spiegelachse zu starken, teilweise bizarr aussehenden Artefakten führen. In den meisten Fällen besteht jedoch keine Diskrepanz zwischen einer Justierung nach dem Augenschein und einer horizontalen Ausrichtung der Augachse.

Alle Bilder hatten eine hohe Auflösung von 1600 x 1600 Pixel. Für die Versuchsdurchführung wurden sie von einem Fotodienst auf Fotopapier in einer Größe von 9 x 9 Zentimeter belichtet. Daraus ergibt sich eine Auflösung von 452 dpi. Durch diese hohe Auflösung sind die Versuchspersonen in der Lage, selbst kleinste Details wahrzunehmen – ein großer Vorteil gegenüber Attraktivitätsuntersuchungen, die am Bildschirm stattfinden, der normalerweise nur eine Auflösung von 72 dpi bietet.

4.4.2.3 Versuchsdurchführung

Die Datenerhebung bestand aus zwei voneinander unabhängigen Experimenten. In einem Experiment bekamen die Versuchspersonen jeweils paarweise Chimären Gesichter, die aus linken Gesichtshälften hergestellt waren, und die dazugehörigen Originalgesichter präsentiert, im anderen Experiment entsprechend das Stimulusmaterial mit den Chimären Gesichtern aus rechten Gesichtshälften.

Die Fotos wurden in einer Fotomappe präsentiert, durch die seitenweise geblättert werden konnte. Dabei konnten die Versuchspersonen stets nur ein Gesichterpaar betrachten und bewerten. Bei der Anordnung der Fotos wurde darauf geachtet, dass die symmetrisch optimierten Fotos gleich häufig auf der linken Seite wie auf der rechten Seite präsentiert wurden, um eventuelle Lerneffekte (z. B. „Das symmetrische Foto ist immer rechts“) oder Präferenzen für eine bestimmte Seite (vgl. *left-field perceptual bias of the observer* (Gilbert & Bakan, 1973; Mertens, Siegmund & Grüsser, 1993)) von vornherein ausschließen zu können. Zusätzlich wurde die Reihenfolge der präsentierten Stimuluspaare für verschiedene Probanden vielfach variiert, obwohl durch die Präsentation von Gesichterpaaren Positionseffekte durch eine bestimmte Reihenfolge der Stimuluspaare kaum vorstellbar sind.

Allen Versuchspersonen wurde die Frage gestellt: „Welches Gesicht ist attraktiver?“ Auch wenn sie keinen Unterschied bzw. keinen Attraktivitätsunterschied zwischen den jeweili-

gen Varianten sahen, mussten die Probanden sich für ein Bild entscheiden (*forced choice*). Es wurden keine zeitlichen Beschränkungen auferlegt. Der Versuchsleiter notierte die Urteile.

4.4.3 Ergebnisse

Zur Überprüfung der Frage, ob durch das Chimären-Verfahren symmetrisch optimierte Gesichter im direkten Paarvergleich attraktiver bewertet wurden als die unveränderten Originalgesichter, wurden in beiden Experimenten jeweils 30 Binomialtests durchgeführt. Aufgrund des gewählten Forced-Choice-Verfahrens liegt die Ratewahrscheinlichkeit, die symmetrische Variante attraktiver zu finden, bei 50 %. Da in beiden Experimenten die Anzahl an Versuchspersonen $N = 64$ betrug, lag ein signifikantes Ergebnis ($p < .05$, zweiseitige Tests) vor, wenn mindestens 41 von 64 (64.1 %) Personen eine Variante attraktiver fanden als die andere; hoch signifikant ($p < .01$) war der Unterschied bei 43 von 64 (67.2 %) Personen. Da in beiden Experimenten 30 Einzeltests durchgeführt wurden, müsste eigentlich eine Adjustierung des α -Niveaus erfolgen, so dass bei einer Bonferroni-Korrektur ein signifikantes Ergebnis erst ab $p = .0016$ vorliegt; dies wäre bei einer Bevorzugung von 45 aus 64 (70.3 %) Personen der Fall. Bei den hier durchgeführten Untersuchungen zum Effekt der Symmetrie ging es in erster Linie nicht um statistische Signifikanz, sondern vielmehr um eine Beurteilung der Effektgröße in Abhängigkeit von der gewählten Untersuchungsmethode. Aus diesem Grund wurden auch alle Prozentwerte nach folgender Umrechnungsformel (Rosenthal & Rubin, 1982) von der relativen Häufigkeit (Binomial effect size display, *BESD*) in ein r umgerechnet:

$$r = 2 * BESD - 1$$

Beispiel: Wenn das symmetrisierte Gesicht von 75 % der Versuchspersonen gegenüber der Originalvariante bevorzugt wird, dann ist $BESD = 0.75$ und dann ist $r = 0.50$. Tabelle 8 listet getrennt für beide Experimente die absolute und relative Häufigkeit der Bevorzugung der symmetrisierten Variante gegenüber der unveränderten Original-Variante sowie die in r umgerechnete Effektgröße auf. Am Ende der Tabelle ist jeweils der Median der erzielten Effektgrößen aufgeführt.

Tabelle 8: Ergebnisse der beiden Paarvergleichs-Experimente mit Chimärengesichtern aus linken und rechten Gesichtshälften im direkten Vergleich mit dem jeweiligen Originalgesicht. Aufgelistet sind absolute und relative Häufigkeiten sowie die in r umgerechneten Effektgrößen. Bei beiden Experimenten war jeweils $N = 64$.

Stimulus-Gesicht	Stimulus-Gruppe	Chimäre links gegenüber Original bevorzugt n (%)	r	Chimäre rechts gegenüber Original bevorzugt n (%)	r
1	Miss	23 (35.9 %)	-0.28	35 (54.7 %)	0.09
2	Miss	21 (32.8 %)	-0.34	30 (46.9 %)	-0.06
3	Miss	8 (12.5 %)	-0.75	29 (45.3 %)	-0.09
4	Miss	27 (42.2 %)	-0.16	29 (45.3 %)	-0.09
5	Miss	16 (25.0 %)	-0.50	45 (70.3 %)	0.41
6	Miss	34 (53.1 %)	0.06	22 (34.4 %)	-0.31
7	Miss	18 (28.1 %)	-0.44	43 (67.2 %)	0.34
8	Miss	34 (53.1 %)	0.06	16 (25.0 %)	-0.50
9	Miss	30 (46.9 %)	-0.06	20 (31.3 %)	-0.38
10	Miss	19 (29.7 %)	-0.41	32 (50.0 %)	0.00
11	Miss	32 (50.0 %)	0.00	32 (50.0 %)	0.00
12	Miss	12 (18.8 %)	-0.63	45 (70.3 %)	0.41
13	Miss	29 (45.3 %)	-0.09	9 (14.1 %)	-0.72
14	Miss	24 (37.5 %)	-0.25	35 (54.7 %)	0.09
15	Miss	24 (37.5 %)	-0.25	29 (45.3 %)	-0.09

Fortsetzung

Fortsetzung von Tabelle 8:

Stimulus-Gesicht	Stimulus-Gruppe	Chimäre links gegenüber Original bevorzugt <i>n</i> (%)	<i>r</i>	Chimäre rechts gegenüber Original bevorzugt <i>n</i> (%)	<i>r</i>
16	Normale	22 (34.4 %)	-0.31	4 (6.3 %)	-0.88
17	Normale	14 (21.9 %)	-0.56	13 (20.3 %)	-0.59
18	Normale	25 (39.1 %)	-0.22	40 (62.5 %)	0.25
19	Normale	9 (14.1 %)	-0.72	49 (76.6 %)	0.53
20	Normale	31 (48.4 %)	-0.03	40 (62.5 %)	0.25
21	Normale	12 (18.8 %)	-0.63	50 (78.1 %)	0.56
22	Normale	18 (28.1 %)	-0.44	36 (56.3 %)	0.13
23	Normale	49 (76.6 %)	0.53	19 (29.7 %)	-0.41
24	Normale	28 (43.8 %)	-0.13	39 (60.9 %)	0.22
25	Normale	8 (12.5 %)	-0.75	56 (87.5 %)	0.75
26	Normale	31 (48.4 %)	-0.03	39 (60.9 %)	0.22
27	Normale	23 (35.9 %)	-0.28	30 (46.9 %)	-0.06
28	Normale	29 (45.3 %)	-0.09	25 (39.1 %)	-0.22
29	Normale	13 (20.3 %)	-0.59	38 (59.4 %)	0.19
30	Normale	45 (70.3 %)	0.41	14 (21.9 %)	-0.56
Median		23.5 (36.7 %)	-0.27	32 (50.0 %)	0.00

Anmerkung: Kursive *r* zeigen einen signifikanten Effekt ($p < .05$) an.

Wie aus der Tabelle 8 zu entnehmen ist, wurde in der Mehrzahl der Fälle das symmetrische Chimärengesicht schlechter bewertet als das jeweilige Originalgesicht. Bei den Chimärengesichtern aus linken Gesichtshälften wurde in 26 von 30 Fällen das Originalgesicht bevorzugt. Bei den Chimärengesichtern aus rechten Gesichtshälften sind die Ergebnisse gemischt: In 8 von 30 Fällen wurde das Original bevorzugt, in 6 Fällen das Chimärengesicht und bei den verbleibenden 16 Paarvergleichen gab es keinen signifikanten Unterschied.

Der Median der Effektgrößen liegt bei Chimärengesichtern aus linken Gesichtshälften bei $r = -0.27$ und bei Chimärengesichtern aus rechten Gesichtshälften bei $r = 0.00$. Zwischen den beiden Gruppen von Stimulusgesichtern (Miss-Germany-Finalistinnen versus „normale“ Frauen) unterscheiden sich die Effektgrößen nicht bzw. unwesentlich (Chimäre links: -0.25 versus -0.28 ; Chimäre rechts -0.06 versus 0.19).

Es gibt jedoch auch eine alternative Form der Datenauswertung, nämlich nach der Methode von Little und Jones (2003, S. 1761), bei der ausgewertet wird, wie häufig jede Versuchsperson die symmetrische Variante gegenüber der Originalvariante bevorzugt hat (= *mean selection ratio*, vgl. auch Jones et al., 2007, Kowner, 1996, sowie Zaidel et al., 1995). Nach der Nullhypothese sollte sie dies bei der Hälfte der präsentierten Gesichterpaare tun (d. h. die A-priori-Wahrscheinlichkeit ist 0.5). Für jede Versuchsperson wird daher eine relative Häufigkeit berechnet. Diese Häufigkeiten werden mit einem Einstichproben-*t*-Test analysiert. Das *N* ist in diesem Fall die Anzahl der Versuchspersonen ($N = 64$) und eine positive mittlere Differenz bedeutet, dass die symmetrische Variante häufiger bevorzugt wurde. Analysiert man die Daten mit dieser Methode, ergibt sich ein hoch signifikantes negatives Ergebnis $t(63) = -8.701$, $p < .001$ für Chimärengesichter aus linken Hälften und ein weiterhin nicht signifikantes Ergebnis für Chimärengesichter aus rechten Hälften: $t(63) = -0.444$, $p = .659$. Standardisiert man nach folgender Formel (Pospeschill, 2006)

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

den *t*-Wert auf eine Effektgröße *r*, um die Effektgröße mit den Ergebnissen der vorangegangenen statistischen Analyse vergleichen zu können, dann erhält man für die Chimärengesichter aus linken Hälften einen stark negativen Effekt von $r = 0.74$ (aufgrund der Formel kann das *r* immer nur positive Werte annehmen). Für die Chimärengesichter aus

rechten Hälften ändert sich dagegen im Vergleich zur ersten Datenanalyse nichts ($r = 0.06$ mit eigentlich negativem Vorzeichen).

Zusammenfassend bleibt festzuhalten, dass der in diesem Experiment untersuchte Effekt der Symmetrisierung durch Chimärengesichter und der Attraktivitätsbewertung im direkten Paarvergleich abhängig vom Stimulusmaterial zu keinen bis schwach negativen Effekten führte. Bei einer statistischen Auswertung (nach Little & Jones, 2003), bei der (durch einen Zwischenschritt der Berechnung von relativen Häufigkeiten) die Urteilsvarianz der Versuchspersonen (künstlich) reduziert wird, erhält man bei Chimärengesichtern aus linken Gesichtshälften sogar einen stark negativen Effekt der Symmetrisierung. Symmetrische Chimärengesichter wurden jedenfalls *nicht* attraktiver bewertet als unveränderte Originalgesichter.

4.5 Experimentelle Manipulation der Symmetrie durch Verwendung von Chimärengesichtern und Attraktivitätsmessung durch Ratings (Untersuchung *Sym 2*)

4.5.1 Stand der Forschung

Die früheste Studie, in der Chimärengesichter verwendet wurden, um den Einfluss von Symmetrie auf die Attraktivitätsbewertung zu untersuchen, stammt von den „Erfindern“ der Durchschnittshypothese Langlois und Roggman. Man darf wohl unterstellen, dass es den Autorinnen bei dieser Methode kein primäres Anliegen war, eine möglichst reliable und valide Methode der Symmetriemessung zu etablieren, denn ihr gesamter Artikel (Langlois et al., 1994) ist eine einzige Verteidigung der Durchschnittshypothese, in dem sie sämtliche Einwände gegen die Bedeutung der Durchschnittlichkeit (z. B. Konfundierung mit Symmetrie) zu entkräften versuchen, die seit ihrem viel zitierten ersten Artikel (Langlois & Roggman, 1990) vorgebracht wurden. So überrascht es auch nicht, dass die Autoren in ihrer Studie nicht nur keinen positiven Effekt der Symmetrie nachweisen konnten, sondern sogar einen stark negativen Einfluss. Sowohl Chimärengesichter aus linken Hälften ($t(47) = 8.27$) wie aus rechten Hälften ($t(47) = 5.6$) wurden in Ratings als deutlich unattraktiver als die Originalgesichter bewertet. Die Autoren verwendeten ein *Between-Design* und berechneten zuerst für jedes Originalgesicht aus den Attraktivitätsurteilen der „Kontrollgruppe“ jeweils einen durchschnittlichen Attraktivitätswert sowie für jedes Chimärengesicht (linke und rechte) aus den Attraktivitätsurteilen beider Experimentalgruppen

einen durchschnittlichen Attraktivitätswert und testeten dann mit jeweils einem *t*-Test für abhängige Stichprobe, ob Chimärengesichter im Durchschnitt anders als Originalgesichter bewertet wurden. Langlois und Kolleginnen folgern aus ihren Ergebnissen:

We conclude from these data that although some degree of symmetry is important to attractiveness, symmetry does not determine perceived attractiveness in a range of normal faces with no craniofacial deformities [...] Only faces that are exceptionally unattractive are likely to be improved by perfect symmetry, and, indeed, perfectly symmetrical faces are not necessarily attractive at all. (Langlois, Roggman & Musselman, 1994, S. 217)

Diese Interpretation der Daten kommt den Autorinnen natürlich entgegen, da sie „ihre“ Durchschnittshypothese verteidigen wollen. Und so findet sich auch kein einziger selbstkritischer Satz in ihrem Artikel, ob die von ihnen verwendete Methode, Chimärengesichter als symmetrische Stimuli zu verwenden, überhaupt valide ist. Dass Chimärengesichter durch Bildverarbeitungsartefakte (doppelte Scheitel etc.) beeinträchtigt sind und meist ein eigenartiges, starres Aussehen haben, thematisieren sie nicht, und illustrierende Bilder als Beispiele für Chimärengesichter sind im Artikel nicht vorhanden. Stattdessen versuchen sie nur, die vorgebrachten Einwände zu entkräften, ihre eigenen Experimente zur Durchschnittshypothese mit gemorphten Gesichtern seien durch Bildverarbeitungsartefakte beeinflusst.

In einer neueren Untersuchung kam dasselbe Untersuchungsdesign wie von Langlois et al. (1994) zur Anwendung (Chimärengesichter, *Between-Design*, Ratings, Auswertung mit *t*-Tests). Die Autoren verwendeten jedoch zusätzlich zu „normalen“ Gesichtern hochattraktive Gesichter von Models (50% der Stimulusgesichter). Es zeigte sich ebenfalls ein stark negativer Effekt der Symmetrie: Perfekt symmetrische Chimärengesichter wurden klar schlechter bewertet als Originalgesichter (Zaidel & Deblieck, 2007). Die Effektstärke lag in derselben Größenordnung wie bei Langlois und Roggman (*t*-Werte zwischen -5.50 und -6.08, bei Chimärengesichtern aus den linken Hälften der schönen Models „nur“ -2.47).

4.5.2 Methode

4.5.2.1 Versuchspersonen

An der Untersuchung *Sym 2* nahmen insgesamt 192 Versuchspersonen teil. 93 Personen waren männlich (48.4 %), 99 weiblich (51.6 %). 50 Personen (26.0 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 11 Jahre, die älteste 87. Der Altersdurchschnitt betrug 33.73 Jahre (*SD* =

15.31). Die 192 Versuchspersonen verteilten sich zu gleichen Teilen auf drei Versuchsbedingungen. Hinsichtlich der erhobenen demografischen Variablen unterschieden sich die Teilstichproben nicht (Geschlecht: $\chi^2(2, N = 192) = 0.375, p = .829$; Alter: $F(2, 191) = 0.530, p = .589$; Studierenden-Status: $\chi^2(2, N = 192) = 3.299, p = .192$).

Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, fanden die Datenerhebungen im öffentlichen Raum statt, z. B. an Bahnhöfen, Fußgängerzonen und Einkaufszentren. Die Teilnahme am Experiment erfolgte freiwillig und wurde nicht entlohnt.

Tabelle 9: Vergleich der demografischen Daten der Versuchspersonen in den drei Versuchsbedingungen

	Originalgesichter (= Kontrollgruppe)	Chimärengesichter aus linken Gesichtshälften	Chimärengesichter aus rechten Gesichtshälften
<i>N</i>	64	64	64
Anzahl Männer	33 (51.6%)	30 (46.9 %)	30 (46.9 %)
Anzahl Frauen	31 (48.4 %)	34 (53.1 %)	34 (53.1 %)
Mittelwert Alter (<i>SD</i>)	32.50 (14.99)	33.45 (15.02)	35.25 (16.00)
Anzahl Studierende	17 (26.6%)	21 (32.8 %)	12 (18.8 %)

4.5.2.2 Stimulusmaterial und Versuchsdurchführung

Das Stimulusmaterial war identisch mit dem Material der Untersuchung *Sym 1*. Die Daten wurden in einem *Between-Subjects Design* erhoben, so dass die Versuchspersonen einer Versuchsbedingung 30 Stimuli einer Variante präsentiert bekamen und vom Vorhandensein anderer Varianten derselben Gesichter nichts wussten. In der Kontrollbedingung bewerteten die Probanden die 30 unveränderten Originalgesichter hinsichtlich ihrer Attraktivität. In den beiden anderen Versuchsbedingungen bewerteten sie die 30 Chimärengesichter, die aus den linken bzw. rechten Gesichtshälften erzeugt waren.

Verwendet wurde dieselbe Ratingskala wie auch in den Experimenten zur Durchschnittshypothese mit folgenden Stufen: 1 = sehr unattraktiv, 2 = ziemlich unattraktiv, 3 = eher unattraktiv, 4 = mittelmäßig attraktiv, 5 = eher attraktiv, 6 = ziemlich attraktiv, 7 = sehr attraktiv. Während des Ratings der Gesichter hatten die Versuchspersonen stets einen Zettel mit der Ratingskala vor sich, um sich die Benennung der Skalenstufen nicht merken zu müssen und sich dadurch ganz auf die präsentierten Gesichter konzentrieren zu können. Die Reihenfolge der präsentierten Stimuli wurde für verschiedene Probanden vielfach variiert, um Positionseffekte zu vermeiden. Es wurden keine zeitlichen Beschränkungen auferlegt. Der Versuchsleiter notierte die Attraktivitätsbewertungen. Die Beurteilerübereinstimmung war bei allen drei Versuchsgruppen mit einem Cronbach's α von jeweils .99 (bei jeweils 64 Versuchspersonen) sehr hoch.

4.5.3 Ergebnisse

Zur Überprüfung der Frage, ob durch das Chimären-Verfahren symmetrisch optimierte Gesichter attraktiver bewertet wurden als die unveränderten Originalgesichter, wurde eine multivariate Varianzanalyse durchgeführt. Die unabhängige Variable ist die Versuchsbedingung in drei Gruppen (Originalgesichter, Chimärengesichter aus linken Gesichtshälften, Chimärengesichter aus rechten Gesichtshälften) und die abhängigen Variablen sind die 30 Stimulusgesichter in ihren jeweils zwei Varianten (original versus symmetrisch optimiert). Der Effekt der Manipulation war hoch signifikant ($p < .001$). Die Tabelle im Anhang B (S. 395) listet Mittelwerte und Streuungen dieser drei Versuchsbedingungen sowie die Signifikanzen der Post-hoc-Tests auf. Auf den ersten Blick ist ersichtlich, dass die unveränderten Originalgesichter in den meisten Fällen attraktiver bewertet wurden als die perfekt symmetrischen Chimärengesichter. Die Chimären aus linken Gesichtshälften wurden nur viermal gegenüber den Originalen bevorzugt, die aus den rechten Hälften ebenfalls nur viermal. Post-hoc-Tests (Scheffé) zeigten, dass der Attraktivitätsunterschied zwischen Originalen und Chimärengesichtern in den meisten Fällen signifikant war ($p < .05$).

Chimärengesichter aus *linken* Gesichtshälften waren in 14 Fällen signifikant unattraktiver als die jeweiligen Originalgesichter, nur in einem Fall war das Originalgesicht signifikant attraktiver. Chimärengesichter aus *rechten* Gesichtshälften waren in 12 Fällen signifikant unattraktiver als die jeweiligen Originalgesichter, nur in 2 Fällen war das Originalgesicht signifikant attraktiver (vgl. Anhang B, S. 395).

In den beiden letzten Spalten von Tabelle 10 sind in Klammern neben den Mittelwertdifferenzen die Effektgrößen ε (= Cohen's d) angeführt. Sie wurden berechnet nach folgender Formel:

$$\varepsilon = \frac{x_1 - x_2}{\hat{\sigma}}$$

Dabei ist x_1 jeweils der Attraktivitätsmittelwert des symmetrisch optimierten Gesichts und x_2 der Attraktivitätsmittelwert des Originalgesichts. Da die Standardabweichungen (die Werte der SD sind aus Platzgründen im Anhang B, S. 395 aufgeführt) bei den einzelnen Versuchsgruppen nicht immer gleich waren, wurde σ grundsätzlich nach folgender Formel (Bortz & Döring, 2009) berechnet:

$$\hat{\sigma} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$$

Da die Standardabweichungen meist um den Wert 1 lagen, besteht zwischen den reinen Mittelwertdifferenzen und Cohen's d meist nur ein geringer Unterschied. Der Median der einzelnen d nach Cohen liegt für Chimärengesichter aus linken Gesichtshälften bei -0.41 und für Chimärengesichter aus rechten Gesichtshälften bei -0.37. Diese beiden Werte könnte man daher als Effektgrößen für das Gesamtergebnis der beiden Symmetrie-Experimente verwenden.

Da Cohen's d jedoch auch Werte größer 1 annehmen kann, wurden diese beiden Effektgrößen nochmals nach folgender Formel (Glass, McGaw & Smith, 1978) standardisiert und in ein r umgerechnet, um sie mit den Ergebnissen der anderen Symmetrie-Untersuchungen besser vergleichen zu können:

$$r = \frac{d}{\sqrt{d^2 + \frac{(n_1 + n_2)^2}{n_1 n_2}}}$$

Da in allen drei Versuchspersonengruppen gleich viele Versuchspersonen waren (jeweils $N = 64$), ergibt sich für den Ausdruck unter Wurzel: $d^2 + 4$.

Nach der Standardisierung auf r ergibt sich als Effektgröße für linke Chimärengesichter ein Wert von $r = .20$ und für rechte Chimärengesichter von $r = .18$. Da aufgrund der For-

mel nur positive Werte resultieren können, der Effekt der Symmetrisierung jedoch negativ ist, müsste vor beiden Werten eigentlich ein Minus stehen.

Es gibt jedoch noch eine zweite Möglichkeit der statistischen Datenauswertung: Man rechnet eine statistische Analyse statt über die 3 x 64 Versuchspersonen (*between groups*) eine Analyse über die 30 Stimulusgesichter in ihren zwei Ausprägungen (symmetrisch versus Original, *within groups*). Hierfür lässt man außer Acht, wie eigentlich der Attraktivitätswert zustande kam – nämlich empirisch durch die Befragung einer Stichprobe von Versuchspersonen – und berücksichtigt nur den Attraktivitätswert eines Stimulusgesichts als Eigenschaft dieses Gesichts. Die Streuung der einzelnen Versuchspersonenurteile bleibt hierbei unberücksichtigt.

Um den Gesamt-Effekt der Symmetrisierung auf diese Weise zu ermitteln, wurde in einem zweiten Schritt mit zwei *t*-Tests (für gepaarte Stichproben, zweiseitig) überprüft, ob die Attraktivität von Chimärengesichtern sich von den Originalgesichtern signifikant unterscheidet.

Tabelle 10: Ergebnisse der beiden Rating-Experimente mit Chimärengesichtern aus linken und rechten Gesichtshälften sowie den jeweiligen Originalgesichtern (= Kontrollgruppe). Aufgelistet sind die Attraktivitätsmittelwerte der jeweiligen Stimuli. Die letzten beiden Spalten listen die Effekte der Symmetrisierung in Form von Mittelwertsdifferenzen auf. In allen drei Experimentalgruppen war jeweils $N = 64$.

Stimulus-Gesicht	Stimulus-Gruppe	M Originale	M Chimäre links	M Chimäre rechts	Differenz Chimäre links - Original (Cohen's d)	Differenz Chimäre rechts - Original (Cohen's d)
1	Miss	3.91	3.27	3.50	-0.64 (-0.64)	-0.41 (-0.34)
2	Miss	4.59	4.22	4.31	-0.38 (-0.35)	-0.28 (-0.26)
3	Miss	4.78	4.19	3.81	-0.59 (-0.47)	-0.97 (-0.82)
4	Miss	4.38	4.47	3.83	0.09 (0.08)	-0.55 (-0.49)
5	Miss	4.33	3.48	4.77	-0.84 (-0.70)	0.44 (0.38)
6	Miss	4.59	4.31	3.80	-0.28 (-0.25)	-0.80 (-0.66)
7	Miss	5.22	4.38	4.88	-0.84 (-0.80)	-0.34 (-0.35)
8	Miss	4.81	4.50	4.28	-0.31 (-0.27)	-0.53 (-0.44)
9	Miss	4.92	4.09	4.20	-0.83 (-0.73)	-0.72 (-0.65)
10	Miss	4.70	3.78	4.19	-0.92 (-0.78)	-0.52 (-0.39)
11	Miss	4.78	4.38	4.27	-0.41 (-0.30)	-0.52 (-0.38)
12	Miss	4.50	3.16	4.73	-1.34 (-1.09)	0.23 (0.19)
13	Miss	5.28	4.88	3.86	-0.41 (-0.33)	-1.42 (-1.16)
14	Miss	5.09	4.42	4.81	-0.67 (-0.56)	-0.28 (-0.25)
15	Miss	5.58	4.31	4.80	-1.27 (-1.04)	-0.78 (-0.59)

Fortsetzung von Tabelle 10:

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i> Originale	<i>M</i> Chimäre links	<i>M</i> Chimäre rechts	Differenz Chimäre links - Original (Cohen's <i>d</i>)	Differenz Chimäre rechts - Original (Cohen's <i>d</i>)
16	Normale	2.83	2.28	1.50	-0.55 (-0.48)	-1.33 (-1.40)
17	Normale	2.64	2.64	1.89	0.00 (0.00)	-0.75 (-0.82)
18	Normale	1.94	1.69	1.67	-0.25 (-0.28)	-0.27 (-0.29)
19	Normale	2.34	1.66	2.31	-0.69 (-0.64)	-0.03 (-0.03)
20	Normale	2.67	2.22	2.13	-0.45 (-0.40)	-0.55 (-0.49)
21	Normale	2.45	1.72	2.59	-0.73 (-0.82)	0.14 (0.13)
22	Normale	2.80	2.02	2.30	-0.78 (-0.76)	-0.50 (-0.46)
23	Normale	2.81	3.05	2.13	0.23 (0.20)	-0.69 (-0.66)
24	Normale	2.53	2.13	2.61	-0.41 (-0.43)	0.08 (0.07)
25	Normale	2.17	1.61	2.83	-0.56 (-0.62)	0.66 (0.60)
26	Normale	2.06	2.31	1.84	0.25 (0.26)	-0.22 (-0.22)
27	Normale	2.80	2.63	2.42	-0.17 (-0.17)	-0.38 (-0.35)
28	Normale	2.44	2.47	2.30	0.03 (0.03)	-0.14 (-0.12)
29	Normale	1.47	1.34	1.30	-0.13 (-0.18)	-0.17 (-0.25)
30	Normale	2.95	3.42	2.03	0.47 (0.46)	-0.92 (-0.93)
<i>M</i>		3.61	3.17	3.20	-0.45 (-0.40)	-0.42 (-0.38)

Anmerkung: Alle Werte sind gerundet. Kursive Mittelwertsdifferenzen und Cohen's *d* zeigen einen signifikanten ($p < .05$) Unterschied im Vergleich zum Originalgesicht an (Post-hoc-Tests nach Scheffé).

Die Ergebnisse zeigen, dass sowohl Chimärengesichter aus linken wie Chimärengesichter aus rechten Gesichtshälften signifikant unattraktiver bewertet werden als die jeweiligen unveränderten Originalgesichter. Chimärengesichter links: $t(29) = -5.697, p < .001$; Chimärengesichter rechts: $t(29) = -4.905, p < .001$. Chimärengesichter aus linken Hälften werden im Durchschnitt um 0.45 Skalenstufen, Chimärengesichter aus rechten Hälften um 0.42 Skalenstufen schlechter als Originalgesichter bewertet. Um den Effekt mit den Ergebnissen der anderen durchgeführten Symmetrie-Untersuchungen vergleichen zu können, wurde nach folgender Formel (Pospeschill, 2006) eine Effektgröße r berechnet:

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

Für Chimärengesichter aus linken Gesichtshälften betrug der Effekt $r = .73$, für Chimärengesichter aus rechten Gesichtshälften $r = .67$. Aufgrund der Formel kann r nur positive Werte annehmen; da aber symmetrisierte Gesichter jeweils schlechter als Originalgesichter bewertet wurden, ist der Effekt der Symmetrie *negativ* – es müsste also eigentlich ein Minus vor den Werten stehen. Die Stärke des negativen Effekts bewegt sich in derselben Größenordnung wie bei der Untersuchung von Langlois, Roggman und Musselman (1994), die ebenfalls Chimärengesichter mit Ratings untersuchten.

In der Studie von Studie von Zaidel et al. (1995) stellten die Autoren fest, dass bei ihrem Stimulusmaterial Chimären aus rechten Hälften besser bewertet wurden als solche aus linken Hälften (obwohl Chimären aus linken Hälften als „mehr lächelnd“ beurteilt wurden). Daher wurden die hier vorliegenden Daten ebenfalls auf solche möglichen Effekte getestet. Ein t -Test für gepaarte Stichproben ergab keinerlei Unterschiede zwischen den mittleren Attraktivitätsbewertungen zwischen „linken“ und „rechten“ Chimärengesichtern ($t(29) = 0.229, p = .821$; Chimäre links: $M = 3.17, SD = 1.10$, Chimäre rechts: $M = 3.20, SD = 1.17$). Bei dieser Art der Analyse ist die Urteilsvarianz der Beurteiler noch nicht einmal berücksichtigt; würde man sie in eine Auswertung mit einbeziehen (in einer MANOVA), gäbe es erst recht keinen signifikanten Unterschied. Es besteht bei vorliegenden Daten also kein nachweisbarer Attraktivitätsunterschied zwischen Chimärengesichtern aus linken bzw. rechten Gesichtshälften.

Wenn Symmetrie – wie die Symmetrie-Hypothese besagt – ein bedeutsamer Faktor für die Attraktivität eines Gesichts ist, dann sollten hochattraktive Gesichter auch sehr symmetrisch sein, „normale“ Gesichter hingegen weniger symmetrisch. Im verwendeten Stimulus-Set waren 15 Endrundenteilnehmerinnen einer Miss-Germany-Wahl und 15 „nor-

male“ junge Frauen vorhanden. Die Miss-Germany-Finalistinnen sollten daher symmetrischere Gesichter haben als die „normalen“ Frauen und von einer zusätzlichen Symmetrisierung im Experiment weniger bzw. gar nicht profitieren, da sie ja schon symmetrisch sind („Decken-Effekt“). Allgemein formuliert lautet daher die Hypothese: „Je unattraktiver das Originalgesicht, desto größer ist der positive Effekt der Symmetrisierung.“

Diese Hypothese wurde für beide Arten von Chimären Gesichtern mit einer Korrelation überprüft. Abbildung 44 zeigt in zwei Streudiagrammen den Zusammenhang.

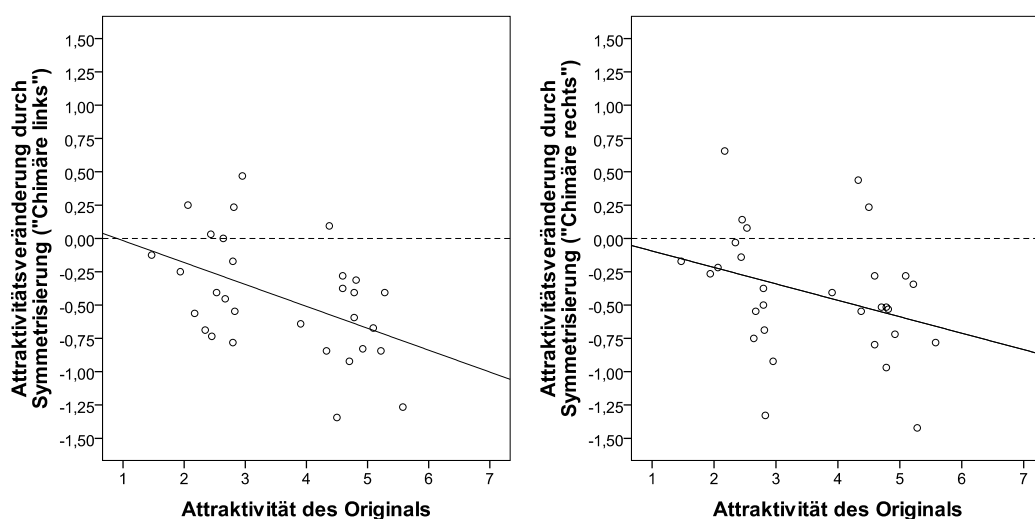


Abbildung 44: Der Zusammenhang zwischen der Attraktivität eines Originalgesichts und der Attraktivitätsveränderung durch das Symmetrisieren mit der Chimären-gesichter-Methode. Das linke Streudiagramm zeigt die Ergebnisse für Chimären-gesichter aus linken Gesichtshälften, das rechte Diagramm für Chimären-gesichter aus rechten Hälften. In beiden Fällen erkennt man, dass diese Art von Symmetrisierung sich auf fast alle Gesichter negativ auswirkt, und zwar tendenziell umso negativer, je attraktiver das Originalgesicht ist.

Ein Blick auf die beiden Diagramme zeigt, dass die Richtung des Zusammenhangs zwar in beiden Fällen vorhanden ist (jeweils negativer Zusammenhang und abfallende Regressionsgerade), jedoch handelt es sich um insgesamt negative Effekte der Symmetrisierung („linke Chimären-gesichter“: $r = -.48$, $p = .008$; „rechte Chimären-gesichter“: $r = -.33$, $p = .076$). Den unattraktiven Gesichtern nützt die Symmetrisierung durch Chimären-gesichter nicht, sondern es schadet ihnen. Den Attraktiveren (die attraktivere Hälfte der

Stichprobe bestand ausschließlich aus Miss-Germany-Finalistinnen) schadet es jedoch noch mehr.

Zusammenfassend bleibt festzuhalten, dass der in diesem Experiment untersuchte Effekt der Symmetrisierung durch Chimärengesichter und der Attraktivitätsbewertung im Rating-Verfahren und *Between-Subjects Design* zu klaren negativen Effekten führte. Symmetrische Chimärengesichter wurden klar als unattraktiver bewertet als unveränderte Originalgesichter. Je nachdem, ob man für die Berechnung der Effektgröße die Varianz der einzelnen Versuchspersonenurteile einbezieht oder weglässt, ergeben sich jedoch unterschiedliche Effekte: Bei Einbeziehung ein schwacher negativer Effekt von $r = -.20$ bzw. $r = -.18$; bei Weglassen ein starker negativer Effekt von $r = -.73$ bzw. $r = -.67$.

4.6 Experimentelle Manipulation der Symmetrie durch Verwendung von Morphings mit Farbwertmittelung („Blends“) und Attraktivitätsmessung durch Paarvergleiche (Untersuchung Sym 3)

4.6.1 Stand der Forschung

Die meisten Untersuchungen, die Morphings mit Farbwertmittelung (*Blends*) zur Untersuchung des Effekts von Symmetrie auf Attraktivität verwendeten, ließen die Attraktivität durch Ratings beurteilen. Eine Ausnahme stellt die Studie von Jones et al. (2007) dar, die versuchten, den Einfluss der Symmetrie vom Effekt der Durchschnittlichkeit zu trennen (Jones, DeBruine & Little, 2007). Dazu erstellen die Autoren von jedem Originalgesicht ein perfekt symmetrisches Blend-Gesicht und näherten sowohl Originalgesicht als auch Blend-Gesicht zu 50 % an die Proportionen eines Durchschnittsgesichts an (= *anticaricatured*). Der paarweise Vergleich erfolgte dabei jedoch nicht zwischen dem unveränderten Originalgesicht und dem symmetrisierten Blend-Gesicht, sondern stattdessen zwischen dem *anticaricatured face* und dem *uncaricatured face*. Insofern sind diese Ergebnisse nicht mit dem hier geplanten Experiment vergleichbar. Für die statistische Analyse verwendeten sie den „Trick“ mit der Reduktion der Versuchspersonen-Urteilsvarianz, bei dem zuerst relative Häufigkeiten der Bevorzugung für eine bestimmte Variante berechnet werden und in einem zweiten Schritt dann ein Ein-Stichproben-*t*-Test (mit der A-priori-Wahrscheinlichkeit von 0.50) gerechnet wird – dasselbe Verfahren, das bereits beispielsweise Little und Jones (2003) verwendeten (zu einer kritischen Diskussion siehe Kapitel 5.2.3).

4.6.2 Methode

4.6.2.1 Versuchspersonen

An der Untersuchung *Sym 3* nahmen insgesamt 64 Versuchspersonen teil. 31 Personen waren männlich (48.4 %), 33 weiblich (51.6 %). 17 Personen (26.6 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 14 Jahre, die älteste 77. Der Altersdurchschnitt betrug 33.70 Jahre ($SD = 16.93$).

Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, fanden die Datenerhebungen im öffentlichen Raum statt, z. B. an Bahnhöfen, Fußgängerzonen und Einkaufszentren. Die Teilnahme am Experiment erfolgte freiwillig und wurde nicht entlohnt.

4.6.2.2 Stimulusmaterial

Als Stimulusmaterial dienten dieselben Fotos der 30 jungen Frauen, die auch die Grundlage für die beiden vorangegangenen Experimente waren. Die Herstellung symmetrischer Varianten erfolgte jedoch hier mit einer anderen Methode, nämlich mit Hilfe von Morphing-Software (*Morphman 4.0*). Dazu wurde zuerst von jedem Gesicht eine Kopie erstellt, die horizontal um eine vertikale Achse gespiegelt wurde. Anschließend wurde jedes Original mit seinem eigenen „Spiegelbild“ zu einem neuen Durchschnittsbild („Blend“) vermorpht. Die Gewichtung der beiden Ausgangsgesichter war dabei 50:50, die Farbwerte wurden ebenfalls gemittelt. Das Resultat ist ein perfekt symmetrisches Durchschnittsgesicht, bei dem (wie auch bei Chimärengesichtern) jeder Bildpunkt in der linken Hälfte des Bildes einen korrespondierenden Bildpunkt mit denselben Farbwerten in der rechten Bildhälfte besitzt. Anders formuliert: Die vertikale Bisymmetrieachse durch die Mitte des Bildes teilt das Bild in zwei identische, jedoch spiegelverkehrte Bildhälften. Der Unterschied dieser „Blends“ zu den Symmetrie-Gesichtern, die in den Chimärenexperimenten verwendet wurden, ist, dass

1. nicht *zwei* verschiedene Gesichter aus der Manipulation entstehen, sondern nur *ein einziges*,
2. das gemorphte Symmetriegesicht in seinen Gesichtsproportionen den Durchschnitt aus den Proportionen der beiden Gesichtshälften darstellt und

3. durch die Neuberechnung (Mittelwertbildung) der Farbwerte ein Weichzeichnungseffekt entsteht, der sich negativ auf die Darstellung der Haare und positiv auf die Haut auswirkt.



Abbildung 45: Illustration des Morph-Prozesses. Das Originalgesicht (links) wird mit seinem eigenen Spiegelbild (Mitte) zu einem neuen Durchschnittsgesicht („Blend“, rechts) verrechnet. Durch die Mittelwertbildung der Farbwerte verwischt die Struktur der Haare und die Haut erscheint feiner. Eventuelle Makel, wie beispielsweise das Muttermal auf der Stirn werden zwar verdoppelt, scheinen jedoch nur noch jeweils zu 50 % durch. Kleine Hautunreinheiten wie Mitesser oder vergrößerte Poren sind dadurch nicht mehr wahrnehmbar.

Alle Bilder hatten wie in den beiden vorangegangenen Experimenten eine hohe Auflösung von 1600 x 1600 Pixel. Für die Versuchsdurchführung wurden sie von einem Fotodienst auf Fotopapier in einer Größe von 9 x 9 Zentimeter belichtet. Daraus ergibt sich eine Auflösung von 452 dpi. Durch diese hohe Auflösung sind die Versuchspersonen in der Lage, selbst kleinste Details wahrzunehmen. Der Unterschied zwischen einem unveränderten Originalgesicht (mit kleinen Makeln und Hautunreinheiten) und einem gemorphten Gesicht mit „weichgezeichneter“ Haut („Blend“) war dadurch beim Stimulusmaterial des Experiments deutlicher sichtbar, als dies hier bei der verkleinerten Darstellung in Abbildung 45 der Fall ist.

4.6.2.3 Versuchsdurchführung

Die Datenerhebung erfolgte analog zu Experiment *Sym 1*, statt der Chimären Gesichter wurden jedoch „Blends“ verwendet. Das Stimulusmaterial wurde in Fotomappen präsentiert, die Position (links oder rechts) der beiden Varianten eines Gesichterpaars wurde variiert und die Reihenfolge der Stimuluspaare für verschiedene Probanden verändert, um etwaige Positionseffekte ausschließen zu können. Die Aufgabenstellung

(„Welches Gesicht ist attraktiver?“) und das Forced-Choice-Verfahren und sonstige Bedingungen waren ebenfalls identisch zum Experiment *Sym 1*.

4.6.3 Ergebnisse

Zur Überprüfung der Frage, ob durch das Morphing-Verfahren mit Neuberechnung der Farbwerte symmetrisch optimierte Gesichter im direkten Paarvergleich attraktiver bewertet wurden als die unveränderten Originalgesichter, wurden in beiden Experimenten jeweils 30 Binomialtests durchgeführt. Da auch in diesem Experiment die Anzahl der Versuchspersonen bei $N = 64$ lag, gelten als Grenzen für signifikante Ergebnisse dieselben absoluten und relativen Häufigkeiten wie in Experiment *Sym 1*. Zur besseren Vergleichbarkeit mit den Ergebnissen der anderen Symmetrie-Experimente wurden auch hier wieder alle Effektgrößen vom Binomial effect size display (*BESD*) nach folgender Formel in ein r umgerechnet (Rosenthal & Rubin, 1982):

$$r = 2 * BESD - 1$$

Tabelle 11 listet getrennt für Miss-Germany-Finalistinnen und „normale“ Frauen die absolute und relative Häufigkeit der Bevorzugung der symmetrisierten „Blend“-Variante gegenüber der unveränderten Original-Variante sowie die in r umgerechnete Effektgröße auf. Am Ende der Tabelle ist jeweils der Median der erzielten Effektgrößen aufgeführt.

Tabelle 11: Ergebnisse des Paarvergleichs-Experiments mit symmetrisierten gemorphten Gesichtern mit Farbwertmittelung („Blends“) im direkten Vergleich mit dem jeweiligen Originalgesicht. Aufgelistet sind absolute und relative Häufigkeiten sowie die in r umgerechneten Effektgrößen. $N = 64$.

Stimulus-Gesicht	Stimulus-Gruppe	„Blend“ gegenüber Original bevorzugt n (%)	r	Stimulus-Gesicht	Stimulus-Gruppe	„Blend“ gegenüber Original bevorzugt n (%)	r
1	Miss	56 (87.5 %)	0.75	16	Normale	52 (81.3 %)	0.63
2	Miss	50 (78.1 %)	0.56	17	Normale	52 (81.3 %)	0.63
3	Miss	46 (71.9 %)	0.44	18	Normale	51 (79.7 %)	0.59
4	Miss	49 (76.6 %)	0.53	19	Normale	57 (89.1 %)	0.78
5	Miss	45 (70.3 %)	0.41	20	Normale	53 (82.8 %)	0.66
6	Miss	40 (62.5 %)	0.25	21	Normale	52 (81.3 %)	0.63
7	Miss	37 (57.8 %)	0.16	22	Normale	43 (67.2 %)	0.34
8	Miss	46 (71.9 %)	0.44	23	Normale	56 (87.5 %)	0.75
9	Miss	49 (76.6 %)	0.53	24	Normale	53 (82.8 %)	0.66
10	Miss	42 (65.6 %)	0.31	25	Normale	47 (73.4 %)	0.47
11	Miss	49 (76.6 %)	0.53	26	Normale	52 (81.3 %)	0.63
12	Miss	42 (65.6 %)	0.31	27	Normale	49 (76.6 %)	0.53
13	Miss	34 (53.1 %)	0.06	28	Normale	47 (73.4 %)	0.47

Fortsetzung

Fortsetzung von Tabelle 11:

Stimulus-Gesicht	Stimulus-Gruppe	„Blend“ gegenüber Original bevorzugt n (%)	<i>r</i>	Stimulus-Gesicht	Stimulus-Gruppe	„Blend“ gegenüber Original bevorzugt n (%)	<i>r</i>
14	Miss	33 (51.6 %)	0.03	29	Normale	47 (73.4 %)	<i>0.47</i>
15	Miss	38 (59.4 %)	0.19	30	Normale	51 (79.7 %)	<i>0.59</i>
Median		45 (70.3 %)	0.41	Median		52 (81.3 %)	0.63

Anmerkung: Kursive *r* zeigen einen signifikanten Effekt ($p < .05$) an.

Wie aus Tabelle 11 zu entnehmen ist, wurde bei allen 30 Gesichtern das durch das Morphing-Verfahren symmetrisch optimierte Gesicht im Vergleich zur unveränderten Originalversion häufiger bevorzugt. In keinem einzigen Fall wurde es negativer bewertet. In 25 Fällen ist der Unterschied statistisch signifikant ($p < .05$), in 5 Fällen (allesamt Miss-Germany-Finalistinnen) nicht. Der Median der Effektgrößen liegt bei Miss-Germany-Finalistinnen bei $r = 0.41$ und bei „normalen“ Frauen bei $r = 0.63$. Der Gesamteffekt (Median über die Effektgrößen aller 30 Gesichter) liegt bei $r = 0.53$.

Der Grund dafür, dass der Effekt der Symmetrisierung bei den „normalen“ Frauengesichtern stärker ausfällt, könne darin liegen, dass sich unter den „normalen“ Frauen einige befanden, die eine unreinere Haut besaßen als die Miss-Germany-Finalistinnen. Daher profitierten sie vom weich zeichnenden Effekt der Farbwertmittelung des Morphing-Verfahrens, so dass die symmetrisierte Variante eine sichtbar gesünder und makelloser aussehende Haut aufweist als das Originalgesicht.

Auch hier gibt es wieder eine alternative Form der Datenanalyse, nämlich nach der Methode von Little und Jones (2003, S. 1761), bei der ausgewertet wird, wie häufig jede Versuchsperson die symmetrische Variante gegenüber der Originalvariante bevorzugt hat. Nach der Nullhypothese sollte sie dies bei der Hälfte der präsentierten Gesichterpaare tun (d. h. die A-priori-Wahrscheinlichkeit ist 0.5). Für jede Versuchsperson wird daher eine relative Häufigkeit berechnet. Diese Häufigkeiten werden mit einem Ein-Stichproben-

t -Test analysiert. Das N ist in diesem Fall die Anzahl der Versuchspersonen ($N = 64$) und eine positive mittlere Differenz bedeutet, dass die symmetrische Variante häufiger bevorzugt wurde. Analysiert man die Daten mit dieser Methode, ergibt sich ein hoch signifikantes Ergebnis $t(63) = 10.473$, $p < .001$. Die Effektstärke, die sich aus dem t -Wert berechnet (siehe Formel in Kapitel 4.4.3), fällt dadurch ebenfalls deutlich größer aus und beträgt dann $r = 0.80$.

Zusammenfassend bleibt festzuhalten, dass der in diesem Experiment untersuchte Effekt der Symmetrisierung durch gemorphte Gesichter mit Farbwertmittelung und der Attraktivitätsbewertung im direkten Paarvergleich zu einem eindeutig positiven, mittelstarken Effekt führte. Bei einer Datenanalyse ohne Berücksichtigung der Versuchspersonen-Urteilsvarianz ist der positive Effekt sogar sehr groß. Symmetrische „Blends“ wurden klar attraktiver bewertet als unveränderte Originalgesichter.

4.7 Experimentelle Manipulation der Symmetrie durch Verwendung von Morphings *mit* Farbwertmittelung („Blends“) und Attraktivitätsmessung durch Ratings (Untersuchung *Sym 4*)

4.7.1 Stand der Forschung

Experimente, bei denen der Effekt der Symmetrisierung im direkten Paarvergleich mit dem unveränderten Originalgesicht gemessen wurde, sind relativ häufig. Seltener finden sich Versuchsdesigns, bei denen der Effekt durch Attraktivitätsratings gemessen wurde (Rhodes, Yoshikawa et al., 2001). Die Studie von Rhodes, Yoshikawa et al. (2001) ist ein Beispiel dafür. Die Autoren verwendeten die Symmetrie von Gesichtern jedoch nur als eine von mehreren unabhängigen Variablen (weitere waren das Geschlecht und die ethnische Gruppe der Beurteiler und der Stimuluspersonen sowie der durch Morphing manipulierte Durchschnittlichkeitsgrad der Stimulusgesichter). Die Symmetrie wurde durch Morphen (mit Farbwertmittelung) mit den Ausgangsgesichtern und ihren jeweiligen Spiegelbildern manipuliert. Verwendet wurde ein *Within-Subjects Design*, bei dem jede Versuchsperson sämtliche Varianten eines Gesichts zu sehen bekam und beurteilen musste. Die Autoren konnten einen hochsignifikanten Effekt der Symmetrisierung nachweisen, der in seiner Effektstärke jedoch sehr gering war. Auf einer Attraktivitätsskala von 1 bis 10 (= Maximum) wurden perfekt symmetrische Gesichter im Durchschnitt mit 4.2 anstatt 4.0 (Originalgesichter) bewertet; das entspricht 2.2 % der gesamten Skala.

Eine ganz ähnliche frühere Untersuchung (Rhodes, Sumich et al., 1999) ebenfalls von einer Forschergruppe um Gillian Rhodes beschränkte sich auf die Untersuchung westlicher Gesichter und westlicher Beurteiler und kam mit der gleichen Methodik zu ähnlichen Ergebnissen. Auf einer Skala von 1 bis 7 (= Maximum) erhielten perfekt symmetrische Gesichter im Durchschnitt den Attraktivitätswert 3.0 und Originalgesichter 2.7. Auch hier war der Effekt hoch signifikant, aber gering, nämlich 5 % der gesamten Skala. Bei beiden Studien ist im Übrigen auffallend, dass selbst solche Stimulusgesichter, die so manipuliert waren, dass sie hoch durchschnittliche Gesichtsproportionen aufwiesen und zugleich auch noch perfekt symmetrisch waren, dennoch im Mittel nicht über eine mittelmäßige Attraktivitätsbewertung hinauskamen (alle Mittelwerte lagen in der unteren Hälfte der Attraktivitätsskala), was auf den Erklärungswert beider Theorien kein besonders gutes Licht wirft.

Eine weitere Untersuchung (Swaddle & Cuthill, 1995) verwendete Blend-Gesichter in fünf verschiedenen Graden der Abstufungen: Originalgesichter und deren Spiegelbilder, perfekt symmetrische Blend-Gesichter sowie „Zwischenformen“, nämlich Morphings von Original mit Spiegelbild, bei denen das Originalgesicht entweder zu 25 % oder zu 75 % (= *intermediate morphs*) durchschien. Die Autoren erhielten einen klar negativen Effekt der Symmetrie, d. h. die perfekt symmetrischen Varianten wurden als unattraktiver als die weniger symmetrischen Varianten bezeichnet. Die Autoren führten für diesen Befund eine überraschende Interpretation an, die argumentativ auf der Unterscheidung zwischen fluktuierender und direktonaler Asymmetrie basiert:

The observed positive relationship between asymmetry and facial attractiveness may be because certain facial features (including those contributing to attractiveness) in fact show directional asymmetry or antisymmetry. Our manipulations thus render naturally asymmetric features symmetrical. This may make symmetric faces less attractive because of the reduction of natural directional asymmetries, perhaps making the faces appear unemotional. (Swaddle & Cuthill, 1995, S. 111)

Diese Argumentation ist ziemlich spekulativ, da die Autoren die fluktuierende bzw. direktonale Asymmetrie nicht unabhängig voneinander messen und quantifizieren konnten. Interessant ist diese Studie in erster Linie deswegen, weil sie als einzige mit dieser Methode einen negativen Effekt der Symmetrisierung fand. In der Metaanalyse von Rhodes (2006) nennt Rhodes diese Studie auch als die einzige Ausnahme und schreibt: „The only exception is a study by Swaddle & Cuthill (1995), but failure to [...] remove blemishes before morphing could have contributed to failure to find a symmetry preference in this study (Rhodes, 2006, S. 206)“. Dieses Argument ist allerdings unlogisch, denn gerade, wenn man Makel (wie Pickel, Muttermale) vor dem Morphen *nicht* entfernt, dann ist der

Effekt der Symmetrisierung aufgrund der Farbwertmittelung ja größer. Bei den symmetrisierten *Blends* scheinen dann Makel nur noch zu 50 % durch, so dass die Gesichter dadurch ja gerade attraktiver und nicht unattraktiver als die Originale wirken müssten.

4.7.2 Methode

4.7.2.1 Versuchspersonen

An der Untersuchung *Sym 4* nahmen insgesamt 128 Versuchspersonen teil. In der Experimentalgruppe, die symmetrisierte „Blends“ bewertete, waren 64 Personen. Als Kontrollgruppe, die Originalgesichter bewertete, wurden die Daten der Kontrollgruppe von Experiment *Sym 2* verwendet (siehe Kapitel 4.5.3).

Von den 64 Personen der Experimentalgruppe waren 28 (43.8 %) männlich und 36 weiblich (56.3 %). 30 Personen (46.9 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 14 Jahre, die älteste 84. Der Altersdurchschnitt betrug 31.90 Jahre ($SD = 16.08$).

Hinsichtlich der erhobenen demografischen Variablen ist die Experimentalgruppe mit der Kontrollgruppe vergleichbar, bei Geschlecht ($\chi^2(1, N = 128) = 0.783, p = .479$) und Alter ($t(126) = -0.600, p = .550$) bestehen keine Unterschiede, allerdings waren in der Experimentalgruppe mehr Personen mit Studierendenstatus ($\chi^2(1, N = 128) = 5.682, p = .027$). Da jedoch Psychologie-Studierende vom Experiment grundsätzlich ausgeschlossen waren, bei den Probanden kein themenspezifisches Vorwissen vorhanden war und die Altersverteilung gleich ist, kann dieser Unterschied vernachlässigt werden.

4.7.2.2 Stimulusmaterial und Versuchsdurchführung

Das Stimulusmaterial war identisch mit dem Material der Untersuchung *Sym 3*, d. h. es wurden die symmetrisch optimierten Gesichter verwendet, die durch das Morphing-Verfahren mit Farbwertmittelung erzeugt worden waren.

Die Versuchsdurchführung war identisch mit der Untersuchung *Sym 2*, d. h. dass die Versuchspersonen der Experimentalbedingung 30 Stimuli der symmetrisch optimierten Variante präsentiert bekamen und auf einer Ratingskala bewerteten. Davon, dass die gezeigten Gesichter auch noch in anderen Varianten existierten (z. B. unveränderte Originalgesichter), wussten die Probanden nichts (*Between-Subjects Design*). Verwendet wurde dieselbe siebenstufige Ratingskala wie in Experiment *Sym 2*. Auch sonst galten

dieselben Randbedingungen wie in Experiment *Sym 2* (siehe Kapitel 4.5.2.2). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .99 (bei 64 Versuchspersonen) sehr hoch.

4.7.3 Ergebnisse

Zur Überprüfung der Frage, ob durch das Morphing-Verfahren mit Farbwertermittlung symmetrisch optimierte Gesichter attraktiver bewertet wurden als die unveränderten Originalgesichter, wurde eine multivariate Varianzanalyse durchgeführt, die ein signifikantes Ergebnis lieferte ($p < .001$). Tabelle 12 listet für jeden Stimulus den erzielten Attraktivitätsmittelwert auf und zeigt für jede Stimulusperson die Attraktivitätsdifferenz zwischen symmetrisch optimiertem „Blend“ und Originalgesicht. Es zeigt sich, dass diese Differenzen in den meisten Fällen negativ sind, d. h. dass die symmetrisierten Varianten meist etwas schlechter bewertet wurden als die Originalgesichter. Nur in vier Fällen erhielten die symmetrischen „Blends“ eine höhere durchschnittliche Attraktivitätsbewertung. Wie jedoch die letzte Spalte von Tabelle 12 zeigt, sind die Attraktivitätsdifferenzen nur in 3 von 30 Fällen statistisch signifikant (allesamt „normale“ Frauen; in zwei Fällen machte die Symmetrisierung unattraktiver und in einem Fall attraktiver).

Tabelle 12: Ergebnisse des Rating-Experiments mit durch das Morphing-Verfahren mit Mittelwertbildung symmetrisch optimierten Gesichtern sowie den jeweiligen Originalgesichtern (= Kontrollgruppe). Aufgelistet sind die Attraktivitätsmittelwerte der jeweiligen Stimuli. Die drittletzte Spalte listet den Effekt der Symmetrisierung in Form von Mittelwertsdifferenzen sowie in Klammern Cohen's d auf. Die beiden letzten Spalten listen F - und p -Werte des Tests für Zwischensubjekteffekte der MANOVA auf. In beiden Versuchsgruppen war jeweils $N = 64$.

Stimulus-Gesicht	Stimulus-Gruppe	M Originale	M "Blends"	Differenz "Blend" – Original (Cohen's d)	F	p
1	Miss	3.91	3.72	-0.19 (-0.17)	0.968	.327
2	Miss	4.59	4.47	-0.13 (-0.11)	0.386	.536
3	Miss	4.78	4.86	0.08 (0.08)	0.188	.665

Fortsetzung

Fortsetzung von Tabelle 12:

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i> Originale	<i>M</i> "Blends"	Differenz "Blend" – Original (Cohen's <i>d</i>)	<i>F</i>	<i>p</i>
4	Miss	4.38	4.17	-0.20 (-0.19)	1.108	.294
5	Miss	4.33	4.25	-0.08 (-0.07)	0.160	.690
6	Miss	4.59	4.36	-0.23 (-0.21)	1.437	.233
7	Miss	5.22	5.00	-0.22 (-0.22)	1.569	.213
8	Miss	4.81	4.78	-0.03 (-0.03)	0.023	.879
9	Miss	4.92	4.70	-0.22 (-0.23)	1.723	.192
10	Miss	4.70	4.39	-0.31 (-0.28)	2.580	.111
11	Miss	4.78	4.66	-0.13 (-0.09)	0.285	.595
12	Miss	4.50	4.17	-0.33 (-0.30)	2.835	.095
13	Miss	5.28	4.94	-0.34 (-0.30)	2.965	.088
14	Miss	5.09	4.98	-0.11 (-0.09)	0.264	.608
15	Miss	5.58	5.22	-0.36 (-0.34)	3.653	.058
16	Normale	2.83	2.63	-0.20 (-0.17)	0.955	.330
17	Normale	2.64	2.22	-0.42 (-0.43)	5.996	.016
18	Normale	1.94	1.92	-0.02 (-0.02)	0.009	.926
19	Normale	2.34	1.97	-0.38 (-0.34)	3.721	.056
20	Normale	2.67	3.03	0.36 (0.31)	3.099	.081

Fortsetzung

Fortsetzung von Tabelle 12:

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i> Originale	<i>M</i> "Blends"	Differenz "Blend" – Original (Cohen's <i>d</i>)	<i>F</i>	<i>p</i>
21	Normale	2.45	2.27	-0.19 (-0.18)	1.071	.303
22	Normale	2.80	2.25	-0.55 (-0.51)	8.353	.005
23	Normale	2.81	2.72	-0.09 (-0.09)	0.256	.614
24	Normale	2.53	2.91	0.38 (0.40)	5.184	.024
25	Normale	2.17	2.13	-0.05 (-0.05)	0.065	.799
26	Normale	2.06	2.03	-0.03 (-0.03)	0.035	.853
27	Normale	2.80	2.59	-0.20 (-0.20)	1.243	.267
28	Normale	2.44	2.80	0.36 (0.31)	3.173	.077
29	Normale	1.47	1.47	0.00 (0.00)	0.000	1.000
30	Normale	2.95	2.80	-0.16 (-0.15)	0.697	.405
<i>M</i>		3.61	3.48	-0.13 (-0.12)		

Anmerkung: Alle Werte sind gerundet.

Wie in Untersuchung *Sym 2* wurde auch hier als Effektgröße Cohen's *d* berechnet. Der Median über alle 30 Cohen's *d* beträgt -.16. Das bedeutet, dass der Effekt der Symmetrisierung nach dieser Methode schwach negativ ist. Da Cohen's *d* auch Werte größer 1 annehmen kann, wurde zur besseren Vergleichbarkeit mit den Ergebnissen der anderen Symmetrie-Untersuchungen das mittlere *d* nach der in Kapitel 4.5.3 angegebenen Formel in ein *r* umgerechnet. Dabei ergibt sich ein äußerst kleiner negativer Effekt von $r = .08$ (Aufgrund der Formel ist der Wert immer positiv und müsste wegen des negativen Effekts der Symmetrisierung eigentlich mit einem Minus notiert werden).

Analog zu Untersuchung *Sym 2* wurden jedoch auch hier mit einem alternativen Verfahren die Daten statistisch ausgewertet. Dabei wurde statt der Analyse über 2 x 64 Versuchspersonen (*between groups*) ein *t*-Test über die 30 Stimulusgesichter in ihren zwei Ausprägungen (symmetrisch versus Original, *within groups*, zweiseitig) gerechnet. Die Varianz der Versuchspersonenurteile bleibt dabei unberücksichtigt; verwendet werden nur die Attraktivitätswerte (= Mittelwerte über 64 Versuchspersonenurteile zu jedem Stimulusgesicht) als Eigenschaft eines jeweiligen Gesichts.

Das Ergebnis des *t*-Tests zeigt, dass Gesichter, die im Morphing-Verfahren mit Farbwert-mittelung symmetrisch optimiert wurden, klar unattraktiver bewertet wurden als unveränderte Originalgesichter $t(29) = -3.337, p = .002$. Im Durchschnitt wurden die symmetrisierten Gesichter um 0.13 Skalenstufen schlechter bewertet. Um den Effekt mit den Ergebnissen der anderen durchgeführten Symmetrie-Untersuchungen vergleichen zu können, wurde nach folgender Formel (Pospeschill, 2006) eine Effektgröße *r* berechnet:

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

Der Effekt beträgt $r = .53$. Da aufgrund der Formel *r* nur positive Werte annehmen kann und da aber symmetrisierte Gesichter schlechter als Originalgesichter bewertet wurden, ist der Effekt der Symmetrie *negativ* – es müsste also eigentlich ein Minus vor dem Wert stehen.

Damit ist die Größe des Effekts exakt gleich groß wie bei der Untersuchung *Sym 3*, bei der Versuchspersonen dasselbe Stimulusmaterial präsentiert wurde – dort jedoch im Paarvergleich. Allerdings ist die Richtung des Effekts genau umgekehrt. Während im Paarvergleichsexperiment symmetrisierte Gesichter *besser* bewertet wurden ($r = .53$) als unveränderte Originalgesichter, wurden sie im Rating-Verfahren *schlechter* ($r = -.53$) bewertet. Dies zeigt, dass die Art der Stimulusdarbietung nicht nur über die Größe des Effekts auf die Attraktivitätsbewertung entscheiden kann, sondern sogar über seine Richtung.

Wenn Symmetrie – wie die Symmetrie-Hypothese besagt – ein bedeutsamer Faktor für die Attraktivität eines Gesichts ist, dann sollten hochattraktive Gesichter auch sehr symmetrisch sein, normale Gesichter hingegen weniger symmetrisch. Im verwendeten Stimulus-Set waren 15 Endrundenteilnehmerinnen einer Miss-Germany-Wahl und 15 „normale“ junge Frauen vorhanden. Die Miss-Germany-Finalistinnen sollten daher symmetrischere

Gesichter haben als die „normalen“ Frauen und von einer zusätzlichen Symmetrisierung im Experiment weniger bzw. gar nicht profitieren, da sie ja schon symmetrisch sind („Decken-Effekt“). Allgemein formuliert lautet daher die Hypothese: „Je unattraktiver das Originalgesicht, desto größer ist der positive Effekt der Symmetrisierung.“

Diese Hypothese wurde für die symmetrisierten „Blend“-Gesichter mit einer Korrelation überprüft. Abbildung 46 zeigt den Zusammenhang.

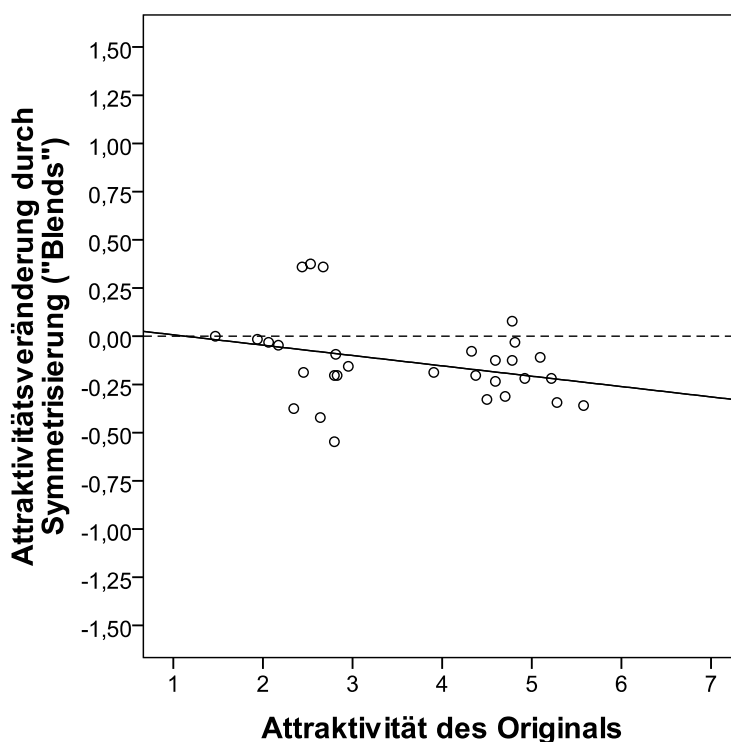


Abbildung 46: Der Zusammenhang zwischen der Attraktivität eines Originalgesichts und der Attraktivitätsveränderung durch das Symmetrisieren mit der *Blend*-Methode. Bei den normalen Frauengesichtern hat die Symmetrisierung keinen Effekt, bei den attraktiven Model-Gesichtern mindert sie die Attraktivität tendenziell.

Ein Blick auf das Diagramm zeigt, dass die Richtung des Zusammenhangs zwar in beiden Fällen vorhanden ist (jeweils negativer Zusammenhang und abfallende Regressionsgerade), der Effekt ist jedoch gering und statistisch nicht signifikant ($r = -.31$, $p = .101$). Vielmehr wurden fast alle Gesichter durch die Symmetrisierung negativer beurteilt. Ins Auge fallen drei Ausreißer unter den „normalen“ Frauen, die von der symmetrischen

Optimierung profitierten. Eine nachträgliche Begutachtung dieser Bilder zeigte, dass zwei dieser Gesichter ein leicht schiefes Lächeln hatten (linker Mundwinkel nach oben gezogen), die dritte Frau, die ihre lockigen Haare nach hinten gebunden hatte, zeigte auf dem Originalfoto eine unvorteilhaft abstehende Strähne, die auf dem symmetrisierten Bild nicht mehr vorhanden war. Eine der drei Frauen mit unreiner Haut profitierte zusätzlich vom Weichzeichnungseffekt der Farbwertmittelung des Morphing-Verfahrens. Solche Post-hoc-Erklärungen sind allerdings nicht unproblematisch, da der Nachweis der Kausalität natürlich nicht mehr erbracht werden kann. Diese drei Gesichter sind jedenfalls für einen Großteil des Effekts verantwortlich; schließt man sie aus der Analyse aus, beträgt die Korrelation zwischen dem Effekt der Symmetrisierung und der Attraktivität des Originalgesichts nur noch $r = -.13$, $p = .512$.

Die Ergebnisse decken sich damit ebenfalls mit dem Befund von Swaddle und Cuthill (1995), die auch zu dem Ergebnis kamen, dass Symmetrisierung (durch *Blend*-Technik) meist einen negativen Effekt hatte und die ebenfalls einen negativen Zusammenhang mit der Attraktivität der Originalgesichter fanden. Je attraktiver das Originalgesicht, desto schlechter wurde das Gesicht nach der symmetrischen Manipulation bewertet ($r = -.74$ für Männergesichter und $r = -.43$ für Frauengesichter, S. 113f). Nur die unattraktivsten Gesichter profitierten bei ihnen von der Symmetrisierung.

Zusammenfassend bleibt festzuhalten, dass der in diesem Experiment untersuchte Effekt der Symmetrisierung durch gemorphte Gesichter mit Farbwertmittelung („Blends“) und der Attraktivitätsbewertung im Rating-Verfahren und Between-Subject-Design zu negativen Effekten führte. Symmetrische „Blend“-Gesichter wurden unattraktiver bewertet als unveränderte Originalgesichter. Je nachdem, ob man für die Berechnung der Effektgröße die Varianz der einzelnen Versuchspersonenurteile einbezieht oder weglässt, ergeben sich jedoch unterschiedliche Effekte: Bei Einbeziehung ein sehr schwacher negativer Effekt von $r = -.08$; bei Weglassen ein mittelstarker negativer Effekt von $r = -.53$.

4.8 Experimentelle Manipulation der Symmetrie durch Verwendung von Morphings *ohne* Farbwertmittelung („Warps“) und Attraktivitätsmessung durch Paarvergleiche (Untersuchung Sym 5)

4.8.1 Stand der Forschung

Die meisten experimentellen Studien, die Symmetrie durch Morphing-Technik manipulierten, verwendeten ein Verfahren, bei dem gleichzeitig auch alle Farbmittelwerte gemittelt wurden (Blends). Durch diese Methode werden die Gesichter perfekt symmetrisch (z. B. jedes Muttermal findet sich in beiden Gesichtshälften), jedoch entsteht gleichzeitig ein Weichzeichnungseffekt, der die Haut glatter und makelloser erscheinen lässt (z. B. jedes Muttermal scheint nun nur noch mit 50 % durch, leichte lokal begrenzte Hautrötungen verschwinden vollständig).

Nur wenige Arbeiten hingegen verwendeten die Morphing-Technik so, dass lediglich Gesichtsproportionen symmetrisiert wurden, die Textur des Originalgesichts jedoch beibehalten wurde (= Warps). In der Metaanalyse von Rhodes, 2006, wird dies leider nicht unterschieden und diese beiden verschiedenen Morphingverfahren wurden in der Kategorie „Blends“ zusammengefasst.

Die am häufigsten zitierte Arbeit, die mit „Warps“ arbeitete, stammt von Perrett et al. (1999). Diese Arbeit wird gerne als Beleg dafür herangezogen, dass es angeblich keinen Unterschied mache, ob man bei der Symmetrisierung von Gesichtern beim Morphen die Farbwerte mittelt oder nicht, da laut Perrett et. al der Effekt auch ohne die Farbwertmittelung belegt sei (Rhodes, 2006).

Perrett et al. konnten einen positiven Einfluss der Symmetrisierung belegen, auch wenn die symmetrischen Stimuli nur in 57.8 % der Fälle bevorzugt wurden. Bei der Datenanalyse wählten die Autoren jedoch ein Verfahren, mit dem sie diesen kleinen Unterschied bei nur 49 Versuchspersonen dennoch statistisch hoch signifikant bekamen ($p < .0005$). Sie berechneten für jeden Beurteiler den Anteil der symmetrisch optimierten Stimuli (von 30 Gesichterpaaren) und zeigten dann mit einem Ein-Stichproben-*t*-Test, dass die Versuchspersonen im Durchschnitt häufiger symmetrische Stimuli bevorzugten als per Zufall (50 %) zu erwarten war (S. 299). Bei 30 einzelnen Binomialtests wären bei einer Versuchspersonenanzahl von nur 49 die allermeisten Einzeltests wohl nicht signifikant ge-

worden, bei einer Adjustierung des α -Fehlers erst recht nicht. Die Autoren berichten jedoch leider nur die zusammengefassten Ergebnisse.

Eine weitere Untersuchung stammt aus derselben Forschergruppe um David Perrett (Little & Jones, 2003). Sie ist allerdings nicht als unabhängige Studie zu betrachten, da sie mit demselben Stimulusmaterial wie Perrett et al. (1999) experimentierte. Die „Neuheit“ bestand darin, dass Little und Jones dieselben Stimuluspaare zusätzlich auch noch auf dem Kopf stehend präsentierten und dabei feststellten, dass die symmetrischen Varianten nun nicht mehr bevorzugt wurden. Nach ihrer Interpretation sei dies ein Beleg gegen die Theorie von Attraktivität als Wahrnehmungserleichterung und spreche für die evolutionäre Sichtweise, dass Symmetrie ein Signal für einen hohen Partnerwert (*mate quality*) sei. Ob Symmetrie tatsächlich attraktiv macht, hinterfragen die Autoren gar nicht mehr, sondern stellen es als Tatsache gleich im ersten Satz ihres Abstracts fest: „Symmetrical human faces are attractive“ (Little & Jones, 2003, S. 1759).

Kaum überraschend, dass die Autoren mit demselben Stimulusmaterial dieselben Ergebnisse erzielten. Die (aufrecht präsentierten) symmetrischen Varianten wurden im Durchschnitt zu 58 % bevorzugt (S. 1761; bei Perrett et al., 1999, zu 57.8 %). Sie griffen bei ihrer Auswertung zum selben statistischen Kunstgriff wie ihre Vorgänger aus demselben Haus und berechneten für jeden Beurteiler den Anteil der symmetrisch optimierten Stimuli (von 28 Gesichterpaaren) und zeigten dann mit einem Ein-Stichproben-*t*-Test, dass die Versuchspersonen im Durchschnitt häufiger symmetrische Stimuli bevorzugten als per Zufall (50 %) zu erwarten war (S. 1761). So erzielten auch sie trotz der eigentlich nur geringfügig besseren Bewertung der Symmetriegesichter (58 %) dennoch ein hoch signifikantes Ergebnis ($p < .001$).

4.8.2 Methode

4.8.2.1 Versuchspersonen

An der Untersuchung *Sym 5* nahmen insgesamt 64 Versuchspersonen teil. 32 Personen waren männlich (50.0 %), 32 weiblich (50.0 %). 8 Personen (12.5 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 15 Jahre, die älteste 79. Der Altersdurchschnitt betrug 34.34 Jahre ($SD = 15.18$).

Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, fanden die Datenerhebungen – wie auch bei den anderen Symmetrie-Untersuchungen – im öffentli-

chen Raum statt, z. B. an Bahnhöfen, Fußgängerzonen und Einkaufszentren. Die Teilnahme am Experiment erfolgte freiwillig und wurde nicht entlohnt.

4.8.2.2 Stimulusmaterial

Als Stimulusmaterial dienten dieselben Fotos der 30 jungen Frauen, die auch die Grundlage für die beiden vorangegangenen Experimente waren. Die Herstellung symmetrischer Varianten erfolgte mit der Morphing-Methode. Im Unterschied zum Stimulusmaterial von Experiment *Sym 3* erfolgte das Vermorphen eines Originalgesichts mit seinem Spiegelbild jedoch *ohne* Mittelung der Farbwerte, d. h. es wurden nur die Proportionen jedes Gesichts symmetrisch optimiert, während die gesamte Textur (also alle Farbwerte) unverändert blieb. Daraus resultieren Gesichter, die in ihren Gesichtsproportionen symmetrisch sind, nicht jedoch in Details der Textur (z. B. symmetrische Frisur, symmetrische Verteilung der Muttermale auf beiden Gesichtshälften). Zudem entsteht – anders als bei den „Blends“ – durch die fehlende Farbwertmittelung kein glättender Weichzeichnungseffekt der Haut. Zum Vergleich der Ergebnisse dieser beiden Morphing-Methoden siehe die Bilder in Abbildung 45 in Kapitel 4.6.2.2.



Abbildung 47: Illustrierung des Morph-Prozesses. Das Originalgesicht (links) wird mit seinem eigenen Spiegelbild (Mitte) zu einem neuen Durchschnittsgesicht („Warp“, rechts) verrechnet. Lediglich die Koordinatenwerte der einzelnen *Facial Landmarks* werden gemittelt, nicht jedoch die Farbwerte der Textur. Dadurch ist das Ergebnis in seinen Proportionen symmetrisch, die Textur des Originalgesichts bleibt jedoch erhalten. Es entstehen keine Artefakte an Haut oder Haaren.

Alle Bilder hatten wie in den beiden vorangegangenen Experimenten eine hohe Auflösung von 1600 x 1600 Pixel. Für die Versuchsdurchführung wurden sie von einem Fotodienst auf Fotopapier in einer Größe von 9 x 9 Zentimeter belichtet. Daraus ergibt sich eine Auflösung von 452 dpi. Durch diese hohe Auflösung sind die Versuchspersonen in der Lage, selbst kleinste Details wahrzunehmen. Die Wahrnehmung von kleinen Asym-

metrien war dadurch beim Stimulusmaterial des Experiments leichter, als dies hier bei der verkleinerten Darstellung in Abbildung 47 der Fall ist.

4.8.2.3 Versuchsdurchführung

Die Datenerhebung erfolgte analog zu den Experimenten *Sym 1* und *Sym 3*, statt der Chimärengesichter bzw. „Blends“ wurden jedoch „Warps“ verwendet. Das Stimulusmaterial wurde in Fotomappen präsentiert, die Position (links oder rechts) der beiden Varianten eines Gesichterpaars wurde variiert und die Reihenfolge der Stimuluspersonenpaare für verschiedene Probanden verändert, um etwaige Positionseffekte ausschließen zu können. Die Aufgabenstellung („Welches Gesicht ist attraktiver?“) und das Forced-Choice-Verfahren und sonstige Bedingungen waren ebenfalls identisch zu den vorangegangenen Paarvergleichsexperimenten.

4.8.3 Ergebnisse

Zur Überprüfung der Frage, ob durch das Morphing-Verfahren ohne Neuberechnung der Farbwerte symmetrisch optimierte Gesichter („Warps“) im direkten Paarvergleich attraktiver bewertet wurden als die unveränderten Originalgesichter, wurden in beiden Experimenten jeweils 30 Binomialtests durchgeführt. Da auch in diesem Experiment die Anzahl der Versuchspersonen bei $N = 64$ lag, gelten als Grenzen für signifikante Ergebnisse dieselben absoluten und relativen Häufigkeiten wie in den Experimenten *Sym 1* und *Sym 3*. Zur besseren Vergleichbarkeit mit den Ergebnissen der anderen Symmetrie-Experimente wurden auch hier wieder alle Effektgrößen vom Binomial effect size display (*BESD*) nach folgender Formel in ein r umgerechnet (Rosenthal & Rubin, 1982):

$$r = 2 * BESD - 1$$

Tabelle 13 listet getrennt für Miss-Germany-Finalistinnen und „normale“ Frauen die absolute und relative Häufigkeit der Bevorzugung der symmetrisierten „Warp“-Variante gegenüber der unveränderten Original-Variante sowie die in r umgerechnete Effektgröße auf. Am Ende der Tabelle ist jeweils der Median der erzielten Effektgrößen aufgeführt.

Tabelle 13: Ergebnisse des Paarvergleichs-Experiments mit symmetrisierten gemorphten Gesichtern *ohne* Farbwertmittelung („Warps“) im direkten Vergleich mit dem jeweiligen Originalgesicht. Aufgelistet sind absolute und relative Häufigkeiten sowie die in r umgerechneten Effektgrößen. $N = 64$.

Stimulus-Gesicht	Stimulus-Gruppe	„Warp“ gegenüber Original bevorzugt n (%)	r	Stimulus-Gesicht	Stimulus-Gruppe	„Warp“ gegenüber Original bevorzugt n (%)	r
1	Miss	45 (70.3 %)	0.41	16	Normale	44 (68.8 %)	0.38
2	Miss	36 (56.3 %)	0.13	17	Normale	35 (54.7 %)	0.09
3	Miss	42 (65.6 %)	0.31	18	Normale	38 (59.4 %)	0.19
4	Miss	34 (53.1 %)	0.06	19	Normale	35 (54.7 %)	0.09
5	Miss	40 (62.5 %)	0.25	20	Normale	34 (53.1 %)	0.06
6	Miss	29 (45.3 %)	-0.09	21	Normale	45 (70.3 %)	0.41
7	Miss	39 (60.9 %)	0.22	22	Normale	34 (53.1 %)	0.06
8	Miss	42 (65.6 %)	0.31	23	Normale	54 (84.4 %)	0.69
9	Miss	39 (60.9 %)	0.22	24	Normale	37 (57.8 %)	0.16
10	Miss	39 (60.9 %)	0.22	25	Normale	40 (62.5 %)	0.25
11	Miss	40 (62.5 %)	0.25	26	Normale	28 (43.8 %)	-0.13
12	Miss	37 (57.8 %)	0.16	27	Normale	44 (68.8 %)	0.38
13	Miss	37 (57.8 %)	0.16	28	Normale	29 (45.3 %)	-0.09

Fortsetzung

Fortsetzung von Tabelle 13:

Stimulus-Gesicht	Stimulus-Gruppe	„Warp“ gegenüber Original bevorzugt n (%)	<i>r</i>	Stimulus-Gesicht	Stimulus-Gruppe	„Warp“ gegenüber Original bevorzugt n (%)	<i>r</i>
14	Miss	30 (46.9 %)	-0.06	29	Normale	33 (51.6 %)	0.03
15	Miss	27 (42.2 %)	-0.16	30	Normale	39 (60.9 %)	0.22
Median		39 (60.9 %)	0.22	Median		37 (57.8 %)	0.16

Anmerkung: Kursive *r* zeigen einen signifikanten Effekt ($p < .05$) an.

Wie aus Tabelle 13 zu entnehmen ist, wurde bei den meisten der 30 Gesichter das durch das Morphing-Verfahren symmetrisch optimierte Gesicht im Vergleich zur unveränderten Originalversion häufiger bevorzugt, nämlich in 25 Fällen. In nur 5 Fällen wurde es negativer bewertet. Jedoch ist der Unterschied nur in 7 Fällen statistisch signifikant ($p < .05$), in der Mehrzahl der Fälle (23) nicht. In den 7 signifikanten Paarvergleichen wurde das symmetrisierte Bild besser bewertet. Der Median der Effektgrößen liegt bei Miss-Germany-Finalistinnen bei $r = 0.22$ und bei „normalen“ Frauen bei $r = 0.16$ und ist damit praktisch gleich. Der Gesamteffekt (Median über die Effektgrößen aller 30 Gesichter) liegt bei $r = 0.17$.

Auch hier gibt es wieder eine alternative Form der Datenauswertung, nämlich nach der Methode von Little und Jones (2003, S. 1761), bei der ausgewertet wird, wie häufig jede Versuchsperson die symmetrische Variante gegenüber der Originalvariante bevorzugt hat. Nach der Nullhypothese sollte sie dies bei der Hälfte der präsentierten Gesichterpaare tun (d. h. die A-priori-Wahrscheinlichkeit ist 0.5). Für jede Versuchsperson wird daher eine relative Häufigkeit berechnet. Diese Häufigkeiten werden mit einem Ein-Stichproben-*t*-Test analysiert. Das *N* ist in diesem Fall die Anzahl der Versuchspersonen ($N = 64$) und eine positive mittlere Differenz bedeutet, dass die symmetrische Variante häufiger bevorzugt wurde. Analysiert man die Daten mit dieser Methode, ergibt sich ein hoch signifikantes Ergebnis $t(63) = 5.208$, $p < .001$. Die Effektstärke, die sich aus dem *t*-Wert berechnet

(siehe Formel in Kapitel 4.4.3), fällt dadurch ebenfalls deutlich größer aus und beträgt dann $r = 0.55$.

Zusammenfassend bleibt festzuhalten, dass der in diesem Experiment untersuchte Effekt der Symmetrisierung durch gemorphte Gesichter *ohne* Farbwertmittelung und der Attraktivitätsbewertung im direkten Paarvergleich zu einem positiven Effekt führte. Symmetrische „Warps“ wurden geringfügig attraktiver bewertet als unveränderte Originalgesichter. Je nachdem, welches statistische Verfahren man aber anwendet, erhält man einmal einen starken positiven Effekt ($r = 0.55$) oder einen sehr schwachen positiven Effekt ($r = 0.17$). Bei der Beurteilung dieser Effektgrößen muss jedoch kritisch im Hinterkopf behalten werden, dass mehr als Dreiviertel der einzelnen Paarvergleiche statistisch nicht signifikant waren. Insofern verzerrt das Verfahren mit dem Ein-Stichproben- t -Test nach Little und Jones (2003) die Effektstärke stark.

4.9 Experimentelle Manipulation der Symmetrie durch Verwendung von Morphings *ohne* Farbwertmittelung („Warps“) und Attraktivitätsmessung durch Ratings (Untersuchung *Sym 6*)

4.9.1 Stand der Forschung

Wie bereits in Kapitel 4.8.1 dargestellt, gibt es nur zwei Studien, die den Effekt der Symmetrisierung mit Warp-Gesichtern (also ohne die Mittelung von Farbwerten) untersucht haben (Perret et al., 1999, Little & Jones, 2003). Diese beiden Untersuchungen sind noch nicht einmal als unabhängig zu betrachten, da sie dasselbe Stimulusmaterial benutzten. Beide Untersuchungen verwendeten allerdings Paarvergleiche als Datenerhebungsmethode. Studien mit Warp-Gesichtern als Stimuli und einem Rating-Verfahren zur Attraktivitätsmessung sind hingegen nicht bekannt.

4.9.2 Methode

4.9.2.1 Versuchspersonen

An der Untersuchung *Sym 6* nahmen insgesamt 128 Versuchspersonen teil. In der Experimentalgruppe, die symmetrisierte „Warps“ bewertete, waren 64 Personen, als Kontrollgruppe, die Originalgesichter bewertete, wurden die Daten der Kontrollgruppe von Experiment *Sym 2* verwendet (siehe Kapitel 4.5.3).

Von den 64 Personen der Experimentalgruppe waren männlich 32 (50.0 %), 32 weiblich (50.0 %). 7 Personen (10.9 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 14 Jahre, die älteste 67. Der Altersdurchschnitt betrug 34.03 Jahre ($SD = 13.88$).

Hinsichtlich der erhobenen demografischen Variablen ist die Experimentalgruppe mit der Kontrollgruppe vergleichbar, bei Geschlecht ($\chi^2(1, N = 128) = 0.031, p = .860$) und Alter ($t(126) = 0.422, p = .674$) bestehen keine Unterschiede, allerdings waren in der Experimentalgruppe weniger Personen mit Studierendenstatus ($\chi^2(1, N = 128) = 5.128, p = .040$). Da jedoch Psychologie-Studierende vom Experiment grundsätzlich ausgeschlossen waren, bei den Probanden kein themenspezifisches Vorwissen vorhanden war und die Altersverteilung gleich ist, kann dieser Unterschied vernachlässigt werden.

4.9.2.2 Stimulusmaterial und Versuchsdurchführung

Das Stimulusmaterial war identisch mit dem Material der Untersuchung *Sym 5*, d. h. es wurden die symmetrisch optimierten Gesichter verwendet, die durch das Morphing-Verfahren *ohne* Farbwertmittelung erzeugt worden waren („Warps“).

Die Versuchsdurchführung war identisch mit den Untersuchungen *Sym 2* und *Sym 4*, d. h. dass die Versuchspersonen der Experimentalbedingung 30 Stimuli der symmetrisch optimierten Variante präsentiert bekamen und auf einer Ratingskala bewerteten. Davon, dass die gezeigten Gesichter auch noch in anderen Varianten existierten (z. B. unveränderte Originalgesichter), wussten die Probanden nichts (*Between-Subjects Design*). Verwendet wurde dieselbe siebenstufige Ratingskala wie in den Experimenten *Sym 2* und *Sym 4*. Auch sonst galten dieselben Randbedingungen wie in Experiment *Sym 2* und *Sym 4* (siehe Kapitel 4.5.2.2 und 4.7.2.2). Die Beurteilerübereinstimmung war mit einem Cronbach's α von .99 (bei 64 Versuchspersonen) sehr hoch.

4.9.3 Ergebnisse

Zur Überprüfung der Frage, ob durch das Morphing-Verfahren *ohne* Farbwertmittelung symmetrisch optimierte Gesichter („Warps“) attraktiver bewertet wurden als die unveränderten Originalgesichter, wurde eine multivariate Varianzanalyse durchgeführt, die ein signifikantes Ergebnis lieferte ($p = .002$). Tabelle 14 listet für jeden Stimulus den erzielten Attraktivitätsmittelwert auf und zeigt für jede Stimulusperson die Attraktivitätsdifferenz zwischen symmetrisch optimiertem „Warp“ und Originalgesicht. Es zeigt sich, dass diese Differenzen in den meisten Fällen (23 Fälle) negativ sind, d. h. dass die symmetrisierten

Varianten meist etwas schlechter bewertet wurden als die Originalgesichter. Nur in 6 Fällen erhielten die symmetrischen „Warps“ eine höhere durchschnittliche Attraktivitätsbewertung (in einem Fall war die Differenz Null). Wie jedoch die letzte Spalte von Tabelle 14 zeigt, sind die Attraktivitätsdifferenzen nur in 6 von 30 Fällen statistisch signifikant. Von diesen 6 signifikanten Ergebnissen wurde 1 Miss-Germany-Finalistin durch die Symmetrisierung besser bewertet und 5 „normale“ Frauen schlechter.

Tabelle 14: Ergebnisse des Rating-Experiments mit durch das Morphing-Verfahren *ohne* Mittelwertbildung symmetrisch optimierten Gesichtern („Warps“) sowie den jeweiligen Originalgesichtern (= Kontrollgruppe). Aufgelistet sind die Attraktivitätsmittelwerte der jeweiligen Stimuli. Die drittletzte Spalte listet den Effekt der Symmetrisierung in Form von Mittelwertsdifferenzen sowie in Klammern Cohen's *d* auf. Die beiden letzten Spalten listen *F*- und *p*-Werte des Tests für Zwischensubjekteffekte der MANOVA auf. In beiden Versuchsgruppen war jeweils *N* = 64.

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i> Originale	<i>M</i> „Warps“	Differenz „Warps“ – Original (Cohen's <i>d</i>)	<i>F</i>	<i>p</i>
1	Miss	3.91	3.75	-0.16 (-0.15)	0.686	.409
2	Miss	4.59	4.56	-0.03 (-0.03)	0.021	.884
3	Miss	4.78	5.19	0.41 (0.41)	5.424	.021
4	Miss	4.38	4.38	0.00 (0.00)	0.000	1.000
5	Miss	4.33	4.03	-0.30 (-0.25)	1.952	.165
6	Miss	4.59	4.25	-0.34 (-0.29)	2.716	.102
7	Miss	5.22	5.02	-0.20 (-0.20)	1.261	.264
8	Miss	4.81	5.00	0.19 (0.17)	0.876	.351
9	Miss	4.92	4.86	-0.06 (-0.06)	0.114	.736
10	Miss	4.70	4.45	-0.25 (-0.24)	1.836	.178

Fortsetzung

Fortsetzung von Tabelle 14:

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i> Originale	<i>M</i> "Warps"	Differenz "Warps" – Original (Cohen's <i>d</i>)	<i>F</i>	<i>p</i>
11	Miss	4.78	4.80	0.02 (0.01)	0.004	.949
12	Miss	4.50	4.52	0.02 (0.01)	0.006	.937
13	Miss	5.28	5.44	0.16 (0.15)	0.690	.408
14	Miss	5.09	4.92	-0.17 (-0.14)	0.654	.420
15	Miss	5.58	5.70	0.13 (0.10)	0.348	.556
16	Normale	2.83	2.30	-0.53 (-0.49)	7.563	.007
17	Normale	2.64	2.30	-0.34 (-0.35)	3.902	.050
18	Normale	1.94	1.66	-0.28 (-0.31)	3.004	.086
19	Normale	2.34	2.05	-0.30 (-0.27)	2.318	.130
20	Normale	2.67	2.66	-0.02 (-0.01)	0.006	.937
21	Normale	2.45	2.44	-0.02 (-0.01)	0.006	.938
22	Normale	2.80	1.84	-0.95 (-0.92)	27.173	<.001
23	Normale	2.81	2.19	-0.63 (-0.54)	9.184	.003
24	Normale	2.53	2.16	-0.38 (-0.38)	4.710	.032
25	Normale	2.17	1.84	-0.33 (-0.33)	3.572	.061
26	Normale	2.06	1.84	-0.22 (-0.22)	1.605	.207
27	Normale	2.80	2.77	-0.03 (-0.03)	0.026	.872

Fortsetzung

Fortsetzung von Tabelle 14:

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i> Originale	<i>M</i> "Warps"	Differenz "Warps" – Original (Cohen's <i>d</i>)	<i>F</i>	<i>p</i>
28	Normale	2.44	1.94	-0.50 (-0.45)	6.482	.012
29	Normale	1.47	1.33	-0.14 (-0.17)	0.906	.343
30	Normale	2.95	2.75	-0.20 (-0.19)	1.148	.286
<i>M</i>		3.61	3.43	-0.18 (-0.17)		

Anmerkung: Alle Werte sind gerundet.

Wie in den Untersuchungen *Sym 2* und *Sym 4* wurde auch hier als Effektgröße Cohen's *d* berechnet. Der Median über alle 30 Cohen's *d* beträgt -.18. Das bedeutet, dass der Effekt der Symmetrisierung nach dieser Methode schwach negativ ist. Da Cohen's *d* auch Werte größer 1 annehmen kann, wurde zur besseren Vergleichbarkeit mit den Ergebnissen der anderen Symmetrie-Untersuchungen das mittlere *d* nach der in Kapitel 4.5.3 angegebenen Formel in ein *r* umgerechnet. Dabei ergibt sich ein äußerst kleiner negativer Effekt von $r = .09$ (Aufgrund der Formel ist der Wert immer positiv und müsste wegen des negativen Effekts der Symmetrisierung eigentlich mit einem Minus notiert werden).

Wie in den Untersuchungen *Sym 2* und *Sym 4* wurden jedoch auch hier mit einem alternativen Verfahren die Daten statistisch ausgewertet. Dabei wurde statt der Analyse über 2 x 64 Versuchspersonen (*between groups*) ein *t*-Test über die 30 Stimulusgesichter in ihren zwei Ausprägungen (symmetrisch versus Original, *within groups*, zweiseitig) gerechnet. Die Varianz der Versuchspersonenurteile bleibt dabei unberücksichtigt; verwendet werden nur die Attraktivitätswerte (= Mittelwerte über 64 Versuchspersonenurteile zu jedem Stimulusgesicht) als Eigenschaft eines jeweiligen Gesichts.

Das Ergebnis des *t*-Tests zeigt, dass Gesichter, die im Morphing-Verfahren *ohne* Farbwertmittelung symmetrisch optimiert wurden, klar unattraktiver bewertet wurden als unveränderte Originalgesichter $t(29) = -3.721$, $p = .001$. Im Durchschnitt wurden die symmetrisierten Gesichter um 0.18 Skalenstufen schlechter bewertet. Um den Effekt mit den

Ergebnissen der anderen durchgeführten Symmetrie-Untersuchungen vergleichen zu können, wurde wieder nach folgender Formel (Pospeschill, 2006) eine Effektgröße r berechnet:

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

Der Effekt beträgt $r = .57$. Da aufgrund der Formel r nur positive Werte annehmen kann und da aber symmetrisierte Gesichter schlechter als Originalgesichter bewertet wurden, ist der Effekt der Symmetrie *negativ* – es müsste also eigentlich ein Minus vor dem Wert stehen.

Wenn Symmetrie – wie die Symmetrie-Hypothese besagt – ein bedeutsamer Faktor für die Attraktivität eines Gesichts ist, dann sollten hochattraktive Gesichter auch sehr symmetrisch sein, normale Gesichter hingegen weniger symmetrisch. Die Miss-Germany-Finalistinnen sollten daher symmetrischere Gesichter haben als die „normalen“ Frauen und von einer zusätzlichen Symmetrisierung im Experiment weniger bzw. gar nicht profitieren, da sie ja schon symmetrisch sind („Decken-Effekt“). Allgemein formuliert lautet daher die Hypothese: „Je unattraktiver das Originalgesicht, desto größer ist der positive Effekt der Symmetrisierung.“

Diese Hypothese wurde für die symmetrisierten „Warp“-Gesichter mit einer Korrelation überprüft. Abbildung 48 zeigt den Zusammenhang. Das Streudiagramm zeigt, dass ein mittelstarker positiver Zusammenhang zwischen der Attraktivitätsveränderung durch Symmetrisierung und der Attraktivität des Originalgesichts besteht. Die Richtung des Effekts ist jedoch genau umgekehrt als von der Hypothese vorhergesagt ($r = .50$, $p = .005$). Die unattraktiveren Gesichter der „normalen“ Frauen hätten von der Symmetrisierung im Vergleich zu den attraktiveren Miss-Germany-Finalistinnen stärker profitieren sollen, profitierten jedoch weniger. Bei den attraktiveren Models gab es ein paar, die symmetrisiert etwas besser bewertet wurden (jedoch alle nicht signifikant), auf die unattraktiveren „normalen“ Frauen hatte dagegen die Symmetrisierung entweder keinen oder einen signifikant negativen Effekt.

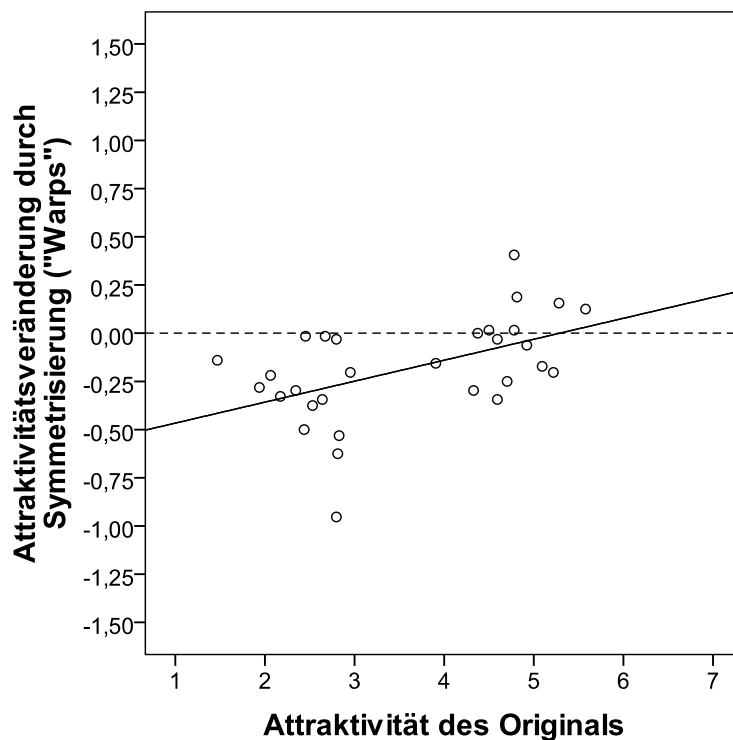


Abbildung 48: Der Zusammenhang zwischen der Attraktivität eines Originalgesichts und der Attraktivitätsveränderung durch das Symmetrisieren mit der *Warp*-Methode. Bei den attraktiven Model-Gesichtern hat die Symmetrisierung keinen Effekt, bei den normalen Frauengesichtern mindert sie die Attraktivität tendenziell.

Zusammenfassend bleibt festzuhalten, dass der in diesem Experiment untersuchte Effekt der Symmetrisierung durch gemorphte Gesichter *ohne* Farbwertmittelung („Warps“) und der Attraktivitätsbewertung im Rating-Verfahren und Between-Subject-Design zu negativen Effekten führte. Symmetrische „Warp“-Gesichter wurden unattraktiver bewertet als unveränderte Originalgesichter. Je nachdem, ob man für die Berechnung der Effektgröße die Varianz der einzelnen Versuchspersonenurteile einbezieht oder weglässt, ergeben sich jedoch unterschiedliche Effekte: Bei Einbeziehung ein sehr schwacher negativer Effekt von $r = -.09$; bei Weglassen ein mittelstarker negativer Effekt von $r = -.57$.

4.10 Korrelatives Versuchsdesign mit direktem Rating der Symmetrie durch Versuchspersonen (Untersuchung *Sym 7*)

4.10.1 Stand der Forschung

Die früheste Studie, in der Versuchspersonen die Symmetrie von Gesichtern durch direkte Ratings beurteilten, stammt von den „Erfindern“ der Durchschnittshypothese Langlois und Roggman. Man darf wohl unterstellen, dass es den Autorinnen bei dieser Methode kein primäres Anliegen war, eine möglichst reliable und valide Methode der Symmetriemessung zu etablieren, denn ihr gesamter Artikel (Langlois et al., 1994) ist eine einzige Verteidigung der Durchschnittshypothese, in welchem sie sämtliche Einwände gegen die Bedeutung der Durchschnittlichkeit (z. B. Konfundierung mit Symmetrie) zu entkräften versuchen, die seit ihrem viel zitierten ersten Artikel (Langlois & Roggman, 1990) vorgebracht wurden. So überrascht es auch nicht, dass die Autoren in ihrer Studie nur eine marginale Korrelation ($r = .11$, n.s.) zwischen Symmetrie-Ratings und Attraktivitätsratings fanden, die auch nicht signifikant wurde.

Die zweite Untersuchung (Zebrowitz, Voinescu & Collins, 1996) mit dieser Methode ist eine Längsschnittstudie zur sozialen Wahrnehmung, die Fotos von Personen aus fünf verschiedenen Lebensabschnitten (von der Kindheit bis in die 50er Jahre) analysierte und den Zusammenhang zwischen Merkmalen des Kindchenschemas wahrgenommener Anständigkeit (*perceived honesty*) erforschte. Eines von mehreren durch Ratings gemessenen Gesichtsmerkmalen war Symmetrie. Zebrowitz et al. fanden einen schwachen signifikanten Zusammenhang ($r = .23$) zwischen Symmetrie-Ratings und Attraktivitätsratings von Gesichtern im jungen Erwachsenenalter zwischen 20 und 30 Jahren, eine Altersklasse also, die üblicherweise für Attraktivitätsstudien verwendet wird (bei Fotos aus Kindheit und Pubertät war der Zusammenhang größer, bei Fotos aus den 30er und 50er Jahren kleiner).

In einer Untersuchung von Rhodes und Kollegen mussten Versuchspersonen (in einem Between-Design, Experiment 1A) entweder die Symmetrie von Gesichtern oder die Attraktivität auf einer Ratingskala beurteilen. Als Stimuli dienten gemorphte und Originalgesichter. Es zeigte sich eine schwache Korrelation von $r = .29$ zwischen Symmetrie- und Attraktivitätsratings (Rhodes et al., 1998).

In einer weiteren Untersuchung (Rhodes, Sumich et al., 1999) ließen die Autoren ebenfalls Versuchspersonen die Symmetrie von Gesichtern (in einem Within-Design) mit Ra-

tingskalen bewerten, jedoch zusammen mit dem Rating von Durchschnittlichkeit und dem Gesichtsausdruck (expression). Sie verwendeten zudem Originalgesichter *und* gemorphte Gesichter. Das eigentliche Ziel war, mittels multipler Regressionsanalysen den Effekt der bewerteten Eigenschaften statistisch zu kontrollieren. Dennoch sind auch die Ergebnisse der einfachen Korrelationen angegeben. Symmetrie-Ratings korrelierten mit $r = .46$ mit den Attraktivitätsratings.

Einen Effekt ähnlicher Größe fanden Simmons, Rhodes et al. (2004). Bei Männergesichtern korrelierten Symmetrie-Ratings und Attraktivitätsratings mit $r = .46$, bei Frauengesichtern $r = .54$. In einer weiteren Studie von Rhodes und Kollegen wurden ebenfalls Direkt-Ratings der Symmetrie von Gesichtern verwendet, diesmal mit dem Ziel, den Zusammenhang mit wahrgenommener Gesundheit (ebenfalls gemessen durch Versuchspersonen-Ratings) zu untersuchen. Die Symmetrie-Ratings und Attraktivitätsratings wurden jedoch aus den Daten der oben beschriebenen früheren Untersuchung (Rhodes, Sumich et al., 1999) verwendet. Insofern kann man diese Untersuchung nicht als eigene Untersuchung werten.

In einer weiteren Untersuchung, die im selben Artikel berichtet wird, wollten sie den Zusammenhang zwischen bewerteter Symmetrie und tatsächlicher Gesundheit (die sie aus vorhandenen medizinischen Daten ableiteten) erforschen. Hierzu ließen sie Symmetrie-Ratings von eigenen Versuchspersonen durchführen und benutzten Attraktivitätsratings derselben Gesichter aus einer früheren Untersuchung (Zebrowitz, Olson & Hoffman, 1993). Sie fanden eine schwache signifikante Korrelation zwischen Symmetrie-Ratings und Attraktivitätsratings ($r = .25$). Ironischerweise konnte jedoch in derselben Studie dieser Zusammenhang mit aus *Facial Landmarks* gemessener Symmetrie nicht gefunden werden ($r = -.11$, n.s.), ebenso wie bereits frühere Studien keinen Zusammenhang zwischen aus *Facial Landmarks* gemessener Symmetrie und Attraktivität nachweisen konnten (Jones & Hill, 1993; Shackelford & Larsen, 1997). Doch anstatt selbstkritisch die Frage zu stellen, ob nicht möglicherweise Symmetrie- und Attraktivitätsratings z. B. aufgrund eines Halo-Effekts der Attraktivität miteinander konfundiert sein könnten, greifen Rhodes und Kollegen die Validität von Symmetrie-Messungen aus *Facial Landmarks* an: „Given the strong evidence that facial symmetry is attractive (Mealey, Bridgestock, & Townsend, 1999; Perrett et al., 1998; Rhodes et al., 1998), failure to find an association between attractiveness and measured asymmetry raises doubts about the validity of the measurements” (Rhodes et al., 2001, S. 40). In dieser Studie hatten sie eine sehr simple Methode der Symmetrie-Messung aus *Facial Landmarks* (nach Grammer und Thornhill, 1994)

verwendet, die neben der fluktuierenden auch die direktionale Asymmetrie misst, eine Methode, die sie später als ungeeignet kritisierten (Simmons, Rhodes et al., 2004).

Tatsächlich ist Gilian Rhodes die wohl einflussreichste Befürworterin der Methode des direkten Symmetrie-Ratings. In ihrer Metaanalyse fasst sie zusammen: „Interestingly, symmetry ratings seem to reflect FA [= fluktuierende Asymmetrie, Anm. d. Verf.], but not DA [= direktionale Asymmetrie, Anm. d. Verf.], and may be a valid proxy of FA (Simmons et al., 2004)“ (Rhodes, 2006, S. 207). Damit widerspricht sie sich selbst, denn in der Studie, die sie zitiert und an der sie selbst beteiligt war, schreibt sie über die fluktuierende Asymmetrie, die dort über *Facial Landmarks* separat gemessen wurde: „Our measures of asymmetry only captured around 20% of the variance in perceived symmetry (Simmons, Rhodes, Peters & Koehler, 2004, S. 867)“. Da stellt sich die Frage, was an nur 20 % Varianzaufklärung valide sein soll. Trotz der von ihr selbst erzielten unbefriedigenden Ergebnisse hinsichtlich des Zusammenhangs von Symmetrie-Ratings mit der direkt über *Facial Landmarks* gemessenen (fluktuierenden) Asymmetrie, verteidigt sie dennoch ihre bevorzugte Methode des Symmetrie-Ratings: „Ratings change systematically when facial symmetry is manipulated, confirming their validity (see, e.g., Rhodes et al. 1999b), and are probably sensitive to more subtle facial asymmetries than are current measurements methods. Therefore, on theoretical grounds, ratings may be preferable to measurements“ (Rhodes, 2006, S. 207).

Von anderen Forschern hingegen wurde die Methode des direkten Ratings mehrfach kritisiert. Scheib et al. (1999) stellten beispielsweise ebenfalls fest, dass die direkten Ratings natürlicher Gesichter hinsichtlich Symmetrie nur sehr schwach und nicht signifikant mit der direkt über *Facial Landmarks* gemessenen Symmetrie korrelierten ($r = .26$, n.s.). Dies freilich könnte ebenso an einer geringen Reliabilität und Validität der Landmark-Vermessungen liegen, die auch vielfach von anderen Autoren kritisiert wurde (Hönekopp et al., 2004; Mealey et al., 1999; Palmer & Strobeck, 2003; Palmer & Strobeck, 1986).

Ein weiteres Ergebnis aus derselben Untersuchung muss zudem stutzig machen. Scheib et al. ließen zusätzlich zu ganzen Gesichtern auch einzelne Gesichtshälften von Versuchspersonen hinsichtlich Symmetrie bewerten. Nur wenige Versuchspersonen (alleamt *psychology undergraduates*) hatten ein Problem mit dieser absurden Aufgabenstellung, die Mehrheit jedoch nicht und alle führten das Symmetrie-Rating durch. „Not surprisingly, a few subjects reported difficulty with rating the symmetry of half faces and completed the task after being instructed to imagine what the full face would look like“ (Scheib et al., 1999, S. 1915). Die Versuchspersonen beurteilen also Symmetrie in Bil-

dern, in denen überhaupt keine Symmetrie zu sehen war, weil es sich nur um einzelne Gesichtshälften handelte. Dies bedeutet, dass die Versuchspersonen (Psychologiestudierende) entweder mehrheitlich überhaupt nicht wussten, was Symmetrie eigentlich bedeutet, oder dass sie völlig unsinnige Aufgabenstellungen brav bearbeiteten, obwohl eine sinnvolle Beantwortung der Fragen gar nicht möglich ist. Möglicherweise verstanden sie unter Symmetrie gar keine Bilateralsymmetrie im eigentlichen Sinn, sondern etwas wie „harmonische Gesichtszüge“ oder Ebenmäßigkeit, was beides mit Attraktivität zusammenhängt.

Leider unterschlagen Scheib et al. (1999) in ihrem Artikel die Ergebnisse der Korrelationen von Symmetriebewertungen einzelner Gesichtshälften und Attraktivität der ganzen Gesichter, ebenso wie sie nicht die Korrelationen zwischen Symmetriebewertungen und Attraktivitätsbewertungen der ganzen Gesichter berichten, obwohl sie diese Daten erhoben hatten. Der Vergleich dieser Korrelationen wäre bei ähnlich hohen Korrelationskoeffizienten eine simple Methode gewesen zu beweisen, dass Versuchspersonen beim Symmetrie-Rating ganzer Gesichter eigentlich nicht Symmetrie bewerten, sondern stattdessen die Attraktivität selbst (bzw. Harmonie oder Ebenmäßigkeit), die man auch an einer einzelnen Gesichtshälfte beurteilen kann.

Insgesamt ist das direkte Rating der Symmetrie von Gesichtern eine relativ häufig angewendete Methode, wohl nicht zuletzt deswegen, weil sie für den Forscher im Vergleich zu den anderen Verfahren am wenigsten Aufwand bedeutet. Gleichzeitig bestehen jedoch Zweifel, ob Versuchspersonen überhaupt in der Lage sind, die Symmetrie eines Gesichts unabhängig von dessen Attraktivität zu bewerten und sich bei der Symmetrie-Beurteilung nicht von der Attraktivität eines Gesichts im Sinne eines Halo-Effekts beeinflussen lassen.

4.10.2 Methode

4.10.2.1 Versuchspersonen

An der Untersuchung *Sym 7* nahmen insgesamt 124 Versuchspersonen teil. In der Experimentalgruppe, die die Originalgesichter direkt hinsichtlich Symmetrie bewertete, waren 60 Personen. Zusätzlich wurden für die Datenauswertung die Daten der 64 Versuchspersonen aus dem Experiment *Sym 2* verwendet, die alle Originalgesichter nach Attraktivität bewertet hatten.

Von den 60 Personen der Experimentalgruppe waren männlich 27 (45.0 %), 33 weiblich (55.0 %). 23 Personen (38.3 %) waren Studierende, jedoch befanden sich keine Psycho-

logiestudierenden in der Stichprobe. Die jüngste Person war 14 Jahre, die älteste 65. Der Altersdurchschnitt betrug 34.25 Jahre ($SD = 13.60$).

Hinsichtlich der erhobenen demografischen Variablen ist die Symmetrie-Bewertungsgruppe mit der Attraktivität-Bewertungsgruppe vergleichbar; bei Geschlecht ($\chi^2(1, N = 124) = 0.534, p = .478$), Alter ($t(122) = -0.679, p = .498$) und Studierendenstatus ($\chi^2(1, N = 124) = 1.963, p = .182$) bestehen keine Unterschiede.

4.10.2.2 Stimulusmaterial und Versuchsdurchführung

Als Stimulusmaterial wurden die 30 Originalgesichter verwendet, die auch in allen anderen Symmetrie-Untersuchungen verwendet wurden. Die Versuchsdurchführung war identisch mit den Attraktivitäts-Ratings der Untersuchungen *Sym 2*, *Sym 4* und *Sym 6*, jedoch mit dem Unterschied, dass die Versuchspersonen nicht die Attraktivität, sondern die Symmetrie jedes Gesichts auf einer siebenstufigen Ratingskala bewerten sollten.

Die Stufen lauteten: 1 = sehr asymmetrisch, 2 = ziemlich asymmetrisch, 3 = eher asymmetrisch, 4 = eher symmetrisch, 5 = ziemlich symmetrisch, 6 = sehr symmetrisch, 7 = völlig symmetrisch. Während des Ratings der Gesichter hatten die Versuchspersonen stets einen Zettel mit der Ratingskala vor sich, um sich die Benennung der Skalenstufen nicht merken zu müssen und sich dadurch ganz auf die präsentierten Gesichter konzentrieren zu können. Ansonsten galten dieselben Randbedingungen wie in den zuvor durchgeführten Symmetrie-Untersuchungen *Sym 2*, *Sym 4* und *Sym 6*.

Die Beurteilerübereinstimmung war mit einem Cronbach's α von .96 (bei 60 Versuchspersonen) bei den Symmetrie-Bewertungen sehr hoch. Dasselbe gilt für die Attraktivitätsbewertungen mit einem Cronbach's α von .99 (bei 64 Versuchspersonen)

4.10.3 Ergebnisse

Um zu überprüfen, ob es einen Zusammenhang zwischen der Attraktivität und der Symmetriebewertung (ermittelt durch direktes Rating der Symmetrie) von Gesichtern gibt, wurde für jedes Gesicht ein mittlerer Symmetrie-Rating-Wert berechnet. Diese Werte wurden als Symmetrie-Indizes der jeweiligen Gesichter verwendet und ins Verhältnis zu den bereits bekannten Attraktivitätsmittelwerten derselben Gesichter gesetzt.

Das Gesicht mit dem niedrigsten mittleren Symmetriewert hatte den Wert 2.87 (entspricht „eher asymmetrisch“), das mit dem höchsten 5.28 (entspricht „ziemlich symmetrisch“).

Der durchschnittliche Symmetriewert über alle Gesichter beträgt 3.93 ($SD = 0.74$). Die genauen Werte finden sich im Anhang C (S. 397).

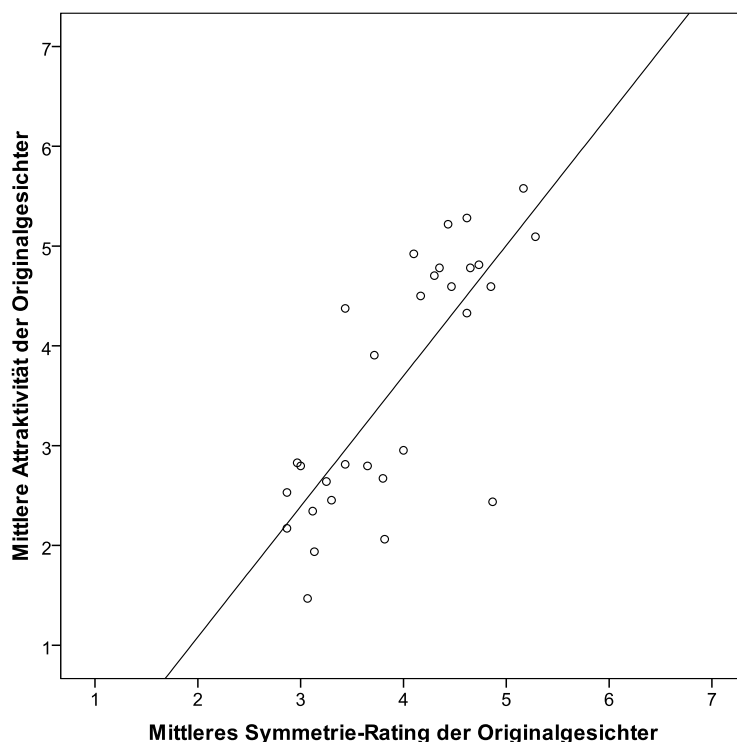


Abbildung 49: Die Korrelation zwischen dem mittleren Symmetrie-Rating der Originalgesichter durch Versuchspersonen und der mittleren Attraktivitätsbewertung durch (eine andere Gruppe von) Versuchspersonen. Je höher die beurteilte Symmetrie, desto höher die Attraktivität ($r = .78$, $p < .001$). Die Richtung der Kausalität ist jedoch unklar. Denkbar ist auch: Je höher die Attraktivität, desto höher die Symmetrie-Beurteilung.

Wie das Streudiagramm illustriert, besteht ein starker positiver Zusammenhang zwischen dem Symmetriemaß (ermittelt durch direktes Bewerten der Symmetrie durch die Vpn) und den mittleren Attraktivitätswerten der Gesichter. Die Korrelation beträgt $r = .78$ ($p < .001$). Je höher die Symmetrie eines Gesichts bewertet wird, desto attraktiver ist es. Da bei Korrelationen die Richtung der Kausalität prinzipiell unklar ist, könnte man natürlich auch sagen: Je attraktiver ein Gesicht ist, desto höher wird seine Symmetrie bewertet. Überhaupt bleibt ungewiss, ob Versuchspersonen überhaupt in der Lage sind, die Symmetrie eines Gesichts unabhängig von dessen Attraktivität zu beurteilen, oder ob das Symmetrie-Urteil womöglich stark vom Attraktivitätseindruck des Gesichts beeinflusst ist.

Falls dies der Fall sein sollte, läge eine Konfundierung zwischen unabhängiger (Symmetrie) und abhängiger Variable (Attraktivität) vor.

Um dies zu überprüfen, wurde nachträglich das im folgenden Kapitel 4.10.4 beschriebene Experiment als Kontrollbedingung durchgeführt.

4.10.4 Zusatzexperiment mit Symmetrie-Rating von „Warp“- Gesichtern als Kontrollbedingung (Untersuchung *Sym 7* *Kontrolle*)

Aufgrund des überraschend hohen positiven Zusammenhangs zwischen Symmetrie-Ratings von Originalgesichtern und deren Attraktivitätsbewertungen und der oben erwähnten methodischen Bedenken hinsichtlich Richtung der Kausalität und der Frage, ob Versuchspersonen überhaupt in der Lage sind, die Symmetrie eines Gesichts zu bewerten, wurde nachträglich ein weiteres Experiment durchgeführt.

Hierbei wurde den Versuchspersonen quasi eine „Falle“ gestellt: Sie sollten die Symmetrie von Gesichtern bewerten, hatten also dieselbe Aufgabe wie in Untersuchung *Sym 7*, bekamen jedoch ohne ihr Wissen Warp-Gesichter als Stimulusmaterial. Da Warp-Gesichter aufgrund ihrer Manipulation bereits alle perfekt symmetrisch sind, ist die Aufgabe eigentlich sinnlos. Wenn es keine Symmetrieunterschiede zwischen den einzelnen Gesichtern gibt, sollten im Durchschnitt alle Gesichter dieselbe (hohe) Symmetrie-Bewertung bekommen. Eine Korrelation zwischen den mittleren Symmetrie-Bewertungen und den mittleren Attraktivitätsbewertungen der Originalgesichter müsste folglich Null sein. Ist die Korrelation jedoch positiv, ist dies ein Beleg dafür, dass Versuchspersonen sich bei der Bewertung der Symmetrie der (symmetrischen) Warp-Gesichter offensichtlich von der Attraktivität der zu bewertenden Gesichter blenden lassen (→ „Halo-Effekt“ der Attraktivität). Lediglich die Differenz zwischen dem Korrelationskoeffizienten in Untersuchung *Sym 7* ($r = .78$) und dem Korrelationskoeffizienten in dieser Kontrollbedingung könnte auf den Effekt der Symmetrie zurückgeführt werden. Der Rest sind methodische Artefakte, nämlich ein Halo-Effekt der Attraktivität oder auch ein Erwartungs-Effekt der Versuchspersonen, d. h. dass diese selbst daran glauben, dass Symmetrie für Attraktivität wichtig ist und deswegen abhängig von der Attraktivität der zu bewertenden Gesichter unterschiedliche Symmetrie-Bewertungen vergeben.

Die Grundüberlegungen dieser zusätzlichen Untersuchung als Kontrollbedingung im Überblick:

- Wenn Versuchspersonen tatsächlich wie von ihnen verlangt *nur* die Symmetrie von Gesichtern bewerten, dann gibt es bei der Symmetrie-Bewertung von perfekt symmetrischen Warp-Gesichtern im Durchschnitt keine Symmetrie-Unterschiede zwischen den Gesichtern.
- Wenn es bei symmetrischen Warp-Gesichtern keine Symmetrie-Unterschiede gibt, dann ist die Korrelation dieser Symmetrie-Bewertungen mit den Attraktivitätsbewertungen der Originalgesichter (sowie sämtlicher anderer Variablen) Null.
- Im Umkehrschluss bedeutet dies:
- Gibt es eine positive Korrelation zwischen Symmetrie-Bewertungen der Warp-Gesichter und Attraktivität der Originalgesichter, dann beeinflusst die Attraktivität eines symmetrischen Warp-Gesichts das Symmetrie-Urteil der Versuchspersonen.
- Ist die Korrelation zwischen Symmetrie-Bewertung der symmetrischen Warp-Gesichter und der Attraktivität der Originalgesichter genauso hoch wie die Korrelation der Symmetrie-Bewertung der *Originalgesichter* und der Attraktivität der Originalgesichter, nämlich $r = .78$, dann bedeutet dies, dass
 - bereits bei Untersuchung *Sym 7* die Versuchspersonen (absichtlich oder unabsichtlich) nicht die Symmetrie der Originalgesichter beurteilt haben, sondern deren Attraktivität,
 - Versuchspersonen offensichtlich nicht in der Lage sind, Symmetrie-Bewertungen unabhängig von der Attraktivität eines Gesichts abzugeben.

Versuchspersonen

An der Zusatz-Untersuchung als Kontrollbedingung zur Untersuchung *Sym 7* nahmen 64 Versuchspersonen teil. Davon waren 34 (53.1 %) männlich und 30 weiblich (46.9 %). 21 Personen (32.8 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 19 Jahre, die älteste 84. Der Altersdurchschnitt betrug 35.44 Jahre ($SD = 14.47$).

Die Beurteilerübereinstimmung war mit einem Cronbach's α von .92 (bei 64 Versuchspersonen) sehr hoch. Dies ist durchaus überraschend, wenn man bedenkt, dass die Versuchspersonen eine eigentlich sinnlose Instruktion zu befolgen hatten (nämlich perfekt symmetrische Gesichter nach Symmetrie zu bewerten) und die Aufgabenschwierigkeit aufgrund der objektiv nicht vorhandenen Symmetrieunterschiede zwischen den einzelnen Gesichterstimuli sehr hoch war.

Hinsichtlich der erhobenen demografischen Variablen unterscheidet sich diese Stichprobe nicht von der Stichprobe in Untersuchung *Sym 7* (Symmetrie-Bewertungen der Originalgesichter) und auch nicht von der Stichprobe für die Attraktivitätsbewertungen der Originalgesichter: Geschlecht $\chi^2(2, N = 188) = 0.913, p = .634$, Alter $F(2, 185) = 0.675, p = .510$ und Studierendenstatus ($\chi^2(2, N = 188) = 1.963, p = .375$).

Stimulusmaterial

Als Stimulusmaterial wurden die 30 Warp-Gesichter verwendet, die bereits bei den Untersuchungen *Sym 5* und *Sym 6* verwendet worden waren. Die Versuchsdurchführung war identisch mit dem Vorgehen bei Untersuchung *Sym 7*, bei dem *Originalgesichter* hinsichtlich Symmetrie bewertet worden waren.

Ergebnisse

Die Datenauswertung erfolgte analog zu Untersuchung *Sym 7*. Für jedes Warp-Gesicht wurde ein mittlerer Symmetrie-Rating-Wert berechnet. Diese Werte wurden als Symmetrie-Indizes der jeweiligen Gesichter verwendet und ins Verhältnis zu den bereits bekannten Attraktivitätsmittelwerten der Originalgesichter gesetzt.

Das Gesicht mit dem niedrigsten mittleren Symmetriewert hatte den Wert 3.67 (entspricht „eher symmetrisch“), das mit dem höchsten 5.45 (entspricht „ziemlich symmetrisch“). Der durchschnittliche Symmetriewert über alle Gesichter beträgt 4.60 ($SD = 0.52$). Die genauen Werte finden sich im Anhang D (S. 398).

Wie eine erste Analyse zeigt, besteht zwischen den Symmetrie-Ratings der Originalgesichter und den Symmetrie-Ratings der symmetrischen Warp-Gesichter eine Korrelation von $r = .75, p < .001$, obwohl dieser Zusammenhang aufgrund der nicht vorhandenen Symmetrie-Unterschiede bei den symmetrisierten Warp-Gesichtern eigentlich Null sein müsste. Dies ist ein klarer Hinweis darauf, dass die Versuchspersonen offensichtlich etwas anderes als die Symmetrie der Gesichter bewerteten – höchstwahrscheinlich die Attraktivität.

Die zweite Analyse ist die Korrelation der Symmetrie-Ratings mit den Attraktivitätsratings der Originalgesichter, die das folgende Streudiagramm illustriert:

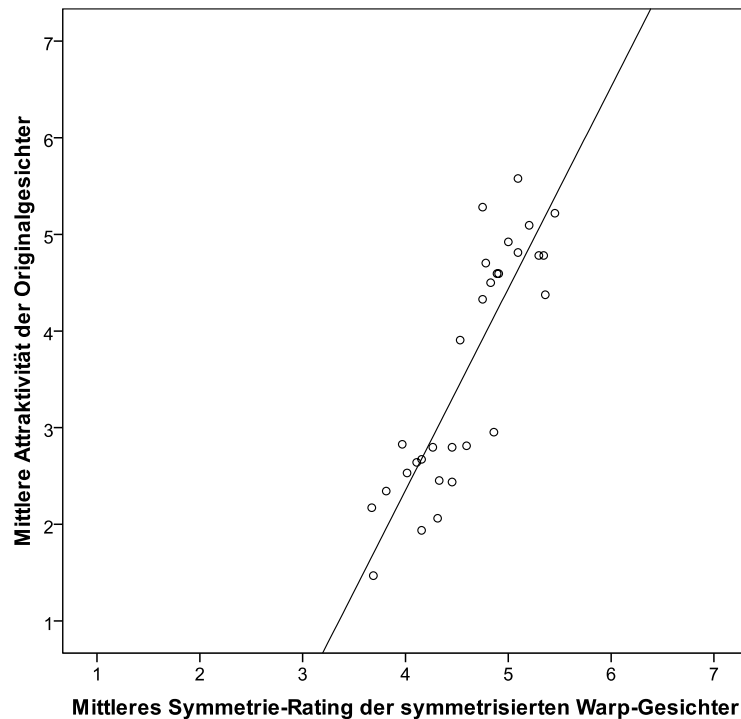


Abbildung 50: Die Korrelation zwischen dem mittleren Symmetrie-Rating der perfekt symmetrischen Warp-Gesichter durch Versuchspersonen und der mittleren Attraktivitätsbewertung der Originalgesichter. Die Warp-Gesichter weisen objektiv keine Asymmetrien auf und sind alle gleich perfekt symmetrisch. Dennoch wird deren Symmetrie von den Versuchspersonen unterschiedlich beurteilt. Je attraktiver ein Gesicht, desto höher wird seine Symmetrie eingeschätzt ($r = .87, p < .001$). In diesem Fall ist die Richtung der Kausalität eindeutig: Attraktivität bedingt Symmetrie-Urteil.

Wie die Abbildung 50 zeigt, besteht ein starker positiver Zusammenhang zwischen dem Symmetriemaß (ermittelt durch direktes Bewerten der Symmetrie durch die Vpn) und den mittleren Attraktivitätswerten der Gesichter. Die Korrelation beträgt $r = .87$ ($p < .001$). Je höher die Symmetrie eines Gesichts bewertet wird, desto attraktiver ist es. Damit ist dieser Zusammenhang sogar noch etwas größer als in der Untersuchung *Sym 7*, in der die Versuchspersonen die Symmetrie der *Originalgesichter* direkt bewerteten. Um zu überprüfen, ob sich die beiden Korrelationskoeffizienten signifikant unterscheiden, wurden die Werte nach Fisher Z-transformiert und ein Test nach Gleichheit der Korrelationskoeffizienten nach Fisher gerechnet. Der Unterschied erwies sich als nicht signifikant ($Z = 1.56, p = .118$).

Diskussion

Hieraus folgt, dass es offensichtlich völlig egal ist, ob man Versuchspersonen asymmetrische Originalgesichter oder perfekt symmetrische Warp-Gesichter für Symmetrie-Ratings vorlegt. Die Ergebnisse sind die gleichen. Sie bewerten ganz offenkundig nicht die Symmetrie, sondern etwas anderes, was sehr hoch (zwischen .78 und .87) mit Attraktivität korreliert. Es spricht viel dafür, dass dieses Etwas die Attraktivität selbst ist.

Erwähnt werden sollte auch, dass keine einzige Versuchsperson den Versuch abbrach oder sagte, dass sie die Symmetrie nicht beurteilen könne, weil die Gesichter alle gleich symmetrisch seien oder die Sinnlosigkeit der Aufgabe erkannte. Die Versuchspersonen hatten augenscheinlich keine Mühe mit der Aufgabenstellung und gaben zügig ihre Bewertungen ab.

Diese Untersuchung belegt klar, dass die in der Metaanalyse von Rhodes (2006) so gelobte Methode des direkten Ratings von Symmetrie durch Versuchspersonen, die angeblich so messempfindlich ist, in Wahrheit Humbug ist. Versuchspersonen sind nicht in der Lage, die Symmetrie eines Gesichts unabhängig von dessen Attraktivität zu bewerten. Damit sind abhängige und unabhängige Variable hochgradig miteinander konfundiert. Was Rhodes und Kollegen in ihren Untersuchungen gemessen haben und für die Symmetrie eines Gesichts hielten, ist nichts weiter als der seit 30 Jahren bekannte Halo-Effekt der Attraktivität (Dion, Berscheid & Walster, 1972). „Was schön ist, ist auch gut“. Selbstverständlich ist es gut, wenn ein Gesicht symmetrisch ist – wer würde das Gegenteil behaupten?! Wenn ein Gesicht attraktiv aussieht, dann schreiben Versuchspersonen ihm auch die Eigenschaft Symmetrie zu. Je schöner ein Gesicht ist, desto symmetrischer wird es also beurteilt – und zwar völlig unabhängig davon, wie symmetrisch oder asymmetrisch es in Wahrheit ist. Damit ist die in Kapitel 4.10.3 aufgeworfene Frage nach der Richtung der Kausalität eindeutig beantwortet: Nicht symmetrische Gesichter werden attraktiver beurteilt, sondern attraktive Gesichter werden symmetrischer beurteilt. Die Attraktivität bedingt das Symmetrie-Urteil.

4.11 Korrelatives Versuchsdesign mit Versuchspersonen-Ratings der Ähnlichkeit von separaten Gesichtshälften als Maß für Symmetrie (Untersuchung *Sym 8*)

4.11.1 Stand der Forschung

Die beurteilte Ähnlichkeit von separaten Gesichtshälften wurde bislang noch niemals als Maß für die Symmetrie eines Gesichts verwendet, obwohl diese Methode eigentlich unkomplizierter ist als die Beurteilung der Ähnlichkeit von Chimärengesichtern, für die es zumindest einige Studien als Vorbild gibt (Hönekopp et al., 2004; Mealey et al., 1999; Penton-Voak, Jones et al., 2001). Denn anders als bei Chimärengesichtern ergibt sich bei separaten Gesichtshälften nicht das Problem, dass Gesichter durch die Manipulation eigenartig und künstlich aussehen können. Sicher ist eine Beurteilung von separaten Hälften für die Versuchspersonen schwieriger, da ein Gesicht nicht ganzheitlich betrachtet werden kann. Andererseits kann man jedoch nicht das Argument gelten lassen, die Beurteilung sei *zu* schwierig, zumindest nicht, wenn man die Auffassung vertritt, dass Versuchspersonen auch in der Lage sind, die Symmetrie eines natürlichen Gesichts direkt zu erfassen und zu raten. Der einzige Unterschied zu dieser Methode besteht schließlich darin, dass bei der Präsentation von separaten Gesichtshälften eine Hälfte horizontal gespiegelt wurde. Ansonsten bleibt die Aufgabe gleich.

4.11.2 Methode

4.11.2.1 Versuchspersonen

An der Untersuchung *Sym 8* nahmen insgesamt 128 Versuchspersonen teil. In der Gruppe, die die Gesichtshälften hinsichtlich Ähnlichkeit bewertete, waren 64 Personen. Zusätzlich wurden für die Datenauswertung die Daten der 64 Versuchspersonen aus dem Experiment *Sym 2* verwendet, die alle Originalgesichter nach Attraktivität bewertet hatten.

Von den 64 Personen der Gruppe mit den Ähnlichkeitsbewertungen waren 30 (46.9 %) männlich und 34 (53.1 %) weiblich. 18 Personen (28.1 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 14 Jahre, die älteste 84. Der Altersdurchschnitt betrug 34.08 Jahre ($SD = 14.83$).

Hinsichtlich der erhobenen demografischen Variablen ist die Ähnlichkeits-Bewertungsgruppe mit der Attraktivität-Bewertungsgruppe vergleichbar; bei Geschlecht ($\chi^2(1, N =$

128) = 0.281, $p = .596$), Alter ($t(126) = -0.599$, $p = .550$) und Studierendenstatus ($\chi^2(1, N = 128) = 0.039$, $p = .843$) bestehen keine Unterschiede.

4.11.2.2 Stimulusmaterial und Versuchsdurchführung

Als Grundlage für das Stimulusmaterial dienten die 30 Originalgesichter, die auch in allen anderen Symmetrie-Untersuchungen verwendet wurden. Diesmal jedoch wurden nur Gesichtshälften präsentiert. Die Aufgabe der Versuchspersonen war es, die Ähnlichkeit der beiden Gesichtshälften zu beurteilen. Der zu Grunde liegende Gedanke dahinter lautet: Je ähnlicher sich die jeweiligen Gesichtshälften sind, desto symmetrischer ist das Gesicht. Der empirisch über Versuchspersonenbefragung ermittelte Ähnlichkeitswert wird als Symmetrie-Index verwendet.

Um zu erreichen, dass die Probanden auch tatsächlich die Gesichtshälften genau betrachteten und miteinander verglichen, wurde eine Gesichtshälfte des Stimulusgesichts horizontal gespiegelt und so präsentiert wie auf Abbildung 51 illustriert.



Abbildung 51: Beispiel für das verwendete Stimulusmaterial. Ähnlich wie bei der Erzeugung der Chimärengesichter wurde jedes Stimulusgesicht exakt in der Mitte geteilt. Die linke Gesichtshälfte wurde horizontal gespiegelt und wie in der Abbildung präsentiert, um zu verhindern, dass die Probanden die beiden Gesichtshälften als ein einziges Gesicht ganzheitlich wahrnahmen.

Durch das Trennen der Gesichtshälften und die gleichzeitige Spiegelung einer Hälfte wurde zudem verhindert, dass die Personen das Gesicht ganzheitlich erfassen konnten,

da in diesem Fall ein stärkerer Halo-Effekt der Attraktivität eines Gesichts wahrscheinlicher gewesen wäre. Alle Fotos mit Gesichtshälften hatten eine hohe Auflösung von 1600 x 800 Pixel. Für die Versuchsdurchführung wurden sie von einem Fotodienst auf Fotopapier in einer Größe von 9 x 4.5 Zentimeter belichtet. Daraus ergibt sich eine Auflösung von 452 dpi. Bei der paarweisen Präsentation der Gesichtshälften erfolgte die Anordnung (links oder rechts) der jeweiligen Bilder per Zufall. Zudem sah die Hälfte der Versuchspersonen eine spiegelverkehrte Version des Stimulusmaterials. Das bedeutet, dass bei der einen Hälfte der Versuchspersonen der Eindruck entstand, sie hätten es mit lauter rechten Gesichtshälften zu tun und bei der anderen Hälfte der Versuchspersonen der Eindruck, sie hätten lauter linke Gesichtshälften vor sich.

Die Versuchsdurchführung war identisch mit den Attraktivitäts-Ratings der Untersuchungen *Sym 2*, *Sym 4* und *Sym 6*, jedoch mit dem Unterschied, dass die Versuchspersonen nicht die Attraktivität, sondern die *Ähnlichkeit* der jeweiligen Gesichtshälften auf einer siebenstufigen Ratingskala bewerten sollten.

Die Frage an die Versuchspersonen lautete: „Wie ähnlich sind die Gesichtshälften?“ Die Stufen der Ratingskala lauteten: 1 = sehr unähnlich, 2 = ziemlich unähnlich, 3 = eher unähnlich, 4 = eher ähnlich, 5 = ziemlich ähnlich, 6 = sehr ähnlich, 7 = völlig gleich. Während des Ratings der Gesichter hatten die Versuchspersonen stets einen Zettel mit der Ratingskala vor sich, um sich die Benennung der Skalenstufen nicht merken zu müssen und sich dadurch ganz auf die präsentierten Gesichtshälften konzentrieren zu können. Ansonsten galten dieselben Randbedingungen wie in den zuvor durchgeführten Symmetrie-Untersuchungen *Sym 2*, *Sym 4* und *Sym 6*. Die Beurteilerübereinstimmung der Ähnlichkeitsbewertung war mit einem Cronbach's α von .92 (bei 64 Versuchspersonen) sehr hoch.

4.11.3 Ergebnisse

Um zu überprüfen, ob es einen Zusammenhang zwischen der Attraktivität und der Symmetriebewertung (ermittelt durch Ähnlichkeits-Ratings der Gesichtshälften) von Gesichtern gibt, wurde für jedes Gesicht ein mittlerer Ähnlichkeits-Rating-Wert berechnet. Diese Werte wurden als Symmetrie-Indizes der jeweiligen Gesichter verwendet und ins Verhältnis zu den bereits bekannten Attraktivitätsmittelwerten derselben Gesichter gesetzt.

Das Gesicht mit dem niedrigsten mittleren Ähnlichkeits-Rating (= Symmetrie-Index) hatte den Wert 3.02 (entspricht „eher unähnlich“), das mit dem höchsten 5.47 (entspricht „ziem-

lich ähnlich“). Der durchschnittliche Symmetriewert über alle Gesichter beträgt 4.37 ($SD = 0.54$). Die genauen Werte finden sich im Anhang E (S. 399)

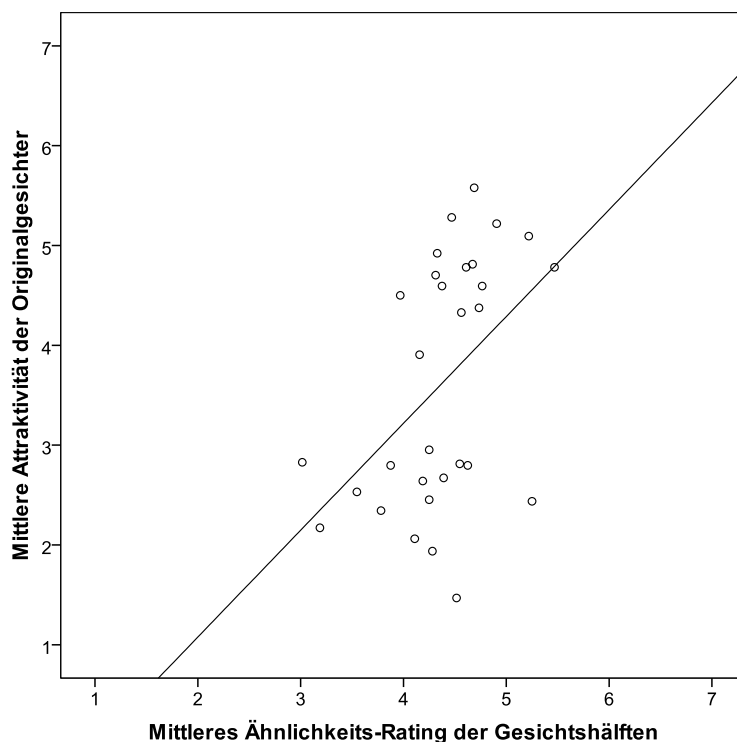


Abbildung 52: Die Korrelation zwischen dem mittleren Ähnlichkeits-Rating der Gesichtshälften als Maß für Symmetrie und der Attraktivität der Originalgesichter. Der Zusammenhang ist positiv und mittelstark ausgeprägt ($r = .47$, $p = .009$). Je ähnlicher die Gesichtshälften eines Originalgesichts (und damit je symmetrischer das Originalgesicht), desto attraktiver wird dieses Originalgesicht (von einer anderen Stichprobe von Versuchspersonen) bewertet.

Wie das Streudiagramm illustriert, besteht ein mittelstarker positiver Zusammenhang zwischen dem Symmetriemaß (ermittelt durch Ähnlichkeitsbewertungen der Gesichtshälften) und den mittleren Attraktivitätswerten der Originalgesichter. Die Korrelation beträgt $r = .47$ ($p = .009$). Je ähnlicher die Gesichtshälften eines Originalgesichts (und damit je symmetrischer das Originalgesicht), desto attraktiver wird dieses Originalgesicht (von einer anderen Stichprobe von Versuchspersonen) bewertet.

Bei einem so hohen positiven Zusammenhang, der ja höher ausfällt als bei fast allen experimentellen Verfahren der Experimente *Sym 1* bis *Sym 6* (mit Ausnahme der Effekt-

größe der Paarvergleiche mit „Blends“; Auswertungen jeweils mit Berücksichtigung der Urteilsvarianz) sollte man jedoch kritisch hinterfragen, inwiefern nicht auch hier ein Erwartungs-Effekt der Versuchspersonen gemessen wurde. Gewiss ist die Aufgabenstellung der Ähnlichkeitsbewertung von Gesichtshälften bei weitem nicht so reaktiv wie die Aufgabenstellung des direkten Symmetrie-Ratings in Untersuchung *Sym 7*. Dennoch kann man auch hier nicht von vornherein ausschließen, dass (zumindest einige) Versuchspersonen während der Bewertungen implizit die Symmetrie-Hypothese im Kopf hatten. In dem Moment, in dem eine Versuchsperson der Meinung ist, dass es gut ist, wenn zwei Gesichtshälften gleich sind, besteht bereits die Möglichkeit, dass die Daten durch das Attraktivitätsstereotyp verfälscht werden. Zwar ist es schwieriger, aus zwei einzelnen Gesichtshälften (von denen die eine ja gespiegelt präsentiert wurde) die Attraktivität des gesamten Gesichts zu beurteilen, aber es ist dennoch möglich. Wenn also beim Betrachten von zwei Gesichtshälften bei einer Versuchsperson der Eindruck entsteht, dass dieses Gesicht attraktiv ist, könnte es im Sinne des Halo-Effekts sein, dass dieser Eindruck auf die Ähnlichkeitsbewertung dieser Gesichtshälften ausstrahlt.

Man könnte die Möglichkeit eines Halo-Effekts überprüfen, indem man von jedem Gesicht nur eine Hälfte nach Attraktivität (von einer neuen Stichprobe von Versuchspersonen) bewerten lässt. Dieses Attraktivitätsurteil wäre unbeeinflusst von der Symmetrie eines Gesichts, da man ja für die Beurteilung der Symmetrie beide Gesichtshälften braucht. Bei der Korrelation von Ähnlichkeitswerten der Gesichtshälften mit der Attraktivität der Originalgesichter könnte man dann mit einer Partialkorrelation die Attraktivität der einzelnen Hälften herausrechnen. Übrig bliebe dann der Netto-Effekt der Symmetrie (= Ähnlichkeit der Hälften) auf die Attraktivitätsbewertung der Originalgesichter (man müsste sich allerdings entscheiden, ob man nur linke oder nur rechte Hälften präsentiert oder beide und daraus einen mittleren Attraktivitätswert berechnet).

Durch diese Partialkorrelation lässt sich die Frage beantworten, was vom Zusammenhang zwischen Symmetrie und Attraktivität von $r = .47$ noch übrig bleibt, wenn der Einfluss aller übrigen Gesichtsmerkmale (außer Symmetrie) statistisch kontrolliert wird. Aufgrund des hohen Aufwands, den eine zusätzliche Datenerhebung zur Attraktivitätsermittlung von einzelnen Gesichtshälften bedeutet hätte, wurde ein sehr ähnliches alternatives Verfahren gewählt: Aus dem Experiment *Sym 6* lagen ja bereits Attraktivitätsbewertungen von perfekt symmetrischen Warp-Gesichtern vor. Diese Daten sind für eine statistische Kontrolle innerhalb einer Partialkorrelation genauso gut geeignet, denn da alle Warp-Gesichter gleich perfekt symmetrisch sind, können Attraktivitätsunterschiede zwischen den Gesichtern nicht durch Unterschiede in der Symmetrie bedingt sein, son-

dern sind ausschließlich auf andere Gesichtsmerkmale oder Attraktivitätsmerkmale zurückzuführen, die nichts mit Symmetrie zu tun haben. Wenn man also den Einfluss dieser sonstigen Attraktivitätsmerkmale aus der Korrelation zwischen Ähnlichkeitsbewertung der Gesichtshälften (= Symmetrie) und Attraktivität der Originalgesichter herausrechnet, dann bleibt der Netto-Effekt übrig, den die Symmetrie auf die Attraktivitätsbewertung hat.

In diesem Fall bleibt dann nur noch ein Zusammenhang von $r = .12$ übrig, der statistisch nicht signifikant ist ($p = .546$). Das bedeutet, dass ganz offensichtlich (wie bereits beim Direkt-Rating von Symmetrie in Untersuchung *Sym 7*) Einstellungen, Erwartungen oder gar konkrete eigene Hypothesen der Versuchspersonen die Messergebnisse beeinflussen. Offensichtlich existiert so etwas wie ein „Symmetrie-Stereotyp“: Wenn Menschen daran glauben, dass Symmetrie ein Gesicht attraktiv macht, dann ist die Wahrscheinlichkeit groß in empirischen Untersuchungen, die in ihrer Methodik nicht von vornherein robust gegenüber solchen Einstellungseffekten sind, auch einen solchen Effekt zu finden.

Gewiss ist die Interpretation von Partialkorrelationen nicht ganz einfach, da ja bereits einfache Korrelationen hinsichtlich Kausalitäten nicht eindeutig zu interpretieren sind. Leidenschaftliche Anhänger der Symmetrie-Hypothese dürften sich von diesen Daten wahrscheinlich wenig beeindrucken lassen. Es gäbe jedoch eine Möglichkeit, mit der sich klar beweisen lässt, dass Versuchspersonen die Ähnlichkeit von Gesichtshälften nicht unabhängig von der Attraktivität dieser Gesichtshälften beurteilen können.

Um dies nachzuweisen müsste man den Versuchspersonen wie in der Kontrolluntersuchung zur Untersuchung *Sym 7* (Direkt-Rating von symmetrischen Warp-Gesichtern) wieder eine „Falle“ stellen: Man könnte ihnen Gesichtshälften zur Beurteilung vorlegen, die völlig identisch sind (nur eine Hälfte gespiegelt). Dieses Experiment könnte man entweder mit lauter linken oder mit lauter rechten Gesichtshälften machen. Alternativ könnte man es auch mit in der Gesichtsmitte geteilten symmetrischen Warp-Gesichtern machen, was den Vorteil hätte, dass hier nur die Gesichtsproportionen identisch, die Farbwerte aber verschieden wären. Im Durchschnitt müssten dann alle Gesichter erstens sehr hohe Ähnlichkeitswerte und zweitens alle dieselben Ähnlichkeitswerte bekommen; bei den mittleren Ähnlichkeitsbewertungen dürfte es dann zwischen den Gesichtern keine Varianz geben und die Korrelation mit den Attraktivitätswerten dieser Gesichter müsste Null sein. Wahrscheinlich jedoch werden nicht alle Gesichtshälften als gleichermaßen ähnlich eingestuft werden, sondern attraktive Hälften werden als ähnlicher bezeichnet als unattraktive Hälften (obwohl sie in Wahrheit alle identisch sind). Wenn dieser Zusammenhang in der Größenordnung von $r = .47$ (die Korrelationskoeffizienten sich nicht signifikant unter-

scheiden), dann ist dies ein eindeutiger Beweis dafür, dass das Ähnlichkeitsurteil der Versuchspersonen mit der Attraktivität der zu bewertenden Gesichtshälften konfundiert und damit die gesamte Methode der Symmetrie-Bestimmung fragwürdig ist. Diesen definitiven Nachweis tatsächlich zu erbringen muss jedoch künftigen Untersuchungen vorbehalten bleiben.

Zuletzt muss noch die Verwendung von Gesichtshälften als Stimulusmaterial kritisch hinterfragt werden. Bei der Bewertung von separaten Gesichtshälften ist es zweifelhaft, inwieweit Menschen aufgrund des Umstandes, dass Gesichterwahrnehmung holistisch erfolgt, überhaupt in der Lage sind, halbe Gesichter angemessen zu erkennen. Die Aufgabe, bestimmte asymmetriebedingte Unterschiede in den Hälften zu erkennen, sollte bei separaten Gesichtshälften jedenfalls schwieriger sein, ähnlich wie bei Gesichtern, die auf dem Kopf stehend präsentiert werden (Little & Jones, 2003). Zudem ist es eine naheliegende Annahme, dass bei der Präsentation von separaten Hälften Texturunterschiede (z. B. Pickel, Muttermale) stärker in den Fokus der Aufmerksamkeit rücken, weil es durch das Trennen der Gesichtshälften (und Spiegeln einer Hälfte) schwieriger geworden ist, Unregelmäßigkeiten der Proportion zu beurteilen. Texturunterschiede der Haut (oder gar der Haare) sind jedoch gerade *nicht* gemeint, wenn Attraktivitätsforscher bzw. Biologen von (fluktuierender) Asymmetrie sprechen. Stattdessen basiert die von evolutionsbiologisch orientierten Attraktivitätsforschern vertretende Theorie, dass hohe Symmetrie aus einer starken Entwicklungsstabilität (*developmental stability*) während der Ontogenese resultiert und ein Indikator für hohe Robustheit gegen Umweltstressoren ist (Gangestad & Thornhill, 1997; Gangestad et al., 1994; Palmer & Strobeck, 1986; Thornhill & Gangestad, 1999; Thornhill & Moller, 1997) auf der Auffassung von Asymmetrien als Abweichungen von perfekt symmetrischen *Gesichtsproportionen*, die durch ein unregelmäßiges Schädelwachstum während Kindheit und Pubertät entstehen. Diese Einschränkung stellt jedoch hauptsächlich für biologisch orientierte Forscher ein methodisches Problem dar, da die Art von Symmetrie (d. h. Ähnlichkeit der Gesichtshälften), die Laien höchstwahrscheinlich beurteilen, wenn man ihnen separate Gesichtshälften zur Beurteilung vorlegt, eine ganz andere Art von Symmetrie ist, als die, für die sich Evolutionsbiologen eigentlich interessieren.

Insgesamt bleibt festzustellen, dass in dieser Untersuchung trotz eines signifikanten positiven Zusammenhangs von $r = .47$ zwischen Symmetrie und Attraktivität *kein* klarer Nachweis erbracht werden kann, dass Symmetrie einen Einfluss auf die Attraktivität eines Gesichts hat. Vielmehr gibt es aufgrund der gerechneten Partialkorrelationen starke Anzeichen dafür, dass sich die Versuchspersonen von der Attraktivität eines Gesichts, die

sie auch bei Präsentation von zwei separaten Gesichtshälften erkennen können, blenden lassen und den attraktiven Gesichtshälften auch eine höhere Ähnlichkeit zusprechen.

4.12 Korrelatives Versuchsdesign mit Versuchspersonen-Ratings der Ähnlichkeit von Chimärengesichtern als Maß für Symmetrie (Untersuchung *Sym 9*)

4.12.1 Stand der Forschung

Die in Untersuchung *Sym 9* durchgeführte Studie gehört wie die beiden vorangegangenen ebenfalls zu den nicht-experimentellen Verfahren. Die Symmetrie eines Gesichts wird also nicht manipuliert und somit entfällt auch das Risiko, Stimuli zu produzieren, die auf irgendeine Art unnatürlich aussehen, wodurch das Urteilsverhalten der Versuchspersonen beeinflusst werden könnte. Stattdessen wird auch hier wieder versucht, für alle (natürlichen, unveränderten) Gesichter jeweils einen Indexwert für Symmetrie zu erhalten, um dann diese Symmetrie-Indexwerte mit den separat ermittelten Attraktivitätswerten dieser Gesichter zu korrelieren.

Die Grundidee dieser Methode liegt darin, jeweils zwei Chimärengesichter (die aus demselben Originalgesicht erzeugt wurden) gleichzeitig zu präsentieren und Versuchspersonen um ein Ähnlichkeitsurteil zu bitten. Der Grundgedanke lautet: Je ähnlicher die Chimärengesichter bewertet werden, desto ähnlicher sind die beiden Gesichtshälften, desto symmetrischer ist das Originalgesicht. Wie auch bei der vorangegangenen Untersuchung (*Sym 8*) wird die bewertete Ähnlichkeit der Gesichtshälften als Symmetriemaß verwendet, nur mit dem Unterschied, dass hier keine separaten Gesichtshälften präsentiert werden, sondern jeweils vollständige Gesichter, was der holistischen Wahrnehmung von Gesichtern entgegenkommt.

Dieses Verfahren wurde beispielsweise von Mealey et al. (1999) verwendet, die in ihrer Untersuchung zusätzlich noch mit Fotos von eineiigen Zwillingen arbeiteten und untersuchten, ob im direkten Vergleich innerhalb eines Zwillingspaars der symmetrischere als attraktiver beurteilt wurde. Die Autoren begründeten die Wahl ihrer Methode mit der Konfundierung von Symmetrie und Attraktivität in den meisten anderen Studien, die mit anderen Methoden arbeiteten: „One problem is that most studies have been unable to disentangle the effects of FA from those of other forms of asymmetry or from other correlates of attractiveness, such as mean trait size, apparent age, texture and color of hair and

skin, or overall facial shape (Langlois et al., 1994; Perrett et al., 1994; Swaddle & Cuthill, 1995; Thornhill & Gangestad, 1996)” (Mealey et al., 1999, S. 152).

In der Metaanalyse von Rhodes (2006) wurde auf diese Methode leider nicht näher eingegangen. Bei der Analyse der Daten fasste Rhodes Studien mit dieser Methode mit solchen Studien zusammen, die Symmetrie durch direktes Befragen der Versuchspersonen im Direkt-Rating ermitteln ließen (*Sym 7*), so wie sie selbst es häufig praktiziert (Kategorie: „Rating“). Dennoch besteht selbstverständlich ein großer Unterschied zwischen dem (plumpen und suggestiven) direkten Fragen nach Symmetrie und der Quantifizierung des Ausmaßes der Symmetrie über Ähnlichkeitsbewertungen.

Mealey et al. (1999) berichten einen starken Effekt der Symmetrie auf die Attraktivitätsbewertung ($r = .67$), allerdings basieren diese Daten auf *Differenz-Scores* von Symmetrie- und Attraktivitätsbewertung der jeweiligen Zwillinge, nicht auf absoluten Bewertungen dieser Gesichter. Deshalb ist dieses Ergebnis mit anderen Studien schwer vergleichbar, da die Effektgröße durch den relativen Vergleich innerhalb der Zwillingspaare vergrößert wird.

In einer weiteren Untersuchung (Penton-Voak, Jones et al., 2001) mit Männergesichtern zeigte sich, dass die Ähnlichkeitsratings von Chimärengesichtern mittelstark mit den aus *Facial Landmarks* gemessenen Symmetrie-Indizes korrelierten (Spearman's $r_s = .48$), jedoch nur schwach und nicht signifikant mit den Attraktivitätsbewertungen der Originalgesichter ($r_s = .22$).

Ähnlichkeitsbewertungen von Chimärengesichtern benutzten auch Hönekopp et al. (2004) in ihrer methodisch sehr präzisen und gewissenhaften Untersuchung. Als Besonderheit machten sie von jeder Stimulusperson zwei Aufnahmen und konstruierten aus jedem dieser Fotos jeweils zwei Chimärengesichter. Durch den Vergleich dieser verschiedenen Chimärengesichter-Paare kontrollierten sie gleichzeitig die Reliabilität dieses Verfahrens und erhielten vernichtend niedrige Werte: „Raters agreed about the similarity of the pairs (average Cronbach's $\alpha = .86$). But when we computed the average similarity rating for each of the two portraits of the ten women, the correlation between the first and the second pictures' similarity ratings was $r_{10} = -.12$ (Hönekopp et al., 2004, S. 157). Die Autoren führen die schlechten Ergebnisse auf minimale Abweichungen von einer perfekt geraden Kopfhaltung bei der Aufnahme des Fotos aufgrund von geringfügigen Kopffrotationen zurück, die dazu führen, dass die Gesichter nicht perfekt frontal abgelichtet sind. Bei der Konstruktion von Chimärengesichtern führt dies dann zu Bildverarbeitungsartefakten. Das gleiche gilt jedoch auch für das Vermessen von Gesichtern über *Facial*

Landmarks, die nach dieser Studie ebenfalls an dem Problem krankten, dass Effekte der Symmetrie (sofern sie überhaupt nachgewiesen werden können) sich lediglich auf das Foto von einer Person beziehen und noch nicht einmal auf dieselbe Person (in Realität und 3D) generalisieren lassen.

4.12.2 Methode

4.12.2.1 Versuchspersonen

An der Untersuchung *Sym 9* nahmen insgesamt 128 Versuchspersonen teil. In der Gruppe, die die Chimärengesichter hinsichtlich Ähnlichkeit bewertete, waren 64 Personen. Zusätzlich wurden für die Datenauswertung die Daten der 64 Versuchspersonen aus dem Experiment *Sym 2* verwendet, die alle Originalgesichter nach Attraktivität bewertet hatten.

Von den 64 Personen der Gruppe mit den Ähnlichkeitsbewertungen waren 28 männlich (43.8 %) und 36 weiblich (56.3 %). 7 Personen (10.9 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 14 Jahre, die älteste 65. Der Altersdurchschnitt betrug 31.33 Jahre ($SD = 11.51$).

Hinsichtlich der erhobenen demografischen Variablen gab es bei der Ähnlichkeitsbewertung einen geringeren Anteil an Studierenden (10.9 %) als bei der Attraktivitäts-Bewertungsgruppe (26.6 %) ($\chi^2(1, N = 128) = 5.128, p = .024$), hinsichtlich Geschlecht ($\chi^2(1, N = 128) = 0.783, p = .376$) und Alter ($t(126) = 0.496, p = .621$) gab es keine Unterschiede.

4.12.2.2 Stimulusmaterial und Versuchsdurchführung

Als Grundlage für das Stimulusmaterial dienten die 30 Originalgesichter, die auch in allen anderen Symmetrie-Untersuchungen verwendet wurden. Das Untersuchungsdesign war analog zur Untersuchung *Sym 8*, jedoch mit dem Unterschied, dass nicht die Ähnlichkeit von Gesichtshälften, sondern die Ähnlichkeit von Chimärengesichtern bewertet wurde. Auch hier wird der empirisch über Versuchspersonenbefragung ermittelte Ähnlichkeitswert als Symmetrie-Index verwendet. Analog zur vorangegangenen Untersuchung (*Sym 8*) lautet auch hier der zu Grunde liegende Gedanke: Je ähnlicher sich die jeweiligen Chimärengesichter sind, desto symmetrischer ist das Gesicht.

Der Unterschied zur vorangegangenen Untersuchung, bei der separate Gesichtshälften präsentiert wurden, ist, dass hier ausschließlich ganze Gesichter präsentiert werden. Da

Gesichter holistisch wahrgenommen werden, kommt diese Versuchsanordnung der menschlichen Wahrnehmung entgegen und ist damit weniger artifiziell. Andererseits hat sie den Nachteil, dass der Versuchsperson bewusst wird, dass hier etwas an den Gesichtern manipuliert wurde. Zudem spielen auch hier wieder Bildverarbeitungsartefakte eine Rolle, die bei Chimärengesichtern unvermeidlich sind (siehe dazu die Diskussion in Kapitel 5.2.1.1) und die bei der Beurteilung der Ähnlichkeit einen Störfaktor darstellen könnten.



Abbildung 53: Beispiel für das verwendete Stimulusmaterial. Links ein Chimärengesicht aus zwei linken Gesichtshälften, rechts ein Chimärengesicht aus zwei rechten Gesichtshälften. Die Frage, die den Probanden gestellt wurde lautete: „Wie ähnlich sind die ‚Zwillinge‘?“

Alle Fotos mit Chimärengesichtern hatten eine hohe Auflösung von 1600 x 1600 Pixel. Für die Versuchsdurchführung wurden sie von einem Fotodienst auf Fotopapier in einer Größe von 9 x 9 Zentimeter belichtet. Daraus ergibt sich eine Auflösung von 452 dpi. Bei der paarweisen Präsentation der Chimärengesichter erfolgte die Anordnung (links oder rechts) der jeweiligen Bilder per Zufall.

Die Versuchsdurchführung war identisch mit den Ähnlichkeits-Ratings der Untersuchung *Sym 8*, nur mit dem Unterschied, dass die Versuchspersonen nicht die Ähnlichkeit von separaten Gesichtshälften, sondern von Chimärengesichtern auf einer siebenstufigen Ratingskala bewerten sollten.

Bei der Instruktion wurde den Versuchspersonen gesagt, sie sollten sich vorstellen, es handele sich bei den gezeigten Gesichtern um eineiige Zwillinge. Auch eineiige Zwillinge sähen ja nicht völlig identisch aus, sondern würden sich in kleinen Details unterscheiden.

Die Frage an die Versuchspersonen lautete daher: „Wie ähnlich sind die ‚Zwillinge‘?“ Die Stufen der Ratingskala waren (analog zur Untersuchung *Sym 8*): 1 = sehr unähnlich, 2 = ziemlich unähnlich, 3 = eher unähnlich, 4 = eher ähnlich, 5 = ziemlich ähnlich, 6 = sehr ähnlich, 7 = völlig gleich. Während des Ratings der Gesichter hatten die Versuchspersonen stets einen Zettel mit der Ratingskala vor sich, um sich die Benennung der Skalenstufen nicht merken zu müssen und sich dadurch ganz auf die präsentierten Chimärengesichter konzentrieren zu können. Ansonsten galten dieselben Randbedingungen wie in den zuvor durchgeführten Symmetrie-Untersuchungen *Sym 2*, *Sym 4*, *Sym 6*, *Sym 7* und *Sym 8*. Die Beurteilerübereinstimmung bei der Ähnlichkeitsbewertung war mit einem Cronbach's α von .96 (bei 64 Versuchspersonen) sehr hoch.

4.12.3 Ergebnisse

Um zu überprüfen, ob es einen Zusammenhang zwischen der Attraktivität und der Symmetriebewertung (ermittelt durch Ähnlichkeits-Ratings der Chimärengesichter) von Gesichtern gibt, wurde für jedes Gesicht ein mittlerer Ähnlichkeits-Rating-Wert berechnet. Diese Werte wurden als Symmetrie-Indizes der jeweiligen Gesichter verwendet und ins Verhältnis zu den bereits bekannten Attraktivitätsmittelwerten derselben Gesichter gesetzt.

Das Gesicht mit dem niedrigsten mittleren Ähnlichkeits-Rating (= Symmetrie-Index) hatte den Wert 3.81 (entspricht „eher ähnlich“), das mit dem höchsten 5.72 (entspricht „sehr ähnlich“). Der durchschnittliche Symmetriewert über alle Gesichter beträgt 4.99 ($SD = 0.56$). Die genauen Werte finden sich im Anhang F (S. 400).

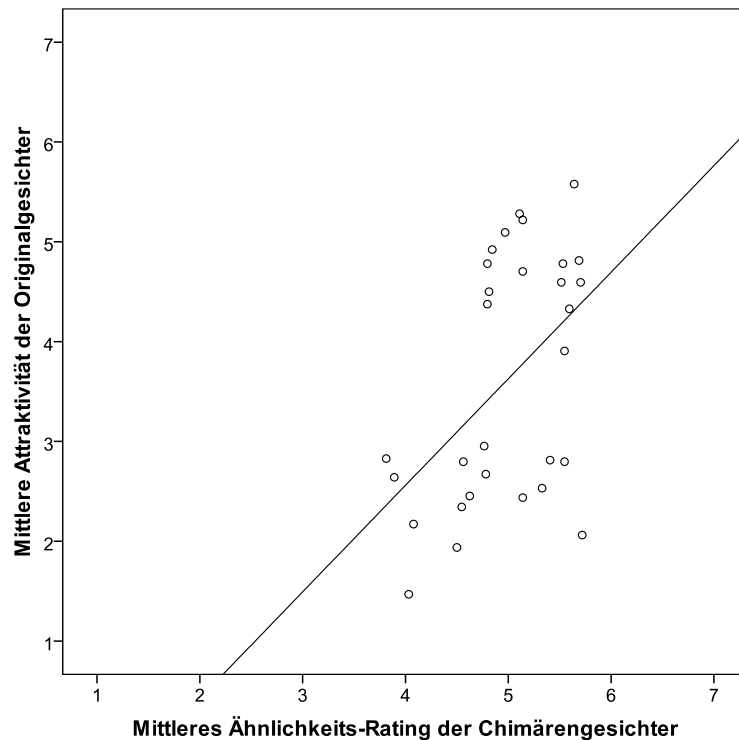


Abbildung 54: Die Korrelation zwischen dem mittleren Ähnlichkeits-Rating der Chimärengesichter als Maß für Symmetrie und der Attraktivität der Originalgesichter. Der Zusammenhang ist positiv und mittelstark ausgeprägt ($r = .48$, $p = .007$). Je ähnlicher die Chimärengesichter (d. h. je ähnlicher die Gesichtshälften und damit je symmetrischer das Originalgesicht), desto attraktiver wird ein Originalgesicht (von einer anderen Stichprobe von Versuchspersonen) bewertet.

Wie das Streudiagramm illustriert, besteht ein mittelstarker positiver Zusammenhang zwischen dem Symmetriemaß (ermittelt durch Ähnlichkeitsbewertungen der Chimärengesichter) und den mittleren Attraktivitätswerten der Originalgesichter. Die Korrelation beträgt $r = .48$ ($p = .007$). Je ähnlicher die Chimärengesichter (d. h. je ähnlicher die Gesichtshälften und damit je symmetrischer das Originalgesicht), desto attraktiver wird ein Originalgesicht (von einer anderen Stichprobe von Versuchspersonen) bewertet.

Der gefundene Zusammenhang zwischen Symmetrie-Maß und Attraktivität ist damit genau gleich groß bei der Untersuchung *Sym 8* unter Verwendung von Gesichtshälften. Der Zusammenhang ist auch größer als der in der Studie von Penton-Voak et al., die lediglich einen schwachen Zusammenhang von $r_s = .22$ fanden (Penton-Voak, Jones et al., 2001). Mit der Studie von Mealey et al. (1999), die einen stärkeren Effekt der Symmetrie auf die

Attraktivitätsbewertung ($r = .67$) berichtete, ist das Ergebnis allerdings kaum vergleichbar, da diese Daten auf *Differenz-Scores* von Symmetrie- und Attraktivitätsbewertung von „echten“ eineiigen Zwillinge basieren, nicht auf absoluten Bewertungen dieser Gesichter.

Wie auch bei der vorangegangenen Untersuchung (*Sym 8*) stellt sich auch hier die Frage, inwieweit das Ergebnis von einer Erwartungshaltung der Versuchspersonen beeinflusst sein könnte. Beurteilen Probanden zwei Chimärengesichter als ähnlicher, wenn diese attraktiv sind? Theoretisch könnte ein solcher Effekt geringer ausfallen als bei der Beurteilung von Gesichtshälften, da den Probanden in der Instruktion gesagt wurde, sie sollten die Ähnlichkeit von eineiigen Zwillingen beurteilen. Durch die Art der Instruktion wird stärker vom eigentlichen Untersuchungsziel (Symmetrie und Attraktivität) abgelenkt.

Aus diesem Grund wurde auch in dieser Untersuchung dieselbe Analyse gerechnet wie bei der Untersuchung *Sym 8*. Mit einer Partialkorrelation lässt sich die Frage beantworten, was vom Zusammenhang zwischen Symmetrie und Attraktivität von $r = .47$ noch übrig bleibt, wenn der Einfluss aller übrigen Gesichtsmerkmale (außer Symmetrie) statistisch kontrolliert wird. Wie in der vorherigen Untersuchung (*Sym 8*) wurden auch hier wieder die Attraktivitätsbewertungen von perfekt symmetrischen Warp-Gesichtern als Kontrollvariable verwendet. Da alle Warp-Gesichter gleich perfekt symmetrisch sind, können Attraktivitätsunterschiede zwischen den Gesichtern nicht durch Unterschiede in der Symmetrie bedingt sein, sondern sind ausschließlich auf andere Gesichtsmerkmale oder Attraktivitätsmerkmale zurückzuführen, die nichts mit Symmetrie zu tun haben. Wenn man also den Einfluss dieser sonstigen Attraktivitätsmerkmale aus der Korrelation zwischen Ähnlichkeitsbewertung der Gesichtshälften (= Symmetrie) und Attraktivität der Originalgesichter herausrechnet, dann bleibt der Netto-Effekt übrig, den die Symmetrie auf die Attraktivitätsbewertung hat (für genauere Erläuterungen siehe Kapitel 4.11.3).

In diesem Fall bleibt (statt $r = .48$) dann noch ein Zusammenhang von $r = .35$ übrig, der statistisch knapp *nicht* signifikant ist ($p = .064$). Der reduzierte Zusammenhang bedeutet, dass ganz offensichtlich (wie bereits bei der Ähnlichkeitsbeurteilung von Gesichtshälften) Einstellungen, Erwartungen oder gar konkrete eigene Hypothesen der Versuchspersonen die Messergebnisse beeinflussen. Allerdings ist das Ausmaß der Beeinflussung weniger stark als bei der Ähnlichkeitsbeurteilung von Gesichtshälften, wo bei gleich großer ursprünglicher Korrelation bei der Partialkorrelation nur noch ein r von $.12$ übrig blieb, das weit von Signifikanz entfernt war.

Unabhängig von Einstellungseffekten von Versuchspersonen, die auf eine zumindest leichte Konfundierung zwischen Symmetrie-Messung (= UV) und Attraktivität (= AV) hin-

weisen, sind die besonderen Probleme des verwendeten Stimulusmaterials im Auge zu behalten. Bereits in Kapitel 4.2 wurde auf die grundsätzliche Problematik von Artefakten bei der Verwendung von Chimären Gesichtern hingewiesen. Zwar werden die Chimären gesichter hier nicht hinsichtlich ihrer Attraktivität bewertet, sondern nur nach ihrer Ähnlichkeit. Dennoch ist es als kritisch anzusehen, dass durch Bildverarbeitungsartefakte Chimären gesichter entstehen können, die recht unterschiedlich aussehen (was zu einem niedrigen gemessenen Symmetrie-Wert führt), obwohl das Gesicht in Wirklichkeit gar nicht besonders asymmetrisch ist, sondern z. B. beim Fotografieren mit leicht gedrehter Kopfhaltung aufgenommen wurde. Bildverarbeitungsartefakte bei Chimären gesichtern beeinträchtigen generell die Validität der Ähnlichkeitsbewertungen und damit der Quantifizierung des Ausmaßes von Symmetrie bzw. Asymmetrie. Für eine detaillierte Diskussion der methodischen Probleme von Chimären gesichtern sei an dieser Stelle jedoch auf Kapitel 5.2.1.1 verwiesen.

5 Zusammenfassende Diskussion der Ergebnisse aller neun Untersuchungen zum Einfluss der Symmetrie

5.1 Tabellarischer Überblick über die Ergebnisse aller Symmetrie-Untersuchungen

Die Frage, die allen hier durchgeführten Symmetrie-Untersuchungen zu Grunde lag, lautete: „Macht Symmetrie Gesichter attraktiv, und wenn ja, wie groß ist der Effekt?“ Simpler kann eine Fragestellung kaum formuliert sein – und verwirrender kann eine Antwort kaum ausfallen. Neun verschiedene Untersuchungen wurden durchgeführt und 19 verschiedene Ergebnisse gefunden. Von stark negativen Effekten der Symmetrie über Nulleffekte bis hin zu stark positiven Effekten war jedes denkbare Ergebnis dabei. Tabelle 15 fasst sämtliche Ergebnisse in Form von Effektgrößen zusammen.

Tabelle 15: Überblick über die Effektgrößen (standardisiert nach r) aller neun Untersuchungen zur Symmetrie-Hypothese in Abhängigkeit von der Berücksichtigung der Varianz von Versuchspersonen-Urteilen bei der Datenanalyse

	Effektgröße r mit Vpn-Urteils- Varianz	Effektgröße r ohne Vpn- Urteils-Varianz
Experiment 1a, Paarvergleich (Chimärengesichter aus linken Hälften)	-.27	-.74
Experiment 1b, Paarvergleich (Chimärengesichter aus rechten Hälften)	.00	-.08
Experiment 2a, Rating (Chimärengesichter aus linken Hälften)	-.20	-.73
Experiment 2b, Rating (Chimärengesichter aus rechten Hälften)	-.18	-.67
Experiment 3, Paarvergleich („Blends“)	.53	.80
Experiment 4, Rating („Blends“)	-.08	-.53
Experiment 5, Paarvergleich („Warps“)	.17	.55
Experiment 6, Rating („Warps“)	-.09	-.57
Untersuchung 7, Direktes Symmetrie-Rating	–	.78
Untersuchung 8, Ähnlichkeits-Rating von Gesichtshälften als Symmetrie-Maß	–	.47
Untersuchung 9, Ähnlichkeits-Rating von Chimärengesichtern als Symmetrie-Maß	–	.48

Für jemanden, der einfach nur wissen möchte, ob nun Symmetrie attraktiv macht oder nicht (z. B. vor dem Hintergrund eines ästhetischen Eingriffs) ist dieses Durcheinander an Effekten in verschiedene Richtungen ein Desaster. Denn es ist geeignet, fundamentale

Zweifel an der Wissenschaftlichkeit sämtlicher bisher durchgeführter Untersuchungen zu diesem Thema zu wecken. Dass Ergebnisse *auch* von der verwendeten Untersuchungsmethode abhängen, ist für jeden wissenschaftlich gebildeten Menschen (egal in welcher Disziplin) keine Überraschung. Hier jedoch scheinen die Ergebnisse und sogar die Richtung des Effekts in extrem starkem Ausmaß von der gewählten Untersuchungsmethode abzuhängen. Sind also alle Ergebnisse nur Methodenartefakte? Oder sind nur manche Ergebnisse Methodenartefakte, andere jedoch verlässlich und aussagekräftig? Welche Methoden sind gut und welche schlecht? Im Folgenden sollen diese Fragen detailliert diskutiert werden. Vorab sei jedoch noch einmal betont, dass hier selbstverständlich keine per se unsinnigen Methoden angewandt wurden. Jede Methode lässt sich inhaltlich begründen und für jede Methode – egal ob es sich um Stimulusmaterial, Versuchsdesign, Symmetrie-Messung oder statistische Datenauswertung handelt – gibt es (teilweise sehr zahlreiche) Vorbilder in der Forschungsliteratur. Sie sind in renommierten Zeitschriften veröffentlicht, von angesehenen Attraktivitätsforschern durchgeführt und werden nach wie vor gerade auch von sehr einflussreichen Attraktivitätsforschern angewandt und gegen Kritik verteidigt.

Die heterogene Befundlage ist zudem auch nicht neu, wie man sich von einem Blick in den Anhang der Metaanalyse von Rhodes (2006) leicht überzeugen kann. Was jedoch die Besonderheit und den Wert der hier durchgeführten Untersuchungen ausmacht, ist der Umstand, dass diesmal die unterschiedlichen Ergebnisse eindeutig auf die verwendeten Methoden zurückzuführen sind. Vergleicht man hingegen ganz verschiedene Studien, dann unterscheiden diese sich selbstverständlich hinsichtlich vieler Variablen, z. B. Geschlecht, Alter und ethnische Gruppe von Gesichtern sowie Beurteilern, ob Stimulusgesichter dasselbe Geschlecht hatten wie die Beurteiler oder das entgegengesetzte, in welcher Kultur die Untersuchung durchgeführt wurde, wie groß die Stichproben Stimulusgesichter bzw. Beurteilerpersonen waren, die Qualität der verwendeten Fotos oder die Attraktivität der Stimuluspersonen.

Bei den hier durchgeführten Untersuchungen hingegen wurden all diese Parameter konstant gehalten. Es wurden immer dieselben Stimuluspersonen als Ausgangsmaterial für die Stimulusbilder verwendet, und die Versuchspersonenstichproben sind in Zusammensetzung und Größe gleich. Unterschiedlich wirkende Störfaktoren können daher als Erklärung für die widersprüchlichen Ergebnisse nicht mehr herhalten. Die Ursache muss also in den Untersuchungsmethoden selbst liegen. Dies ist das bedeutendste Ergebnis und der eigentliche Erkenntnisgewinn der hier durchgeführten Untersuchungen. In ihrer Gesamtheit sind sie weit mehr als neun Einzeluntersuchungen, die der Frage nachgehen,

ob Symmetrie Gesichter attraktiv macht oder nicht. Es handelt sich um eine Methodenstudie, deren Wert darin besteht, dass durch ihre Ergebnisse erstmals bisher durchgeführte Studien zum Thema Symmetrie und Attraktivität einschließlich deren Metaanalysen angemessen beurteilt werden können. Die Besonderheiten, Schwächen und „Tricks“ der einzelnen Methoden zu analysieren und zu verstehen, ist eine unabdingbare Voraussetzung für künftige Untersuchungen mit ähnlicher Fragestellung. Nur dadurch können in Zukunft fragwürdige methodische Praktiken unterbleiben und Fehler vermieden werden, so dass man tatsächlich reliable und valide Daten produziert, deren Ergebnisse für Theorie und Praxis auch belastbar und aussagekräftig sind.

Wie im Folgenden aufgezeigt wird, sind diese Ergebnisse überaus lehrreich und ihre Implikationen reichen über den hier untersuchten Kontext von Symmetrie und Attraktivität weit hinaus, weil es sich um methodische Aspekte oder Probleme handelt, die in ähnlicher Form in vielen psychologischen oder sozialwissenschaftlichen Fragestellungen auftreten. In mancher Hinsicht sind sie sogar spannender als die ursprüngliche Frage nach der Relevanz von Symmetrie für die Attraktivität eines Gesichts.

5.2 Methodische Reflexion der Symmetrie-Untersuchungen

5.2.1 Effekte des Stimulusmaterials: Chimärengesichter, Morphing-Artefakte und Originalgesichter

Wie in so vielen Untersuchungen im Bereich der Attraktivitätsforschung zeigte sich auch hier, dass die Art des Stimulusmaterials einen ganz erheblichen Einfluss auf die erzielten Ergebnisse hat. Über manche Einflüsse kann auch mit dieser Untersuchungsreihe keine Aussage getroffen werden, da diese Merkmale hier nicht verändert wurden, z. B. welche Rolle das Geschlecht, das Alter und die ethnische Gruppe der Stimulusgesichter spielt. Zu folgenden Merkmalen lassen sich jedoch in der Zusammenschau klare Regeln ableiten.

5.2.1.1 Effekte von Chimärengesichtern

Bei den Experimenten mit den Chimärengesichtern als perfekt symmetrische Alternativen zu den Originalgesichtern zeigte sich insgesamt ein negativer Einfluss der Symmetrie – oder besser gesagt – der Präsentation der Chimärengesichter. Sie wurden meist negativer beurteilt. Angesichts der manchmal seltsam und starr wirkenden Gesichter ist dieses

Ergebnis nicht sonderlich überraschend. Bildverarbeitungsartefakte sind bei dieser Methode unvermeidlich, zumindest bei manchen Gesichtern. Die Bildverarbeitungsartefakte einiger weniger Gesichter können jedoch das Gesamtergebnis eines Experiments bereits in eine bestimmte Richtung verzerren und dadurch einen schwachen negativen Effekt (hier zwischen $-.18$ und $-.27$) verursachen. Da es bei Publikationen nicht üblich ist, das gesamte Stimulusmaterial offen zu legen, ist es dem Leser nicht möglich, sich selbst ein Urteil von der Qualität des Stimulusmaterials zu bilden. Wenn überhaupt, dann muss er sich mit einem vom Autor ausgewählten Beispiel begnügen. Dass ein Autor kaum mit unvoreilhaften und abschreckenden Beispielen für das verwendete Stimulusmaterial einen Artikel illustriert, sondern dafür lieber die Positivbeispiele herauspickt, liegt auf der Hand.

Die Bildverarbeitungsartefakte bei Chimärengesichtern lassen sich grob in vier Kategorien unterteilen

1. Artefakte durch unnatürliche symmetrische Verdoppelungen der Textur (z. B. Muttermale, Scheitel, Locken)
2. Artefakte, die aus einem gedrehten Gesicht bei der Aufnahme des Fotos resultieren und zu vermeintlich ungleich breiten Gesichtshälften führen
3. Artefakte durch einen geneigten Hals bei der Aufnahme des Fotos, die Chimären mit zu dickem oder zu dünnem Hals produzieren
4. Artefakte durch tatsächliche Asymmetrien im Originalgesicht um die Symmetrieachse herum (z. B. schiefe Nase oder seitlich verschobener Mund)

Die folgenden drei Abbildungen zeigen, welche typischen Artefakte bei Chimärengesichtern entstehen.



Abbildung 55: Beispiel für Morphing-Artefakte, die aus einem gedrehten Gesicht bei der Aufnahme des Fotos resultieren (Bild links). Die linke Gesichtshälfte ist dadurch auf dem Foto kleiner abgebildet, obwohl sie es in Wirklichkeit nicht ist. Bei Verdopplung dieser Hälfte führt dies zu einem unnatürlich schmal wirkenden Gesicht. Bei der rechten Gesichtshälfte ist der Effekt umgekehrt. Unnatürlich wirken auch die symmetrischen Muttermale an der Oberlippe.



Abbildung 56: Ein eigentlich sehr symmetrisches Originalgesicht (Bild links), bei dem der Hals zum Aufnahmezeitpunkt jedoch in einer leicht schiefen Position war. Dies führt zu Chimärengesichtern mit einem entweder unnatürlich dünnen Hals oder einem breiten Hals, der die gesamte Person übergewichtig wirken lässt.



Abbildung 57: Das fotografierte Originalgesicht (Bild links) weist tatsächlich eine Asymmetrie auf. Der Mund ist zur rechten Gesichtshälfte hin verschoben. Dies macht es unmöglich, das Bild beim Setzen der Symmetrieachse fehlerfrei zu teilen. Ginge die Achse exakt durch die Nase, würde der Mund nicht mittig geteilt und es entstünde ein seltsames Artefakt an der Oberlippe. Lässt man hingegen die Symmetrieachse durch die Mitte der Oberlippe gehen, wird der Nasenrücken nicht mittig geteilt und ist bei den Chimärengesichtern einmal zu dünn (Bild Mitte) und einmal zu dick (Bild rechts). Zusätzlich entsteht beim rechten Bild ein Artefakt aufgrund der dunklen Linie auf dem Nasenrücken und den zwei Glanzpunkten auf der Nasenspitze.

Meist werden nur die unter Punkt (1) genannten Artefakte in Form von Textur-Verdopplungen kritisiert. Sie sind auch unmittelbar einleuchtend und bedürfen keiner weiteren Kommentierung. Weitaus negativer wirken sich jedoch solche unter Punkt (2) genannte Artefakte aus, die daher kommen, weil ein Gesicht nicht perfekt frontal aufgenommen wurde (Abbildung 47 ist ebenfalls ein solches Beispiel). In der Praxis ist es jedoch extrem schwierig, ein Gesicht absolut perfekt frontal zu fotografieren. Schon bei kleinen Abweichungen von der perfekt frontalen Ansicht kommt es daher zu Artefakten bei Chimärengesichtern. Bei der vorliegenden Untersuchungsreihe wurden alle Personen von einem professionellen Fotografen in einem Fotostudio aufgenommen. Dennoch treten solche Unregelmäßigkeiten auf. Unter weniger professionellen Bedingungen wie z. B. mit einer ungleichmäßigen Ausleuchtung des Gesichts werden solche kleinen Fehler noch zusätzlich verstärkt.

Die Problematik einer nicht perfekten frontalen Standardisierung wurde bereits von Hönekopp et al. (2004) thematisiert, der in seiner Untersuchung von jeder Stimulusperson sogar zwei Fotos anfertigte, aus denen er jeweils zwei Chimärengesichter erzeugte. Durch winzige Unterschiede bei der Kopfhaltung während der Aufnahme des Fotos resultierten unterschiedliche Chimärengesichter. Hönekopp und Kollegen nutzten dieses Ma-

terial um die Reliabilität der Chimärengesichter-Methode zu überprüfen und stellen er-
nüchtert fest:

Much to our surprise, neither the measurement of horizontal facial asymmetry nor the asymmetry measurement via chimeric faces provided results that were reliable across different pictures of the same person. This probably resulted from head turning, although we had taken great care to control head position. While there is an easy remedy for head tilting (namely rotating the pictures accordingly) there is no counter measure for head turning after the picture was taken. (Hönekopp et al, 2004, S. 161)

Diese methodischen Probleme haben massive Konsequenzen auf die Validität des Symmetrie-Konzepts. Denn geforscht wird schließlich nicht mit den Gesichtern selbst, sondern mit zweidimensionalen Abbildungen von Gesichtern. Treten jedoch Fehler bei der Reduktion der dreidimensionalen Information auf eine zweidimensionale Ebene auf, dann arbeitet man mit scheinbar asymmetrischen Gesichtern, obwohl diese in Wirklichkeit möglicherweise hoch symmetrisch sind. Als Gedankenexperiment ließe sich eine solche Untersuchung mit perfekt symmetrischen Köpfen von Schaufensterpuppen vorstellen, bei denen eine perfekt standardisierte Aufnahme ebenfalls nicht gelingt. Die methodischen Probleme beschränken sich jedoch nicht nur auf Chimärengesichter: Auch wenn man über *Facial Landmarks* Asymmetrie messen und damit quantifizieren will, kommt man über solche Bilder zu falschen Ergebnissen.

Die unter Punkt (3) genannten Artefakte, die durch einen bei der Aufnahme geneigten Hals entstehen, wurden in der Forschungsliteratur bislang ebenfalls nicht thematisiert. Ist ein Hals bei einem Chimärengesicht zu dünn, führt dies zu Abwertungen beim Attraktivitätsurteil, weil ein unnatürlicher Gesamteindruck entsteht. Ist der Hals zu dick, entsteht leicht der Eindruck von Übergewicht (vgl. Abbildung 56, S. 209). (Bei Männergesichtern ist aufgrund des sexuellen Dimorphismus mit etwas anderen Effekten zu rechnen: Ein breiter Hals / Nacken könnte hier männlicher und damit attraktiver wirken – dafür sollte ein zu dünner Hals umso negativer wirken.) Somit beeinflusst der Hals, also ein Körperteil, der eigentlich gar nicht mehr zum Gesicht gehört, stark die Attraktivitätsbewertung des Gesichts und beeinträchtigt dadurch die Validität der Ergebnisse.

Punkt (4), Artefakte durch tatsächliche Asymmetrien im Originalgesicht um die Symmetrieachse herum, sind das am schwierigsten zu verstehende Problem bei Chimärengesichtern, da sie doch scheinbar genau das sichtbar machen, worum es geht, nämlich die Asymmetrie. Das Problem ist hier jedoch, dass sich bei Asymmetrien, die in der Gesichtsmitte liegen, eine Symmetrieachse nicht präzise bestimmen lässt – Abbildung 57 (S. 210) illustriert dies. Das gezeigte Gesicht ist gerade fotografiert, die Augen liegen auf

gleicher Höhe. Würde man jedoch die Achse so setzen, dass sie durch die Mitte der Nase verlief, dann würde die Oberlippe neben der Mitte des Philtrums geteilt und es entstünde ein Chimären Gesicht (aus zwei rechten Hälften) mit einem künstlichen, zusätzlichen Maximum zwischen den *crista philtri*. Nimmt man hingegen die Mitte des Philtrums als Referenz für die Symmetrieachse, entstehen Artefakte am Nasenrücken (vgl. Abbildung 57, S. 210, mittleres und rechtes Bild). Egal, wie man es macht, man produziert bei solchen Ausgangsvoraussetzungen immer Chimären Gesichter mit sehr störenden Artefakten. Ähnliche Probleme entstehen, wenn ein Originalgesicht eine schiefe Nase besitzt. Die Nasenspitze liegt dann neben der Symmetrieachse und das Resultat sind dann Chimären Gesichter mit einer grotesk breiten Nasenspitze („Knollennase“) bzw. ohne echte Nasenspitze.

Auch wenn all diese genannten Artefakte keineswegs bei allen Chimären Gesichtern auftreten, so sind sie dennoch stets bei einigen vorhanden, was ausreicht, um die kleinen negativen Effekte zu verursachen, die bei der Verwendung von Chimären Gesichtern gemessen wurden. Daher sind Chimären Gesichter für Attraktivitätsbeurteilungen schlecht geeignet, denn das vielfach seltsame Aussehen schlägt sich sehr schnell negativ in der Attraktivitätsbewertung nieder. Dies zeigt auch die Untersuchung von Rhodes, Roberts und Simmons (1999), die Chimären Gesichter wie Originalgesichter hinsichtlich *strangeness* beurteilen ließen und damit zeigen konnten, dass Chimären Gesichter als seltsam empfunden werden (S. 286). Es wird also bei Chimären Gesichtern weniger der Effekt perfekter Symmetrie gemessen, sondern vielmehr der Effekt von störenden Bildverarbeitungsartefakten. Damit fehlt es dieser Methode an Validität. Mit Abstrichen gilt dies auch für die Untersuchung *Sym 9*, bei der die Messung der Symmetrie über Ähnlichkeitsurteile der beiden Chimären Gesichter stattfand.

Zwar gab es auch Versuche (Kowner, 1996b), die Chimären Gesichter-Methode zu verbessern und dadurch Bildverarbeitungsartefakte zu reduzieren (z. B. durch Kopieren der Frisur des Originalgesichts auf die Chimären Gesichter oder das nachträgliche Retuschieren von Texturartefakten wie doppelten Muttermalen auf einer Gesichtshälfte), doch lösen sie lediglich einen Teil der angesprochenen Probleme. Letztlich ist aber durch die mittlerweile verfügbare Morphing-Technik diese über 100 Jahre alte Methode (Hallevorden, 1902) der Symmetrisierung von Gesichtern überholt.

5.2.1.2 Effekte von Morphing-Artefakten

In den Kapiteln 4.2, 4.6.1 und 4.7.1 wurde bereits dargelegt, dass die meisten Studien, die Symmetrie durch die Verwendung von Morphing-Software manipulierten, eine Morphing-Technik verwendeten, die nicht nur Gesichtsproportionen, sondern auch die Textur eines Gesichts symmetrisierte. Die dadurch entstehenden Bildverarbeitungsartefakte in Form einer makelloseren Haut wurden bereits ausführlich erläutert.

Der Vorteil der hier durchgeführten Experimente ist, dass nun belegt werden kann, wie groß der Effekt auf den Attraktivitätszugewinn ist, der durch die verschiedenen Morphing-Techniken bei der Symmetrisierung entsteht. In Experiment *Sym 3* wurden symmetrisierte „Blends“ (symmetrisierte Gesichter mit Farbwertmittelung) und in Experiment *Sym 5* symmetrisierte „Warps“ (symmetrisierte Gesichter ohne Farbwertmittelung) jeweils in Paarvergleichsexperimenten mit den Originalgesichtern verglichen.

„Blends“ mit makelloserer Haut wurden im direkten Vergleich klar besser bewertet als Originalgesichter ($r = .53$), „Warps“ (bei denen nur die Gesichtsproportionen symmetrisiert waren) jedoch nur geringfügig besser ($r = .17$). Dies legt den Schluss nahe, dass die Differenz zwischen $.53$ und $.17$ auf den Effekt der Morphing-Technik zurückzuführen ist, denn es ist das einzige Merkmal, in dem sich die beiden Experimente unterscheiden. Man könnte auch sagen, dass der Effekt durch die Artefakte der makelloseren Haut größer ist als der eigentliche Effekt der reinen Symmetrisierung der Gesichtsproportionen. Daraus folgt auch, dass frühere Untersuchungen (Rhodes et al., 1998; Rhodes, Roberts et al., 1999; Rhodes, Sumich et al., 1999; Rhodes, Yoshikawa et al., 2001; Swaddle & Cuthill, 1995), bei denen das „Blend“-Verfahren zur Symmetrisierung verwendet wurde, in ihrer Aussagekraft vorsichtig zu beurteilen sind. Denn bei ihnen ist der Einfluss der Symmetrie konfundiert mit einem weiteren und bedeutsamen Parameter, nämlich einer glatten und makellosen Haut. Der wahre Einfluss der Symmetrie wurde in diesen Studien überschätzt.

5.2.1.3 Effekte der Ausgangsattraktivität der Originalgesichter

Die durchgeführten Experimente machen deutlich, dass hinsichtlich des Einflusses von Symmetrie die Attraktivität der Originalgesichter einen Einfluss auf die erzielten Ergebnisse hat. Normalerweise werden für Attraktivitätsuntersuchungen Gesichter von „normalen“ jungen Menschen (meist Studierende) im Alter zwischen 18 und 30 Jahren verwendet. Bei den hier durchgeführten Untersuchungen wurden hingegen Fotos von 15 „normalen“

jungen Frauen und 15 Fotos von Miss-Germany-Finalistinnen verwendet. Die Daten aus Experiment 2, bei dem unter anderem alle Originalgesichter nach Attraktivität bewertet wurden, bestätigen die naheliegende Annahme, dass die fotografierten Models tatsächlich attraktiver bewertet wurden als die „normalen“ Frauen. Selbst das unattraktivste Model bekam einen höheren mittleren Attraktivitätswert als die attraktivste „normale“ Frau. Dies überrascht selbstverständlich nicht.

Auffallend ist jedoch, dass die Attraktivität der Originalgesichter die festgestellten Effekte hinsichtlich Symmetrie beeinflusst, aber anders als von der Theorie vorausgesagt. Die Streudiagramme (vgl. Abbildung 44, S. 150, Abbildung 46, S. 164 und Abbildung 48, S. 178) illustrieren diesen Effekt. Die 15 Punkte der jeweils linken Hälfte des Diagramms zeigen die Ergebnisse der weniger attraktiven „Normalgesichter“, die 15 Punkte der jeweils rechten Diagrammhälfte die Ergebnisse der attraktiveren Modelgesichter.

Abbildung 44 links (= Attraktivitätsveränderung bei Chimärengesichtern aus *linken* Gesichtshälften, S. 150) zeigt, dass die weniger attraktiven Gesichter durch eine Symmetrisierung mittels Chimären-Verfahren geringfügig unattraktiver werden, attraktivere Gesichter hingegen verlieren stark an Attraktivität. Abbildung 44 rechts (= Attraktivitätsveränderung bei Chimärengesichtern aus *rechten* Gesichtshälften, S. 150) zeigt einen ähnlichen Effekt, wenn auch nicht so stark. Die Symmetrie-Theorie setzt etwas anderes voraus: Wenn Symmetrie attraktiv macht, dann sind unattraktive Gesichter auch deswegen unattraktiv, weil sie asymmetrischer sind. Macht man sie experimentell perfekt symmetrisch, dann sollten sie attraktiver werden. Hochattraktive Gesichter hingegen sind laut Theorie schon sehr symmetrisch, daher sollte eine künstliche perfekte Symmetrisierung keine Veränderung bewirken.

Die Ergebnisse widerlegen diese Vorhersage. Tatsächlich wirkt sich eine Symmetrisierung durch das Chimären-Verfahren tendenziell umso Attraktivität reduzierender aus, je attraktiver das Originalgesicht ist. Hierfür gibt es eine recht banale Interpretation: Je attraktiver ein Gesicht ist, desto größer ist das Risiko, dass man es durch eine Manipulation unattraktiver macht. Da die Chimären-Technik ein ziemlich schlechtes Symmetrisierungsverfahren ist, ist bei einem attraktiven Modelgesicht die Gefahr größer, dass Bildverarbeitungsartefakte sich negativ auswirken. Bei einem attraktiven Gesicht, das sehr ästhetische Proportionen besitzt, jedoch bei der Aufnahme leicht gedreht fotografiert wurde, macht sich dies dann in zwei Chimärengesichtern bemerkbar, die entweder zu schmal oder zu breit sind (vgl. Kapitel 5.2.1.1). Bei einem weniger attraktiven Gesicht hingegen, das vom Ideal stärker abweicht, besteht bei demselben Störeinfluss bei der Aufnahme

sogar eher die Möglichkeit, dass sogar eines der beiden Chimärengesichter an Attraktivität gewinnt; hinzu kommen bei den deutlich unattraktiven Gesichtern Boden-Effekte, wenn der Attraktivitätsmittelwert beim Originalgesicht ohnehin schon am unteren Ende der Skala ist.

Dass dieser Effekt tatsächlich aus Bildverarbeitungsartefakten des Chimären-Verfahrens resultiert, wird auch deutlich, wenn man die Streuung der Attraktivitätsveränderung bei den Chimären-Experimenten vergleicht mit der Streuung der Attraktivitätsveränderung bei den analogen Experimenten mit „Blends“ bzw. „Warps“. Bei diesen fällt sie nämlich weit aus kleiner aus (man vergleiche dazu die Abbildung 44, S. 150 mit Abbildung 46, S. 164 und Abbildung 48, S. 178; die Skalierung der y-Achse ist stets konstant gehalten).

Beim Experiment mit „Blends“ lässt sich kein Einfluss der Ausgangsattraktivität des Stimulusmaterials nachweisen – im Experiment mit „Warp-Gesichtern“ hingegen wieder schon. Auf attraktivere Gesichter (Models) hatte die Symmetrisierung der Gesichtsproportionen keinen Einfluss (= hypothesenkonform), auf unattraktivere („normale“ Frauen) hingegen einen negativen Effekt. Die Symmetrie-Hypothese sagt jedoch einen positiven Effekt voraus (siehe oben!). Eine Erklärung hierfür ist spekulativ.

Fakt ist jedoch, dass die Attraktivität der Originalgesichter einen Einfluss auf die erzielten Ergebnisse hat. Wer Symmetrie-Untersuchungen mit normalen Stimuluspersonen durchführt, kommt zu anderen Resultaten, als jemand, der dieselben Untersuchungen mit hochattraktiven Gesichtern durchführt – und die Unterschiede sind dabei noch nicht einmal konform mit der Symmetrie-Hypothese, sondern widersprechen ihr.

Würde man mit unattraktiven Gesichtern arbeiten, wären wiederum nochmals andere Effekte nicht auszuschließen. Sehr plausibel ist beispielsweise die Annahme, dass bei Verwendung älterer Stimuluspersonen, bei denen manche deutlich sichtbare Falten besitzen, entsprechende Effekte (bzw. Methoden-Artefakte) bei Anwendung der „Blend“-Technik anzunehmen sind. Denn diese Symmetrisierungs-Technik führt zu einem Weichzeichnungseffekt der Haut, wodurch diejenigen Gesichter am stärksten profitieren, die im Original die meisten Falten besitzen und aufgrund dieser Falten unattraktiver bewertet werden. In diesem Falle würde die Ausgangsattraktivität der Originalgesichter mit dem verwendeten Symmetrisierungsverfahren interagieren.

Obwohl also die Attraktivität des verwendeten Stimulusmaterials offenkundig eine wichtige Rolle spielt, wird dies bei vielen Symmetrie-Studien überhaupt nicht thematisiert, nämlich bei all denen, die Symmetrie lediglich im relativen Paarvergleich untersucht haben,

jedoch die absolute Attraktivität der verwendeten Stimuli gar nicht gemessen haben (z. B. durch Verwendung von Ratingskalen) und deswegen auch gar keine Aussage dazu treffen können.

Zusammenfassend bleibt festzuhalten, dass das verwendete Stimulusmaterial einen wesentlichen Einfluss auf die gemessenen Ergebnisse hat. Hält man andere zusätzliche Einflussfaktoren wie das Versuchsdesign und die Auswertungsmethode konstant und betrachtet nur die Experimente, die sich lediglich in der Art der Symmetrie-Manipulation unterscheiden, dann lässt sich die Bedeutung dieses Parameters abschätzen. Bei den Experimenten 1, 3 und 5 handelt es sich jeweils um Paarvergleichsexperimente; sie unterscheiden sich nur in der Art und Weise, wie Symmetrie manipuliert wurde. Unter Einbeziehung der Versuchspersonen-Urteilsvarianz liegen die gefundenen Effekte zwischen $r = -.27$ und $r = .53$; ohne Berücksichtigung der Urteilsvarianz ist die Spannbreite sogar noch größer, nämlich zwischen $r = -.74$ und $r = .80$.

Vergleicht man die Ergebnisse aus den Rating-Experimenten (Experimente 2, 4 und 6), sind die Unterschiede des Stimulusmaterials nicht so ausgeprägt. Allerdings ergibt sich überall ein negativer (!) Einfluss der Symmetrie. Der Bereich der Effektstärken liegt zwischen $r = -.20$ und $r = -.08$ (mit Urteilsvarianz) bzw. $r = -.73$ und $r = -.53$ (ohne Urteilsvarianz). In diesen Zahlen nicht berücksichtigt sind die Effekte der Attraktivität der Originalgesichter, also des Ausgangsmaterials, die einen zusätzlichen Einfluss auf die Experimente zur Symmetrie-Hypothese haben. Daraus folgt, dass sämtliche bisher durchgeführten Studien zur Symmetrie von Gesichtern, bei denen Symmetrie mit einer der genannten drei Methoden manipuliert wurde, zu einem mehr oder weniger großen Teil auf methodischen Artefakten beruhen.

5.2.2 Effekte des Versuchsdesigns: Paarvergleich, Rating, Between-Subjects Design und korrelative Verfahren

Zusätzlich zum Effekt des Stimulusmaterials konnte klar gezeigt werden, dass die Art des Versuchsdesigns einen erheblichen Einfluss auf die erzielten Ergebnisse hat. Wurden in den ersten sechs Experimenten die symmetrisierten Gesichter im direkten Paarvergleich mit den Originalgesichtern präsentiert, dann wurden die symmetrischen Gesichter entweder attraktiver beurteilt (*Sym 3* $r = .53$ und *Sym 5* $r = .17$) oder gleich attraktiv (*Sym 1b* $r = .00$, jeweils Auswertungsverfahren mit Versuchspersonen-Urteilsvarianz) beurteilt. Experiment *Sym 1a* stellt hier eine Ausnahme dar ($r = -.27$). Wurden hingegen die symmetrisierten Gesichter im Between-Design von verschiedenen Urteilergruppen auf ihre Attrak-

tivität auf einer Ratingskala bewertet, dann schnitten die symmetrisierten Varianten immer schlechter ab als die Originalvarianten (Experiment *Sym 2a* $r = -.20$, *Sym 2b* $r = -.18$, *Sym 4* $r = -.08$ und *Sym 6* $r = -.09$). Betrachtet man die Effektgrößen ohne Versuchspersonen-Urteilsvarianz, dann sind alle Effekte noch stärker (in die positive bzw. negative Richtung) ausgeprägt (vgl. Tabelle 15, S. 205).

Offensichtlich kommt es also darauf an, ob eine Versuchsperson einen direkten Vergleich hat und zwei hochgradig ähnliche Bilder auf möglicherweise winzige Attraktivitätsunterschiede vergleicht, oder ob eine Versuchsperson ein absolutes Urteil abgibt, wie das beim Between-Design der Fall ist. Nicht untersucht wurde in dieser Reihe, wie es sich auswirkt, wenn Probanden zwar Ratingskalen benutzen, hierbei jedoch ein Within-Subject-Design verwendet wird, dass also jede Versuchsperson immer beide Varianten einer Stimuluspersion (Original und symmetrisiert) zu Gesicht bekommt. Dies zu testen hätte noch einmal drei zusätzliche Experimente erfordert und bei zwei möglichen Datenauswertungsmethoden (mit und ohne Urteilsvarianz) insgesamt weitere sechs Ergebnisse produziert. Diese Art der Datenerhebung wurde in der Vergangenheit relativ häufig angewandt (Rhodes et al., 1998; Rhodes, Sumich et al., 1999; Rhodes, Yoshikawa et al., 2001; Simmons et al., 2004) und steht methodisch zwischen Rating im Between-Design und Paarvergleich. Einerseits geben die Probanden ein (vermeintlich) absolutes Urteil auf einer Skala ab. Andererseits jedoch haben sie durchaus einen relativen Vergleich zwischen den einzelnen Varianten eines Gesichts, sie können sich an bestimmte Gesichter erinnern, sie können leichter erkennen, was manipuliert wurde, erraten, was die Hypothese des Experiments ist und dadurch suggestiv beeinflusst werden.

Aus methodischer Sicht ist ein Between-Subject-Design das weniger reaktive und extern validere Verfahren, um den Effekt der Symmetrie auf Attraktivität zu messen. Es ist weniger reaktiv, denn die Probanden haben dabei eine weit geringere Chance zu erraten, was bei dieser Datenerhebung eigentlich gemessen wird (v. a. bei ökologisch validen Stimuli wie „Warp“-Gesichtern). Bei einem Paarvergleich werden sie durch die Art der Darbietung der Stimuli geradezu darauf gestoßen, den Unterschied zwischen den beiden Varianten eines Gesichts zu erkennen und möglicherweise auch, die Systematik der Variationen zu verstehen. Die ganze Aufgabe ähnelt eher einem Fehlersuchbild. Die Wahrscheinlichkeit ist daher groß, dass sie dabei auch die Hypothese erraten und durch diese Erwartungshaltung ihr Urteilsverhalten beeinflusst wird.

Das *Between-Subjects Design* ist zudem extern valider als der Paarvergleich, weil es prinzipiell auch eine Aussage darüber erlaubt, wie wichtig Symmetrie überhaupt für die

Attraktivität eines Gesichts ist. Wäre Symmetrie ein sehr wichtiges Kriterium, dann sollten alle symmetrisierten Gesichter nach der symmetrischen Optimierung sehr hohe Attraktivitätsratings erzielen. Dem ist jedoch nicht so. Die Attraktivitätsmittelwerte der symmetrisierten Varianten unterscheiden sich nur unwesentlich von den Originalgesichtern – und obendrein sind sie im Durchschnitt sogar niedriger. Gerade unter einem angewandten Aspekt ist dies die entscheidende Information. Ein Patient, der sich an den kleinen Asymmetrien in seinem Gesicht stört und der sich mit dem Gedanken trägt, mit Hilfe eines ästhetischen Eingriffs diese Asymmetrien zu beseitigen (z. B. durch den Einsatz von *facial fillers*), ist von diesem Vorhaben klar abzuraten. Seine Absicht ist es ja, von seiner sozialen Umwelt als attraktiver wahrgenommen zu werden. Dieses Ziel wird aber nicht erreicht. Seine Mitmenschen werden sein Gesicht auch nie direkt in einem A-B-Vergleich beurteilen können, denn sein Gesicht ist *entweder* unverändert mit Asymmetrien *oder* symmetrisch aufgrund von medizinischen Eingriffen, aber nicht beides zugleich.

Es ist also keineswegs so, dass der Paarvergleich einfach nur ein sensitiveres Testverfahren ist, das kleine Unterschiede leichter nachweisen kann als ein Between-Design mit Ratingskalen. Er produziert zugleich auch Effekte, die in eine ganz andere Richtung gehen: Statt schwach negativer Effekte (wie beim Between-Design) erzielt der Paarvergleich schwache bis mittelstarke positive Effekte. Benutzt man zusätzlich noch statistische Kunstgriffe wie das Eliminieren von Versuchspersonen-Urteilsvarianz (wie z. B. Perrett et al., 1999 oder Jones et al., 2003), erzeugt man sogar mittlere (*Sym 5* $r = .55$) bis starke (*Sym 3* $r = .80$) Effekte.

Wenn man nun annimmt, dass das Rating im Between-Design ein „ehrlicheres“ und methodisch korrekteres Messverfahren als der direkte Paarvergleich ist, dann stellt sich trotzdem immer noch die Frage, warum eigentlich symmetrisierte Gesichter geringfügig negativer bewertet wurden als Originalgesichter. Macht Symmetrie in Wahrheit *unattraktiver*? Hierzu muss zunächst betont werden, dass die mittleren Effektgrößen (*Sym 4* $r = -.08$ und *Sym 6* $r = -.09$) nur sehr gering sind (die Experimente mit Chimärengesichtern werden hier nicht betrachtet, da das Stimulusmaterial ohnehin nicht valide ist). Bei Experiment *Sym 4* („Blends“) waren nur 3 von 30 Tests auf Zwischensubjekteffekte der MANOVA überhaupt statistisch signifikant, bei Experiment 6 („Warps“) nur 6 von 30 Tests (jeweils teilweise auch noch in verschiedene Richtungen). Insofern könnte man also die geringen negativen Effekte ebenso gut zu einem Großteil als Zufallsrauschen abtun. Nach dieser Interpretation ist es also schlichtweg egal, ob ein Gesicht in der Originalvariante oder der symmetrisierten Variante präsentiert wird – die Unterschiede sind so gering, dass sie auf das Attraktivitätsurteil praktisch keinen Einfluss haben.

Möglicherweise gibt es jedoch einen winzigen tatsächlich negativen Einfluss der Symmetrisierung. Es könnte sein, dass den Versuchspersonen das völlig Symmetrische tatsächlich weniger gefällt als das „natürlich“ asymmetrische. Doch es ist denkbar, dass dies gar nicht an der Symmetrie an sich liegt, sondern an der Art und Weise, wie das Gesicht auf dem Foto abgebildet ist. Da in Kapitel 5.2.1.1 bereits festgestellt wurde, dass manche Personen nicht perfekt frontal fotografiert wurden, sondern den Kopf auf dem Bild minimal gedreht hatten, könnte es auch sein, dass diese leicht gedrehte Kopfposition auf die Beurteiler geringfügig angenehmer wirkt als der absolut perfekt frontale Anblick, den alle symmetrisierten Varianten bieten. Dies ist jedoch spekulativ und anhand der Daten weder zu beweisen noch zu widerlegen. Denkbar wäre es jedoch, dass eine perfekt frontale Ansicht eines Gesichts (trotz perfekter Symmetrie) auf den Betrachter unangenehmer wirkt als dasselbe Gesicht mit einer minimal gedrehten Kopfposition jedoch mit geringfügigen natürlichen Asymmetrien. Der Effekt der Kopfposition könnte experimentell getestet werden, indem man beispielsweise perfekt symmetrische Köpfe von Schaufensterpuppen bzw. computergenerierte, perfekt symmetrische Avatare mit geringfügig veränderten Kopfpositionen fotografiert und nach Attraktivität bewerten lässt. Die Gesichter wären in diesem Fall immer exakt identisch, die Symmetrie der Gesichter ebenfalls, folglich müssen eventuelle Attraktivitätsunterschiede durch den Aufnahmewinkel entstanden sein.

In der Untersuchungsreihe kamen neben den sechs Experimenten auch drei korrelative Verfahren zum Einsatz. Dabei wurde auf unterschiedliche Arten für jedes Originalgesicht ein Symmetriewert ermittelt und alle Symmetriewerte dann mit den dazugehörigen Attraktivitätswerten dieser Gesichter korreliert. Alle drei Verfahren ermittelten einen positiven Einfluss der Symmetrie auf die Attraktivitätsbewertung. Die Untersuchungen *Sym 8* und *Sym 9*, die Symmetrie auf der Grundlage von Ähnlichkeitsurteilen von Chimären-gesichtern bzw. Gesichtshälften ermittelten, erzielten einen schwachen Effekt ($r = .48$ bzw. $r = .47$), Untersuchung *Sym 7* (direktes Rating der Symmetrie durch Versuchspersonen) hingegen einen starken Effekt ($r = .78$).

Gerade dieses direkte Rating der Symmetrie durch Versuchspersonen ist jedoch sehr kritisch zu sehen. Denn es deutet viel darauf hin, dass die Versuchspersonen bei der Aufgabe, die Symmetrie eines Gesichts zu bewerten, in Wirklichkeit hauptsächlich dessen Attraktivität bewerten. Offensichtlich gilt *nicht* der Satz: „Wenn ein Gesicht symmetrisch ist, dann wird es attraktiver bewertet“, sondern stattdessen: „Wenn ein Gesicht attraktiver ist, dann wird es symmetrischer bewertet“. Ursache und Wirkung sind in Wahrheit genau anders herum als von Attraktivitätsforschern wie Rhodes permanent behauptet.

Es reicht nämlich nicht aus, einfach nur die Korrelation zwischen den Attraktivitätsbewertungen der Originalgesichter und deren Symmetrie-Ratings zu betrachten. Denn die Korrelationen zwischen Symmetrie-Ratings der Originalgesichter und Attraktivitätsbewertungen der symmetrisierten Varianten sind genauso hoch. Bei den symmetrisierten Varianten (Chimärengesichter, „Blends“ und „Warps“) gibt es jedoch gar keine Varianz der Symmetrie – alle sind gleich perfekt symmetrisch. Die Attraktivitätswerte können also nicht aufgrund von Symmetrie-Unterschieden zustande gekommen sein, sondern aufgrund irgendwelcher anderer Attraktivitätsmerkmale, die gar nichts mit Symmetrie zu tun haben. Dennoch korrelieren sie hoch (zwischen $r = .75$ und $r = .80$) mit der angeblichen Symmetrie der Originalgesichter (vgl. Tabelle 16). Hier müsste jedoch eigentlich jeweils ein Zusammenhang von Null vorhanden sein!

Die gleiche Argumentation gilt auch für die anderen beiden Methoden der Symmetriermittlung, nämlich Ähnlichkeitsbewertungen von Gesichtshälften bzw. Chimärengesichtern. Diese Maße dürften eigentlich nicht mit den Attraktivitätsbewertungen der symmetrisierten Varianten korrelieren. Sie tun es aber dennoch und zwar in derselben Höhe wie bei den Attraktivitätsbewertungen der Originalgesichter.

Freilich ist auch bereits der extrem hohe Zusammenhang zwischen Attraktivität der Originale und Attraktivität der symmetrisierten „Blend“- bzw. „Warp“-Gesichter von jeweils $r = .99$, $p < .001$ ($r^2 = .98$) ein deutlicher Hinweis, dass mindestens 98 % der Attraktivitätsvarianz durch gemeinsame Attraktivitätsmerkmale erklärt werden, die kausal nichts mit Symmetrie zu tun haben können, und deswegen nur maximal 2 % der Attraktivitätsvarianz für Symmetrie verbleiben könnte.

Tabelle 16: Korrelationsmatrix. Der Zusammenhang (r) zwischen der Attraktivität der Originalgesichter und der Attraktivität symmetrisierter Varianten (Chimären-gesichter, Blends und Warps). Zusätzlich die Korrelationen zwischen verschiedenen Symmetrie-Maßen (Symmetrie-Rating Originale, Ähnlichkeitsrating Gesichtshälften und Ähnlichkeitsrating Chimären-gesichter) und der Attraktivität symmetrisierter Gesichter.

	Originale	ChiL	ChiR	Blend	Warp
Originale	–				
Chimäre Links	.94***	–			
Chimäre Rechts	.93***	.81***	–		
Blends	.99***	.94***	.92***	–	
Warps	.99***	.93***	.92***	.98***	–
Symmetrie-Rating (Originale)	.78***	.77***	.75***	.80***	.76***
Symmetrie-Rating (Warps, = Kontrolle)	.87***	.92***	.80***	.88***	.87***
Ähnlichkeitsrating Gesichtshälften	.47**	.55**	.45*	.50**	.46*
Ähnlichkeitsrating Chimären-gesichter	.48**	.51**	.51**	.51**	.44*

Anmerkung: * = $p < .05$, ** = $p < .01$, *** = $p < .001$

Da also Unterschiede in Attraktivitätsbewertungen von symmetrisierten Gesichtern ausschließlich von Attraktivitätsmerkmalen beeinflusst sein können, die nichts mit Symmetrie zu tun haben (da ja alle Stimuli gleichermaßen perfekt symmetrisch sind), besteht die Möglichkeit, aus der Korrelation zwischen Symmetrie-Rating des Originalgesichts und Attraktivitätsbewertung des Originalgesichts die Attraktivitätsbewertung für ein solches symmetrisiertes Gesicht heraus zu partialisieren. Tut man dies, so erhält man für alle vier Partialkorrelationen ein nicht signifikantes Ergebnis. Beispiel: Unter statistischer Kontrolle

der Attraktivitätsbewertungen für symmetrische „Warp“-Gesichter bleibt dann noch ein Zusammenhang von $r = .26$, $p = .174$ übrig und unter Kontrolle der Werte der „Blend“-Gesichter ein Zusammenhang von $r = -.13$, $p = .505$ (Chimärengesichter sind wie bereits dargelegt ohnehin keine valide Methode, doch auch hier bleibt kein signifikanter Zusammenhang mehr übrig).

Die vorhandenen hohen Korrelationen zwischen Symmetrie-Ratings von Originalgesichtern und Attraktivitätsbewertungen von gleichermaßen perfekt symmetrischen Gesichtern, die eigentlich gar nicht existieren dürften, sind bereits ein sehr deutliches Indiz dafür, dass eine starke Konfundierung zwischen unabhängiger Variable (Symmetrie) und abhängiger Variable (Attraktivität) vorliegt.

Einen klaren Beweis für die methodische Unsinnigkeit des direkten Symmetrie-Ratings liefert jedoch die nachträglich durchgeführte Untersuchung bei Gesichtern, die alle perfekt symmetrisch waren (Warp-Gesichter) und von Versuchspersonen direkt nach Symmetrie geratet wurden. Da es dort keine Symmetrieunterschiede gab, hätten eigentlich alle Stimuli im Durchschnitt denselben Symmetriewert bekommen müssen und die Varianz der Variable Symmetrie hätte Null sein müssen. Tatsächlich bekamen sie jedoch unterschiedliche Symmetriebewertungen. Es zeigte sich: Je attraktiver ein Gesicht ist, desto höher sein Symmetrie-Rating, obwohl in Wirklichkeit alle Gesichter gleich symmetrisch waren ($r = .87$, $p < .001$). Der Zusammenhang ist damit sogar noch etwas höher als der Zusammenhang zwischen Symmetriebewertung der asymmetrischen Originalgesichter und Attraktivität ($r = .78$, der Test auf Unterschiede zwischen den Korrelationskoeffizienten ist statistisch nicht signifikant). Hieraus folgt, dass es offensichtlich völlig egal ist, ob man Versuchspersonen asymmetrische Originalgesichter oder perfekt symmetrische Warp-Gesichter für Symmetrie-Ratings vorlegt. Die Ergebnisse sind die gleichen. Die Beurteiler bewerten ganz offenkundig nicht die Symmetrie, sondern die Attraktivität selbst. Es spricht alles dafür, dass hier das altbekannte Attraktivitätsstereotyp („Was schön ist, ist auch gut“) in Form eines Halo-Effekts der Attraktivität auf die Symmetrie-Bewertung vorliegt.

Die Ergebnisse dieser zusätzlichen Kontrolluntersuchung zeigen eindeutig, dass das direkte Rating der Symmetrie ein Verfahren ist, bei dem abhängige und unabhängige Variable hochgradig konfundiert sind. Es ist kaum nachvollziehbar, dass eine so renommierte Attraktivitätsforscherin wie Gilian Rhodes (noch dazu in einer Metaanalyse) eine so invalide Methode als besonders positiv herausstellen kann: „Ratings change systematically when facial symmetry is manipulated, confirming their validity (see, e.g., Rhodes et

al. 1999b), and are probably sensitive to more subtle facial asymmetries than are current measurements methods. Therefore, on theoretical grounds, ratings may be preferable to measurements“ (Rhodes, 2006, S. 207). Vor dem Hintergrund der oben diskutierten Daten dürfte wohl eher die Schlussfolgerung angemessen sein, dass es sich beim direkten Rating der Symmetrie um die am wenigsten valide Methode zur Untersuchung des Einflusses von Symmetrie auf Attraktivität handelt.

Etwas positiver zu bewerten sind hingegen die beiden korrelativen Methoden, die Symmetrie über Ähnlichkeitsbewertungen erfassen. Sie ergaben gleich große positive Effektstärken ($r = .48$ und $r = .47$). Ihr Vorzug liegt in einer (im Vergleich zum direkten Rating bzw. zu Paarvergleichen) geringeren Reaktivität. Es ist für Versuchspersonen schwerer zu durchschauen, dass sich die eigentliche Fragestellung nicht mit Ähnlichkeitsurteilen, sondern in Wirklichkeit mit Symmetrie und Attraktivität befasst. Dennoch gibt es auch hier, insbesondere bei der Ähnlichkeitsbewertung von Gesichtshälften, einen Halo-Effekt der Attraktivität, wie die Ergebnisse der Partialkorrelationen (vgl. Kapitel 4.11.3 und 4.12.3) zeigen.

Zudem gelten für die Bewertung der Ähnlichkeit von Chimärengesichtern teilweise ähnliche Einwände, wie sie bereits bei den Attraktivitätsbewertungen von Chimärengesichtern vorgebracht wurden (vgl. die vier Hauptartefakte von Chimärengesichtern in Kapitel 5.2.1.1) und die hier nicht noch einmal wiederholt werden sollen. Bei der Bewertung von separaten Gesichtshälften ist hingegen fraglich, inwieweit Menschen aufgrund des Umstandes, dass Gesichterwahrnehmung holistisch erfolgt, überhaupt in der Lage sind, halbe Gesichter angemessen zu erkennen. Die Aufgabe, bestimmte asymmetriebedingte Unterschiede in den Hälften zu erkennen, sollte bei separaten Gesichtshälften jedenfalls schwieriger sein, ähnlich wie bei Gesichtern, die auf dem Kopf stehend präsentiert werden (Little & Jones, 2003). Zudem ist es eine naheliegende Annahme, dass bei der Präsentation von separaten Hälften Texturunterschiede (z. B. Pickel, Muttermale) stärker in den Fokus der Aufmerksamkeit rücken, weil es durch das Trennen der Gesichtshälften (und Spiegeln einer Hälfte) schwieriger geworden ist, Unregelmäßigkeiten der Proportion zu beurteilen. Texturunterschiede der Haut (oder gar der Haare) sind jedoch gerade *nicht* gemeint, wenn Attraktivitätsforscher bzw. Biologen von (fluktuierender) Asymmetrie sprechen.

Zu diesen praktischen, konkreten Einschränkungen der beiden Ähnlichkeitsbewertungsmethoden kommt natürlich noch aus theoretischer Sicht die Einschränkung, dass von Korrelationen generell nicht auf Kausalität geschlossen werden kann. Es wäre z. B. auch

denkbar, dass das Merkmal der Symmetrie mit einer weiteren Variablen, beispielsweise einer gesünder aussehenden Haut, korreliert ist und dass es in Wirklichkeit diese Drittvariable ist, die zu einem höheren Attraktivitätsurteil führt – erste empirische Hinweise dafür gibt es zumindest (Jones et al., 2004). Wenn hohe Symmetrie tatsächlich aus einer starken Entwicklungsstabilität (developmental stability) während der Ontogenese resultiert und ein Indikator für hohe Robustheit gegen Umweltstressoren ist (Gangestad & Thornhill, 1997; Gangestad et al., 1994; Palmer & Strobeck, 1986; Thornhill & Gangestad, 1999; Thornhill & Moller, 1997), dann ist ein solcher Wirkmechanismus sogar sehr plausibel. Den Attraktivitätsforscher würde dies jedoch zu falschen Schlussfolgerungen führen, denn er würde ein Merkmal – nämlich Symmetrie – als kausal verantwortlich für Attraktivität annehmen, das in Wirklichkeit jedoch nicht oder nur eingeschränkt verantwortlich ist.

Aus theoretischer Sicht ist hier selbstverständlich ein experimentelles Design überlegen, doch unter praktischen Aspekten gibt es auch dort zahlreiche Probleme. Die oben dargelegten Ausführungen haben deutlich gezeigt, dass man durchaus streng experimentell arbeiten und dennoch Ergebnisse erzielen kann, die aufgrund der angewandten experimentellen Manipulationen so voller Methodenartefakte sind (z. B. Chimären gesichter, „Blends“), dass die Ergebnisse kaum Aussagekraft besitzen. Insofern kann man die auf Ähnlichkeitsbewertungen basierenden korrelativen Verfahren einigen experimentellen Verfahren als durchaus ebenbürtig bezeichnen – wenn auch als ebenbürtig auf einem teilweise wenig befriedigenden methodischen Niveau.

5.2.3 Effekte der Datenauswertung: Der Umgang mit Urteilsvarianzen

Ein Blick auf die Zusammenfassung der Ergebnisse aus den Symmetrie-Experimenten macht deutlich, dass es einen großen Unterschied macht, mit welchem statistischen Verfahren die erhobenen Daten ausgewertet werden. Berücksichtigt man die Varianz von Versuchspersonen-Urteilen, dann sind die Beträge der ermittelten Effekte kleiner (d. h. sie liegen näher am Wert Null), berücksichtigt man hingegen die Urteilsvarianzen *nicht*, sind die Beträge der Effekte größer. Grundsätzlich klingt diese Tendenz zwar nicht überraschend, bemerkenswert sind jedoch die deutlichen Unterschiede bei den ermittelten Effekten. Es macht einen erheblichen Unterschied, ob eine Effektstärke $r = .17$ oder $r = .55$ (Sym Experiment 5) beträgt, ob sie $r = -.09$ beträgt oder $r = -.57$ (Sym Experiment 6), $r = -.20$ oder $r = -.73$ (Sym Experiment 2a). Insbesondere unter praktischen Aspekten spielt dies eine Rolle. Während man eine Effektstärke von $r = .17$ unter einem anwendungsorientierten Standpunkt als unbedeutend abtun kann, ist ein Effekt von $r = .55$ so

bedeutsam, dass man ihm Beachtung schenken muss. Wenn sich jedoch beide Ergebnisse aus ein und derselben Datenlage ableiten lassen, dann wirft dies zu Recht methodische Zweifel auf.

Solche Zweifel entstehen gewiss nur bei dem, der sich bewusst ist, dass es diese beiden Möglichkeiten der Datenauswertung gibt. In keiner Publikation zum Einfluss der Symmetrie werden jedoch beide Verfahren gleichzeitig verwendet oder zumindest die Problematik des Umgangs mit Urteilsvarianzen thematisiert. Üblicherweise wird das Verfahren gewählt, das größere Effektstärken erzielt.

Die gewählte Auswertungsmethode hat jedoch nicht nur Konsequenzen für die Effektstärke, sondern auch auf die Signifikanz. Wie in den Ergebnisteilen der jeweiligen Experimente dargelegt wurde, wird ein Großteil der Ergebnisse überhaupt nicht signifikant, wenn man bei der Auswertung Urteilsvarianzen berücksichtigt.

Als Beispiel sei das Paarvergleichsexperiment mit den „Warp“-Gesichtern herausgegriffen, weil die Verwendung von „Warps“ die valideste Art der Symmetrie-Manipulation ist. Nur in 7 von 30 Paarvergleichen ($N = 64$) wurde die symmetrische Variante signifikant besser bewertet als die Originalvariante, in 23 Fällen gibt es keinen signifikanten Unterschied. Hierbei ist eine Adjustierung des α -Niveaus noch nicht einmal vorgenommen. Die mittlere Effektgröße (umgerechnet nach r) von .17 sagt deswegen kaum etwas aus, da es nur der Median aus Effekten ist, die aus überwiegend (77 %) nicht signifikanten Ergebnissen stammen. Bei der Auswertungsmethode *ohne* Berücksichtigung der Urteilsvarianz durch Verwendung eines Ein-Stichproben- t -Tests, wie sie beispielsweise Little und Jones (2003, S. 1761) für genau das gleiche Experiment verwendeten, ergibt sich hingegen ein Testwert von $t(63) = 5.208$, $p = .000002$, standardisiert nach r ein Effekt von $r = .55$, also ein mittel großer Effekt mit einem höchst signifikanten Ergebnis. Für die anderen Paarvergleichsexperimente sind die Ergebnisse ähnlich und sollen hier nicht noch einmal wiederholt werden, da sie in den jeweiligen Ergebnisteilen ausführlich dargelegt wurden.

Bei Anwendung der einen Statistik wird geschlussfolgert, dass es praktisch keinen Unterschied zwischen symmetrischen und originalen Gesichtern gibt, dass sie praktisch gleich attraktiv sind und ein marginaler Attraktivitätsvorteil der symmetrischen Gesichter im Zufallsrauschen größtenteils statistisch nicht signifikanter Ergebnisse untergeht. Bei Anwendung der anderen Statistik kommt man hingegen zum Schluss, dass es einen mittelstarken attraktivitätssteigernden Effekt der Symmetrie gibt, und das mit einer „Irrtumswahrscheinlichkeit“, die über jeden Zweifel erhaben ist ($p = .000002$).

Das gleiche Problem wie bei den Paarvergleichsexperimenten gibt es auch bei den Experimenten mit Rating-Verfahren. Als Beispiel – stellvertretend für alle drei Rating-Experimente – seien hier wieder die Ergebnisse des Experiments mit den „Warp“-Gesichtern aufgeführt. Berücksichtigt man die Urteilsvarianzen der Versuchspersonen und rechnet man eine MANOVA, so ist das *Gesamtergebnis* zwar signifikant, aber bei den Post-hoc-Tests sind nur 6 von 30 Mittelwertsvergleiche statistisch signifikant. Das heißt, bei 23 Stimulusgesichtern spielt es keine Rolle, ob sie symmetrisiert oder als Originale präsentiert werden, sie bekommen immer die gleichen Attraktivitätsbewertungen. Der Median über alle 30 Cohen's *d* beträgt $-.18$. Das bedeutet, dass der Effekt der Symmetrisierung nach dieser Methode schwach negativ ist. Standardisiert nach *r* ergibt sich ein Wert von $r = -.09$.

Verwendet man jedoch statt der Analyse über 2×64 Versuchspersonen (*between groups*) einen *t*-Test über die 30 Stimulusgesichter in ihren zwei Ausprägungen (symmetrisch versus Original, *within groups*, zweiseitig), dann bleibt dadurch die Varianz der Versuchspersonenurteile unberücksichtigt und es werden nur die Attraktivitätswerte (= Mittelwerte über 64 Versuchspersonenurteile zu jedem Stimulusgesicht) als Eigenschaft eines jeweiligen Gesichts verwendet. Das Ergebnis beträgt dann $t(29) = -3.721$, $p = .0008$, der *t*-Wert umgerechnet nach *r* beträgt dann $r = -.57$.

Auch beim Rating-Verfahren gibt es also ebenfalls einmal einen marginal negativen Effekt aus überwiegend nicht signifikanten Einzelergebnissen, aus dem man den Schluss ziehen muss, dass es bei den meisten Gesichtern keine Rolle spielt, ob sie symmetrisiert oder im Original präsentiert werden, und im anderen Fall ein mittelstarkes, negatives, hoch signifikantes Ergebnis, aus dem man folgern muss, dass Symmetrie Gesichter deutlich *unattraktiver* macht. Was stimmt also nun? Welches Auswertungsverfahren ist das korrekte? Welches ist besser? Oder sind es nur zwei Sichtweisen auf dieselbe Wahrheit? Welche Konsequenzen ergeben sich daraus für die eigentliche Fragestellung nach der Bedeutung von Symmetrie für Attraktivität?

Bei beiden Auswertungsverfahren gibt es Argumente dafür und dagegen. Für das Einbeziehen der Urteilsvarianz spricht, dass es ein konservatives, vorsichtiges Verfahren ist, das die Daten quasi in Rohform auswertet und sie nicht zuvor „aufbereitet“. Bei einem Paarvergleich ist es die natürliche und folgerichtige Art, die erhobenen Daten mit einem Test auf Binomialverteilung auszuwerten. Führt man nicht *einen* Paarvergleich durch, sondern 30, ist es konsequent 30 Binomialtests zu rechnen, dabei jedoch das α -Niveau entsprechend anzupassen. Bereits ein einziger Paarvergleich mit zwei Gesichtervarian-

ten (symmetrisch – original) der einer Stichprobe von Versuchspersonen zur Beurteilung vorgelegt wird, ist ein Experiment. Das Präsentieren von 30 Paarvergleichen kann man insofern als Wiederholung dieses Experiments sehen, bei dem eigentlich das Gleiche Ergebnis herauskommen sollte. Ähnlich verhält es sich bei den Rating-Experimenten. Es gibt zwei Experimentalbedingungen (symmetrisch und original), die im Between-Design durch zwei Gruppen von Versuchspersonen realisiert werden. Jede Person beurteilt die Attraktivität von 30 Gesichtern. Folglich gibt es 30 abhängige Variablen und somit ist es konsequent, diese Daten mit einem multivariaten Verfahren mit anschließenden Post-hoc-Tests auszuwerten.

Gegen dieses Verfahren spricht, dass dadurch statistische Power verschenkt wird und dass es bei einer zu kleinen Stichprobe dazu führen kann, dass ein Effekt der Symmetrie nicht nachgewiesen werden kann, obwohl er vorhanden ist. Möchte man also die Nullhypothese beweisen (also dass Symmetrie *keine* Rolle für die Attraktivität spielt), dann muss man nur eine Stichprobe wählen, die klein genug ist, und erhält dadurch keine oder nur wenige signifikante Ergebnisse. Bei der vorliegenden Untersuchungsreihe wurden jedoch genau aus diesem Grund für jedes Experiment genau 64 Versuchspersonen getestet, um dem Vorwurf vorzubeugen, über eine Veränderung der Probandenzahl Ergebnisse in die eine oder andere Richtung provozieren zu wollen.

Die Vor- und Nachteile des Verfahrens *ohne* Einbeziehung der Urteilsvarianz verhalten sich invers zu den Vor- und Nachteilen der Auswertung *mit* Urteilsvarianz. Für das Weglassen der Urteilsvarianz spricht die hohe Empfindlichkeit des Verfahrens. Selbst winzig kleine Attraktivitätsunterschiede können dadurch noch nachgewiesen werden. Wenn man argumentiert, dass es vor allem darauf ankomme, das „Signal im Rauschen der Daten“ nachzuweisen, dann kann man dieses Vorgehen rechtfertigen. Zusätzlich kann man natürlich auch andere Studien zitieren, die es ebenso machen.

Die inhaltliche Begründung für das Weglassen der Urteilervarianz ist, dass man Attraktivität auf der Ebene der Gesichter betrachtet. Man betrachtet Attraktivität also als eine Eigenschaft, die den Dingen selbst anhaftet, wie z. B. die Masse, die Größe oder das Volumen eines Objekts. Wenn also eine Gesichtsvariante A beispielsweise den Attraktivitätswert 4.3 bekommt und die Variante B den Wert 4.2, dann wird A als attraktiver als B bezeichnet. Dass ein Attraktivitätsunterschied von 0.1 auf einer Skala von 1 bis 7, Streuungen von in der Regel über 1 und einer Urteileranzahl von 64 Personen statistisch überhaupt nicht signifikant ist, spielt bei dieser Betrachtung keine Rolle mehr. Wie

die Zahlen, die die Gesichtereigenschaft Attraktivität ausdrücken, zu Stande kamen, und dass sie aus Befragungen von Menschen resultieren, ist dabei „vergessen“.

Aus methodischer Sicht ist dies jedoch bedenklich, denn man tut dabei so, als sei die Messung von Attraktivität überhaupt kein Problem. So einfach ist es jedoch nicht. Man könnte die Art und Weise, wie die Attraktivitätswerte entstanden, dann ausblenden, wenn die Messung hoch reliabel wäre. Das Ermitteln der Größe eines Objekts ist zwar gewiss auch mit einem Messfehler behaftet. Misst man beispielsweise die Länge eines Gesichts dreimal hintereinander mit einem Maßband, wird man wahrscheinlich auch nicht exakt zum selben Ergebnis kommen. Benutzen verschiedene Personen dasselbe Maßband wird es ebenfalls kleine Unterschiede in den Messergebnissen geben. Doch diese Unterschiede (statistisch quantifizierbar durch die Standardabweichung) sind sehr klein und spielen bei den Größenunterschieden zwischen verschiedenen Gesichtern keine Rolle. Bei Attraktivitätsbewertungen ist die Urteilerübereinstimmung zwischen den Versuchspersonen viel geringer. Sie unterscheiden sich in der Strenge ihres Attraktivitätskriteriums, in der Ausnutzung des Spektrums der Attraktivitätsskala und natürlich in ihren subjektiven Präferenzen (Henss, 1992; Hönekopp, 2006; Hönekopp, Becker & Oswald, 2006). Die Messung des Attraktivitätswerts eines Gesichts ist daher viel weniger reliabel.

Das Problem des Ausblendens der Urteilervarianz ist weniger die statistische Analyse selbst, die nur noch mit zusammengefassten (z. B. gemittelten) Daten rechnet. Auch die resultierende Interpretation („Gesichter mit der Eigenschaft A sind attraktiver als Gesichter mit der Eigenschaft B“) lässt sich nicht kritisieren, da sie formal korrekt ist. Schwierig sind jedoch die Missverständnisse, die daraus leicht entstehen können. Denn die Aussage bezieht sich auf die Ebene der Gesichter, nicht auf die Ebene der Versuchspersonenurteile. Denn auf der Ebene der Versuchspersonenurteile gibt es erhebliche Meinungsunterschiede, die bei der Ermittlung des Ergebnisses ignoriert wurden. Die eigentliche Effektstärke ist daher in Wirklichkeit weit kleiner und die Ergebnisse weniger eindeutig als die Statistik suggeriert.

Es liegt auf der Hand, dass ein signifikantes, eindeutiges Ergebnis mit einer großen Effektstärke im Interesse der meisten Attraktivitätsforscher ist, schon deswegen, weil nicht signifikante Ergebnisse nur schwer publizierbar sind. Insofern ist es auch nicht verwunderlich, dass solche Methoden beliebt sind, bei denen Urteilsvarianzen vor der statistischen Analyse durch Mittelwertbildung (oder anderweitiges Zusammenfassen, z. B. bei Paarvergleichsexperimenten Bilden eines Prozentwertes pro Urteiler für eine anschließende Analyse der Prozentwerte mit einem Ein-Stichproben-*t*-Test) eliminiert werden.

Auch wenn eine Datenanalyse ohne Berücksichtigung der Urteilervarianzen selbstverständlich nicht falsch ist, so fällt es angesichts der Vorteile für den Versuchsleiter doch schwer, darin nicht einen statistischen Trick zu sehen.

Vielleicht ist eine Synthese dieser beiden konträren Sichtweisen dadurch möglich, indem man stärker auf den eigentlichen Zweck der Fragestellung fokussiert. Für eine grundlagenorientierte Untersuchung, bei der es tatsächlich „nur“ darum geht, ein „Signal im Rauschen“ nachzuweisen, wenn es „nur“ um das Überprüfen einer Theorie in einem akademischen Kontext geht, mag es angemessen sein, eine möglichst empfindliche Auswertungsmethode zu verwenden, d. h. die Urteilsvarianz *nicht* einzubeziehen. Die Symmetrie-Hypothese hat ja einen evolutionsbiologischen Hintergrund und postuliert, dass Symmetrie deswegen als attraktiv empfunden wird, weil sie eine hohe genetische Qualität eines potenziellen Fortpflanzungspartners signalisiert (Gangestad & Thornhill, 1997; Gangestad et al., 1994; Palmer & Strobeck, 1986; Thornhill & Gangestad, 1999; Thornhill & Moller, 1997) und es für ein Individuum einen evolutionären Fitnessvorteil bedeute, sich für einen solchen Partner zu entscheiden. Unabhängig von der Frage, ob diese Theorie nun stimmt oder nicht, muss man zugestehen, dass aus evolutionsbiologischer Sicht tatsächlich auch ein verschwindend kleiner Effekt im Sinne eines Fitnessvorteils evolutionär wirksam sein kann. Über einen Zeitraum von Zehntausenden bis Millionen von Jahren kann dies tatsächlich dazu führen, dass sich solch eine Eigenschaft im Genpool einer Spezies durchsetzt. Aus evolutionsbiologischer Sicht kann ein Effekt gar nicht so klein sein, als dass er evolutionär nicht doch wirksam sein könnte. Dementsprechend ist es nachvollziehbar, dass vor solch einem Hintergrund auch zu den empfindlichsten Methoden gegriffen wird, die vorhanden sind, um einen solchen kleinen Effekt noch nachweisen zu können.

Anders sieht dies jedoch aus einer alltagsnahen oder angewandten Perspektive aus, z. B. für eine Person, die das Ziel hat, die Attraktivität eines Gesichts zu erhöhen (des eigenen oder das eines Kunden oder Patienten). Hier steht eher die Frage im Vordergrund, wie sich die Veränderung eines Gesichts, beispielsweise durch Maßnahmen der ästhetischen Medizin, auf die Attraktivitätsbewertung auswirkt. Einen solchen Eingriff kann man durch die Verwendung von Vorher-Nachher-Bildern einfach evaluieren, auch für ein einzelnes Gesicht. Bei dieser Art der Datenerhebung spielt dann die Urteilsvarianz der Versuchspersonen wieder eine Rolle. Da bei einem normalen Gesicht Asymmetrien eher unauffällig sind, werden bei einer Evaluation des Eingriffs auch keine großen Veränderungen der Attraktivitätsbewertung zu erwarten sein. Viele Beurteiler werden zwischen Vorher- und Nachher-Bild gar keinen Unterschied feststellen (z. B. bei einem Paarver-

gleich), andere werden die Nuancen von Symmetrie bzw. Asymmetrie zwar erkennen, doch sie werden sich kaum auf die Gesamtattraktivität auswirken, da andere Merkmale, die ebenfalls Attraktivität beeinflussen, eine weit größere Wirkung haben. Dieses Urteilsverhalten wird durch eine Datenanalyse, die die Urteilervarianz berücksichtigt, korrekt erfasst, und auch die ermittelten kleineren Effektstärken spiegeln die Realität besser wider.

Zuletzt ist auch noch auf eine Besonderheit der *korrelativen* Untersuchungen (*Sym 7* bis *9*) hinzuweisen. Wie aus Tabelle 15 (S. 40) zu entnehmen ist, sind hier nur Ergebnisse vorhanden, bei denen die Varianz von Versuchspersonenurteilen *nicht* enthalten ist. Dies liegt daran, dass in allen drei Experimenten eine Korrelation über alle Stimulusgesichter ($N = 30$) gerechnet wurde, in die nur Mittelwerte eingingen: Zum einen der Attraktivitätsmittelwert jedes Gesichts und zum anderen ein gemittelter Wert aus Symmetrie- (bzw. Ähnlichkeits)urteilen. Alle diese Urteile wurden durch Versuchspersonenratings erhoben und unterliegen damit denselben Problemen, die bereits oben bei den Attraktivitätsratings innerhalb der Experimente diskutiert wurden: Die Varianz innerhalb der Urteile geht bei einer bloßen Korrelation der Mittelwerte verloren.

Eine Möglichkeit der gezielten Beeinflussung der Ergebnisse liegt in der Anzahl der befragten Versuchspersonen, die in diesem Fall auch die Größe des erzielten Effekts beeinflusst, nicht nur die Signifikanz des Ergebnisses. Die erzielte Effektgröße steigt hier mit zunehmender Anzahl der Versuchspersonen, weil dadurch die Reliabilität steigt. Bei Befragung einer großen Stichprobe nähert sich der Mittelwert des Ratings zunehmend an den „wahren“ Wert der Eigenschaft an. Bei einer kleinen Stichprobe hingegen fallen Zufallseinflüsse (z. B. individuelle Vorlieben der Probanden) stärker ins Gewicht. Nachweisen lässt sich dies, indem man für die drei Experimente den Datensatz in unterschiedlich große Teilstichproben unterteilt, aus diesen Teilstichproben jeweils Mittelwerte berechnet und diese dann miteinander korreliert. Tabelle 17 zeigt für das Experiment *Sym 7* (Direktes Rating der Symmetrie) die Ergebnisse der Korrelation, simuliert für unterschiedlich große Stichproben.

Tabelle 17: Simulation der Korrelationskoeffizienten (= Effektstärken) für *Sym 7* unter der Annahme unterschiedlich großer Versuchspersonen-Stichproben

		Mittelwert für Attraktivitäts-Rating, berechnet aus ...				
		<i>N</i> = 4	<i>N</i> = 8	<i>N</i> = 16	<i>N</i> = 32	<i>N</i> = 64
Mittelwert für Symmetrie-Rating, berechnet aus ...	<i>N</i> = 4	.54	.57	.60	.61	.61
	<i>N</i> = 8	.69	.72	.75	.73	.75
	<i>N</i> = 16	.71	.74	.76	.74	.77
	<i>N</i> = 32	.71	.72	.75	.72	.76
	<i>N</i> = 60	.72	.75	.76	.75	.78

Anmerkung: Die Bildung der Teilstichproben erfolgte in der Reihenfolge, in der die Daten erhoben wurden, d. h. bei *N* = 4 sind es die ersten 4 Datensätze der Gesamtstichprobe, bei *N* = 8 die ersten 8 Datensätze usw. Für alle Korrelationen ist $p < .01$. Das *N* der Korrelationen ist in allen 30, da es immer 30 Gesichter sind.

Die Tabelle zeigt, dass sich allein über die Stichprobengröße die Effektgröße, also die Höhe der Korrelation beeinflussen lässt (und eben nicht nur die Signifikanz). Bei einer sehr kleinen Stichprobe von jeweils vier Versuchspersonen beträgt der Zusammenhang zwischen Symmetrie-Rating und Attraktivität $r = .54$, $p = .002$, bei einer der hier verwendeten Stichproben von 60 und 64 Versuchspersonen hingegen $r = .78$, $p < .001$. Dass in diesem Fall auch bei einer so winzigen Gesamtstichprobe von acht Beurteilern die Korrelation trotzdem so hoch ausfällt, ist wohl eher Zufall – bei anderen kleinen zufälligen Teilstichproben könnte der Zusammenhang noch weit schlechter ausfallen.

Für Forscher mit der Absicht, die Nullhypothese zu beweisen, also dass Symmetrie für Attraktivität *keine* Rolle spielt, bietet der Umstand, dass die Effektstärke mit kleiner Probandenzahl aufgrund der geringeren Reliabilität sinkt, auch Möglichkeiten zur Manipulation der Ergebnisse. Er muss dafür nur eine kleinere Stichprobe untersuchen. Etablierte Standards, wie groß eine solche Stichprobe sein sollte, gibt es nicht. Angegeben wird lediglich ein Conbach's α , das mit steigender Rater-Zahl ebenfalls steigt. Es liegt häufig über $\alpha = .90$ (Rhodes et al., 1998; Rhodes, Yoshikawa et al., 2001), manchmal auch nur über $\alpha = .79$ (Rhodes, Zebrowitz et al., 2001), doch auch ein sehr mäßiger Wert von z. B.

$\alpha = .63$ (bei 111 Beurteilern!) (Simmons et al., 2004) wird von den Autoren nicht als problematisch erachtet (bei allen hier durchgeführten Untersuchungen liegen sie bei über .92, meist bei über .98); trotzdem werden auf Grundlage dieser Daten Mittelwerte berechnet, die Korrelation nur mit diesen Daten gerechnet und die Problematik einer geringen Reliabilität nicht diskutiert. Dass Cronbach's α ohnehin kein aussagekräftiges Maß für Urteilerübereinstimmung bei Attraktivitätsbeurteilungen ist, wurde schon früh von Henss (1992) und später von Hönekopp (2006) klar nachgewiesen.

Prinzipiell gäbe es auch noch eine andere Möglichkeit, wie man den Zusammenhang zwischen Symmetrie-Rating und Attraktivitätsrating korrelativ auswerten könnte, ohne dabei die Urteilsvarianz zu eliminieren. Man könnte dafür in einem *Within-Subjects Design* jede Versuchsperson die Symmetrie jedes Gesichts sowie in einem zweiten Durchgang (nach längerer Pause) die Attraktivität jedes Gesichts bewerten lassen. Anschließend berechnet man für jede Versuchsperson separat den individuellen Zusammenhang zwischen dem eigenen Symmetrie-Rating und dem eigenen Attraktivitätsrating und erhält dadurch so viele Korrelationskoeffizienten wie man Versuchspersonen in der Stichprobe hat. Das Gesamtergebnis ist dann der Median der Korrelationskoeffizienten. Mit dieser Methode wäre der zu erwartende Effekt der Symmetrie deutlich kleiner, da zufällig verteilte Fehler um den „wahren“ Wert sich nicht durch eine Mittelwertbildung gegenseitig aufheben. Bisher wurde diese Methode noch bei keiner Untersuchung zur Symmetrie verwendet, obwohl es durchaus Studien mit *Within-Subjects Design* und direkten Symmetrie-Ratings gibt (Rhodes et al., 1998). Da jedoch bei dieser Methode Effektstärken kleiner ausfallen und Ergebnisse schwerer signifikant werden, ist es nicht besonders wahrscheinlich, dass dieses Verfahren häufig zur Anwendung kommt.

Hinsichtlich der Ergebnisse der korrelativen Studien muss man also feststellen, dass aufgrund des Wegfalls von Urteilsvarianz die Effektstärken größer erscheinen, als sie in Wirklichkeit sind. Das bedeutet für die ohnehin schon geringen Effektstärken von $r = .47$ und $r = .48$ in den Studien *Sym 8* und *9* (basierend auf Ähnlichkeitsratings), dass der wahre Beitrag der Symmetrie zur Attraktivität in Wirklichkeit noch einmal kleiner wäre und dann gegen Null geht. Für den großen Effekt von $r = .78$ aus der Untersuchung *Sym 7* bedeutet es, dass der Effekt deutlich kleiner (vielleicht mittelgroß), jedoch immer noch klar vorhanden wäre. Dies betrifft jedoch nur die Konsequenzen hinsichtlich Datenauswertung – dass die Art der Datenerhebung (direktes Beurteilen der Symmetrie eines Gesichts) eine invalide Methode ist, wurde bereits in den Kapiteln 4.10.4 und 5.2.2 ausführlich dargelegt.

Zusammenfassend bleibt festzuhalten, dass die hoch empfindliche Auswertungsmethode *ohne* Berücksichtigung der Urteiler-Varianz eher in einem akademischen Kontext und bei der Überprüfung von theoretischen Fragestellungen angemessen ist. Die weniger empfindliche Auswertungsmethode *mit* Berücksichtigung der Urteiler-Varianz hingegen kommt ohne statistische Kunstgriffe wie einem vorgeschalteten Zwischenschritt mit Mittelwertbildungen aus und sie analysiert *alle* erhobenen Daten direkt anstatt nur einer groben Zusammenfassung dieser Daten. Sie spiegelt nicht durch beeindruckend niedrige Signifikanzwerte eine Sicherheit vor, die in Wirklichkeit nur auf zahlreichen kleinen, aber nicht signifikanten Einzeleffekten beruht und birgt dadurch auch ein geringeres Risiko, Methodenartefakte für echte Effekte zu halten. Zudem berücksichtigt sie die Tatsache, dass die Attraktivität eines Gesichts keine absolut präzise bestimmbare Eigenschaft ist, sondern dass sie ein Gesamturteil einer großen Menge von Beurteilerpersonen ist, die sich zu einem gewissen Anteil auch in ihren Ansichten unterscheiden. Damit wird sie auch der eigentlichen Natur von Schönheit eher gerecht, die keine Eigenschaft ist, die den Objekten selbst anhaftet, sondern eine Wahrnehmung und eine Bewertung, die im Kopf von Menschen entsteht.

5.3 Fazit zum Effekt der Symmetrie auf die Attraktivität von Gesichtern und Konsequenzen für künftige Studien

Die durchgeführte Studienreihe mit neun verschiedenen Untersuchungen zum Effekt der Symmetrie auf die Attraktivität von Gesichtern hat klar gezeigt, dass alle Untersuchungsergebnisse zu diesem Thema in Richtung, Größe und Signifikanz ihrer Effekte stark von der jeweils verwendeten Untersuchungsmethode abhängen. Dies ist das wichtigste Ergebnis der Untersuchungsreihe. Es ist besonders aussagekräftig, weil bei allen Studien Gesichter derselben 30 Stimuluspersonen verwendet wurden und alle Versuchspersonenstichproben hinsichtlich Größe und Zusammensetzung gleich waren. Die erzielten Unterschiede können daher nur auf die verschiedenen Methoden zurückzuführen sein.

Die Unterschiede bei Richtung und Größe der Effektstärke sind teilweise so erheblich, dass sich der Eindruck aufdrängt, dass man es hier mehr mit Artefakten als mit eigentlichen Effekten der Symmetrie zu tun hat. Tabelle 15 (S. 205) fasst die Effektstärken der einzelnen Untersuchungen auf einen Blick zusammen.

In Kapitel 5.2 wurden ausführlich Vor- und Nachteile der einzelnen Methoden diskutiert. Es wurde dargelegt, wie die Art des Stimulusmaterials, die Art des Versuchsdesigns und

die Art der Datenanalyse Richtung, Effektstärke und Signifikanz der Ergebnisse beeinflussen. Für künftige Untersuchungen zum Einfluss von Symmetrie auf Attraktivität bleibt dies nicht ohne Folgen. Jeder, der eine Untersuchung zu diesem Thema plant, muss sich selbstverständlich für irgendeine Methode entscheiden. Er kann es jedoch künftig nicht mehr mit dem bloßen Hinweis auf bestimmte andere Autoren tun, die diese Methode ebenfalls verwendet haben, denn jede der hier beschriebenen Methoden wurde schon von irgendwem einmal verwendet. Künftig ist auch eine inhaltliche Begründung notwendig, warum man sich für eine bestimmte Methode entscheidet, denn es ist nun mit der vorliegenden Arbeit klar belegt, welche Effekte durch welche Methode begünstigt werden.

Durch einen Blick auf Tabelle 15 (S. 205) kann ein (unseriöser) Forscher, der beabsichtigt, einen Effekt in bestimmter Richtung (positiv oder negativ) und Größe (Nulleffekt, klein, mittel, groß) zu produzieren, vorinformieren und danach eine entsprechende Methode auswählen. Insofern beinhalten die Ergebnisse dieser Untersuchungsreihe auch ein gewisses Missbrauchspotenzial. Andererseits kann und muss künftig aber auch von ihm verlangt werden, dass er seine Wahl begründet. Damit kommt er nicht um eine Auseinandersetzung mit den Besonderheiten der einzelnen Methoden herum, die im Kapitel 5.2 ausführlich dargelegt wurden.

Gleichzeitig wurde aber auch in der methodischen Diskussion (Kapitel 5.2) klar herausgearbeitet, dass nicht alle Methoden gleichwertig sind. Es gibt bessere und schlechtere Methoden, und manche sind sogar so schlecht, dass man künftig besser ganz auf ihre Verwendung verzichten sollte. Im Folgenden sollen überblicksartig die wichtigsten Konsequenzen und Forderungen zusammengefasst werden. Auf eine detaillierte Begründung muss an dieser Stelle aus Platzgründen verzichtet werden – sie findet sich jedoch in den jeweiligen Kapiteln, auf die verwiesen wird.

Die Verwendung von Chimärengesichtern als Stimuli für Attraktivitätsbewertungen ist ökologisch invalide und sollte künftig unterbleiben. Symmetrisierte Chimärengesichter sind im Durchschnitt unattraktiver, weil sich bei den meisten Gesichterfotos Bildverarbeitungsartefakte nicht vermeiden lassen. Diese Artefakte lassen sich in vier Kategorien einteilen, nämlich (1) Artefakte durch unnatürliche symmetrische Verdoppelungen der Textur, (2) Artefakte, die aus einem gedrehten Gesicht bei der Aufnahme des Fotos resultieren, (3) Artefakte durch einen geneigten Hals bei der Aufnahme des Fotos und (4) Artefakte durch tatsächliche Asymmetrien im Originalgesicht um die Symmetrieachse herum. Es wird daher bei Chimärengesichtern weniger der Effekt perfekter Symmetrie

gemessen, sondern vielmehr der Effekt von störenden Bildverarbeitungsartefakten (vgl. Kapitel 5.2.1.1).

Die Verwendung von „Blends“ als Stimuli für Attraktivitätsbewertungen ist ökologisch invalide und sollte künftig unterbleiben. Stattdessen sollten „Warps“ verwendet werden, weil bei „Warps“ nur die Gesichtsproportionen symmetrisch optimiert werden, während bei „Blends“ auch die Farbwerte der Textur gemittelt werden. Durch die Farbwertmittlung entsteht ein Weichzeichnungseffekt, der die Haut makelloser erscheinen lässt. Der Effekt der Symmetrisierung ist daher konfundiert mit dem Effekt einer makelloseren Haut (vgl. Kapitel 5.2.1.2).

Das verwendete Stimulusmaterial sollte das gesamte Spektrum menschlicher Attraktivität widerspiegeln, auch im hoch attraktiven Bereich. Werden wie üblich nur Stimuli von niedriger oder mittelmäßiger Attraktivität verwendet (erkennbar an den absoluten Werten von Attraktivitätsratings), sind Schlussfolgerungen über den eigentlichen Untersuchungsgegenstand, nämlich (hohe) Attraktivität, problematisch und Ergebnisse damit ökologisch weniger valide. Zudem hat der Vergleich der Attraktivitätsbewertungen von „normalen“ Frauengesichtern und Modelgesichtern gezeigt, dass die Ausgangsattraktivität eines Originalgesichts mit der unabhängigen Variablen Symmetrie bei einer experimentellen Manipulation interagiert (vgl. Kapitel 5.2.1.3).

Es sollte ein *Between-Subjects Design* mit Ratingskalen verwendet werden, denn dieses ist weniger reaktiv und ökologisch valider als ein Paarvergleichsexperiment. Beim *Between-Subjects Design* ist es für die Probanden viel schwieriger zu erraten, was die Hypothese des Experiments ist, so dass es unwahrscheinlicher ist, dass ihre Erwartungshaltung ihr Urteilsverhalten beeinflusst. Zudem wird ein absoluter Attraktivitätswert gemessen, der etwas über den absoluten Effekt auf die Attraktivitätsveränderung durch die experimentelle Manipulation aussagt, was v. a. aus angewandter Perspektive wichtig ist. Die geringere Power eines *Between-Subjects Designs* im Vergleich zu einem Paarvergleichsexperiment sollte durch eine entsprechend größere Versuchspersonenanzahl ausgeglichen werden (vgl. Kapitel 5.2.2).

Ein direktes Rating von Symmetrie durch Versuchspersonen ist strikt abzulehnen, denn Versuchspersonen sind nicht in der Lage, eine Symmetriebewertung unabhängig von der Attraktivität eines Gesichts vorzunehmen, wodurch das Symmetrie-Urteil in hohem Maße mit dem Attraktivitätsurteil konfundiert ist. Der klarste Beleg dafür, dass Probanden beim Symmetrie-Rating eigentlich die *Attraktivität* eines Gesichts bewerten, ist, dass sie auch bei perfekt symmetrischen „Warp“-Gesichtern Symmetrie-

Unterschiede wahrzunehmen glauben, obwohl in Wirklichkeit keine vorhanden sind. Diese Symmetrie-Urteile perfekt symmetrischer Gesichter korrelieren mit $r = .75$ mit den Symmetrie-Urteilen der Originalgesichter und mit $r = .87$ mit Attraktivität, obwohl der Zusammenhang in beiden Fällen Null sein müsste (vgl. Kapitel 4.10.4).

Eine Bestimmung von Symmetrie durch Ähnlichkeitsbewertungen separater Gesichtshälften ist möglicherweise kritisch, da die Validität dieses Verfahrens nicht klar beurteilt werden kann. Es bleibt unklar, inwieweit Versuchspersonen diese Aufgabe überhaupt zu leisten im Stande sind. Zudem ist zu vermuten, dass bei der Präsentation von separaten Hälften Texturunterschiede (z. B. Pickel, Muttermale) stärker in den Fokus der Aufmerksamkeit rücken, weil es durch das Trennen der Gesichtshälften (und Spiegeln einer Hälfte) schwieriger ist, Unregelmäßigkeiten der Proportion zu beurteilen. Da Texturunterschiede zwischen Gesichtshälften jedoch in der Attraktivitätsforschung nicht als (fluktuierende) Asymmetrie verstanden werden, stellt dies für die Validität der Methode ein Problem dar.

Bei der Datenauswertung sollten Verfahren verwendet werden, die die Urteilsvarianz berücksichtigen. Zwar ist diese Methode weniger empfindlich als Methoden *ohne* Berücksichtigung der Urteilsvarianz, aber dafür kommt sie (1) ohne statistische Kunstgriffe wie einem vorgeschalteten Zwischenschritt mit Mittelwertbildungen aus und sie analysiert (2) *alle* erhobenen Daten direkt anstatt nur einer groben Zusammenfassung dieser Daten. Sie spiegelt (3) nicht durch beeindruckend niedrige Signifikanzwerte eine Sicherheit vor, die in Wirklichkeit nur auf zahlreichen kleinen, aber nicht signifikanten Einzeleffekten beruht und birgt dadurch (4) auch ein geringeres Risiko, Methodenartefakte für echte Effekte zu halten. Aufgrund der geringeren Teststärke sind jedoch entsprechend große Versuchspersonen-Stichproben nötig, um einen möglichen vorhandenen Effekt auch nachweisen zu können.

Wenn man für künftige Untersuchungen zur Symmetrie alle diese Empfehlungen berücksichtigt, dann ist die Palette an verwendbaren Methoden stark reduziert. Die Methode, gegen die sich am wenigsten einwenden lässt, ist die experimentelle Manipulation von Symmetrie mit dem „Warp“-Verfahren (*Sym 5* und *6*). Dem Rating-Verfahren im *Between-Subjects Design* ist dabei gegenüber dem Paarvergleichsexperiment der Vorzug zu geben, in jedem Fall aber sollte die Urteilsvarianz bei der Datenauswertung einbezogen werden (z. B. Auswertung wie in *Sym 6* mit MANOVA).

Betrachtet man die Ergebnisse, die hier mit diesen beiden Verfahren erzielt wurden, dann ist von einem Effekt der Symmetrie auf die Attraktivität auszugehen, der im Bereich zwi-

schen $r = .17$ und $r = -.09$ liegt – eine Effektstärke also, die marginal ist und überdies auf größtenteils nicht signifikanten Einzelergebnissen beruht. Wenn es also überhaupt möglich ist, aus der Vielzahl an verschiedenen Untersuchungen und zum Teil völlig widersprüchlichen Ergebnissen dennoch zu einer abschließenden Schlussfolgerung zu kommen, dann ist es, basierend auf den Ergebnissen dieser beiden methodisch sinnvollsten Experimente (Sym 5 und 6), diese:

Symmetrie spielt für die Attraktivität eines Gesichts praktisch keine Rolle.

Es spricht vieles dafür, dass ganz normale Asymmetrien in einem normalen Ausmaß, wie sie in einem normalen Gesicht natürlicherweise vorkommen, keinen Einfluss auf die Attraktivitätsbewertung haben, sondern einfach egal sind. *Starke* Asymmetrien mögen gewiss Attraktivität senken, doch dies ist erstens trivial und zweitens nicht Gegenstand der Symmetrie-Hypothese. Denn diese behauptet nicht, dass Asymmetrie ein Gesicht unattraktiv macht, sondern dass Symmetrie ein Gesicht attraktiv macht. Im Übrigen ließen sich Dutzende von Eigenschaften aufzählen, die ein Gesicht hässlicher machen oder es gar entstellen können. Schönheit lässt sich jedoch nicht aus dem Nicht-Vorhandensein eines entstellenden Merkmals erklären, sonst wären schließlich alle Menschen schön bzw. gleich schön, die dieses entstellende Merkmal nicht besitzen.

Vielleicht ist es auch endlich an der Zeit, die Symmetrie-Hypothese als widerlegt anzusehen. In der durchgeführten Untersuchungsreihe kranken alle Methoden, die zu mittelstarken oder starken Effekten führten, an schweren methodischen Mängeln. Die Mehrzahl der Methoden führte ohnehin zu *negativen* Effekten, die jedoch ebenfalls teilweise methodische Schwächen aufweisen (z. B. statistische Tricks mit der Eliminierung von Urteilsvarianz), und die Effektstärken der validesten Methoden sind marginal.

Doch vermutlich wird es nicht dazu kommen. Seit über 20 Jahren werden Forschungsergebnisse produziert, die den Einfluss von Symmetrie auf Attraktivität zu belegen scheinen, in Wirklichkeit jedoch zum Großteil nur das Resultat von Methodenartefakten sind. Spätestens seit der Metaanalyse von Rhodes (2006) gilt es als Konsens, dass Symmetrie einen Einfluss auf die Gesichtsattraktivität hat (mittlere Effektstärke $R = .25$, S. 202). Es dürfte nicht einfach sein, die herrschende Meinung dadurch zu ändern, indem man Ergebnisse nachweist, die zu einem Null-Effekt gekommen sind. Vermutlich dürfte es ebenso schwer sein, ein solches nicht signifikantes Ergebnis überhaupt publiziert zu bekommen. Umso wichtiger ist jedoch die kritische Auseinandersetzung mit den jeweiligen Methoden, die bisher zum Einsatz kamen. Unter den 23 Studien, die Rhodes ausgewertet hat, befindet sich keine einzige, die einer strengen methodischen Betrachtung Stand hält!

Selbst Experimente mit qualitativ hochwertigem Stimulusmaterial (Verwendung von „Warps“), kommen nicht ohne den statistischen Trick der Urteilsvarianz-Eliminierung aus (Jones et al., 2007; Little & Jones, 2003; Perrett et al., 1999), da andernfalls bei den gefundenen Effektstärken ihre Ergebnisse statistisch nicht signifikant geworden wären.

Aus diesem Grund ist Rhodes' Metaanalyse mit einer gewissen Skepsis zu beurteilen. Ihr unbestreitbarer Wert besteht darin, die bis dato verfügbaren Studien zum Thema Symmetrie und Attraktivität zusammengetragen zu haben. Bei der kritischen Würdigung der verwendeten Methodik dieser Studien bleibt Rhodes Analyse jedoch sehr oberflächlich und verkennt zahlreiche grundlegende Schwächen dieser Studien völlig. Bei einer Wissenschaftlerin, die selber in zahlreichen eigenen Untersuchungen kein Problem damit hatte, die Symmetrie von Gesichtern durch Befragen von Versuchspersonen zu messen – die mit Abstand naivste und invalideste aller untersuchten Methoden – wundert dieses mangelnde methodische Problembewusstsein nicht.

Aufgrund des herrschenden Konsenses in der Szene der Attraktivitätsforscher, dass Symmetrie attraktiv macht, und des Umstands, dass nicht-signifikante Ergebnisse und Nulleffekte nur sehr schwer publizierbar sind, erscheint es unwahrscheinlich, dass in Zukunft durch eine größere Anzahl von Studien mit negativem Befund die Hypothese vom Einfluss der Symmetrie aufgegeben wird. Dies ist wohl nur möglich, wenn die bisher verwendeten Methoden, mit denen bisherige Untersuchungen durchgeführt wurden, auf den Prüfstand kommen und kritisch hinterfragt werden.

Wahrscheinlicher ist es jedoch, dass die Forscherszene allmählich das Interesse verliert zu untersuchen, *ob* Symmetrie attraktiv macht, da dies ja vermeintlich schon klar sei. Stattdessen verlagert sich das Forschungsinteresse bereits seit einigen Jahren zunehmend auf die Frage, *warum* Symmetrie attraktiv macht. Mit dem *Warum* ist bei der großen Mehrheit der Forscher die ultimate Ursache gemeint, also nicht die unmittelbaren (proximaten) Ursachen auf der Ebene der Stimulusverarbeitung, sondern der evolutionsbiologische Hintergrund. Symmetrie sei als Attraktivitätsmerkmal ein Signal für „gute Gene“ (*good genes*) und folglich müsse es einen Zusammenhang mit Gesundheit, Langlebigkeit oder Fruchtbarkeit geben (Grammer, Fink, Moller & Manning, 2005; Grammer, Fink, Moller & Thornhill, 2003; Thornhill & Gangestad, 1999). Wie eine Metaanalyse von Weeden und Sabini zeigt, konnte ein solcher Zusammenhang jedoch bislang nicht nachgewiesen werden (Pawlowski, Boothroyd, Perrett & Kluska, 2008; Weeden & Sabini, 2005).

Stattdessen verweisen Attraktivitätsforscher auf angebliche Teilerfolge, z. B. dass es Zusammenhänge mit *wahrgenommener* Gesundheit (*perceived health*) gebe, wenn auch nicht mit *tatsächlicher* Gesundheit – was jedoch eher auf einem Halo-Effekt der Schönheit beruhen dürfte (Kalick et al., 1998). Einige verweisen auch auf Korrelationen zwischen Attraktivität und Langlebigkeit (Henderson & Anglin, 2003) (jedoch ohne den Beleg eines direkten Zusammenhangs mit den angeblichen Komponenten von Attraktivität wie Symmetrie oder auch Durchschnittlichkeit). Ein zusätzliches Kernproblem dieser Art von Forschung ist, dass man bei der Untersuchung von Zusammenhängen von Symmetrie (als objektiv vorhandene Eigenschaft von Personen) mit anderen Parametern die Symmetrie auch objektiv, reliabel und valide messen muss – experimentelle Bildmanipulationen mit Computersoftware (z. B. Symmetrisierung mit Morphing-Software) sind hier prinzipiell nicht möglich. Damit steht man wieder vor dem Problem, dass eine Quantifizierung von Symmetrie methodisch sehr schwierig ist, wie die Untersuchungen *Sym 7* bis *9* gezeigt haben. Und auch bei der Erfassung von Asymmetrie durch *Facial Landmarks* gibt es erhebliche methodische Probleme, die seit langem bekannt sind und in den Überblicksartikeln von Palmer und Strobeck (1986, 2003) ausführlich dargelegt wurden.

Bei all dem ist selbstverständlich der angewandte Aspekt der Attraktivitätsforschung schon längst aus dem Auge verloren. Die Forschung kreist um sich selbst, repliziert mit fragwürdigen Methoden immer dieselben Artefakte, bis diese nach einiger Zeit von der Mehrheit der Forscher als Tatsachen angesehen werden. Aus diesem Grund sagen die meisten Attraktivitätsstudien zur Gesichtssymmetrie weniger darüber aus, was ein Gesicht attraktiv macht, sondern mehr darüber, wie Forschung im akademischen Betrieb funktioniert.

6 Untersuchungen zur Hypothese des sexuellen Dimorphismus

6.1 Stand der Forschung zum Einfluss des sexuellen Dimorphismus

6.1.1 Begriffsdefinition des sexuellen Dimorphismus

Neben Durchschnittshypothese und Symmetriehypothese ist die dritte wichtige Hypothese zur Attraktivität von Gesichtern die Hypothese des sexuellen Dimorphismus (auch *sex-hormone-markers theory*). Sie besagt, dass ein Frauengesicht dann attraktiv ist, wenn es

besonders weiblich aussieht, und ein Männergesicht attraktiv ist, wenn es besonders männlich aussieht. Die Geschlechtstypikalität ist demzufolge ein Kriterium für Attraktivität. Damit widerspricht die Theorie auch der Durchschnittshypothese.

Typisch männliche Merkmale sind solche, deren Wachstum ab Beginn der Pubertät dem Einfluss von Testosteron unterliegt, v. a. ein ausgeprägtes Kinn, ein kräftiger Unterkiefer, ausgeprägte Wangenknochen, ausgeprägte Augenbrauenbögen, eine große Mittelzone des Gesichts (von den Augen bis zu den Nasenflügeln) und eine protrusive (nach vorne dringende) Nase (Enlow, 1989; Farkas, 1987; Thornhill & Gangestad, 1996).

Typisch weibliche Merkmale sind dementsprechend die gegenteiligen Merkmale. Bei Frauen wird die Ausprägung dieser Merkmale durch Östrogen gehemmt. Gleichzeitig gilt Östrogen als verantwortlich für die Ausprägung voller Lippen, die ein typisch weibliches Merkmal sind (Symons, 1995; Thornhill & Moller, 1997).

Eng verwandt mit der Hypothese des sexuellen Dimorphismus ist die Kindchenschema-Hypothese (*facial neoteny*), nach der Frauengesichter attraktiver sind, wenn ihre Gesichtsproportionen denen eines Kindes ähneln (Johnston & Franklin, 1993; Jones, 1995). Typisch kindliche Merkmale (Eibl-Eibesfeldt, 1997; Enlow, 1989; Lorenz, 1943) decken sich weitgehend mit typisch femininen, mit dem Unterschied, dass das kindliche Gesicht feminine Merkmale in übersteigerter Form beinhaltet. Es gibt jedoch auch einige Unterschiede: So haben Kinder zusätzlich rundliche Augen und Pausbacken; die Wangenknochen sind nicht ausgeprägt und die Lippen nicht voll (Berry & Brownlow, 1989; Berry & McArthur, 1985; Zebrowitz, 1997).

6.1.2 Hintergrund der Theorie des sexuellen Dimorphismus

Eine theoretische Begründung dafür, warum ein geschlechtstypisches Aussehen zu besseren Attraktivitätsbewertungen führt, liefert die Prototypentheorie. Ein sehr feminin aussehendes Frauengesicht beispielsweise sei gleichzeitig ein besonders guter Prototyp für die Kategorie Frauengesicht. Dasselbe gilt analog für Männergesichter. Die Anhänger der Prototypentheorie argumentieren, dass das prototypische Aussehen zu einer Wahrnehmungserleichterung führe, die von den Menschen als positiv erlebt werde (Winkielman et al., 2006; Winkielman et al., 2003), weshalb auch geschlechtstypisch aussehende Gesichter als attraktiver empfunden würden (für eine Diskussion siehe die Metaanalyse von Rhodes, 2006). Ein Vorteil dieser Erklärung ist, dass sie sowohl das männliche als auch das weibliche Attraktivitätsideal durch denselben Mechanismus erklärt.

Allerdings spielt die Prototypentheorie zur Erklärung des sexuellen Dimorphismus unter Attraktivitätsforschern praktisch keine Rolle und wird eher aus Höflichkeit oder der Vollständigkeit halber erwähnt. Die Diskussion wird hingegen von der evolutionspsychologischen Sichtweise dominiert, die in Gesichtsmerkmalen letztlich ein Signal für Fruchtbarkeit, Gesundheit, Parasitenresistenz oder allgemein „gute Gene“ sieht. Aufgrund der Tatsache, dass geschlechtsdimorphe Merkmale, also sekundäre Geschlechtsmerkmale, durch Hormone (Testosteron, Östrogen) verursacht werden, ist ein biologischer Hintergrund zunächst durchaus naheliegend.

Die einflussreichste Theorie ist dabei die Immunkompetenz-Handicap-Hypothese (*immunocompetence-handicap hypothesis*). Sie basiert auf dem Umstand, dass Testosteron zwar einerseits bei Männern die Ausprägung von sekundären Geschlechtsmerkmalen fördert, andererseits jedoch gleichzeitig das Immunsystem unterdrückt (Alexander & Stimson, 1988; Grossman, 1985; Peters, 2000; Wedekind, 1992; Zuk, Johnsen & Maclarty, 1995). Dadurch sollen Männer anfälliger für Infektionskrankheiten und Parasiten sein als Frauen (Folstad, Nilssen, Halvorsen & Andersen, 1989; Moller, Christe & Lux, 1999; Poulin, 1996; Thornhill & Gangestad, 1993; Thornhill & Moller, 1997) – wobei solche Belege meist nur aus Studien an Tieren stammen. Anstatt jedoch zu schlussfolgern, dass Testosteron-Marker (also ein typisch männliches Aussehen bei Männern) als negativer und damit unattraktiver empfunden werden müssten, da sie ja schließlich ein Indikator für Krankheitsanfälligkeit sind, drehen Evolutionsbiologen den Spieß um und argumentieren mit der Handicap-Hypothese (Folstad & Karter, 1992; Zahavi, 1975): Nur besonders gesunde Individuen mit einem (genetisch bedingten) besonders guten Immunsystem können es sich leisten, solch kostspielige Testosteron-Marker wie einen kräftigen Unterkiefer oder ein breites Kinn auszubilden. Wenn Gesundheit also eine genetische Basis hat, dann sei es für Frauen/Weibchen evolutionär adaptiv, solche Partner als Geschlechtspartner zu bevorzugen, die besonders prächtig ausgebildete sekundäre Geschlechtsmerkmale besitzen, da dadurch die Wahrscheinlichkeit steige, dass auch ihr Nachwuchs die genetische Voraussetzung für gute Gesundheit besitze. Problematisch an dieser Theorie sind drei Aspekte:

(1) Studien, die eine höhere Krankheitsanfälligkeit durch Testosteron fanden, basieren meist auf Untersuchungen mit Tieren (Moller et al., 1999). Trotz zahlreicher Versuche, den Zusammenhang zwischen sexuellem Dimorphismus und tatsächlicher (nicht wahrgenommener!) Gesundheit beim Menschen nachzuweisen, gibt es keine überzeugenden empirischen Belege, wie eine Metaanalyse zeigt (Weeden & Sabini, 2005).

(2) Die Handicap-Hypothese ist aus wissenschaftstheoretischer Sicht grundsätzlich problematisch, weil derselbe Zusammenhang (d. h. ein niedriger Testosteronspiegel bedingt eine bessere Immunabwehr), unterschiedlich interpretiert wird: Bei Frauen soll ein niedriger Testosteronspiegel gut sein, weil er direkt bessere Krankheitsresistenz bedingt. Bei Männern hingegen soll hingegen das Gegenteil dessen gut sein, nämlich ein hoher Testosteronspiegel, jedoch mit demselben Ziel, nämlich ebenfalls einer besseren Krankheitsresistenz. Wenn die Handicap-Hypothese wahr ist, warum sollte sie dann nur für Männer gelten und nicht auch für Frauen?

(3) Die Immunkompetenz-Hypothese wurde ursprünglich für Männer (bzw. männliche Tiere) aufgestellt. Ihre Befürworter gestehen die beschränkte Erklärungskraft für Frauen (bzw. weibliche Tiere) ein (Getty, 2002; Rhodes, Chan, Zebrowitz & Simmons, 2003). „Currently, there is no general prediction for the relationship between health and trait expression“ (Rhodes, Chan et al., 2003, S. 93).

Aus diesem Grund wird als Erklärung für die Präferenz von typisch weiblichen Gesichtszügen bei Frauen auch der Zusammenhang mit erhöhter Fruchtbarkeit angeführt (Symons, 1995; Thornhill & Gangestad, 1996). Ein hoher Östrogenspiegel in der Pubertät verursacht die Ausprägung von femininen Gesichtszügen. Der Östrogenspiegel in der Pubertät korreliert mit dem Östrogenspiegel im Erwachsenenalter und da ein hoher Östrogenspiegel die Wahrscheinlichkeit einer Schwangerschaft erhöhe, seien feminine Gesichtszüge ein Signal für Fruchtbarkeit bei Frauen (Johnston & Franklin, 1993; Symons, 1995; Thornhill & Moller, 1997).

6.1.3 Methoden zur Untersuchung der Theorie des sexuellen Dimorphismus

Wie auch bei der Untersuchung der Symmetrie gibt es methodisch mehrere Herangehensweisen: Die Geschlechtstypikalität von Gesichtern

- kann mit Hilfe von *Facial Landmarks* anthropometrisch vermessen werden (Cunningham, 1986; Cunningham et al., 1995; Johnston & Franklin, 1993; Jones & Hill, 1993; Koehler, Simmons, Rhodes & Peters, 2004),
- kann mit Hilfe von Computerprogrammen (z. B. Morphing-Software) experimentell manipuliert werden (Johnston et al., 2001; Perrett et al., 1998; Rhodes, Hickford & Jeffery, 2000), oder

- durch Versuchspersonen (z. B. mittels einer direkten Ratingskala) beurteilt werden (Bruce, Burton & Dench, 1994; Dunkle & Francis, 1990; Koehler et al., 2004; O'Toole et al., 1998; Rhodes et al., 2003).

Die Qualität dieser drei grundlegenden Methoden unterscheidet sich deutlich. Untersuchungen mit direkten Ratings von Maskulinität oder Femininität haben mit demselben Validitätsproblem zu kämpfen wie die bereits oben erwähnten direkten Ratings von Symmetrie, bei denen mit der Zusatzuntersuchung *Sym 7 Kontrolle* in Kapitel 4.10.4 klar nachgewiesen wurde, dass sie methodisch unsinnig sind. Man kann sich nie sicher sein, ob man hiermit tatsächlich z. B. die Maskulinität eines Gesichts misst oder vielmehr ein in der Bevölkerung vorhandenes Stereotyp, wonach ein Mann idealerweise männlich auszusehen hat. Es könnte daher durchaus sein, dass attraktiven Männergesichtern aufgrund dieses Klischees eine gesteigerte Männlichkeit zugeschrieben wird, obwohl ihre Gesichtszüge – anatomisch vermessen – möglicherweise gar nicht besonders männlich sind. Analoges gilt für Frauengesichter und Femininität. In diesem Fall wären UV und AV konfundiert. So überrascht es nicht, dass alle bisherigen Studien, bei denen die Geschlechtstypikalität durch Ratings von Originalgesichtern gemessen wurde, positive Belege für den Einfluss des Geschlechtsdimorphismus fanden.

Zu gegenteiligen Schlussfolgerungen gelangen hingegen bei Männergesichtern alle Studien, bei denen Männlichkeit (z. B. mit Morphing-Software) experimentell manipuliert wurde. Diese Methode führte stets zu dem Ergebnis: Je männlicher ein Männergesicht, desto *unattraktiver* ist es (zur Übersicht siehe Metaanalyse von Rhodes, 2006). Auch hier muss man jedoch einen kritischen Blick auf die Methodik werfen: In der Regel werden hypermaskuline Männergesichter durch die sogenannte Karikatur-Technik erzeugt, d. h. es werden bei einem normalen Männergesicht die Abweichungen zu einem männlichen Durchschnittsgesicht durch Morphing (Verzerrung der Gesichtsproportionen) zusätzlich verstärkt. Diese Methode simuliert jedoch nicht die komplexen, nichtlinearen Prozesse des durch Testosteron geförderten Schädelwachstums (Enlow, 1966, 1979; Grumbach, 2000; Johnston et al., 2001; Tanner, 1978) und führt daher (je nach Stärke des Morphings) zu unrealistisch und seltsam aussehenden Gesichtern, die im Extremfall tatsächlich wie Karikaturen aussehen und deswegen als unattraktiver bewertet werden.

Diese beiden Methoden zeigen klar: Wenn nicht nur die Größe, sondern auch noch die Richtung des erzielten Effekts davon abhängt, mit welcher Methode man eine Fragestellung untersucht hat, dann ist mindestens eine dieser Methoden fragwürdig. Möglicherweise kann dies die Vermessung von *Facial Landmarks* klären. Hierbei ist das Attraktivitäts-

urteil nicht von ungeeignetem (da schlecht manipuliertem) Stimulusmaterial negativ beeinflusst und die Vermessung von Gesichtern ist zumindest objektiv und reliabel. Zu beachten ist hier allerdings, dass es trotz objektiver anthropometrischer Messpunkte keine einheitlichen Standards gibt, welche Merkmale genau vermessen werden, ob und wie sie gewichtet werden und ab wann man ein Gesicht als typisch feminin oder maskulin bezeichnet. Das Problem ist hier ähnlich wie bei der Vermessung von Symmetrie mittels *Facial Landmarks*, bei der es ebenfalls verschiedene Operationalisierungsverfahren gibt.

6.1.4 Empirische Befunde zum Einfluss femininer Gesichtszüge bei Frauengesichtern

Untersuchungen zur Hypothese des sexuellen Dimorphismus gibt es deutlich weniger als zu den beiden anderen bedeutenden Hypothesen. Hinzu kommt, dass es hier offensichtlich nicht egal ist, ob das Stimulusmaterial aus Frauen- oder Männergesichtern besteht. Denn auch, wenn sich beispielsweise bestätigt, dass Frauengesichter attraktiv sind, wenn sie besonders weiblich aussehen, folgt daraus keineswegs, dass Männergesichter attraktiv sind, wenn sie besonders männlich aussehen. Während die Theorie bei Männergesichtern aufgrund uneinheitlicher empirischer Befunde sehr umstritten ist, gibt es hingegen mehrere Bestätigungen der Hypothese für Frauengesichter:

Einige Studien verwendeten dazu anthropometrische Vermessungen von weiblichen Originalgesichtern (Cunningham, 1986; Cunningham et al., 1995; Johnston & Franklin, 1993; Jones & Hill, 1993; Koehler et al., 2004), während andere nur weniger aussagekräftige Femininitäts-Ratings durch Versuchspersonen als Maß für Femininität nutzten (Bruce et al., 1994; Dunkle & Francis, 1990; Koehler et al., 2004; O'Toole et al., 1998; Rhodes et al., 2003). Fast alle (mit einer Ausnahme von Koehler et al., 2004) fanden einen positiven Zusammenhang zwischen Femininität und Attraktivität – allerdings mit sehr unterschiedlichen Effektgrößen.

Dasselbe gilt für Studien, die mit Computerprogrammen (z. B. Morphing-Software) Femininität experimentell manipulierten (Johnston et al., 2001; Perrett et al., 1998; Rhodes et al., 2000). Bei einigen Experimenten wurden auch feminine Eigenschaften übersteigert (z. B. durch die Karikatur-Technik), was zu einem weiteren Anstieg der Attraktivitätsbewertungen führte (Johnston & Franklin, 1993; Perrett et al., 1998; Rhodes et al., 2000). Eine besonders interessante Untersuchung stammt von Johnston und Franklin (1993), die eine spezielle Software entwickelten, mit der Versuchspersonen mit einer Art Herstellungsmethode ihr persönliches weibliches Attraktivitätsideal visualisieren konnten. Die auf

diese Weise hergestellten Frauenbilder wiesen im Vergleich zu einem durchschnittlichen Frauengesicht deutlich femininere Attribute sowie typische Kindchenschemamerkmale auf. Ein Ergebnis dieser Studie war, dass das in dieser Studie generierte ideale Frauengesicht einen sehr kleinen, zierlichen Unterkiefer hatte, der die Größe des Unterkiefers (*nasion-chin distance*) eines durchschnittlichen 11jährigen Mädchens hatte (Johnston & Franklin, 1993).

6.1.5 Empirische Befunde zum Einfluss maskuliner Gesichtszüge bei Männergesichtern

Empirische Befunde für die Attraktivität steigernde Wirkung von maskulinen Gesichtszügen bei Männern sind sehr uneinheitlich. Man muss bei der Bewertung dieser Befunde zwingend berücksichtigen, mit welcher Methode sie gewonnen wurden:

Die eine Gruppe von Studien manipulierte dazu experimentell maskuline Merkmale und testete sie z. B. gegen unveränderte Originalgesichter (Ishi, Gyoba, Kamachi, Mukaida & Akamatsu, 2004; Little, Burt, Penton-Voak & Perrett, 2001; Little & Hancock, 2002; Perrett et al., 1998; Rhodes et al., 2000; Swaddle & Reiersen, 2002). Solche Studien kommen durchweg zu dem Ergebnis, dass maskuline Merkmale *unattraktiver* (!) machen (in der Metaanalyse von Rhodes, 2006, ist die Effektstärke dieser Art von Studien negativ; $r = -0.47$). Die einzigen Ausnahmen sind ein Experiment mit Gesichtern als Strichzeichnungen, also einem wenig realistischen Stimulusmaterial, bei dem maskulinisierte Gesichter (dickere Augenbrauen, schmalere Lippen, ein breites Kinn und schmale Augen) gegenüber femininisierten bevorzugt wurden (Keating, 1985) sowie eine weitere Untersuchung von Johnston et al. (2001), die zu den am häufigsten zitierten Untersuchungen der Attraktivitätsforschung zählt.

Die andere Gruppe von Untersuchungen veränderte die Maskulinität nicht, sondern präsentierte entweder unveränderte Fotos oder gemorphte Gesichter und ermittelte dann die Maskulinität der Gesichter. Wie bei der Ermittlung von Symmetrie gibt es auch hier wieder zwei grundlegende Verfahren, nämlich einerseits das Vermessen von *Facial Landmarks* (Cunningham et al., 1990; Grammer & Thornhill, 1994; Koehler et al., 2004; Penton-Voak, Jones et al., 2001; Scheib et al., 1999) und andererseits die Bewertung der Maskulinität durch Versuchspersonen mittels Ratingskalen (Bruce et al., 1994; Cunningham et al., 1990; Dunkle & Francis, 1990; Koehler et al., 2004; Little & Hancock, 2002; Neave, Laing, Fink & Manning, 2003; O'Toole et al., 1998; Rennels, Bronstad & Langlois, 2008; Rhodes et al., 2003). Studien, die mit Gesichtern arbeiteten, bei denen die Maskulinität

nicht manipuliert war, kamen zu ganz anderen Ergebnissen, nämlich zu einem positiven Effekt von maskulinen Gesichtsmerkmalen. In der Metaanalyse von Rhodes betrug der positive Effekt $r = 0.35$.

Wenn verschiedene Methoden zuverlässig zu gegensätzlichen Ergebnissen kommen, liegt der Verdacht nahe, dass offensichtlich (zumindest auch) methodische Artefakte zu einem Resultat beitragen. Bei der Verstärkung maskuliner Merkmale durch die Karikatur-Technik muss beispielsweise die Validität dieser Bildverarbeitungstechnik kritisch hinterfragt werden. Denn wenn die Gesichtszüge eines individuellen Männergesichts von einem Durchschnittsmännergesicht oder einem Durchschnittsfrauengesicht entfernt werden, kann dies leicht dazu führen, dass Abweichungen von einem prototypischen Normalgesicht (= Durchschnittsgesicht) dadurch ebenfalls verstärkt werden, wodurch das Gesicht unattraktiver wird, ohne dass dies etwas mit Testosteronmarkern zu tun haben muss. Denn generell ist die Wahrscheinlichkeit groß, durch die Karikatur-Methode Gesichter unattraktiver zu machen (auch wenn man die Ergebnisse bei hyperfemininen weiblichen Gesichtern als Gegenargument nehmen mag). Zum Zweiten ist es fraglich, inwieweit der Algorithmus der geometrischen Interpolation einer Grafik-Software, bei dem ein zweidimensionales Foto verzerrt wird, die komplexen, nichtlinearen testosteronbedingten Wachstumsprozesse eines dreidimensionalen Schädels überhaupt adäquat simuliert (Enlow, 1966, 1979; Grumbach, 2000; Johnston et al., 2001; Meyer & Quong, 1999; Tanner, 1978). Rhodes (2006) zog in ihrer Metaanalyse den Schluss: „Reported preferences for feminized male faces appear to be an artefact of using sex continua that do not adequately capture sexual dimorphism” (p. 211).

Ein dritter Kritikpunkt, der zu Recht mehrfach geäußert wurde (Johnston et al., 2001; Meyer & Quong, 1999; Rhodes, 2006; Swaddle & Reiersen, 2002), betrifft den Umstand, dass Experimente, bei denen Maskulinität manipuliert wurde, häufig nicht mit Originalgesichtern, sondern mit gemorphten Durchschnittsgesichtern gemacht wurde (d. h. ein Männer-Durchschnittsgesicht wurde von den Proportionen eines Frauendurchschnittsgesichts weg verzerrt). Bei gemorphten Männergesichtern sind jedoch typisch männliche Merkmale (grobe Hauttextur, Bartwuchs, Bartstoppeln, aber auch testosteronbedingte kantige, markante Merkmale v. a. des Kinns und des Unterkiefers) nur unzureichend wiedergegeben. Gemorphte Männergesichter erscheinen generell weiblicher, weicher, sanfter und mit weniger Ecken und Kanten. Ein Blick auf das männliche Durchschnittsgesicht (Abbildung 6, S. 21, in Kapitel 2.3.2) bestätigt dies. Experimente mit *sex continua*, bei denen gleitende Übergänge zwischen einem femininisierten und einem hypermaskulinisierten Männer-Durchschnittsgesicht visualisiert werden, leiden damit von vornherein

an einer geringen Validität durch ein unrealistisches, artifizielles Stimulusmaterial. Die Präferenz der Versuchspersonen für femininere Gesichtszüge könnte deswegen auch durch die Tatsache verzerrt sein, dass bei maskulinen Versionen die männlichen Gesichtsproportionen nicht mit der sanften weiblichen Hauttextur übereinstimmen. Die Probanden präferieren daher femininere Männer-Versionen, da hier die Diskrepanz zwischen Gesichtsproportionen und Hauttextur geringer ist.

Auf der anderen Seite könnten aber auch die positiven Befunde der Untersuchungen, die Maskulinität nicht experimentell manipuliert, sondern durch Versuchspersonen-Ratings ermittelt haben, auf einem Einstellungseffekt der Probanden beruhen. Wenn Versuchspersonen der *Meinung* sind, dass attraktive Männergesichter auch männlich aussehen, dann ist die Wahrscheinlichkeit hoch, diesen Zusammenhang bei einem Rating der Maskulinität auch zu finden. Die Heuristik lautet dann: „Es ist gut, wenn ein Mann männlich aussieht.“ In Kombination mit dem bekannten Attraktivitätsstereotyp „Was schön ist, ist auch gut“ (Dion et al., 1972) folgt daraus, dass attraktive Männergesichter automatisch als männlicher bezeichnet werden, selbst wenn sie dies objektiv gar nicht sind. Dass solche Einstellungseffekte eine massive Rolle spielen, wurde im Zusatzexperiment der Untersuchung *Sym 7* (Kapitel 4.10.4) am Beispiel der Symmetrie eindeutig belegt, wo attraktive Gesichter von den Versuchspersonen als symmetrischer beurteilt wurden als unattraktive, obwohl in Wirklichkeit alle Personen durch eine Manipulation perfekt symmetrisch waren.

Als objektives Maß könnte man hier *Landmark*-Vermessungen verwenden, allerdings haben die Studien mit dieser Methodik wiederum die Einschränkung, dass es keinen einheitlichen Standard gibt, wie man eigentlich die Maskulinität eines Gesichts misst oder zu einem einzigen Index-Wert verrechnet. So beschränken sich diese Untersuchungen z. B. auch auf wenige Gesichtsmerkmale bzw. kommen zu Ergebnissen (z. B. Cunningham et al., 1990), die über die einzelnen sekundären Geschlechtsmerkmale nicht konsistent sind (beispielsweise erscheint ein markanter Unterkiefer zwar attraktiv, eine große Nase jedoch nicht; beide Merkmale sind jedoch typisch männliche Merkmale und werden durch die Wirkung von Testosteron beeinflusst).

Eine plausible Erklärung für die heterogenen Befunde zu Maskulinität bei Männergesichtern ist auch, dass es möglicherweise einen kurvilinearen Zusammenhang zwischen männlichen Gesichtszügen und Attraktivität gibt. Gar keine maskulinen Züge sind schlecht, stark ausgeprägte jedoch auch. Das Optimum liegt möglicherweise irgendwo in der Mitte. Darauf deuten auch die Ergebnisse von Cunningham (1990) hin, die eine Prä-

ferenz für moderate statt stark ausgeprägte männliche Gesichtszüge erkennen lassen sowie die Studien von Little und Hancock (2002) und O'Toole et al. (1998), die zu dem Ergebnis kamen, dass sowohl Durchschnittlichkeit als auch Maskulinität unabhängig voneinander zu männlicher Attraktivität beitragen. Kurvilineare Zusammenhänge sind jedoch zum einen generell schwierig zu belegen. Zum anderen würde eine weibliche Präferenz für mäßig maskuline Phänotypen aber auch biologisch orientierte Forscher mit ihrer Handicap-Hypothese in Erklärungsnot bringen, da sie mit der Theorie schwer vereinbar ist.

Zusammenfassend kann festgehalten werden, dass es für die Theorie des sexuellen Dimorphismus lediglich für Frauengesichter empirische Bestätigungen gibt – wenn auch mit sehr unterschiedlichen Effektgrößen. Bei Männergesichtern hingegen ist die Befundlage völlig widersprüchlich. Manche Untersuchungen bestätigen die Attraktivität von maskulinen Gesichtsmerkmalen, andere wiederum kommen zu dem gegenteiligen Schluss und postulieren auch bei Männergesichtern typisch feminine Merkmale als attraktiv. Zudem lässt sich ein klarer Einfluss der verwendeten Untersuchungsmethode auf das Ergebnis nachweisen. Studien, die Männlichkeit über *Facial Landmarks* erfassen oder von Beurteilern raten lassen, liefern Argumente *für* die Attraktivität von maskulinen Merkmalen bei Männergesichtern, Studien hingegen, die Maskulinität experimentell manipulierten, liefern Argumente *dagegen*.

6.1.6 Ziel der Untersuchungen zum Einfluss des sexuellen Dimorphismus und Hypothesen

Aus den oben genannten bisherigen empirischen Befunden und methodischen Problemen früherer Untersuchungen wurden für die eigenen geplanten Untersuchungen mehrere Konsequenzen gezogen.

Da es keinen einheitlichen Standard gibt (Rhodes, 2006), wie man anhand von Landmark-Vermessungen den Sexualdimorphismus eines Gesichts quantifizieren und mehrere Gesichtsmkmale zu einem einzigen Maskulinitäts- bzw. Femininitäts-Index-Wert zusammenfassen könnte, wird auf Landmark-Vermessungen realer Gesichter und korrelative Auswertungen verzichtet.

Die alternative Methode für korrelative Verfahren, nämlich die Messung von Femininität bzw. Maskulinität über Versuchspersonen-Ratings wird ebenfalls *nicht* verwendet. Der Grund sind die negativen Erfahrungen aus der zusätzlichen Datenerhebung zur Untersuchung *Sym 7* (vgl. Kapitel 4.10.4). Aufgrund dieser Ergebnisse kann *nicht* davon ausge-

gangen werden, dass Versuchspersonen in der Lage sind, die Femininität bzw. Maskulinität eines Gesichts unabhängig von dessen Attraktivität zu beurteilen und daher von vornherein mit einer Konfundierung von unabhängiger und abhängiger Variable zu rechnen ist. Da bei dieser Kontrollbedingung mit *gleichermaßen perfekt symmetrischen* Gesichtern die attraktiven mit einem so starken Effekt ($r = .87$) als symmetrischer bewertet wurden als die unattraktiven (obwohl die Korrelation eigentlich Null sein müsste), kann auch nicht von einer *kleinen* Konfundierung gesprochen werden. Die Konfundierung ist *extrem stark* und der dort gemessene Zusammenhang genauso groß wie die Symmetrie-Ratings der asymmetrischen Originalgesichter. Daher muss daraus die Schlussfolgerung gezogen werden, dass der scheinbare Effekt der Symmetrie auf die Attraktivität ausschließlich auf Artefakten beruht, d. h. auf einem Einstellungs-Effekt der Versuchspersonen und einem Halo-Effekt der Attraktivität (vgl. Kapitel 4.10.4 und 5.2.2). Wenn man also mit Sicherheit sagen kann, dass eine Methode in dem einen Anwendungsbereich (Symmetrie) komplett invalide ist, dann ist es nicht zu begründen, weshalb man dieselbe Methode in einem ganz ähnlichen Bereich (Femininität bzw. Maskulinität) anwenden sollte.

Damit bleiben nur noch die experimentellen Verfahren übrig, bei denen die Geschlechtstypikalität über Morphing-Software manipuliert wird. Hier zeigen jedoch, wie oben beschrieben, frühere Studien, dass bei Anwendung der Karikatur-Methode bei Männergesichtern offensichtlich Bildverarbeitungsartefakte beim Stimulusmaterial entstehen, die die Ergebnisse des Experiments verzerren. Insbesondere die von Meyer und Quong (1999) vorgebrachten Einwände, die zwischen *geometrical sexuality* und *biological sexuality* differenzieren, entziehen der Karikatur-Technik, die auf geometrischen Interpolationen von Distanzen beruht und damit nicht die tatsächlichen biologisch-medizinischen Wachstumsprozesse des Schädels simuliert, die wissenschaftliche Basis. Aus diesem Grund wird auch auf diese Methode von vornherein verzichtet.

Während das Interpolieren und Vergrößern von Distanzen methodisch kritisch zu sehen ist, ist dies bei einem Reduzieren dieser Distanzen unbedenklich. Denn bei diesem Verfahren können prinzipiell keine Extremwerte resultieren, denn die Extrema sind bereits die beiden Ausgangsgesichter (bzw. Schemata), die für den Morph-Prozess verwendet wurden. Beispiel: Wenn ein Männergesicht an die Proportionen eines durchschnittlichen Frauengesichts angenähert wird, dann ist die stärkste denkbare Verzerrung ein Gesicht mit der Originaltextur des Mannes, das vollständig an die weiblichen Durchschnittsproportionen angenähert ist. Da jedoch durchschnittliche weibliche Proportionen in keiner Weise irgendwie auffällig aussehen und sich weibliche Durchschnittsproportionen von männlichen ohnehin nicht sonderlich stark unterscheiden (vgl. dazu Abbildung 6, S. 21, und

Abbildung 22, S. 49, in den Kapiteln 2.3.2 und 2.7.2), wird das Ergebnis keineswegs sonderbar oder verzerrt aussehen (das Problem ist eher umgekehrt, dass der Effekt der Manipulation relativ gering ausfällt, so dass er unter Umständen – abhängig vom Ausgangsgesicht – nur schwer zu erkennen ist).

Die für die folgende Untersuchungsreihe durchgeführten experimentellen Manipulationen sind daher allesamt Annäherungen an ein Schema, d. h. ein experimentelles Reduzieren von Distanzen zwischen Ausgangsgesicht und Schema anstatt eines Vergrößerns und Interpolierens (Karikieren) dieser Distanzen. Als Schema bietet sich hier in besonderer Weise das Kindchenschema an, da dieses mehrere Vorteile bietet (Hanae & Jiro, 2001; Ishi, Gyoba, Kamachi et al., 2004):

- Ein Kindchenschema lässt sich empirisch durch Durchschnittsbildung aus mehreren fotografierten Kindergesichtern herstellen.
- Das Kindchenschema ist ein theoretisch in sich geschlossenes, einheitliches und bewährtes Konzept (Lorenz, 1943).
- Das Kindchenschema birgt im Gegensatz zur Verwendung spezieller Mann- oder Frau-Schemata (z. B. als besonders weiblich geltende Models oder Schauspielerinnen oder Composites hoch attraktiver Gesichter) nicht die Gefahr, von vornherein mit Attraktivität konfundiert zu sein.
- Das Kindchenschema bildet einen (durchschnittlichen) natürlichen Ausgangspunkt, von dem aus sich jedes Erwachsenengesicht unter dem Einfluss von Testosteron zu seinem individuellen Aussehen entwickelt hat. Durch ein Annähern eines Originalgesichts an kindliche Proportionen können Zwischenstufen erzeugt werden, die ein Aussehen dieses Gesichts simulieren, wenn das Gesichtswachstum unter einem geringeren Einfluss von Testosteron gewachsen wäre. Bei einer nur geringfügigen Annäherung von Männergesichtern an das Kindchenschema wird dadurch automatisch ein feminineres Aussehen simuliert (Eibl-Eibesfeldt, 1997; Enlow, 1979, 1989).
- Bei Verwendung von Morphing-Software kann man erwachsene Gesichter zu den Proportionen des Kindchenschemas *hin* verzerren (d. h. Distanzen verringern), während bei der Karikaturmethode von einem Schema (z. B. Durchschnittsgesicht) *weg* verzerrt wird (d. h. Distanzen vergrößert und interpoliert werden, vgl. auch die Kritik von Rennels et al., 2008, p. 886). Dadurch sehen die Ergebnisse natürlicher und realistischer aus als bei der Karikaturmethode, die die tatsächlichen Wachstumsprozesse nicht valide simuliert (Meyer & Quong, 1999).

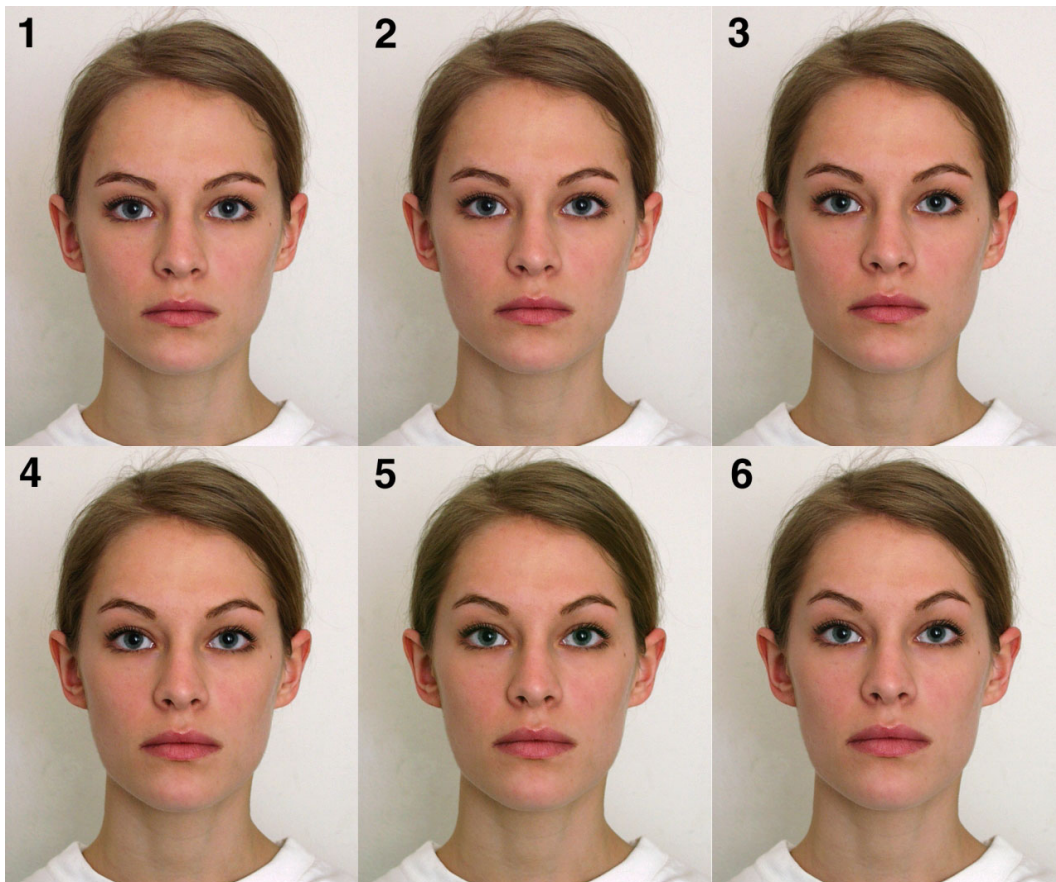


Abbildung 58: Gesicht einer erwachsenen Frau (Bild 6), das mit Morphing-Software schrittweise an die Gesichtsproportionen des Kindchenschemas angenähert wurde. Bei Bild 1 wurde das Kindchenschema mit 50 % gewichtet, bei Bild 2 mit 40 %, bei Bild 3 mit 30 %, bei Bild 4 mit 20 % und bei Bild 5 mit 10%. Es wird die Hypothese aufgestellt, dass Varianten mit einem hohen Kindchenschema-Anteil als attraktiver empfunden werden als das erwachsene Originalgesicht.

Für die Untersuchung des Effekts des sexuellen Dimorphismus auf die Attraktivität von Gesichtern werden drei Experimente durchgeführt, bei denen jeweils Gesichter an die Proportionen des Kindchenschemas angenähert werden.

In einem ersten Experiment (*Dimorph 1*) werden dazu unterschiedliche Frauengesichter in 10-Prozent-Abstufungen an die Proportionen eines durchschnittlichen Kindergesichts angenähert. Unter diesen Ausgangsgesichtern sind auch der weibliche Prototyp für ein attraktives Gesicht („Sexy-Gesicht“, vgl. Abbildung 5, S. 20, in Kapitel 2.3.2) sowie das weibliche Durchschnittsgesicht (vgl. Abbildung 22, S. 49, in Kapitel 2.7.2). Diese beiden

Gesichter sind von besonderem Interesse, weil es beim attraktiven Prototyp, der sehr hohe Attraktivitätsbewertungen erhielt, die Frage ist, ob durch die Manipulation in Richtung Kindchenschema eine weitere Steigerung der Attraktivität noch möglich ist. Beim Durchschnittsgesicht ist dieser Versuch spannend, weil dies ein Argument gegen die etablierte Durchschnittshypothese wäre, und bei einem positiven Befund belegt wäre, dass es Gesichtsproportionen gibt, die systematisch und klar definierbar vom Durchschnitt abweichen und gerade deswegen attraktiv machen.

Im zweiten Experiment (*Dimorph 2*) wird bei Frauengesichtern die Anzahl der verkindlichten Varianten auf eine einzige Version (40 % Kindchenschema) beschränkt und im Paarvergleich gegen die erwachsene Originalvariante getestet. Als zusätzlicher Parameter wird jedoch die Attraktivität des Ausgangsgesichts mit einbezogen. Denn es wäre denkbar, dass die Anwendung desselben Schemas auf Gesichter bei attraktiven Gesichtern zu anderen Effekten führt als bei unattraktiven Gesichtern.

Das dritte Experiment (*Dimorph 3*) ist analog zu Experiment 2, nur dass hier Männergesichter an das Kindchenschema angenähert werden. Nach der Theorie des sexuellen Dimorphismus sollten die Gesichter dadurch unattraktiver werden, da sie an Maskulinität verlieren. Nach den empirischen Befunden der Studien, die experimentell die Maskulinität mit Morphing-Software manipulierten, ist jedoch eher mit dem Gegenteil zu rechnen. Das Verwenden von attraktiven und unattraktiven Männergesichtern wird zeigen, ob es möglicherweise Interaktionen mit der Ausgangsattraktivität eines Originalgesichts gibt. Es wäre durchaus denkbar, dass unattraktive Männergesichter durch die Manipulation an Attraktivität gewinnen, nämlich dann, wenn der Grund ihrer Unattraktivität in testosteronbedingten Gesichtsmerkmalen liegt, die ein Gesicht zugleich unfreundlicher, finsterer oder aggressiver wirken lassen (z. B. dünne Lippen, tief liegende, schmale Augen, tief liegende Augenbrauen (Penton-Voak & Chen, 2004; Perrett et al., 1998)).

6.2 Experimentelle Manipulation des sexuellen Dimorphismus durch stufenweise Annäherung von Frauengesichtern an das Kindchenschema (Untersuchung *Dimorph 1*)

6.2.1 Methode

6.2.1.1 Versuchspersonen

An der Untersuchung *Dimorph 1* nahmen insgesamt 70 Versuchspersonen teil. 37 Personen waren männlich (52.9 %), 33 weiblich (47.1 %). Die jüngste Person war 14 Jahre, die älteste 75. Der Altersdurchschnitt betrug 33.5 Jahre ($SD = 16.0$; von zwei Frauen fehlen Angaben zum Alter). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, fand die Datenerhebung in einem Einkaufszentrum statt. Die Teilnahme am Experiment erfolgte freiwillig und wurde nicht entlohnt. Beim durchschnittlichen Alter ($t(66) = 1.19, p = .238$) sowie bei der Altersverteilung (Levene-Test auf Varianzgleichheit: $p = .786$) gibt es keine signifikanten Unterschiede zwischen den Geschlechtern.

6.2.1.2 Stimulusmaterial

Das Kindchenschema, an das die erwachsenen Frauengesichter durch Morphing-Software angenähert wurden, wurde empirisch erzeugt. Dazu wurden vier Kindergesichter frontal fotografiert und mit Morphing-Software (*Morpher 3.0*) zu einem Durchschnittsgesicht vermorpht. Die fotografierten Kinder waren Mädchen im Alter von 6.5 Jahren, zweimal 5 Jahren und 4 Jahren. Abbildung 59 zeigt die Proportionen dieses Durchschnittskindergesichts, das für die Erzeugung des Stimulusmaterials verwendet wurde.

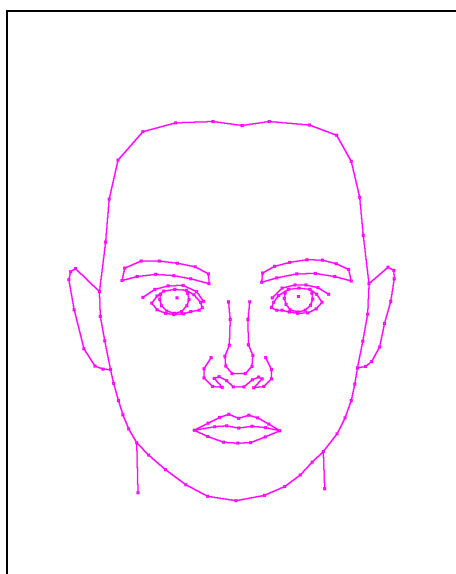


Abbildung 59: Das für das Experiment verwendete Kindchenschema, das erzeugt wurde, indem aus vier Kindern im Kindergartenalter mit Morphing-Software ein Durchschnittsgesicht berechnet wurde.

Anschließend wurden sechs Frauengesichter ausgewählt, von denen verkindlichte Varianten erstellt wurden. Verwendet wurden dazu die zwei weiblichen Originalgesichter, die in der Datenerhebung (Kapitel 2.7.1.1) die höchsten Attraktivitätsbewertungen bekommen hatten. Zusätzlich zwei weitere gemorphte Gesichter (*2-face-morphs*), bei denen Reife-kennzeichen wie ein schmales Gesicht und prominente Wangenknochen deutlich ausgeprägt waren, sowie der attraktive weibliche Prototyp (vgl. Abbildung 26, S. 57) und das weibliche Durchschnittsgesicht, das aus 64 Gesichtern berechnet worden war (vgl. Abbildung 27, S. 58).

Es wurden absichtlich keine unattraktiven Gesichter ausgewählt, da bei einem unattraktiven Gesicht die Möglichkeit bestanden hätte, dass allein die Annäherung an *irgendein* Schema zu einer Attraktivitätserhöhung führt, da im Morph-Prozess etwaige Disproportionierungen des Originalgesichts gemildert wurden. Besonders aussagekräftig ist das Experiment jedoch dann, wenn ein ohnehin schon attraktives Gesicht durch die Annäherung an das Kindchenschema noch attraktiver wird. Deswegen wurden die beiden attraktivsten Originalgesichter ausgewählt sowie der attraktive Prototyp, der in der Untersuchung *Durch 4* mit einem mittleren Attraktivitätswert von keinem anderen Gesicht signifikant übertroffen wurde. Das weibliche Durchschnittsgesicht ist deswegen von Interesse, weil es zum einen durchschnittliche weibliche Gesichtsproportionen eines erwachsenen Frau-

engesichts repräsentiert und zum anderen eine mögliche Attraktivitätserhöhung durch die Manipulation ein Argument gegen die Gültigkeit der Durchschnittshypothese ist.

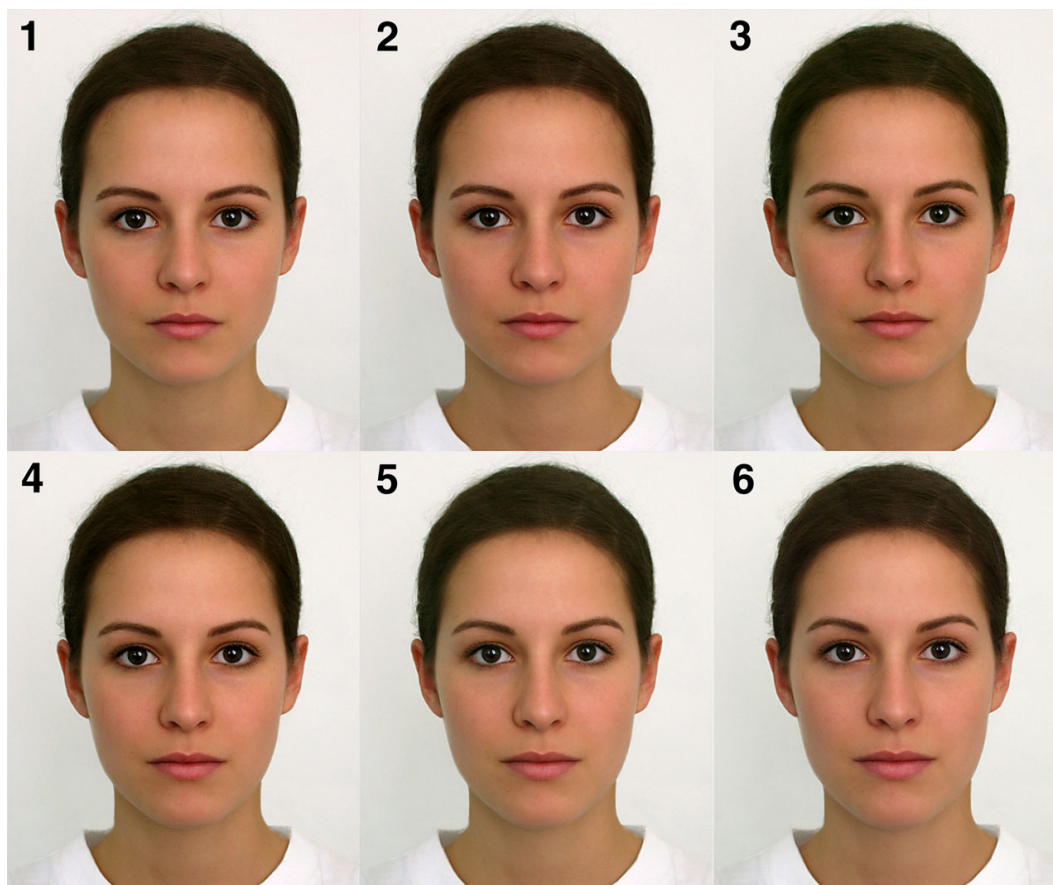


Abbildung 60: Bild 6 zeigt den attraktiven weibliche Prototyp („w(61-64)“, ein Composite aus den vier attraktivsten weiblichen Originalgesichtern aus dem Experiment *Durch 4*), der in der früheren Untersuchung als hochattraktiv bewertet worden war. Dieses Gesicht wurde mit Morphing-Software schrittweise an die Gesichtsproportionen des Kindchenschemas angenähert wurde. Bei Bild 1 wurde das Kindchenschema mit 50 % gewichtet, bei Bild 2 mit 40 %, bei Bild 3 mit 30 %, usw.

Die Annäherung an das Kindchenschema erfolgte durch das Verfahren des „*Warping*“, also durch das Verzerren der Gesichtsproportionen des Ausgangsgesichts *ohne* Neuberechnung (Mittelwertbildung) der Farbwerte. Es ist dasselbe Verfahren wie es auch in den Symmetrie-Experimenten *Sym 5* und *Sym 6* zum Einsatz kam. Der Vorteil dieses Verfahrens wurde dort sowie in Kapitel 2.1.2 („Effekte von Morphing-Artefakten“) ausführlich dargelegt: Es entstehen keine Bildverarbeitungsartefakte wie bei der *Blend*-Technik, die

automatisch bei Übergängen zwischen zwei Gesichtern zu einer glatteren, makelloseren Haut führen. Eine mögliche Attraktivitätserhöhung könnte dadurch nicht mehr eindeutig auf die kindlichen Proportionen des Kindchenschemas zurückgeführt werden, sondern könnte auch einfach nur an der makelloseren Haut der manipulierten Gesichter liegen.

Von jeder Stimulusperson wurden mit der Software *Morpher 3.0* fünf Varianten erstellt, die sich im Ausmaß der Gewichtung des Kindchenschemas unterschieden. Die Annäherung an das Kindchenschema erfolgte in 10%-Schritten. Die stärkste Veränderung wurde durch einen Kindchenschema-Anteil von 50 % erzeugt. Zusammen mit dem unveränderten Originalgesicht als Kontrollbedingung wurden den Versuchspersonen damit jeweils sechs Varianten einer Stimulusperson präsentiert.

6.2.1.3 Versuchsdurchführung

Alle sechs Varianten einer Stimulusperson wurden mit einem Fotodrucker auf Fotopapier gedruckt. Jede einzelne Variante hatte eine Größe von 6 cm x 7.5 cm, die Gesamtgröße betrug damit 18 cm x 15 cm bei einer sehr hohen Auflösung von 300 dpi. Die Schärfe und Bildinformation war damit weit höher als bei vergleichbaren Untersuchungen wie z. B. dem Experiment von Rhodes et al. (2000), die ebenfalls mit Ausdrucken arbeiteten, jedoch nur weniger als einem Viertel der Auflösung (72 dpi, S. 132). Die Varianten wurden mit den Ziffern 1 bis 6 nummeriert (vgl. Abbildung 60, S. 255). Je kleiner die Ziffer, desto größer war der Anteil des Kindchenschemas, das in einer Variante enthalten war. Für die Datenerhebung wurden alle sechs Ausdrücke auf Augenhöhe an eine Pinnwand geheftet. Die Versuchspersonen wurden instruiert, bei jeder 6er-Variation anzugeben, welche Variante sie am attraktivsten fanden. Wenn sie keinen Unterschied sähen oder alle Varianten gleich attraktiv fänden, sollten sie einfach per Zufall ein Bild auswählen. Informationen darüber, worin sich die Varianten unterschieden, wurden nicht gegeben. Jede Versuchsperson beurteilte die Varianten aller sechs Stimuluspersonen. Die Nummern der ausgewählten Varianten wurden vom Versuchsleiter notiert.

6.2.2 Ergebnisse

Für die Auswertung der Daten wurde analysiert, wie häufig bei jeder Stimulusperson eine bestimmte Variante von den Versuchspersonen als die attraktivste ausgewählt worden war. Die Nullhypothese lautet dabei, dass alle Varianten gleich häufig bevorzugt werden, d. h. eine Gleichverteilung vorliegt. Dies wäre der Fall, wenn jede Variante etwa 12 Mal (16.7 % der Fälle, d. h. 1/6) ausgewählt würde. Nach der Theorie des sexuellen Di-

morphismus sollten jedoch solche Varianten mit einem hohen Kindchenschema-Anteil häufiger bevorzugt werden. Tabelle 18 listet die Ergebnisse auf.

Tabelle 18: Absolute und relative Häufigkeiten der Bevorzugung der einzelnen Gesichtervarianten für die Stimuluspersonen (A bis F) in Abhängigkeit vom Kindchenschema-Anteil des Gesichts. Die beiden letzten Zeilen listen die Testwerte und p -Werte der einzelnen χ^2 -Anpassungstests auf.

Anteil Kindchenschema	A w64	B w63	C w45&46	D w61-64	E w35&36	F w1-64	Gesamt
50 % Kind	17 24.3 %	26 37.1 %	7 10.0 %	12 17.1 %	13 18.6 %	14 20.0 %	89 21.2%
40 % Kind	18 25.7 %	15 21.4 %	19 27.1 %	7 10.0 %	13 18.6 %	17 24.3 %	89 21.2 %
30 % Kind	8 11.4 %	15 21.4 %	9 12.9 %	12 17.1 %	9 12.9 %	17 24.3 %	70 16.7 %
20 % Kind	14 20.0 %	9 12.9 %	19 27.1 %	19 27.1 %	13 18.6 %	10 14.3 %	84 20.0 %
10 % Kind	5 7.1 %	2 2.9 %	12 17.1 %	10 14.3 %	14 20.0 %	5 7.1 %	48 11.4 %
0 % Kind (erwachsen)	8 11.4 %	3 4.3 %	4 5.7 %	10 14.3 %	8 11.4 %	7 10.0 %	40 9.5 %
Gesamt	70 100 %	70 100 %	70 100 %	70 100 %	70 100 %	70 100 %	420 100
χ^2	12.457	34.571	16.743	6.971	2.686	11.257	32.886
p	.029	<.001	.005	.223	.748	.047	<.001

Anmerkung: Die A-priori-Wahrscheinlichkeit für eine Variante liegt bei 16.7 %.

Es zeigte sich, dass die erwachsenen Originalgesichter (mit 0 % Kindchenschema-Anteil) bei fast allen Stimulusgesichtern seltener als von der A-priori-Wahrscheinlichkeit zu erwarten als die attraktivste Variante ausgewählt wurden. Eine Ausnahme ist lediglich Gesicht D, der attraktive Prototyp. Zur statistischen Überprüfung wurde für jede Stimulusperson ein χ^2 -Anpassungstest auf Gleichverteilung gerechnet. Bei den Stimuluspersonen A, B, C und F war der Test signifikant, bei den Stimuluspersonen D und E nicht.

Der deutlichste positive Einfluss des Kindchenschemas wurde bei Stimulusperson B (Abbildung 58, S. 251) gefunden: Hier gab es mit 37.1 % eine starke Präferenz für die kindlichste Variante mit 50 % Kindchenschema-Anteil, und auch die beiden Varianten mit 40 und 30 % Kindchenschema-Anteil wurden mit 21.4 % der Nennungen rund doppelt so häufig bevorzugt als dies nach der A-priori-Wahrscheinlichkeit zu erwarten war.

Beim attraktiven Prototyp gab es zwar keine signifikanten Abweichungen von der Gleichverteilung, allerdings wurde die 20%-Kindchenschema-Variante im Vergleich zur erwachsenen Originalversion von doppelt so vielen Probanden als die attraktivste genannt.

Beim Durchschnittsgesicht zeigte sich, dass es selbst hier eine klare Präferenz für Gesichtsproportionen gibt, die eben *nicht* einem weiblichen (erwachsenen) Durchschnittsgesicht entsprechen, sondern kindlichere (und damit femininere) Gesichtszüge aufweisen. Dieser Befund steht klar im Widerspruch zur „Durchschnittshypothese“.

Die Ergebnisse zeigen, dass offensichtlich die Anwendung des Kindchenschemas nicht bei allen Stimulusgesichtern zum gleichen Ergebnis führt, sondern dass es Wechselwirkungen mit anderen Gesichtsmerkmalen gibt, ohne dass im Einzelnen angegeben werden kann, an welchen Merkmalen dies genau liegt. Trotz dieser Unterschiede wurden die Daten dieser sechs Stimuluspersonen für einen besseren Überblick zusammengefasst und es wurde ausgewertet, wie häufig ein bestimmter Anteil des Kindchenschemas über alle Stimuluspersonen als am attraktivsten genannt wurde. Die letzte Spalte in Tabelle 18 (S. 257) listet diese Daten auf. Bei einer A-priori-Wahrscheinlichkeit von 16.7 % schnitt die erwachsene Originalversion klar am schlechtesten ab, während es eine deutlich Präferenz für Gesichtervarianten mit 40 % oder 50 % Kindchenschema-Anteil gibt. Diese beiden Varianten wurden zusammen mehr als doppelt so häufig bevorzugt (42.4 %) wie die beiden Varianten mit dem geringsten Kindchenschema-Anteil (0 % und 10 % Anteil in 20.9 % der Fälle).

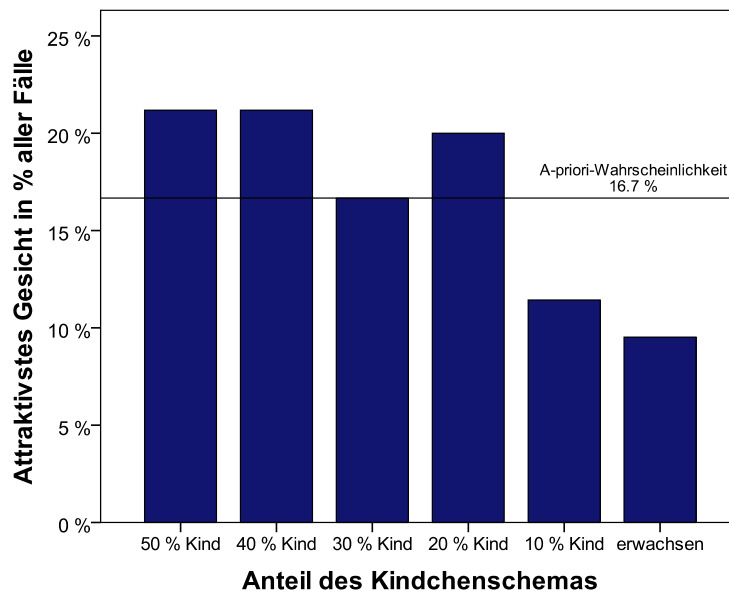


Abbildung 61: Häufigkeitsverteilung in Prozent, wie oft Frauengesichter mit einem bestimmten Kindchenschema-Anteil als attraktivstes Gesicht ausgewählt wurden. Die horizontale Linie markiert die A-priori-Wahrscheinlichkeit von 16.7 %. Am seltensten werden die Varianten des erwachsenen Originalgesichts als am attraktivsten ausgewählt – am häufigsten die Varianten mit einem hohen Kindchenschema-Anteil von 40 und 50 %.

In einer zusätzlichen Analyse wurde geprüft, ob das Geschlecht der Beurteilerpersonen einen Einfluss auf die Ergebnisse hatte. Beispielsweise wäre es denkbar, dass v. a. Männer eher jüngere und kindlichere Frauengesichter bevorzugen, während Frauen selbst reifere und erwachsenere Frauengesichter präferieren. Hierzu wurden (separat für jede Stimulusperson) sechs χ^2 -Unabhängigkeitstests gerechnet. In keinem Fall ergab sich jedoch ein signifikanter Unterschied. Die p -Werte liegen zwischen .093 und .604. Das bedeutet, dass Männer und Frauen sich hinsichtlich ihrer Vorliebe für verkindlichte Gesichter *nicht* unterscheiden.

Außerdem wurde getestet, ob das Alter der Beurteilerpersonen einen Einfluss auf die Präferenz für das Kindchenschema hatte. Hierzu wurden (separat für jede Stimulusperson) sechs Korrelationen gerechnet: Die eine Variable war das Alter der Beurteilerpersonen, die andere der Anteil (Prozentsatz von 0 % bis 50 %) des Kindchenschemas, das in dem ausgewählten Bild enthalten war. Die Rangkorrelationskoeffizienten nach Spearman waren allesamt gering (zwischen -.19 und +.16) und nicht signifikant (p -Werte zwischen .13 und .94). Das bedeutet, dass auch das Alter der Beurteilerpersonen keinen Einfluss

auf die Präferenz für verkindlichte Gesichter hat. Kindlichere Gesichter werden von allen Altersgruppen gleichermaßen bevorzugt.

Zusammenfassend bleibt festzuhalten, dass es unabhängig von Geschlecht oder Alter der Beurteiler eine allgemeine Präferenz für Frauengesichter gibt, die kindlichere Gesichtszüge aufweisen, als dies für erwachsene Frauengesichter normal ist. Die Größe des Effekts hängt jedoch vom konkreten Aussehen des Originalgesichts ab, das an das Kindchenschema angenähert wird. Durchschnittliche weibliche Gesichtsproportionen können durch eine Annäherung an kindliche Proportionen in ihrer Attraktivität gesteigert werden. Das Gleiche gilt für hochattraktive erwachsene „Original-Frauengesichter“. Damit bestätigen die Ergebnisse im Wesentlichen die Theorie des sexuellen Dimorphismus bei Frauengesichtern.

6.3 Paarvergleichsexperiment mit Manipulation des sexuellen Dimorphismus durch Annäherung von Frauengesichtern an das Kindchenschema (Untersuchung *Dimorph 2*)

Aufgrund der Erfahrungen mit dem Experiment *Dimorph 1* wurden zwei Versuchsbedingungen im Experiment *Dimorph 2* verändert:

- Die Anzahl der Alternativen wurde auf zwei beschränkt (eine erwachsene Originalversion und nur eine einzige „verkindlichte“ Variante), weil sich im vorausgegangenen Experiment die Präferenzen der Versuchspersonen bei „verkindlichten“ Varianten auf mehrere Bilder verteilten und somit eine statistische Absicherung des Effekts schwieriger ist. Für das neue Experiment wurden bei der manipulierten Variante einheitlich 40 % Kindchenschema-Anteil festgesetzt, da sich im vorausgegangenen Experiment gezeigt hatte, dass ein hoher Kindchenschema-Anteil keineswegs zu stark ist, sondern von vielen als attraktiv empfunden wird.
- Es wurden sowohl attraktive als auch unattraktive Gesichter als Stimulusmaterial verwendet, um zu untersuchen, inwieweit Wechselwirkungen zwischen der Wirkung des Kindchenschemas und der Ausgangsattraktivität des Originalgesichts bestehen.

6.3.1 Methode

6.3.1.1 Versuchspersonen

An der Untersuchung *Dimorph 2* nahmen insgesamt 104 Versuchspersonen teil, davon waren 55 Personen männlich (52.9 %) und 49 weiblich (47.1 %). Die jüngste Person war 14 Jahre, die älteste 75. Der Altersdurchschnitt betrug 29.4 Jahre ($SD = 12.2$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, fanden die Datenerhebungen an öffentlichen Orten wie Einkaufszentren, Fußgängerzonen oder Bahnhöfen statt. Die Teilnahme am Experiment erfolgte freiwillig und wurde nicht entlohnt. Bei den Versuchspersonen waren ältere Frauen etwas überrepräsentiert, was am etwas höheren Altersdurchschnitt ($M = 31.5$ versus $M = 27.5$) sowie an der größeren Standardabweichung erkennbar ist ($SD = 14.6$ versus $SD = 9.3$). Trotz dieser leichten Ungleichverteilung ist der Altersunterschied zwischen den beiden Geschlechtern *nicht* signifikant ($t(79.658) = -1.561$, $p = .103$, Testwert für ungleiche Varianzen) und kann daher vernachlässigt werden.

6.3.1.2 Stimulusmaterial

Insgesamt wurden zehn Stimulusgesichter mit Morphing-Software an das Kindchenschema angepasst. Die Hälfte der Frauengesichter waren solche, die in der Datenerhebung des Experiments *Durch 4* als unattraktiv bewertet worden waren (Stimulusgesichter A bis E), die andere Hälfte Gesichter, die hohe Attraktivitätsbewertungen bekommen hatten (Stimulusgesichter F bis K, der Buchstabe „J“ wurde ausgelassen). Unter den Gesichtern befanden sich auch der weibliche unattraktive Prototyp, w(1-4), sowie nochmals der attraktive Prototyp aus dem vorangegangenen Experiment, um den dortigen Befund (*kein* Attraktivitätserhöhender Effekt des Kindchenschemas) nochmals durch die andere Versuchsanordnung des Paarvergleichsexperiments zu validieren.

Mit Ausnahme eines sehr attraktiven Originalgesichts handelte es sich um gemorphte Gesichter (*2-face-morphs* und *4-face-morphs*), um auszuschließen, dass ein möglicher Attraktivität steigernder Effekt der Schemaanpassung nur darauf zurückzuführen sein könnte, dass irgendwelche zufälligen Disproportionierungen eines Originalgesichts durch die Annäherung an ein Schema abgemildert und harmonisiert würden. Denn in einem solchen Fall könnte eine Attraktivitätserhöhung nicht eindeutig auf die kindlichen Proportionen zurückgeführt werden, da es auch an der harmonisierenden und ausgleichenden Wirkung eines beliebigen normalen und regelmäßigen Schemas liegen könnte. Denn

erstens gibt es bei Kindergesichtern generell weniger extreme Abweichungen einzelner Gesichtermerkmale von einem normalen, durchschnittlichen (Kinder)gesicht, da Kindern der hormonell bedingte Wachstumsschub in der Pubertät ja noch bevorsteht. Und zweitens ist das verwendete Kindchenschema selbst ja bereits ein Durchschnittsgesicht (aus vier Kindern), wodurch bereits dort Unregelmäßigkeiten herausgemittelt wurden.

Das Kindchenschema, an das die erwachsenen Frauengesichter durch Morphing-Software (*Morpher 3.0*) angenähert wurden, war dasselbe wie in der vorangegangenen Untersuchung *Dimorph 1*. Die Annäherung erfolgte mit dem Warp-Verfahren unter Beibehaltung sämtlicher Farbwerte und mit einer Stärke von 40 %. Die beiden folgenden Abbildungen zeigen Beispiele für ein unattraktives (Abbildung 62) und ein attraktives Stimulusgesicht (Abbildung 63).



Abbildung 62: Unattraktives erwachsenes Frauengesicht (Gesicht D, links) und mit Morphing-Software verkindlichte Variante, bei der das Kindchenschema mit 40 % gewichtet wurde (rechts).



Abbildung 63: Attraktives erwachsenes Frauengesicht (Gesicht F, links) und mit Morphing-Software verkindlichte Variante, bei der das Kindchenschema mit 40 % gewichtet wurde (rechts).

6.3.1.3 Versuchsdurchführung

Die Versuchsdurchführung verlief analog zu den Paarvergleichsexperimenten zur Messung des Effekts der Symmetrie. Alle Bilder wurden von einem Fotodienst auf Fotopapier in einer Größe von 9 x 11 Zentimeter belichtet. Bei einer Bildgröße von 1024 x 1280 Pixeln ergibt sich eine Auflösung von 296 dpi. Für die Versuchsdurchführung wurden die Fotos in einer Fotomappe präsentiert, durch die seitenweise geblättert werden konnte. Dabei konnten die Versuchspersonen stets nur ein Gesichterpaar betrachten und bewerten. Bei der Anordnung der Fotos wurde darauf geachtet, dass die „verkindlichten“ Fotos gleich häufig auf der linken Seite wie auf der rechten Seite präsentiert wurden, um eventuelle Lerneffekte (z. B. „Das rechte Gesicht sieht immer irgendwie jünger aus“) oder Präferenzen für eine bestimmte Seite (vgl. *left-field perceptual bias of the observer* (Gilbert & Bakan, 1973; Mertens et al., 1993)) von vornherein ausschließen zu können. Zusätzlich wurden attraktive Gesichterpaare und unattraktive stets abwechselnd präsentiert, obwohl durch die Präsentation von Gesichterpaaren Positionseffekte durch eine bestimmte Reihenfolge der Stimuluspersonenpaare kaum vorstellbar sind.

Allen Versuchspersonen wurde die Frage gestellt: „Welches Gesicht ist attraktiver?“ Auch wenn sie keinen Unterschied bzw. keinen Attraktivitätsunterschied zwischen den jeweiligen Varianten sahen, mussten die Probanden sich für ein Bild entscheiden (*forced choice*). Es wurden keine zeitlichen Beschränkungen auferlegt. Der Versuchsleiter notierte die Urteile.

6.3.2 Ergebnisse

Zur Überprüfung der Frage, ob die durch das Morphing-Verfahren an das Kindchenschema angenäherten Gesichter im direkten Paarvergleich attraktiver bewertet wurden als die unveränderten erwachsenen Originalgesichter, wurden zehn Binomialtests durchgeführt. Zur besseren Vergleichbarkeit mit den Ergebnissen der anderen Untersuchungen (Durchschnittshypothese und Symmetrie-Hypothese) wurden auch hier wieder alle Effektgrößen vom *Binomial effect size display (BESD)* nach folgender Formel in ein r umgerechnet (Rosenthal & Rubin, 1982):

$$r = 2 * BESD - 1$$

Tabelle 19 listet die absolute und relative Häufigkeit der Bevorzugung der „verkindlichten“ Variante gegenüber der unveränderten erwachsenen Original-Variante sowie die in r umgerechnete Effektgröße auf. Am Ende der Tabelle ist der Median der erzielten Häufigkeiten und Effektgrößen aufgeführt.

Tabelle 19: Ergebnisse des Paarvergleichs-Experiments mit „verkindlichten“ Frauengesichtern (40 % Kindchenschema) im direkten Vergleich mit dem jeweiligen unveränderten erwachsenen Gesicht. Aufgelistet sind absolute und relative Häufigkeiten sowie die in r umgerechneten Effektgrößen. $N = 104$.

Stimulus-Gesicht	Attraktivität des Originalgesichts	Häufigkeit der Bevorzugung der kindlichen Variante n (%)	p	r
A	unattraktiv	65 (62.5 %)	.014	0.25
B	unattraktiv	71 (68.3 %)	<.001	0.37
C	unattraktiv	77 (74.0 %)	<.001	0.48
D	unattraktiv	81 (77.9 %)	<.001	0.56
E	unattraktiv	84 (80.8 %)	<.001	0.62
F	attraktiv	66 (63.5 %)	.008	0.27
G	attraktiv	45 (43.3 %)	.202	-0.13
H	attraktiv	87 (83.7 %)	<.001	0.67
I	attraktiv	68 (65.4 %)	.002	0.31
K	attraktiv	65 (62.5 %)	.014	0.25
Median		69.5 (66.8 %)		0.34

Wie aus **Tabelle 19** zu entnehmen ist, wurde bei fast allen 10 Gesichterpaaren die an das Kindchenschema angenäherte Variante im Vergleich zur unveränderten erwachsenen Originalversion häufiger bevorzugt. Lediglich bei Gesicht G gab es keinen signifikanten Unterschied. Der Median der Präferenzen für die kindliche Variante liegt bei 66.8 %; der Median der nach r standardisierten Effektgrößen liegt bei $r = .34$. Somit gibt es insgesamt einen statistisch klar nachweisbaren, aber schwachen Effekt des Kindchenschemas auf Attraktivität von Frauengesichtern.

Bei den unattraktiven Frauengesichtern (A bis E) ist dieser Effekt größer ($r = 0.48$), bei den attraktiven (F bis K) hingegen kleiner ($r = 0.27$). Man könnte dies als Hinweis darauf interpretieren, dass attraktive Frauengesichter ohnehin bereits mehr kindliche (typisch feminine) Gesichtsproportionen als unattraktive Gesichter enthalten (= „Deckeneffekt“) und deswegen die unattraktiven Gesichter von der Manipulation stärker profitieren. Abbildung 64 illustriert dies anhand eines unattraktiven Frauengesichts und des attraktiven Frauengesichts G, bei dem als einziges kein attraktivitätssteigernder Effekt des Kindchenschemas beobachtet wurde.



Abbildung 64: Obwohl bei allen Stimulusgesichtern die Annäherung an das Kindchenschema immer 40 % betrug, wirkt sich diese Manipulation nicht auf jedes Gesicht gleich aus. Beim linken Gesichterpaar (Gesicht C) ist die Veränderung zum Kindchenschema aufgrund des länglicheren Gesichts und des (in der Vertikalen) langen Kinns deutlicher sichtbar und der Attraktivitätsgewinn deutlich (74 % der Beurteiler bevorzugten die kindliche Version). Beim rechten Bildpaar (Gesicht G) ist der Unterschied viel schwächer ausgeprägt und es gab keinen signifikanten Attraktivitätsunterschied.

Besondere Beachtung verdienen die beiden Prototypen aus der Untersuchung *Durch 4*, nämlich der Prototyp für ein unattraktives bzw. ein attraktives Gesicht (in diesem Experiment die Gesichter E und K, vgl. Abbildung 65). Für beide Prototypen ergab sich ein Attraktivitätssteigernder Effekt des Kindchenschemas, der beim unattraktiven Prototypen deutlich stärker ausfiel. Während im vorhergehenden Experiment (*Dimorph 1*) mit den sechs Auswahlalternativen beim attraktiven Prototypen *kein* signifikanter positiver Effekt des Kindchenschemas vorhanden war, konnte er hier gefunden werden. Haben Versuchspersonen also nur die Wahl zwischen zwei Alternativen (erwachsen oder kindlich), bevorzugen sie mehrheitlich die „verkindlichte“ Version. Der nicht signifikante Befund des vorangegangenen Experiments ist damit wohl auch auf das Versuchsdesign mit sechs Alternativen und der geringeren Power des χ^2 -Anpassungstests bei gleichzeitig weniger Probanden ($N = 70$ statt 104) zurückzuführen. Der hier erzielte Befund zeigt also, dass

auch ein hochattraktives Frauengesicht noch an Attraktivität gewinnen kann, wenn seine Gesichtsproportionen an das Kindchenschema angenähert werden.



Abbildung 65: Linkes Bildpaar: Der unattraktive weibliche Prototyp (E). In der „verkindlichten“ Variante ist der Attraktivitätsgewinn deutlich (80.8 % der Versuchspersonen bevorzugen diese). Rechtes Bildpaar: Der attraktive weibliche Prototyp (K). In der kindlichen Version wird er leicht bevorzugt (62.5 % der Versuchspersonen).

Auch bei den hier vorliegenden Daten gibt es wieder – wie schon bei den Paarvergleichen der Symmetrie-Experimente – anstatt der mehrfachen Binomialtests eine alternative Form der Datenanalyse: Mit der Methode von Little und Jones (2003, S. 1761) wird ausgewertet, wie häufig jede Versuchsperson die „verkindlichte“ Variante gegenüber der Originalvariante bevorzugt hat. Nach der Nullhypothese sollte sie dies bei der Hälfte der präsentierten Gesichterpaare tun (d. h. die A-priori-Wahrscheinlichkeit ist 0.5). Für jede Versuchsperson wird daher eine relative Häufigkeit berechnet. Diese Häufigkeiten werden mit einem Ein-Stichproben- t -Test analysiert. Das N ist in diesem Fall die Anzahl der Versuchspersonen ($N = 104$) und eine positive mittlere Differenz bedeutet, dass die „verkindlichte“ Variante häufiger bevorzugt wurde. Analysiert man die Daten mit dieser Methode, ergibt sich ein hoch signifikantes Ergebnis $t(103) = 8.972, p < .001$. Die Effektstärke, die sich aus dem t -Wert berechnet (siehe Formel in Kapitel 4.4.3), fällt dadurch ebenfalls deutlich größer aus und beträgt dann $r = 0.75$. Mit diesem Auswertungsverfahren fällt also der Attraktivität erhöhende Effekt der Annäherung von Frauengesichtern an das Kindchenschema weit größer aus als mit der konservativeren Auswertung mit einzelnen Binomialtests, bei der der Median über alle Effektgrößen nur $r = 0.34$ beträgt. Dass diese Art der Datenanalyse aufgrund des künstlichen Unterdrückens von statistischem Rauschen problematisch ist und zu einer Überschätzung der tatsächlichen Effekte führt,

wurde bereits in Kapitel 5.2.3 ausführlich diskutiert und soll daher an dieser Stelle nicht wiederholt werden.

In einer abschließenden Analyse wurde mit zusätzlichen Tests überprüft, ob Geschlecht oder Alter der Versuchspersonen einen Einfluss auf das Urteilsverhalten hatten. Für den Test auf Geschlechtsunterschiede wurden zehn χ^2 -Tests gerechnet. Es zeigte sich, dass Frauen die „verkindlichten“ Gesichter noch etwas stärker bevorzugten als Männer. Bei drei Stimulusgesichtern war dieser Unterschied signifikant ($p < .05$), bei drei weiteren zwar nicht signifikant, aber mit einem $p < .10$. Daher wurde zusätzlich eine Analyse gerechnet, bei der die im obigen Absatz erwähnten „relativen Häufigkeiten der Bevorzugung für die kindliche Variante pro Versuchsperson“ für die beiden Geschlechter verglichen wurden. Ein t -Test für unabhängige Stichproben zeigte, dass Frauen im Durchschnitt häufiger als Männer die kindliche Version gegenüber der erwachsenen präferieren ($t(102) = 2.741, p = .007$).

Für diese Analyse, die lediglich zuvor zusammengefasste Daten analysiert, gelten selbstverständlich dieselben Einschränkungen wie oben erwähnt, so dass man diesen kleinen Geschlechtsunterschied nicht überbewerten sollte. Interessant sind diese Ergebnisse dennoch, jedoch in anderer Hinsicht. Falls es bei den Beurteilern tatsächlich einen Geschlechtsunterschied geben sollte, dann in der Art, dass Frauen noch stärker als Männer kindliche Frauengesichter bevorzugen. Sicher widerlegen kann man die gegenteilige Hypothese, dass Männer eine stärkere Präferenz für kindliche (oder jünger aussehende) Frauengesichter hätten. Dies wäre nämlich die klassisch feministische Interpretation (Wolf, 1991), nach der erstens eine patriarchalisch geprägte Gesellschaft (und damit die Männer) definiert, was bei Frauen als attraktiv gilt und zweitens Männer eher kindlich, naiv und harmlos aussehende Partnerinnen bevorzugen würden, weil sie von ihnen annehmen, dass sie sie leichter kontrollieren, unterdrücken und beherrschen könnten. Diese Sichtweise kann – wie gesagt – aufgrund der Daten klar zurückgewiesen werden. Wenn es überhaupt einen Geschlechtsunterschied bei der Präferenz kindlich aussehender Frauengesichter gibt, dann sind es die Frauen, die ein kindliches Aussehen besonders bevorzugen.

Eine Überprüfung des Einflusses des Alters der Versuchspersonen auf das Urteilsverhalten ergab keinerlei Zusammenhänge. Bei zehn χ^2 -Tests zwischen Alterskategorie (Altersklassen: bis 25 Jahre, 26 bis 50 Jahre und 51 Jahre und älter) und gewählter Variante ergab sich für keines der Stimulusgesichter ein Zusammenhang (p -Werte zwischen .203 und .874). Auch eine Korrelation zwischen Alter und der relativen Häufigkeit (pro Ver-

suchperson), mit der die kindliche Variante ausgewählt wurde, zeigte keinerlei Zusammenhang ($r = .01$, $p = .915$). Damit kann gesagt werden, dass eine Präferenz für „verkindlichte“ Frauengesichter vollkommen unabhängig vom Alter der Beurteilerpersonen ist.

Zusammenfassend bleibt festzuhalten, dass Frauengesichter insgesamt attraktiver beurteilt werden, wenn ihre Gesichtsproportionen an die Proportionen eines Kindergesichts angenähert werden. Der Effekt ist jedoch klein, der Median der nach r standardisierten Effektgrößen liegt bei $r = .34$. Bei einer Datenanalyse, die die Urteilsvarianz der Versuchspersonen *nicht* berücksichtigt, ist der Effekt dagegen groß ($r = 0.75$). Bei den unattraktiven Frauengesichtern ist dieser Effekt größer ($r = 0.48$), bei den attraktiven hingegen kleiner ($r = 0.27$). Selbst hochattraktive Frauengesichter wie der attraktive weibliche Prototyp können durch eine Annäherung der Gesichtsproportionen an das Kindchenschema noch etwas attraktiver gemacht werden. Damit lässt sich die Theorie des sexuellen Dimorphismus bei Frauengesichtern statistisch klar belegen, wenn auch die Effektgrößen eher gering sind.

6.4 Paarvergleichsexperiment mit Manipulation des sexuellen Dimorphismus durch Annäherung von Männergesichtern an das Kindchenschema (Untersuchung *Dimorph 3*)

6.4.1 Methode

6.4.1.1 Versuchspersonen

Am Experiment zur Untersuchung des Einflusses des Kindchenschemas bei Männergesichtern (Untersuchung *Dimorph 3*) nahmen insgesamt 99 Versuchspersonen teil. Davon waren 46.1 % Männer und 53.9 % Frauen. Die jüngste Person war 14 Jahre, die älteste 81. Der Altersdurchschnitt betrug 32.6 Jahre ($SD = 15.4$). Von 10 Personen fehlen Alters- und Geschlechtsangaben. Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, fanden die Datenerhebungen an öffentlichen Orten wie Einkaufszentren, Fußgängerzonen oder Bahnhöfen statt. Die Teilnahme am Experiment erfolgte freiwillig und wurde nicht entlohnt. Beim durchschnittlichen Alter ($t(87) = -0.155$, $p = .877$) sowie bei der Altersverteilung (Levene-Test auf Varianzgleichheit: $p = .138$) gibt es keine signifikanten Unterschiede zwischen den Geschlechtern.

6.4.1.2 Stimulusmaterial und Versuchsdurchführung

Das gesamte Experiment erfolgte analog zur vorangegangenen Untersuchung (*Dimorph 2*), nur dass anstatt Frauengesichter Männergesichter an das Kindchenschema angenähert wurden. Auch hier wurden fünf unattraktive Gesichter (A bis E) und fünf attraktive (F bis K) verwendet (Grundlage der Attraktivitätsbewertungen waren die Daten aus der Untersuchung *Durch 1*). Unter den Gesichtern befanden sich auch hier die beiden gemorpheten Prototypen für ein unattraktives bzw. attraktives Männergesicht.

Mit Ausnahme eines sehr attraktiven Originalgesichts (ein männliches Fotomodel, das von allen männlichen Originalgesichtern der Voruntersuchung von *Durch 1* die höchsten Attraktivitätsbewertungen bekommen hatte) handelte es sich um gemorphte Gesichter (*2-face-morphs* und *4-face-morphs*).

Die Annäherung an das Kindchenschema erfolgte wie bei den Frauengesichtern mit dem Warp-Verfahren unter Beibehaltung sämtlicher Farbwerte und mit einer Stärke von 40 %. Die beiden folgenden Abbildungen zeigen Beispiele für ein unattraktives (Abbildung 66) und ein attraktives Stimulusgesicht (Abbildung 67).



Abbildung 66: Unattraktives erwachsenes Männergesicht (D, links) und mit Morphing-Software verkindlichte Variante, bei der das Kindchenschema mit 40 % gewichtet wurde (rechts).

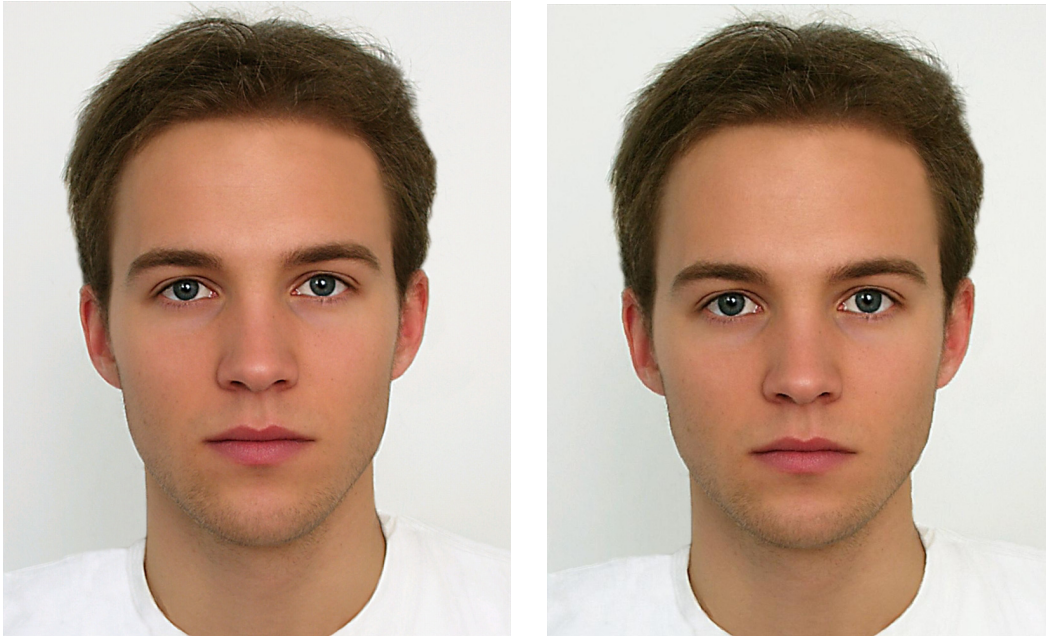


Abbildung 67: Attraktives erwachsenes Männergesicht (K, links) und mit Morphing-Software verkindlichte Variante, bei der das Kindchenschema mit 40 % gewichtet wurde (rechts).

Die Versuchsdurchführung verlief analog zum Experiment mit den „verkindlichten“ Frauengesichtern. Alle Bilder hatten dieselbe Größe und Auflösung und wurden paarweise in Fotomappen in ausbalancierter Reihenfolge präsentiert. Die Versuchspersonen mussten angeben, welche Bildversion sie attraktiver fanden (*forced choice*).

6.4.2 Ergebnisse

Nach der Theorie des sexuellen Dimorphismus sollten Männergesichter an Attraktivität einbüßen, wenn ihre Gesichtsproportionen an das Kindchenschema angenähert werden, da sie durch diese Manipulation weniger maskulin aussehen. Ob dies tatsächlich zutrifft, wurde durch zehn Binomialtests berechnet. Wie im Experiment mit den Frauengesichtern wurden auch hier wieder alle Effektgrößen vom Binomial effect size display (*BESD*) nach dem Verfahren von Rosenthal und Rubin (1982) in ein r umgerechnet. Tabelle 20 listet die absolute und relative Häufigkeit der Bevorzugung der „verkindlichten“ Variante gegenüber der unveränderten erwachsenen Original-Variante des Männergesichts sowie die in r umgerechnete Effektgröße auf. Am Ende der Tabelle ist der Median der erzielten Häufigkeiten und Effektgrößen aufgeführt.

Tabelle 20: Ergebnisse des Paarvergleichs-Experiments mit „verkindlichten“ Männergesichtern (40 % Kindchenschema) im direkten Vergleich mit dem jeweiligen unveränderten erwachsenen Gesicht. Aufgelistet sind absolute und relative Häufigkeiten sowie die in r umgerechneten Effektgrößen. $N = 104$.

Stimulus-Gesicht	Attraktivität des Originalgesichts	Häufigkeit der Bevorzugung der kindlichen Variante n (%)	p	r
A	unattraktiv	77 (77.8 %)	<.001	0.56
B	unattraktiv	80 (80.8 %)	<.001	0.62
C	unattraktiv	62 (62.6 %)	.015	0.25
D	unattraktiv	71 (71.7 %)	<.001	0.43
E	unattraktiv	38 (38.4 %)	.027	-0.23
F	attraktiv	46 (46.5 %)	.547	-0.07
G	attraktiv	59 (59.6 %)	.070	0.19
H	attraktiv	66 (66.7 %)	.001	0.33
I	attraktiv	53 (53.5 %)	.547	0.07
K	attraktiv	28 (28.3 %)	<.001	-0.43
Median		60.5 (61.1 %)		0.22

Anmerkung: **Fett** gedruckte p -Werte zeigen eine signifikante Bestätigung der Theorie des sexuellen Dimorphismus, *kursiv* gedruckte p -Werte zeigen einen signifikanten Widerspruch zur Theorie.

Wie ein Blick auf Tabelle 20 zeigt, fallen die Ergebnisse uneinheitlich aus. Von den 10 Binomialtests sind 7 statistisch signifikant, allerdings in unterschiedliche Richtungen. Von den 5 unattraktiven Männergesichtern wurde nur eines in der kindlichen Variante als unattraktiver bewertet, die übrigen 4 Männer jedoch signifikant attraktiver – also im Widerspruch zur Theorie des sexuellen Dimorphismus. Bei den 5 attraktiven Männergesichtern wurde eines in der „verkindlichten“ Variante signifikant schlechter und eines signifi-

kant besser bewertet – bei den übrigen 3 Gesichtern gab es keinen Unterschied. Die signifikanten Attraktivitätsunterschiede bewegen sich hinsichtlich der Effektgröße in einem mittleren Bereich.

Von einer Bestätigung der Theorie des sexuellen Dimorphismus kann bei den Männergesichtern nicht gesprochen werden, da nur bei 2 von 10 Gesichtern ein signifikanter Effekt in der vorhergesagten Richtung auftrat. Stattdessen überwiegen die Hinweise gegen die Theorie, da bei 6 von 10 Gesichtern die kindlicher und femininer aussehenden Männergesichter als attraktiver bezeichnet wurden. Insbesondere die unattraktiven Männergesichter scheinen von der „Verkindlichung“ ihres Gesichts zu profitieren. Der Median der Effektgrößen beträgt hier $r = .43$. Bei den attraktiven Männergesichtern gibt es hingegen keinen erkennbaren Trend ($r = .07$).

Auch bei den hier vorliegenden Daten gibt es wieder – wie schon bei den Paarvergleichen der Symmetrie-Experimente und dem vorangegangenen Experiment mit den verkindlichten Frauengesichtern – anstatt der mehrfachen Binomialtests eine alternative Form der Datenanalyse nach Little und Jones (2003, S. 1761). Für jede Versuchsperson wird dabei eine relative Häufigkeit der Bevorzugung der kindlichen Version berechnet. Diese Häufigkeiten werden mit einem Ein-Stichproben- t -Test analysiert und gegen den Wert von 0.5 (= A-priori-Wahrscheinlichkeit) getestet. Analysiert man die Daten mit dieser Methode, ergibt sich ein hoch signifikantes Ergebnis $t(98) = 3.488$, $p = .001$. Die Effektstärke, die sich aus dem t -Wert berechnet (siehe Formel in Kapitel 4.4.3), fällt dadurch ebenfalls deutlich größer aus und beträgt dann $r = 0.40$ (anstatt 0.22). Mit diesem Auswertungsverfahren müsste man also den Schluss ziehen, dass sich insgesamt eine Annäherung von Männergesichtern an das Kindchenschema Attraktivität erhöhend auswirkt, die Theorie des sexuellen Dimorphismus für Männergesichter also widerlegt ist. Der erzielte Effekt ist dabei schwach bis mittel und das Ergebnis hoch signifikant. Eine kritische Diskussion dieser Auswertungsmethode findet sich in Kapitel 5.2.3.



Abbildung 68: Die Stimuluspersonen B (linkes Paar) und C (rechtes Paar). Bei unattraktiven Männergesichtern wirkte sich das Verzerren der Gesichtsproportionen in Richtung Kindchenschema fast immer attraktivitätserhöhend aus – ein Widerspruch zur Theorie des sexuellen Dimorphismus.

In einer abschließenden Analyse wurde mit zusätzlichen Tests überprüft, ob Geschlecht oder Alter der Versuchspersonen einen Einfluss auf das Urteilsverhalten hatten. Für den Test auf Geschlechtsunterschiede wurden zehn χ^2 -Tests gerechnet. Es ergaben sich keinerlei signifikante Unterschiede (p -Werte zwischen .084 und .652). Auch die analog zu den Frauengesichtern durchgeführte Analyse der „relativen Häufigkeiten der Bevorzugung für die kindliche Variante pro Versuchsperson“ zeigte keinen Geschlechtsunterschied (t -Test für unabhängige Stichproben: $t(87) = -1.073$, $p = .286$). Frauen haben also hinsichtlich der Maskulinität eines Männergesichts das gleiche Attraktivitätsempfinden wie Männer.

Eine Überprüfung des Einflusses des Alters der Versuchspersonen auf das Urteilsverhalten zeigte eine schwache Tendenz: Ältere Personen bevorzugten etwas häufiger kindlich aussehende Männergesichter. Bei zehn χ^2 -Tests zwischen Alterskategorie (Altersklassen: bis 25 Jahre, 26 bis 50 Jahre und 51 Jahre und älter) und gewählter Variante ergab sich nur für das attraktive Stimulusgesicht F ein signifikanter Unterschied ($p = .002$). Bei den attraktiven Gesichter I und K zeigte sich eine ähnliche, jedoch knapp nicht signifikante Tendenz ($p = .082$ und $p = .059$). Die Übrigen p -Werte liegen zwischen .123 und .638.

Eine Korrelation zwischen Alter und der relativen Häufigkeit (pro Versuchsperson), mit der die kindliche Variante ausgewählt wurde, zeigte ebenfalls einen schwachen, aber knapp nicht signifikanten Zusammenhang ($r = .20$, $p = .061$) in derselben Richtung. Damit kann gesagt werden, dass es keinen eindeutigen Beleg dafür gibt, dass das Alter der Versuchspersonen einen Einfluss auf die Präferenz von Maskulinität bei Männergesichtern hat. Dies müssten weitere Untersuchungen mit einer größeren Stichprobe klären. Wenn

es jedoch einen Alterseffekt geben sollte, dann ist er sehr schwach und dahin gehend, dass ältere Personen im Vergleich zu jüngeren Männergesichter bevorzugen, die weniger maskulin aussehen.

Zusammenfassend bleibt festzuhalten, dass bei Männergesichtern die Ergebnisse zur Wirkung geschlechtstypischer Merkmale nicht eindeutig sind. Von einer Bestätigung der Theorie des sexuellen Dimorphismus kann bei den Männergesichtern nicht gesprochen werden, da nur bei 2 von 10 Gesichtern ein signifikanter Effekt in der vorhergesagten Richtung auftrat. Stattdessen überwiegen die Hinweise gegen die Theorie, da bei 5 von 10 Gesichtern die kindlicher und femininer aussehenden Männergesichter signifikant als attraktiver bezeichnet wurden. Insbesondere die unattraktiven Männergesichter scheinen von der „Verkindlichung“ ihres Gesichts zu profitieren. Der Median der Effektgrößen beträgt hier $r = .43$. Bei den attraktiven Männergesichtern gibt es hingegen keinen erkennbaren Trend ($r = .07$). Der Median über alle Männergesichter liegt bei $r = .22$, d. h. dass „verkindlichte“ Männergesichter etwas attraktiver bewertet werden. Bei einer Datenanalyse, die die Urteilsvarianz der Versuchspersonen *nicht* berücksichtigt, ist dieser Effekt noch größer ($r = 0.40$). Das Geschlecht der Beurteiler hat keinen Einfluss auf die Ergebnisse und das Alter kann vernachlässigt werden. Insgesamt kann man nach den Ergebnissen dieses Experiments die Theorie des sexuellen Dimorphismus bei Männergesichtern eher als widerlegt ansehen.

7 Zusammenfassende Diskussion der Ergebnisse zum Einfluss des sexuellen Dimorphismus

7.1 Vergleich der Ergebnisse mit bisherigen Studien zum sexuellen Dimorphismus bei Frauengesichtern

Bei den beiden durchgeführten Experimenten *Dimorph 1* und *Dimorph 2* mit den manipulierten *Frauengesichtern* lässt sich die Theorie des sexuellen Dimorphismus klar bestätigen. Beim ersten Experiment mit den 6er-Variationen der in Richtung Kindchenschema manipulierten Gesichter gab es unabhängig von Geschlecht oder Alter der Beurteiler eine allgemeine Präferenz für Frauengesichter, die kindlichere (d. h. femininere) Gesichtszüge aufweisen, als dies für erwachsene Frauengesichter normal ist. Die Größe des Effekts

hängt jedoch vom konkreten Aussehen des Originalgesichts ab, das an das Kindchenschema angenähert wird. Das zweite Experiment mit verkindlichten Frauengesichtern im Paarvergleichsverfahren bestätigt die Ergebnisse des ersten Experiments, zeigt jedoch auch, dass der Effekt eher klein ist, denn der Median der nach r standardisierten Effektgrößen liegt bei $r = .34$. Bei den unattraktiven Frauengesichtern ist dieser Effekt größer ($r = 0.48$), bei den attraktiven hingegen kleiner ($r = 0.27$).

Die Ergebnisse bestätigen damit die Befunde bisheriger Studien, bei denen die Femininität eines Gesichts ebenfalls experimentell manipuliert (Johnston et al., 2001; Perrett et al., 1998; Rhodes et al., 2000) bzw. überzeichnet wurde (z. B. durch die Karikatur-Technik), was zu einem weiteren Anstieg der Attraktivitätsbewertungen führte (Johnston & Franklin, 1993; Perrett et al., 1998; Rhodes et al., 2000). Auch die Attraktivitätssteigerung als Folge einer moderaten Verkindlichung durch Annäherung an die Proportionen eines durchschnittlichen Kindergesichts konnte bestätigt werden (Ishi, Gyoba & Kamachi, 2004).

Die Ergebnisse decken sich auch mit anderen Studien, die Femininität nicht experimentell manipulierten, sondern durch anthropometrische Vermessungen von weiblichen Originalgesichtern erfassten (Cunningham, 1986; Cunningham et al., 1995; Johnston & Franklin, 1993; Jones & Hill, 1993; Koehler et al., 2004). Zu einem ähnlichen Ergebnis kommen Studien, die Femininitäts-Ratings durch Versuchspersonen als Maß für Femininität nutzten (Bruce et al., 1994; Dunkle & Francis, 1990; Koehler et al., 2004; O'Toole et al., 1998; Rhodes et al., 2003). Diese Methode ist allerdings kritisch zu bewerten, da es Versuchspersonen sehr schwer fallen dürfte, die Femininität eines Frauengesichts zu bewerten, ohne sich dabei gleichzeitig von der Attraktivität dieses Gesichts beeinflussen zu lassen (vgl. auch die Ergebnisse des Zusatzexperiments zur Symmetrie-Untersuchung *Sym 7 Kontrolle*, Kapitel 4.10.4, bei der eindeutig nachgewiesen wurde, dass Symmetrie-Ratings von Gesichtern mit deren Attraktivitätsbewertungen hochgradig konfundiert sind). Fast alle diese Untersuchungen (mit einer Ausnahme einer Teiluntersuchung von Koehler et al., 2004) fanden einen positiven Zusammenhang zwischen Femininität und Attraktivität – wenn auch mit sehr unterschiedlichen Effektgrößen.

Hervorzuheben ist aus dem Experiment *Dimorph 1* das Ergebnis mit dem manipulierten weiblichen Durchschnittsgesicht: Die höheren Attraktivitätsbewertungen der Versionen mit einem höheren Anteil von kindlichen Gesichtsproportionen zeigen, dass die Proportionen eines durchschnittlichen erwachsenen Frauengesichts offensichtlich nicht ideal attraktiv sind (im Gegensatz zur Behauptung der Durchschnittshypothese!), sondern dass kindli-

chere Gesichtsproportionen als attraktiver empfunden werden. Hier drängt sich der Vergleich mit den Studien auf, die als Stimulusmaterial ebenfalls ein weibliches Durchschnittsgesicht verwendet haben, das sie dann stufenlos auf einem Kontinuum (*sex continuum*) maskuliner bzw. femininer gemacht haben (Johnston et al., 2001; Perrett et al., 1998; Rhodes et al., 2000). Sie fanden ebenfalls einen positiven Effekt femininer Proportionen im Gegensatz zu durchschnittlichen Proportionen. Gleichzeitig zeigen die Ergebnisse der anderen fünf manipulierten Stimulusgesichter aus dem Experiment *Dimorph 1* aber auch, dass das Ergebnis des Durchschnittsgesichts keineswegs auf alle anderen Stimulusgesichter generalisierbar ist. Denn es wurde ebenfalls dargestellt, dass das Treatment mit dem Ausgangsgesicht interagiert, d. h. dieselbe Manipulation bei unterschiedlichen Gesichtern zu unterschiedlichen Effekten führt. Bei anderen Gesichtern war der positive Effekt der Verkindlichung noch viel deutlicher – selbst bei hoch attraktiven Originalgesichtern (z. B. Gesicht B, Abbildung 58, S. 251).

Interaktionseffekte wurden auch im Experiment *Dimorph 2* gefunden. Hier war der Attraktivitätsgewinn durch die Manipulation in Richtung Kindchenschema bei unattraktiven Gesichtern größer als bei attraktiven. Dies zeigt, dass die oben genannten Experimente mit *sex-continua* und nur einem einzigen manipulierten Durchschnittsgesicht wenig aussagekräftig sind, da bei Verwendung eines anderen Stimulusgesichts vermutlich ein abweichendes Ergebnis herausgekommen wäre. Das Ergebnis dieser Untersuchung ist daher nicht auf andere Gesichter generalisierbar, daran ändert auch die Tatsache nichts, dass es sich bei dem manipulierten Frauengesicht um ein Durchschnittsgesicht handelt. Im Gegenteil, diese Composites besitzen ja eben etliche Eigenschaften (v. a. in ihrer Textur), die sie alles andere als durchschnittlich erscheinen lassen. Von daher sind sie sehr artifizielle und daher ökologisch wenig valide Stimuli, um die Wirkung von kindlich-femininen Merkmalen auf die Attraktivität eines Gesichts zu untersuchen. In den hier durchgeführten beiden Experimenten mit Frauengesichtern wurden immerhin 16 Stimulusgesichter manipuliert – doch auch das ist eigentlich immer noch viel zu wenig, um diese Ergebnisse auf alle Frauengesichter zu verallgemeinern.

7.2 Vergleich der Ergebnisse mit bisherigen Studien zum sexuellen Dimorphismus bei Männergesichtern

Die Ergebnisse zur Überprüfung der Theorie des sexuellen Dimorphismus bei den Männergesichtern können die Theorie *nicht* bestätigen, da nur bei 2 von 10 Gesichtern ein

signifikanter Effekt in der vorhergesagten Richtung auftrat. Stattdessen überwiegen die Hinweise *gegen* die Theorie, da bei 6 von 10 Gesichtern die kindlicher und femininer aussehenden Männergesichter als attraktiver bezeichnet wurden. Insbesondere die unattraktiven Männergesichter scheinen von der „Verkindlichung“ ihres Gesichts zu profitieren. Der Median der Effektgrößen beträgt hier $r = .43$, d. h. kindliche und damit weniger maskuline Männergesichter werden attraktiver bewertet. Bei den attraktiven Männergesichtern gibt es hingegen keinen erkennbaren Trend ($r = .07$).

Die Ergebnisse sind damit ein Spiegelbild der ebenfalls sehr heterogenen Befundlage bisheriger Studien. Am besten vergleichbar sind die Daten mit solchen Untersuchungen, die geschlechtstypisches Aussehen ebenfalls experimentell mit Morphing-Software manipuliert und den Effekt auf die Attraktivität gemessen haben (Ishi, Gyoba, Kamachi et al., 2004; Little et al., 2001; Little & Hancock, 2002; Perrett et al., 1998; Rhodes et al., 2000; Swaddle & Reiersen, 2002). Solche Studien kommen durchweg zu dem Ergebnis, dass maskuline Merkmale *unattraktiver* (!) machen (in der Metaanalyse von Rhodes, 2006, ist die Effektstärke dieser Art von Studien negativ; $r = -0.47$). Zwei Ausnahmen gibt es jedoch: Ein Experiment von Keating (1985), das jedoch mit Gesichtern als Strichzeichnungen arbeitete, bei dem maskulinisierte Gesichter (dickere Augenbrauen, schmalere Lippen, ein breites Kinn und schmale Augen) gegenüber femininisierten bevorzugt wurden. Aufgrund des wenig realistischen Stimulusmaterials der Strichzeichnungen ist die externe Validität des Experiments jedoch gering. Die zweite Ausnahme ist das Experiment von Johnston et al. (2001), das zu den am häufigsten zitierten Untersuchungen der Attraktivitätsforschung zählt, das jedoch einige methodische Mängel besitzt, die im folgenden Kapitel 7.3 ausführlich dargelegt werden.

Anders hingegen die Studien, bei denen Maskulinität *nicht* manipuliert war, sondern entweder durch das Vermessen von *Facial Landmarks* unveränderter Originalgesichter ermittelt und dann mit Attraktivitätsbewertungen korreliert wurde (Cunningham et al., 1990; Grammer & Thornhill, 1994; Koehler et al., 2004; Penton-Voak, Jones et al., 2001; Scheib et al., 1999). Sie fanden positive Effekte für Maskulinität. Gleiches gilt für die umstrittene Operationalisierung von Maskulinität mittels Ratingskalen durch Versuchspersonen (Bruce et al., 1994; Cunningham et al., 1990; Dunkle & Francis, 1990; Koehler et al., 2004; Little & Hancock, 2002; Neave et al., 2003; O'Toole et al., 1998; Rennels et al., 2008; Rhodes et al., 2003) – diese Studien fanden ebenfalls positive Effekte. In der Metaanalyse von Rhodes (2006) betrug der positive Effekt von Maskulinität bei Studien, die mit Gesichtern arbeiteten, bei denen die Maskulinität *nicht* manipuliert war, $r = 0.35$.

Da bei den hier vorliegenden drei Untersuchungen sexueller Dimorphismus experimentell mit Morphing-Software manipuliert wurde und die verwendete Methode (v. a. bei männlichen Gesichtern, vgl. Kapitel 6.1.5 „Empirische Befunde zum Einfluss maskuliner Gesichtszüge bei Männergesichtern“) einen so offensichtlichen Einfluss auf das Ergebnis hat, ist es wenig zielführend, die hier gewonnenen Ergebnisse mit den Ergebnissen aus Studien zu vergleichen, die korrelativ mit der Vermessung von *Facial Landmarks* oder Ratings der Geschlechtstypikalität arbeiteten. Sinnvoller erscheint es, die hier dargestellten Ergebnisse mit anderen experimentellen Studien zu vergleichen. Hierzu werden diejenigen drei Studien herausgegriffen und ausführlicher diskutiert, die mit Abstand am häufigsten zitiert werden und mit einer vergleichbaren Methodik arbeiteten.

7.3 Vergleich der Ergebnisse mit den drei meist zitierten Experimenten zum sexuellen Dimorphismus

Die drei am häufigsten zitierten experimentellen Arbeiten zum sexuellen Dimorphismus sind die Studien von Perrett et al. (1998), Johnston et al. (2001) und Rhodes et al. (2000). Sie sollen im Folgenden detaillierter dargestellt und v. a. hinsichtlich ihrer Methodik kritisch diskutiert werden, da sie nämlich alle drei an teils erheblichen methodischen Schwächen leiden. Im Anschluss daran werden die eigenen Untersuchungen mit diesen drei Untersuchungen verglichen.

Die bekannteste Studie, die den sexuellen Dimorphismus von Gesichtern bei beiden Geschlechtern experimentell manipulierte, ist die Untersuchung von Perrett et al. (1998). Sie ist zugleich die erste dieser Art und schaffte es sogar in das hoch renommierte Journal *Nature*. Das Fazit dieses Artikels ist, dass nicht sexuell dimorphe Merkmale an sich, sondern feminine (neotene) Merkmale Gesichter attraktiv machen, denn sowohl weibliche als auch männliche Durchschnittsgesichter (sowohl von britischen als auch von japanischen Stimuluspersonen) profitierten von einer Femininisierung ihrer Gesichtsproportionen durch Morphing-Software. Perrett et al. näherten für ihr Experiment das weibliche Durchschnittsgesicht stufenlos (bis 50 %) an die männlichen Durchschnittsproportionen an (= *masculinized*) und zum anderen verzerrten sie es (durch die Karikatur-Technik) ebenfalls stufenlos (bis 50 %) von den männlichen Durchschnittsproportionen weg (= *feminized*) – analog verfahren sie mit Männergesichtern. Die Messung der Präferenz erfolgte durch die sogenannte *Herstellungsmethode*, bei der die Versuchspersonen durch Betätigung eines Reglers das für sie attraktivste Maß an Feminität bzw. Maskulinität am

präsentierten Gesicht auf einem Bildschirm einstellen konnten. Problematisch an dieser Untersuchung sind zwei Aspekte, nämlich (1) die dargebotenen Stimuli selbst und (2) die Art der Datenauswertung.

(1) Wie sich der Leser mit einem Blick auf die Abbildung 2 (S. 885) im Artikel von Perrett et al. (1998) selbst überzeugen kann, weisen die manipulierten Gesichter nur äußerst geringe Unterschiede auf, selbst zwischen den beiden Extremversionen der maskulinierten und femininisierten Version eines Gesichts. Das einzige geschlechtstypische Merkmal, das sich klar erkennbar verändert, ist die Dicke der Augenbrauen. Selbst ein Merkmal wie die Fülle der Lippen verändert sich nicht mit der Manipulation. Der auffälligste Unterschied hingegen ist ein Artefakt des Stimulusmaterials: Denn Perrett und Kollegen präsentierten die Stimuli nicht als vollständige Gesichter mit Hals und Frisur, sondern schnitten die Gesichter unterhalb des Unterkiefers und oberhalb des Haaransatzes ab (obwohl die Größe des Hirnschädels, d. h. die Höhe des Scheitelpunkts einen wichtigen Einfluss auf die Attraktivität eines Gesichts hat, vgl. Johnston & Franklin, 1993). Die Ohren fehlen ebenfalls. Umgeben sind diese „Restgesichter“ von einer Art schwarzem Passpartout, so dass die gezeigten Gesichter etwas an Eier erinnern. Bei allen maskulinierten Gesichtern wird ein erheblicher Teil der Stirnpartie einfach abgeschnitten, was angeblich ein maskulines Aussehen darstellen soll, in Wirklichkeit aufgrund der fehlenden Frisur und fehlenden optischen Informationen über die Kopfform einfach nur seltsam aussieht. Die Manipulation geschlechtstypischer Merkmale war bei Perrett et al. aufgrund fehlender Unterschiede offenkundig nicht valide und die Reizdarbietung war aufgrund der beschnittenen Stirnpartien von einem erheblichen Störfaktor beeinträchtigt.

Auch die „niedrigere“ Stirn ist kein valides Merkmal, denn wie aus der Abbildung 1 (Perrett et al. 1998, S. 885) mit den unveränderten Durchschnitts-Composites ersichtlich ist, resultiert die untere niedrigere Haarlinie (*hairline*) bei den Männern gar nicht aus dem Haaransatz, sondern aus einzelnen Stimuluspersonen, denen während der Aufnahme des Fotos die Haare teilweise die Stirn verdeckten (– diese Haarlinie ist zudem völlig schief). Sie spiegelt *nicht* einen durchschnittlichen niedrigeren Haaransatz bei den Männern wider und erst recht *nicht* einen proportional zur Gesichtslänge kleineren Hirnschädel (was ein Reifekennzeichen und das Gegenteil zum Kindchenschema wäre). Denn erstens ist der Hirnschädel aufgrund der fehlenden Frisur bei den präsentierten Stimuli gar nicht sichtbar und zweitens sieht man auf Abbildung 1 sowohl bei den britischen als auch bei den japanischen Durchschnittsgesichtern, dass die männlichen (und nicht die weiblichen) Composites aufgrund der höher liegenden oberen Haarlinie einen größer wirkenden Hirnschädel besitzen als die weiblichen Composites. Dies freilich sind Artefak-

te durch den Morphprozess, die jedoch durch mehr Sorgfalt und Mühe beim Erstellen der Composites leicht zu verhindern gewesen wären (vgl. dazu auch die Erläuterungen in Kapitel 2.3.1.3 bei der Beschreibung zur Erstellung der Durchschnittsgesichter für die in der vorliegenden Untersuchung verwendeten gemorphten Gesichter). Morphing-Artefakte sind auch die verschwommenen und schiefen Übergänge zwischen Haaren und Stirnpartie, die durch mehr Sorgfalt bei der Fotoaufnahme und sorgfältiges Retuschieren nach jedem paarweisen Vermorphen ebenfalls vermeidbar gewesen wären. Dass dem Leser von Perrett et al. jedoch diese klassischen Morphing-Artefakte als geschlechtstypische Unterschiede in der Gesichtsform verkauft werden, ist nicht mangelnde Sorgfalt, sondern schlicht unseriös. Aufgrund der beschriebenen Artefakte beim Stimulusmaterial ist es daher auch nicht sonderlich erstaunlich, dass die Versuchspersonen bei ansonsten praktisch nicht vorhandenen optischen Unterschieden diejenigen Gesichtsausschnitte bevorzugten, bei denen weniger von der Stirnpartie abgeschnitten war.

(2) Ein zweiter Kritikpunkt ist die Art der Datenauswertung. Perrett et al. (1998) stellen als Nullhypothese auf, dass im Durchschnitt über alle Versuchspersonen genau diejenige Gesichtsversion eingestellt wird, die die Proportionen des Durchschnittsgesichts besitzen, die Version also, die weder maskulinisiert noch femininisiert wurde. Die Alternativhypothese hingegen lautet, dass im Durchschnitt irgendeine andere Version ausgewählt wird. Für die Datenanalyse werten die Autoren nicht etwa Häufigkeitsverteilungen aus, sondern sie berechnen den durchschnittlich eingestellten Verzerrungsgrad des gezeigten Durchschnittsgesichts. Der Wertebereich reicht dabei von -50 % bis +50 % (mit vermutlich 1%-Abstufungen – das geben die Autoren nicht an). In einem Ein-Stichproben-*t*-Test berechnen sie dann, ob der durchschnittliche gewählte Verzerrungsgrad signifikant von 0 % (= Nullhypothese) abweicht. Problematisch an diesem Vorgehen ist, dass ein solcher Test sehr leicht statistisch signifikant werden kann, z. B. wenn es einige wenige Ausreißer in dieselbe Richtung gibt oder wenn die Mehrheit der Versuchspersonen eine Bildvariante einstellt, die nur geringfügig von der 0%-Verzerrung abweicht, jedoch in die gleiche Richtung (z. B. in Richtung nicht verstümmelter Stirnpartie) geht. In diesem Fall wird dann eine Bildvariante als die signifikant attraktivste dargestellt, die sich in ihrem Aussehen optisch praktisch gar nicht von der 0%-Verzerrung unterscheidet (so z. B. beim japanischen Frauen-Composite: „mean 10.2%; $t_{49} = 2.3$, $P = 0.027$ “, Perrett et al., 1998, S. 885). Ob jedoch die durchschnittliche Verzerrung tatsächlich eine Bildvariante bezeichnet, die attraktiver als die 0%-Verzerrung des Durchschnittsgesichts ist, müsste eigentlich erst durch eine weitere Datenerhebung validiert werden, indem man paarweise diese ermittelte Bildvariante gegen die unveränderte Variante des Durchschnittsgesichts testet. Dass dabei immer noch ein signifikanter Attraktivitätsunterschied herauskäme, ist angesichts

der kaum wahrnehmbaren Unterschiede selbst zwischen den jeweiligen Extremvarianten ziemlich unwahrscheinlich. Leider drucken Perrett und Kollegen die Bildvarianten, die das vermeintliche Attraktivitätsideal darstellen, nicht ab, denn dann könnte jeder Leser selbst ausprobieren, ob er denn einen Unterschied zwischen dem unveränderten Gesicht und seiner 10%-Verzerrung erkennen kann. Aus statistischer Sicht wäre es zudem auch sehr interessant gewesen, die Häufigkeitsverteilungen für die präferierten Bildvarianten zu erfahren (schließlich ist Normalverteilung Voraussetzung für einen *t*-Test). Leider berichten Perrett und Kollegen in ihrem *nature*-Artikel jedoch nur Mittelwerte und noch nicht einmal eine einzige Standardabweichung oder einen einzigen Standardfehler. Insgesamt muss man daher festhalten, dass der viel zitierte Artikel von Perrett et al. (1998) unter erheblichen methodischen Mängeln leidet und es daher fraglich ist, ob die berichteten Ergebnisse mehr sind als bloße Methodenartefakte.

Die zweite wichtige und viel zitierte Arbeit, bei der experimentell der sexuelle Dimorphismus eines Composites manipuliert wurde, stammt von Johnston et al. (2001). Hier wurde Versuchspersonen (ausschließlich Frauen) ein 40 Sekunden langer Film (aus 1200 Einzelbildern) präsentiert, bei dem sich ein hypermaskulines Männergesicht allmählich in ein hyperfemines Frauengesicht verwandelt. Die Probandinnen sollten den Film an der Stelle anhalten, an der ihnen das Männergesicht am besten gefiel (analog für weibliche Gesichter). Das präsentierte Stimulusmaterial sieht realistisch und natürlich aus und besitzt eine weit höhere Qualität als die von Perrett et al. (1998) verwendeten Stimuli. Johnston et al. fanden heraus, dass ein Männergesicht attraktiver wirkt, wenn es maskulinere Gesichtsm Merkmale besitzt als ein durchschnittliches Männergesicht und ist damit die einzige Studie, die bei einem Männergesicht den sexuellen Dimorphismus experimentell manipuliert hat und die Vorhersage dieser Theorie tatsächlich nachweisen konnte. Dieses Ergebnis steht somit auch im Widerspruch zum Ergebnis des hier durchgeführten Experiments, bei dem das Gegenteil von Maskulinität, nämlich kindlich-feminine Merkmale, bei Männern besser bewertet wurden. Genau wie in der Studie von Perrett et al. (1998) gibt es aber auch hier die gleichen methodischen Schwachstellen, nämlich (1) das Stimulusmaterial selbst und (2) die Art der Datenauswertung.

(1) Das Ergebnis von Johnston et al. (2001) erscheint in einem anderen Licht, wenn man sich vor Augen hält, auf welche Weise das Reizmaterial entstand. Zuerst morphten die Autoren aus jeweils 16 Männer- und 16 Frauengesichtern ein männliches und ein weibliches Durchschnittsgesicht. Dann vermorphten sie diese beiden Durchschnittsgesichter und stellten daraus einen Filmabschnitt her, bei dem sich das männliche Composite (unter Neuberechnung der Farbwerte) langsam in das weibliche verwandelt; die Mitte dieses

Films ist gleichsam eine androgyne Mischung aus Mann und Frau. In einem nächsten Schritt stellten sie mit einer selbst entwickelten Art Phantombildsoftware namens *FacePrints* ein typisch maskulines und ein typisch feminines Composite-Gesicht her, indem sie Versuchspersonen instruierten, mit Hilfe dieser Software ein sehr männlich bzw. sehr weiblich aussehendes Gesicht zu generieren. Leider geben Johnston et al. in ihrem Methodenteil dem Leser keine genauen Informationen darüber, wie genau sie zu ihren geschlechtstypischen Gesichtern kamen und verweisen nur auf einen eigenen älteren Artikel (Johnston & Franklin, 1993), in dem allgemein beschrieben ist, nach welchem Prinzip die Software funktioniert. Der bestehende Film wurde dann „erweitert“, indem gleitende Übergänge zwischen dem Durchschnittsgesicht eines Geschlechts und seiner mit der Phantombildsoftware hergestellten Version gemorpht wurden (vgl. Johnston et al., 2001, Abbildung 1, S. 255). In einem letzten Schritt wurde der Film noch einmal erweitert, indem mit Hilfe der Karikatur-Technik die geschlechtstypischen Gesichter zusätzlich noch übertrieben wurden.

Aus methodischer Sicht lautet bei der Arbeit von Johnston und Kollegen die entscheidende Frage: Wie kommt das geschlechtstypische männliche Gesicht zustande? Schließlich sieht es deutlich anders als das männliche Durchschnittsgesicht aus. Was ist der Beleg dafür, dass bei den Phantombildern tatsächlich nur die Maskulinität optimiert wurde und nicht etwa auch gleichzeitig die Attraktivität? Denn wenn man einer Versuchsperson die Instruktion gibt, ein typisch männlich aussehendes Gesicht zu kreieren, ist es plausibel, dass sie mit der Phantombildsoftware gleichzeitig auch ein Gesicht herstellt, das nicht nur männlich, sondern auch attraktiv aussieht. Das Ergebnis ist jedoch nicht ein Beleg dafür, dass es tatsächlich einen Zusammenhang zwischen typisch maskulinen Gesichtszügen und Attraktivität bei Männergesichtern gibt, sondern lediglich, dass die Versuchsperson *der Meinung* ist, dass es einen solchen Zusammenhang gibt. Man erfasst damit nicht die tatsächliche Ursache von Attraktivität, sondern möglicherweise nur ein Vorurteil über Attraktivität. Genau dies ist schließlich auch der Einwand, der gegen die Studien vorgebracht wurde, bei denen Attraktivitätsratings von Gesichtern mit Maskulinitätsratings korreliert wurden und die durchweg zu positiven Zusammenhängen kamen (Bruce et al., 1994; Cunningham et al., 1990; Dunkle & Francis, 1990; Koehler et al., 2004; Little & Hancock, 2002; Neave et al., 2003; O'Toole et al., 1998; Rennels et al., 2008; Rhodes et al., 2003). Letztendlich stößt man hier wieder auf das altbekannte Attraktivitätsstereotyp („Was schön ist, ist auch gut“, Dion et al., 1972) und die triviale Einstellung, dass es gut ist, wenn ein Mann männlich (und nicht weiblich) aussieht (vgl. dazu auch die Erläuterungen in Kapitel 6.1.5 „Empirische Befunde zum Einfluss maskuliner Gesichtszüge bei Männergesichtern“). Ein ganz ähnliches Problem sind Symmetrie-Ratings von Gesicht-

tern, die ausführlich in Kapitel 4.10.4 untersucht wurden und deren fehlende Validität klar nachgewiesen wurde.

Anders gefragt: Wie kommt es, dass beim Stimulusmaterial von Johnston und Kollegen mit zunehmender Männlichkeit des Gesichts zwar der Unterkiefer markant und kantig sowie der Nacken breit und kräftig werden, gleichzeitig (selbst bei der hypermaskulinen Karikatur) die Lippen dennoch *nicht* schmal und die Nase *nicht* groß ist, obwohl beides typisch männliche Merkmale sind? Obwohl also in diesem Experiment das Stimulusmaterial hochwertig aussieht, bleibt dennoch der Verdacht, dass bei der für den Leser nicht nachvollziehbaren Generierung des maskulinen Phantombilds, bereits Maskulinität und Attraktivität miteinander konfundiert sind.

(2) Bei der Datenauswertung bestehen dieselben methodischen Probleme wie bei Perrett et al. (1998). Johnston et al. zeigten einen 40 Sekunden langen Film, der aus 1200 Einzelbildern (*frames*) bestand, und berechneten dann jeweils über alle Versuchspersonen den durchschnittlichen *frame*. Genau wie Perrett et al. geben auch sie lediglich Mittelwerte und keine Standardabweichungen oder Standardfehler an und sie machen keine Angaben über die Verteilung der Daten. Ansonsten gelten für diese Art der Datenauswertung dieselben Kritikpunkte, wie bereits oben bei der Studie von Perrett et al. (1998) dargelegt, die deswegen an dieser Stelle nicht noch einmal wiederholt werden sollen. In ihrem Methodenteil schreiben Johnston et al. (2000, S. 257), dass sie die von den Versuchspersonen hergestellten männlichen Gesichter (darunter u. a. ein ideal attraktives Männergesicht und ein durchschnittliches Männergesicht) auch auf Ratingskalen bezüglich unterschiedlicher Eigenschaften (v. a. zugeschriebene Charaktereigenschaften) bewerten ließen. Eine dieser Eigenschaften war die Attraktivität. Leider berichten die Autoren nicht die Ergebnisse dieser Attraktivitätsratings, sondern berichten nur durchschnittliche *frames*. Die Ratings wären eine gute Möglichkeit gewesen, die Operationalisierung von Attraktivität über durchschnittliche *frame*-Nummern eines Films zu validieren. Da sie diese Ergebnisse nicht nennen, bleibt der Verdacht, dass möglicherweise nur bei den durchschnittlichen *frame*-Nummern signifikante Effekte gefunden wurden, bei den direkten Attraktivitätsratings, die eine viel größere Aussagekraft besitzen, jedoch nicht.

Aufgrund (1) der fehlenden Nachvollziehbarkeit, auf welche Weise das hypermaskuline und das hyperfeminine Gesicht mit der Phantombildsoftware erzeugt wurde, und ob nicht bereits an dieser Stelle sexueller Dimorphismus mit Attraktivität konfundiert ist, (2) aufgrund der Schwierigkeit, diese Studie ohne die Verwendung der sehr speziellen Phan-

tombild-Software *FacePrint* zu replizieren und (3) aufgrund der methodischen Probleme mit der Mittelung von *frame*-Nummern eines Films sind die von Johnston et al. gezogenen Schlüsse unter Vorbehalt zu sehen. Problematisch ist es auch, die dort dargestellten *t*-Werte (basierend auf durchschnittlichen *frame*-Nummern) in eine Effektgröße *R* umzurechnen, wie dies Rhodes (2006) in ihrer Metaanalyse tut (dies gilt auch für die beiden anderen experimentellen Studien von Perrett et al., 1998, und Rhodes et al., 2000, mit *sex-continua*, bei denen nur ein einzelnes Gesicht stufenweise mit Morphing-Software verzerrt wurde). Denn die Gefahr ist dabei groß, dass der eigentliche Effekt der Manipulation auf die Attraktivität des Gesichts stark überschätzt wird. Die von Rhodes so berechneten positiven Effektgrößen für sexuellen Dimorphismus dieser Studie von $R = 0.79$ für das Frauengesicht und $R = 0.64$ für das Männergesicht lassen sich kaum mehr sinnvoll interpretieren.

Die dritte wichtige und viel zitierte experimentelle Untersuchung zum sexuellen Dimorphismus ist die Studie von Rhodes et al. (2000), die ihrerseits als Versuch konzipiert war, die Ergebnisse von Perrett et al. (1998) zu replizieren. Auf den ersten Blick hat sie dies auch erreicht. Die Autorinnen berichten einen generell positiven Effekt femininer Merkmale, d. h. ein durchschnittliches weibliches Composite wurde durch Übertreibung femininer Merkmale (mit Hilfe der Karikatur-Technik) attraktiver, und ein männliches durchschnittliches Composite durch Annäherung an weibliche Durchschnittsproportionen ebenso. Bei einer genaueren Betrachtung weist die Studie jedoch eklatante methodische Mängel auf. So näherten Rhodes und Kolleginnen zum einen das weibliche Durchschnittsgesicht in 25%-Schritten (bis 100 %) an die männlichen Durchschnittsproportionen an (= *masculinized*) und zum anderen verzerrten sie es (durch die Karikatur-Technik) in 25%-Schritten von den männlichen Durchschnittsproportionen weg (= *female-exaggerated*). Analog verfahren sie mit dem männlichen Durchschnitts-Composite. Es resultierten daraus insgesamt jeweils neun Varianten (acht experimentell verzerrte Gesichter und eine „Kontrollbedingung“, nämlich unveränderte Durchschnittsgesichter mit 0 % Verzerrung). In einem *9-alternatives-forced-choice*-Verfahren sollten die Versuchspersonen diejenige Gesichtsvariante auswählen, die sie am attraktivsten fanden.

Für die statistische Datenanalyse werteten Rhodes et al. jedoch nicht die Häufigkeiten dieser Bevorzugungen an sich aus, sondern stattdessen Zahlen, die sie den jeweiligen Bild-Varianten zugeordnet hatten. Beispiele: Dem Frauengesicht, das zu 100 % an die männlichen Gesichtsproportionen angenähert war, wurde der Zahlenwert „-100“ zugeordnet; dem Frauengesicht, das zu 50 % von den männlichen Proportionen weg verzerrt wurde, wurde der Zahlenwert „+50“ zugeordnet. Für die Analyse wurde dann der durch-

schnittliche Prozentsatz der Verzerrung der ausgewählten Bilder berechnet („mean distortion level of most attractive image“, Rhodes et al., 2000, S. 135) und in einem Ein-Stichproben-*t*-Test auf einen signifikanten Unterschied zu 0 % getestet, weil das unveränderte Durchschnittsgesicht eine Verzerrung von 0 % hatte. Mit dieser Methode fanden sie, dass bei dem weiblichen Durchschnittsgesicht eine *positive distortion* von $M = 36\%$ (d. h. in Richtung feminin) und beim männlichen Durchschnittsgesicht eine *negative distortion* von $M = -33\%$ (d. h. weg vom maskulinen, hin in Richtung feminin) am attraktivsten bewertet wurden (p jeweils $<.001$).

Diese Art der Datenauswertung ist jedoch falsch, weil gegen eine falsche Nullhypothese getestet wird. Die korrekte Nullhypothese würde lauten, dass alle neuen Gesichtsvarianten eines Geschlechts gleich häufig wie die attraktivsten ausgewählt werden, bei neun Auswahlmöglichkeiten ist also die A-priori-Wahrscheinlichkeit $1/9$ (= 11.11 %). Eine korrekte Methode, diese Hypothese zu überprüfen, wäre ein χ^2 -Anpassungstest auf Gleichverteilung, so wie er auch bei der vorliegenden Arbeit im Experiment *Dimorph 1* (vgl. Kapitel 6.2.2) gerechnet wurde. Stattdessen ordnen jedoch Rhodes und Kolleginnen verschiedenen aussehenden Gesichtern (Nomalskalenniveau) Zahlen auf Intervallskalenniveau (!) zu und berechnen daraus einen Mittelwert.

Wie unsinnig diese Methode ist, zeigt ein Gedankenexperiment anhand des Stimulusmaterials aus dem eigenen Experiment *Dimorph 1*, bei dem ebenfalls das weibliche Durchschnittsgesicht (Gesicht F, Kapitel 6.2.2) mit Morphing-Software manipuliert wurde. Hier gab es folgende sechs Gesichtsvarianten: unverändertes Durchschnittsgesicht (= 0% Kindchenschemaanteil), sowie die Varianten mit 10 %, 20 %, 30 %, 40 % und 50 % Kindchenschemaanteil. Angenommen, es gäbe keinerlei Attraktivitätsunterschiede zwischen den sechs Bildvarianten (z. B. weil die Manipulation so schwach wäre, dass überhaupt keine Unterschiede zwischen den Varianten erkennbar wären), dann würden im Forced-Choice-Verfahren insgesamt alle Varianten gleich häufig bevorzugt, d. h. die beobachtete Verteilung läge genau bei der A-priori-Wahrscheinlichkeit von $1/6$ (16.7 %) und das Ergebnis wäre nicht signifikant. Berechnet man jedoch basierend auf dieser Gleichverteilung den durchschnittlichen Prozentwert, mit dem an das Kindchenschema angenähert wurde (*distortion*), dann kommt man auf einen mittleren Prozentwert von 25 %, d. h. $(0 + 10 + 20 + 30 + 40 + 50) : 6$. Testet man diesen Mittelwert mit einem Ein-Stichproben-*t*-Test gegen den Wert von 0 %, erhält man (selbst bei einer sehr kleinen Stichprobe) selbstverständlich ein hochsignifikantes Ergebnis. Die Schlussfolgerung, dass im Mittel (!) Gesichter mit einem 25%-Anteil an kindlichen Proportionen bevorzugt würden, wäre dennoch unsinnig, da in Wirklichkeit keine der Varianten besonders präferiert wird. Würde

man tatsächlich eine solche angenommene Gleichverteilung der Daten mit der Methode von Rhodes et al. (2000) auswerten, dann erhielte man bei einer etwa gleichen Stichprobengröße ($N = 72$, jede Variante genau 12 Mal bevorzugt) ein höchst signifikantes Ergebnis ($t(71) = 12.335$, $p < .001$).

Die Konsequenzen dieses statistischen Fehlers sind geradezu absurd: je weniger unterscheidbar die einzelnen Varianten sind und je mehr Varianten es gibt, desto leichter ist es, die (falsch formulierte) Nullhypothese zu widerlegen! Übrigens kann man auf diese Weise auch jeden beliebigen Prozentsatz an Kindchenschema-Anteil (quasi die Effektgröße) statistisch signifikant bekommen, nämlich indem man die Prozentabstufungen nicht mit 0 %, 10 %, 20 %, 30 % usw. wählt, sondern z. B. mit 0 %, 1 %, 2 %, 3 %, 4 %, 5 % und 6 %. In diesem Fall wäre der mittlere Prozentsatz an Verzerrung 2.5 % und die beobachtete Verteilung der Daten ganz sicher gleichförmig, da absolut keine Unterschiede zwischen den Varianten wahrnehmbar wären. Das Ergebnis des t -Tests wäre jedoch dasselbe und man würde dann analog den Schluss daraus ziehen, dass der ideale Anteil an Kindchenschema-Proportionen in einem Frauengesicht nicht bei 25 %, sondern bei 2.5 % liegt ($p < .001$).

Genau diese Fehler begehen jedoch Rhodes und Kolleginnen. Sie interpretieren den gemittelten *distortion level* als Effekt und rechnen einen falschen statistischen Test. Eine Verteilung der Präferenzen für die einzelnen Bildvarianten liefern sie dem Leser nicht mit. Nun könnte man zu deren Verteidigung einwenden, dass die Bildmanipulationen ja in zwei Richtungen durchgeführt wurden, nämlich einerseits hin zu den Proportionen des anderen Geschlechts und andererseits die Überzeichnung der geschlechtstypischen Merkmale. Wenn alle Gesichtsvarianten die gleiche Wahrscheinlichkeit hätten, ausgewählt zu werden, dann sollte sich im Durchschnitt eine *distortion* von 0 % (= Durchschnittsgesicht) ergeben. Eine mittlere Bevorzugung von z. B. +36 % würde dann bedeuten, dass (im Mittel) die geschlechtstypischen Proportionen attraktiver sind. Das Problem ist jedoch, dass die einzelnen Gesichter keine gleich große Wahrscheinlichkeit haben, ausgewählt zu werden. Davon kann man sich leicht überzeugen, indem man einen Blick auf das abgedruckte Stimulusmaterial des Artikels wirft (Rhodes et al., 2000, S. 133f). Bei den Verzerrungen auf ein Schema *hin* sind nur äußerst geringe Unterschiede wahrnehmbar, dagegen bei den Verzerrungen von einem Schema *weg* (Karikatur-Technik) sind die Unterschiede deutlicher sichtbar. Hier hat man es offensichtlich wieder mit einem Artefakt der Morphing-Technik zu tun und dem Problem, dass die Karikatur-Methode zu keinen validen Resultaten führt, was bereits Meyer und Quong (1999) als Reaktion auf das ähnliche Experiment von Perrett et al. (1999) kritisiert hatten.

Voraussetzung, um überhaupt einen Mittelwert über die verschiedenen Prozentgrade an *distortion* berechnen zu dürfen, wäre Intervallskalenniveau, also gleiche Abstände zwischen den Veränderungen in den Bildvarianten. Diese Abstände sind jedoch augenscheinlich *nicht* gleich groß, sondern bei den Bildvarianten, die durch die Karikaturtechnik entstanden sind, größer. Dieser Einwand ließe sich übrigens auch gegen Johnston et al., 2001, anführen, die einen 40 Sekunden langen Film aus fünf einzelnen Filmsequenzen montierten, bei dem jedoch offen bleibt, ob mit fortschreitender Zeit, also Anzahl der *frames* pro Sekunde, die Veränderung des Gesichts von hypermaskulin nach hyperfeminin als linear empfunden wird, oder ob es im Film Phasen gibt, in denen sich das Gesicht wenig verändert und dann wiederum Phasen, in denen sich die Veränderung schnell vollzieht; in diesem Fall wäre auch kein Intervallskalenniveau gegeben.

Ironischerweise liefern Rhodes et al. (2000) jedoch gleich selbst den empirischen Beweis dafür mit, dass (zumindest bei Frauengesichtern) die Ergebnisse ihres Experiments (bei ihnen das Experiment 2) auf einem methodischen Artefakt beruhen. Denn in ihrem Experiment 1 karikierten sie das weibliche Durchschnittsgesicht mit einer Gewichtung von 50 % in Richtung Femininität und testeten die beiden Varianten im Paarvergleichsexperiment mit 128 Versuchspersonen, die die attraktivere Variante auswählen sollten. Trotz der großen Versuchspersonenanzahl ergab sich im Binomialtest kein signifikanter Unterschied. Ein Blick auf das abgedruckte Stimulusmaterial (Figure 1B, S. 130) macht deutlich, warum. Die beiden Gesichter sehen praktisch gleich aus. Während Rhodes et al. also aus ihrem zweiten Experiment den Schluss ziehen, dass eine mittlere *distortion* von 36 % in Richtung Femininität am attraktivsten ist ($p < .001$), ist bei der 50%-Karikatur (trotz stärkerer Veränderung) praktisch kein Unterschied zum Original sichtbar und das Ergebnis nicht signifikant. Im ersten Experiment werteten die Autorinnen also statistisch korrekt eine Häufigkeitsverteilung auf Basis eines Präferenzurteils aus, im zweiten Experiment hingegen berechneten sie einen unsinnigen Mittelwert aus technischen Werten, die sich auf einen Algorithmus einer Grafiksoftware beziehen, unterstellten fälschlicherweise ein Intervallskalenniveau und testeten gegen eine falsch formulierte Nullhypothese. Aus diesem Grund sind die Ergebnisse ihres zweiten Experiments praktisch wertlos.

Die Ergebnisse ihres ersten Experiments mit den 50%-Karikaturen und dem Paarvergleichsexperiment sind zwar methodisch korrekt, stehen jedoch unter dem generellen Vorbehalt gegen die Manipulation von Stimuli mit der Karikaturmethode, da der Algorithmus der geometrischen Interpolation einer Grafik-Software, bei dem ein zweidimensionales Foto verzerrt wird, die komplexen, nichtlinearen testosteronbedingten Wachstumsprozesse eines dreidimensionalen Schädels nicht adäquat simuliert (Enlow, 1966, 1979;

Grumbach, 2000; Johnston et al., 2001; Meyer & Quong, 1999; Rennels et al., 2008; Tanner, 1978) und daher auch sexuell dimorphe Gesichtsmerkmale nicht realistisch und valide verändern kann.

Betrachtet man die beschriebenen drei experimentellen Untersuchungen zum sexuellen Dimorphismus in der Zusammenschau, werden einige Unterschiede zu den hier durchgeführten Experimenten deutlich. Sie lassen sich in fünf Aspekte einteilen, die im Folgenden erläutert werden.

1. Generalisierbarkeit der Ergebnisse vom verwendeten Stimulusmaterial auf normale Gesichter
2. Mögliche Konfundierungen zwischen sexuellem Dimorphismus und Attraktivität beim verwendeten Stimulusmaterial
3. Verwendete statistische Methoden bei der Datenanalyse
4. Stärke (= Quantität) der Manipulation des sexuellen Dimorphismus
5. Art (= Qualität) der Manipulation des sexuellen Dimorphismus

(1) Generalisierbarkeit der Ergebnisse vom verwendeten Stimulusmaterial auf normale Gesichter: Alle drei früheren Untersuchungen arbeiteten mit *sex continua* eines einzigen manipulierten Gesichts pro Geschlecht. Bei Perrett et al. (1998) und Johnston et al. (2001) wurden Filmsequenzen gezeigt, bei Rhodes (1998) eine Bilderserie aus neun Gesichtsvarianten. Grundlage war jedoch jeweils ein Durchschnittsgesicht, das je nach Experiment aus 16 bis 30 Originalgesichtern berechnet war. Diese Composites sind mit den typischen Artefakten behaftet, die die Verwendung von Morphing-Software mit sich bringt. Das heißt, sie besitzen eine glatte, makellose Haut, unharmonische Gesichtsproportionen oder Asymmetrien fehlen vollständig, und die Textur der Haare geht verloren (v. a. bei Rhodes et al., wo sogar aufgrund fehlender Sorgfalt beim Morphen die Kontur der Frisur verschwimmt). Bei Perrett und Kollegen waren die Artefakte sogar so groß, dass sie es offensichtlich für das kleinere Übel hielten, mit gemorphten Bildern von Gesichtern zu experimentieren, bei denen sie zuvor Frisur, Hals und Ohren abgeschnitten hatten. Bei einem solchen Stimulusmaterial ist es fraglich, inwieweit gewonnene Ergebnisse eigentlich für die Grundgesamtheit der Gesichter repräsentativ sind, auf die Autoren ihre Ergebnisse generalisieren möchten, nämlich auf normale, natürliche Gesichter.

Bei den eigenen hier durchgeführten Experimenten hingegen wurde die Manipulation durch Morphing bei vielen unterschiedlichen Gesichtern durchgeführt und die Ergebnisse zeigen, dass es keineswegs egal ist, mit welchen Ausgangsgesichtern experimentiert

wird. Eines dieser Gesichter war auch das weibliche Durchschnittsgesicht (Untersuchung *Dimorph 1*). Gleichzeitig zeigen die Ergebnisse der anderen fünf manipulierten Stimulusgesichter aus derselben Untersuchung aber auch, dass das Ergebnis des Durchschnittsgesichts keineswegs auf alle anderen Stimulusgesichter übertragbar ist. Denn es wurde ebenfalls dargestellt, dass das Treatment mit dem Ausgangsgesicht interagiert, d. h. dieselbe Manipulation bei unterschiedlichen Gesichtern zu unterschiedlichen Effekten führt. Dies erkennt man auch an den unterschiedlichen Effektgrößen in Abhängigkeit von der Ausgangsattraktivität der Gesichter, sowohl bei Männer- als auch Frauengesichtern (vgl. Kapitel 6.3.2 und 6.4.2).

Hinzu kommt, dass bei den eigenen Untersuchungen alle Gesichter-Stimuli natürlich und realistisch aussahen – sowohl vor als auch nach der Manipulation und unabhängig davon, ob Originalgesichter verwendet wurden oder gemorphte Gesichter (alle gemorphten Gesichter waren *2-face-* oder *4-face-morphs*; eine Ausnahme bildet selbstverständlich das erwähnte weibliche Durchschnittsgesicht). Ähnliches lässt sich nur über das Experiment von Johnston et al. (2001) sagen; das Stimulusmaterial von Rhodes et al. (2000) und Perrett et al. (1998) leidet hingegen an starken Morphing-Artefakten bzw. unrealistisch wirkenden „beschnittenen“ Gesichtern.

(2) Mögliche Konfundierungen zwischen sexuellem Dimorphismus und Attraktivität beim verwendeten Stimulusmaterial: Bei der Studie von Johnston et al. (2001) besteht der naheliegende Verdacht, dass bereits bei der Erzeugung der geschlechtstypischen Versionen des männlichen bzw. weiblichen Gesichts die manipulierte unabhängige Variable (= sexueller Dimorphismus) mit der zu untersuchenden abhängigen Variablen (= Attraktivität) konfundiert sind. Der Grund dafür ist, wie oben ausführlich dargelegt, dass aus dem Artikel nicht hervorgeht, auf welche Weise die geschlechtstypisch aussehenden Varianten mit der Phantombild-Software erzeugt wurden. Wenn Versuchspersonen instruiert werden, ein typisch männlich aussehendes Männergesicht oder ein typisch weiblich aussehendes Frauengesicht zu erzeugen, ist damit zu rechnen, dass sie gleichzeitig auch ein attraktives Gesicht herstellen, da es in der *Meinung* der Versuchspersonen (möglicherweise als Vorurteil, v. a. bei Männergesichtern) einen Zusammenhang zwischen geschlechtstypischem Aussehen und Attraktivität gibt. Dies zeigen diejenigen korrelativen Untersuchungen, bei denen sexueller Dimorphismus durch Versuchspersonen-Ratings operationalisiert wurde. (Bruce et al., 1994; Cunningham et al., 1990; Dunkle & Francis, 1990; Koehler et al., 2004; Little & Hancock, 2002; Neave et al., 2003; O'Toole et al., 1998; Rennels et al., 2008; Rhodes et al., 2003).

Ein besonders hervorzuhebendes Experiment ist in diesem Zusammenhang die Studie von Rennels, Bronstad und Langlois (2008). Hier wurde maskulines Aussehen eines Durchschnittsmännergesichts auf zwei Weisen manipuliert: Erstens durch die (umstrittene) Karikatur-Technik und zweitens, indem eine Stichprobe von Original-Männern Gesichtern auf einer bipolaren Ratingskala (*very feminine* – *very masculine*) bewertet wurde. Anschließend wurden aus den jeweils 32 maskulinsten und 32 femininsten Gesichtern Composites erzeugt, die jeweils in Paarvergleichen mit dem Durchschnittsmännern Gesicht hinsichtlich Attraktivität getestet wurden. Das Resultat: Wurde die Karikatur-Technik zur Stimulusgenerierung verwendet, dann schnitten die maskulinen Gesichtsvarianten schlechter und die femininisierten besser ab (d. h. konsistent mit anderen Studien, die diese Methode eingesetzt haben). Basierten die Durchschnittsgesichter hingegen auf Originalgesichtern, die als besonders maskulin bzw. besonders feminin beurteilt worden waren, kehrte sich das Ergebnis um und die maskulinen Männer-Varianten schnitten am besten und die femininen Männervarianten am schlechtesten ab (Rennels et al. 2008, S. 889).

Eine plausible Erklärung für diesen Effekt ist, dass die Versuchspersonen beim Rating der Maskulinität von der Attraktivität der Männern Gesichter beeinflusst worden waren. Attraktivere Männern Gesichter bekamen höhere Maskulinitäts-Ratings (= Halo-Effekt der Attraktivität). Tatsächlich berichten die Autoren auch in ihrem Methodenteil, dass – wie nicht anders zu erwarten – Maskulinitäts-Ratings und Attraktivitäts-Ratings der Männern Gesichter hochsignifikant mit $r = .36$ korreliert waren. Aufgrund dieser Korrelation ist es auch nicht überraschend, dass ein Composite aus solchen Männern Gesichtern, die als die maskulinsten bezeichnet (!) werden (was nicht heißt, dass sie objektiv auch diese Gesichtsmerkmale haben), auch attraktiver bewertet wird als ein Durchschnittsmännern Gesicht oder ein Composite, das aus Original-Männern Gesichtern erstellt wurde, die alle als besonders feminin bezeichnet wurden. Wenn stattdessen Maskulinität zuvor über die Vermessung von *Facial Landmarks* quantifiziert und dann aus *diesen* maskulinsten Gesichtern ein Composite berechnet worden wäre, dann wären die Ergebnisse überzeugender und frei vom Verdacht möglicher Konfundierungen.

Bei der Studie von Johnston et al. (2001) muss daher davon ausgegangen werden, dass bei den geschlechtstypischen Varianten aus der Phantombild-Software derselbe Effekt eingetreten ist wie beim Composite von Rennels et al., das auf der Grundlage von Maskulinitäts-Ratings erstellt wurde, und dass damit unabhängige und abhängige Variable miteinander konfundiert sind. Bei den eigenen Experimenten hingegen kann eine solche Konfundierung ausgeschlossen werden, da hier mit dem Kindchenschema experi-

mentiert wurde. Das Schema selbst basiert auf einem Durchschnittsgesicht von Kindergartenkindern. Es wurden vier Kinder fotografiert und aus allen vier das Durchschnittsgesicht erzeugt. Die Auswahl der Kinder erfolgte ohne Einfluss des Versuchsleiters, sondern durch die Erzieherinnen des kooperierenden Kindergartens.

(3) Verwendete statistische Methoden bei der Datenanalyse: Bei den Experimenten von Perrett et al. (1998) und Johnston et al. (2001) wurden die mittleren *frame*-Nummern einer Filmsequenz ausgewertet, indem aus diesen Nummern Mittelwerte berechnet wurden und diese dann in einem Ein-Stichproben-*t*-Test statistisch gegen die *frame*-Nummer des Durchschnittsgesichts getestet wurden. Problematisch ist hier v. a.

- dass Mittelwerte berechnet werden, ohne auch nur ein Wort über Standardabweichungen zu verlieren,
- dass keine Angaben über die Verteilungsform der Daten gemacht werden (Normalverteilung oder Gleichverteilung),
- dass mit den *frame*-Nummern (oder Prozent-Verzerrungs-Werten bei Perrett et al., 1998) Intervallskalenniveau unterstellt wird, obwohl die sichtbaren Unterschiede zwischen den Gesichtervarianten *nicht* gleich sind, da *Warping* auf ein Schema *hin* zu geringeren Verzerrungen führt als *Warping* von einem Schema *weg* (= Karikatur),
- dass aufgrund einer geringfügigen allgemeinen Tendenz in eine bestimmte Richtung oder aufgrund weniger Ausreißer eine solche Bildvariante (*frame*) resultieren kann, die zwar basierend auf gemittelten *frame*-Nummern signifikant verschieden von der *frame*-Nummer des Durchschnittsgesichts ist, die jedoch optisch dem Durchschnittsgesicht so ähnlich ist, dass sich in einem direkten Paarvergleich (ganz zu schweigen von einem Attraktivitätsrating) kein Attraktivitätsunterschied nachweisen ließe.

Die statistische Datenanalyse von Rhodes et al. (2000) in ihrem Experiment 2 mit dem *9-alternatives-forced-choice*-Verfahren ist hingegen schlicht unsinnig und falsch, wie oben ausführlich dargelegt wurde. Korrekt wäre hier ein χ^2 -Anpassungstest auf Gleichverteilung gewesen, wie er hier für das Experiment *Dimorph 1* gerechnet wurde. Für die anderen beiden selbst durchgeführten Experimente wurden für die Paarvergleiche Binomialtests verwendet. Der Vorteil der Paarvergleiche („Welches Gesicht ist attraktiver, A oder B?“) ist, dass das Ergebnis klar und eindeutig interpretierbar ist, im Gegensatz zu Mittelwerten aus *frame*-Nummern eines Films, bei denen die Gefahr von Methodenartefakten groß ist.

(4) Stärke (= Quantität) der Manipulation des sexuellen Dimorphismus: In der Studie von Johnston et al. (2001) fällt positiv auf, dass beim Stimulusmaterial die Bandbreite zwischen dem hypermaskulinen und dem hyperfemininen Gesicht sehr groß ist. Auch wenn es bei den geschlechtstypischen Versionen Konfundierungen mit Attraktivität geben mag, so sind zumindest die geschlechtstypischen Eigenarten der Gesichter deutlich erkennbar. Anders hingegen bei Perrett et al. (1998) und Rhodes et al. (2000): Wie man sich selbst am abgedruckten Stimulusmaterial der Artikel überzeugen kann, sind die sichtbaren Unterschiede zwischen einem Durchschnittsgesicht und dessen beiden Variationen Richtung maskulin bzw. feminin marginal (– umso komischer wirkt es, wenn Rhodes et al. die Gesichter als „*supermale*“ und „*superfemale*“ bezeichnen). Die Reize decken nur einen kleinen Teil der Bandbreite des Stimulusmaterials von Johnston et al. ab. Angesichts der kaum wahrnehmbaren Unterschiede erstaunt es auch nicht, dass die Autoren zu statistischen Tricks bei der Datenauswertung greifen, um Effekte statistisch signifikant zu bekommen.

Bei den hier durchgeführten eigenen Experimenten sind die sichtbaren Unterschiede im Stimulusmaterial hingegen größer, da die Ausgangsgesichter an die Durchschnittsproportionen eines kleinen Kindes angenähert wurden. Die Stärke der Annäherung war 40 % bzw. bis maximal 50 % (beim Experiment *Dimorph 1*), also vergleichbar mit den maximalen 50%-Verzerrungen bei Perrett et al. (1998). Da aber kindliche Gesichtsproportionen beispielsweise viel weiter als durchschnittliche weibliche Gesichtsproportionen von einem durchschnittlichen Männergesicht entfernt sind, ist der verzerrende Effekt auf das Ausgangsgesicht viel größer, auch weit größer als die 100%-Verzerrungen bei Rhodes et al. (2000), da sich der Wert von 100 % ja auf die geringen Distanzen der *Facial Landmarks* zwischen männlichem und weiblichem Durchschnittsgesicht bezieht. Obwohl die optischen Unterschiede bei den eigenen Experimenten deutlicher waren, waren dennoch die erzielten Effektgrößen eher gering. Der Median der nach r standardisierten Effektgrößen liegt bei Frauengesichtern bei $r = .34$ und bei Männergesichtern bei $r = .22$ (mit einigen nicht signifikanten Ergebnissen bei attraktiven Männergesichtern). Umso auffallender sind trotz des schwachen Treatments die hochsignifikanten Unterschiede bei Perrett et al. (1998) und Rhodes et al. (2000) – Effektgrößen geben sie nicht an.

(5) Art (= Qualität) der Manipulation des sexuellen Dimorphismus: Ein letzter wichtiger Unterschied zwischen den hier durchgeführten Experimenten und den anderen diskutierten früheren Untersuchungen ist, dass der sexuelle Dimorphismus nicht durch die Reduzierung bzw. Übertreibung der Unterschiede zwischen einem Durchschnittsfrauengesicht und einem Durchschnittsmännergesicht hergestellt wurde, sondern durch Verwendung

des Kindchenschemas. Der Grund für diese Entscheidung war, dass die Karikaturmethode vielfach kritisiert wurde, da der Algorithmus einer Morphing-Software die komplexen, nichtlinearen testosteronbedingten Wachstumsprozesse eines dreidimensionalen Schädels nicht adäquat simuliert (Enlow, 1966, 1979; Grumbach, 2000; Johnston et al., 2001; Meyer & Quong, 1999; Rennels et al., 2008; Tanner, 1978) und daher auch sexuell dimorphe Gesichtsmerkmale nicht realistisch und valide verändern kann. Als Konsequenz musste hier auch auf eine Überprüfung des Effekts von maskulinen Merkmalen bei Männergesichtern verzichtet werden. Stattdessen wurden Männergesichter lediglich durch Annäherung an das Kindchenschema femininisiert. Die Verwendung des Kindchenschemas und das erzielte Resultat, dass offensichtlich nicht der sexuelle Dimorphismus selbst, sondern vielmehr neotene Gesichtszüge generell attraktiv zu sein scheinen, sind so wichtige Aspekte, dass sie im folgenden Kapitel 7.4 vertieft diskutiert werden.

7.4 Methodische Reflexion zur Verwendung des Kindchenschemas

Die Theorie des sexuellen Dimorphismus postuliert, dass attraktiv das ist, was geschlechtstypisch aussieht. Sie macht also Aussagen über *beide* Geschlechter und beruht damit auf zwei Säulen. Die Vorhersagen über weibliche Attraktivität konnten mit den beiden hier durchgeführten Experimenten bestätigt werden, die Vorhersagen über männliche Attraktivität jedoch keineswegs. Vielmehr deuten die Daten auf das Gegenteil hin, nämlich dass es zumindest einige typisch männliche Merkmale geben muss, die die Attraktivität eines Gesichts senken. Die Ergebnisse mit den manipulierten Männergesichtern passen somit zur heterogenen Befundlage des Standes der Forschung.

Der Umstand, dass geschlechtstypische Merkmale empirisch eindeutig nachweisbar nur *eines* von beiden Geschlechtern attraktiv machen, nämlich nur Frauen, nicht jedoch Männer, bringt die Theorie des sexuellen Dimorphismus ins Wanken. Denn es lässt sich dadurch nicht mehr länger argumentieren, dass generell *geschlechtstypische* Merkmale attraktiv machen, sondern lediglich typisch *feminine* Merkmale – und zwar bei Frauen klar und bei Männern unter bestimmten Umständen ein bisschen.

In den hier durchgeführten drei Experimenten wurde die Femininität durch Anwendung des Kindchenschemas manipuliert. Dies ist aus theoretischer Sicht ein großer Vorteil, weil damit auf ein biologisch gut fundiertes Konzept zurückgegriffen wird, das nicht von vornherein mit einzelnen Merkmalen konfundiert ist, die (kulturell) als attraktiv definiert sind (beispielsweise gezupfte, schmale Augenbrauen oder lippenstiftrote Lippen als Attribute

von Weiblichkeit). Deswegen lautet die entscheidende Frage: Warum macht das Kindchenschema Gesichter (v. a. Frauen, aber auch manche Männer) attraktiv?

Daran anschließend ist die Frage zu diskutieren, ob tatsächlich *alle* Merkmale des Kindchenschemas attraktivitätssteigernd wirken, oder nur *manche*. Bei einer so komplexen Merkmalskonfiguration könne es schließlich leicht sein, dass nur einige wenige Merkmale kausal für eine erhöhte Attraktivität verantwortlich sind, während andere typisch kindliche Merkmale gar keine Rolle spielen und möglicherweise sogar andere eher attraktivitäts-senkend wirken. Was durch die drei Experimente untersucht wurde, ist lediglich der Gesamt- oder Nettoeffekt. Auch wenn sich die folgenden Überlegungen nicht anhand der erhobenen Daten beweisen lassen, ist es dennoch angebracht, die Wirkung des Kindchenschemas kritisch zu hinterfragen, um mit der Akzeptanz des Kindchenschemas als ganzheitliche komplexe Merkmalskonstellation nicht automatisch und pauschal auch jedes Einzelmerkmal dieses Schemas als Attraktivitätskriterium zu postulieren.

7.4.1 Attraktivitätserhöhende Merkmale des Kindchenschemas

Die Ursachen für die attraktivitätssteigernde Wirkung des Kindchenschemas bei Frauengesichtern und mit Abstrichen auch bei Männergesichtern lassen sich auf drei mögliche Erklärungen zurückführen:

1. Das Kindchenschema als verkappte Durchschnittshypothese
2. Der verjüngende Effekt des Kindchenschemas
3. Der Effekt des Kindchenschemas auf eine positivere soziale Stereotypisierung

(1) Das Kindchenschema als verkappte Durchschnittshypothese: Diese Sichtweise zielt sehr nüchtern auf die verwendete Methodik ab. Bei den hier durchgeführten Experimenten wurden die erwachsenen Stimulusgesichter an ein Schema angenähert, das selbst ein Durchschnittsgesicht ist. Zwar wurden lediglich *vier* Kindergesichter zu einem Durchschnittskindergesicht verrechnet, doch die Ergebnisse der Untersuchungen zur Durchschnittshypothese (Kapitel 2, vgl. Streudiagramme von Abbildung 3, S. 18, Abbildung 8, S. 26, Abbildung 12, S. 32, Abbildung 19, S. 46, Abbildung 24, S. 55 und Abbildung 28, S. 61) zeigen, dass auch schon ein Composite aus nur vier Originalgesichtern zu einem kleinen Attraktivitätsgewinn führen kann. Da dieses Kinderdurchschnittsgesicht jedoch nur als Schema verwendet wurde, d. h. die Erwachsenengesichter durch das *Warping*-Verfahren lediglich in ihren Proportionen (ohne die Veränderung von Farbwerten und ohne Weichzeichnungseffekte) an dieses Durchschnittsgesicht angenähert wurden, ist dieser Effekt als sehr gering einzustufen. Am ehesten geben noch die Experimente zur

Durchschnittshypothese mit den Schattenprofilen einen Eindruck von einem möglichen Attraktivitätsgewinn, der allein durch die Durchschnittsbildung aus vier Originalen resultieren kann. Die Ergebnisse der Schattenprofile (Kapitel 2.5.2 und 2.9.2) sprechen jedoch eher dafür, dass der Effekt der Durchschnittsbildung durch den Morph-Prozess zu vernachlässigen ist.

Allerdings gibt es bei Kindergesichtern noch einen zweiten Aspekt. Dadurch, dass Kindergesichtern der starke Wachstumsschub in der Pubertät noch bevorsteht, gibt es bei ihnen weniger unharmonische Abweichungen von einem normalen (Kinder)gesicht. Anders ausgedrückt: Unter Kindergesichtern gibt es eine geringere Varianz als unter Erwachsenengesichtern. Wenn die Verschiedenheit der Erwachsenengesichter größer ist, steigt damit auch die Wahrscheinlichkeit, dass einzelne Gesichtsmerkmale deutlich von der Norm eines Erwachsenengesichts abweichen (im Englischen als *distinctive* bezeichnet), z. B. in Form einer zu großen Nase, einem zu langen Kinn oder einem zu schmalen Gesicht. Weichen einzelne Merkmale deutlich von der Norm ab, führt dies mit großer Wahrscheinlichkeit zwar zu einem charakteristischeren, aber gleichzeitig auch zu einem unharmonischeren Aussehen und damit zu geringerer Attraktivität. Dass erwachsene Gesichter in vielen Fällen attraktiver wirken, wenn ihre Proportionen an kindliche Proportionen angenähert werden, könnte also auch etwas damit zu tun haben, dass dadurch eventuelle Disharmonien in einem Gesicht gemildert werden. In diesem Fall läge der positive Effekt der Bildmanipulation nicht an den kindlichen Merkmalen an sich, sondern daran, dass das verwendete Schema weniger unharmonische Proportionen aufweist.

Als Indiz für diese Sichtweise kann man den Umstand heranziehen, dass in den Untersuchungen 2 und 3 bei beiden Geschlechtern der positive Effekt des Kindchemas bei den unattraktiven Gesichtern größer war als bei den attraktiven (unattraktive Frauen: $r = 0.48$, attraktive Frauen: $r = 0.27$; unattraktive Männer: $r = 0.43$, attraktive Männer $r = 0.07$). Auch dass die unattraktiven Männer fast in gleicher Weise von der Manipulation profitierten wie die unattraktiven Frauen, die attraktiven Männer (unter dem Strich) jedoch praktisch gar nicht, kann man als Indiz werten. Die Annäherung an das Schema fügt also nicht etwas Attraktives hinzu, sondern nimmt etwas Unattraktives weg. Da die attraktiven Männer weniger unharmonische Gesichtsproportionen besitzen, profitieren sie kaum von der Bildmanipulation.

Ob es tatsächlich weniger die typisch kindlichen Merkmale, sondern eher der ausgleichende Effekt eines Schemas an sich ist, lässt sich aus den drei hier durchgeführten Experimenten nicht beantworten. Zu testen wäre diese Hypothese, indem man den Effekt

mehrerer Schemata miteinander vergleicht. Dies wurde so in den Experimenten in Kapitel 8.1 und 8.2 durchgeführt, bei denen der Effekt des Kindchenschemas mit dem eines Durchschnittsgesichts sowie einem weiteren Schema als Kontrollbedingung im direkten Vergleich getestet wurde.

(2) Der verjüngende Effekt des Kindchenschemas: An das Kindchenschema angenäherte Gesichter könnten deswegen attraktiver wirken, weil sie jünger aussehen. Obwohl bei den Bildmanipulationen lediglich Gesichtsproportionen verändert wurden und dadurch die Haut der Stimuluspersonen, die den größten Informationsgehalt über das biologische Alter besitzt, unverändert blieb, wirken diese „verkindlichten“ Gesichter dennoch jünger. Dies gilt v. a. für die Frauengesichter, die (abhängig vom Ausgangsgesicht) teilweise aussehen wie 14- oder 15jährige Teenager. Bei den Männergesichtern ist dagegen deutlich Bartwuchs zu erkennen (der ja durch die bloße Veränderung der Gesichtsproportionen gleich bleibt), so dass die Gesichter dadurch älter als die weiblichen Stimuli wirken.

Wenn also junges Aussehen als attraktiv empfunden wird, dann wäre dies grundsätzlich eine Erklärung für die gesteigerte Attraktivität der an das Kindchenschema angenäherten Gesichter. Fraglich ist jedoch, ob der Effekt der Bildmanipulation stark genug ist, denn die Erwachsenengesichter sahen ja keineswegs alt aus, waren alle jünger als 30 Jahre und zeigten keine Alterserscheinungen wie Falten, graue Haare oder erschlafftes Bindegewebe. Andererseits sind jedoch auch die im Paarvergleich gemessenen Effekte ebenfalls nicht überwältigend groß.

Für das jüngere Aussehen als Ursache scheinen auf den ersten Blick die Ergebnisse der Frauengesichter zu passen. Denn bei Frauen ist Attraktivität in stärkerem Maße an eine jugendliche Erscheinung gekoppelt (*double standard of aging*) wie zahlreiche Untersuchungen zeigen (Alley, 1988; Deutsch, Zalski & Clark, 1986; England & McClintock, 2009; Falser, Bendel, Voller & Wegner, 2008; Giesen, 1989; Henss, 1992; Jones, 1995; Teuscher & Teuscher, 2007; Wilcox, 1997). Die Erklärung dieses Phänomens fügt sich gut in die biologische Sichtweise von Attraktivität als Signal für irgendwelche begehrten Eigenschaften des Partners im Kontext von Partnerwahl und Fortpflanzung. Henss (1992) fasst diese Sichtweise gut zusammen:

Erstens ist die Zahl der Nachkommen, die eine Frau haben kann, relativ klein, und nach oben ziemlich eng umschrieben, da Frauen bekanntlich ab einem bestimmten Alter nicht mehr fortpflanzungsfähig sind. Daher waren im Verlauf der Evolution Männer, die sich mit älteren Frauen paarten, gegenüber Männern mit jüngeren Partnerinnen (im Hinblick auf den Fortpflanzungserfolg!) zwangsläufig im Nachteil, oder andersrum gesprochen: Män-

ner, die sich mit jüngeren Frauen paarten, hatten im evolutionären Wettlauf die Nase vorn. (Henss, 1992, S. 218)

Dies mag zwar eine Erklärung für die unterschiedliche Bedeutung von Jugendlichkeit für Männer und Frauen haben, erklärt jedoch noch nicht, warum im Experiment bei den Frauengesichtern die Varianten bevorzugt wurden, die wie Teenager aussahen, obwohl die erwachsenen Varianten ebenfalls jung aussahen (eher wie Anfang oder Mitte Zwanzig). Hier ist eine Differenzierung des Konzepts Partnerwert (*mate value*) nützlich, die zwischen Fruchtbarkeit (*fertility*) und reproduktivem Wert (*reproductive value*) unterscheidet (Grammer, 2000; Symons, 1979; Symons, 1987; Williams, 1975)

Der *reproduktive Wert* einer Person bezieht sich auf ihre zu erwartende künftige Reproduktion. Die *Fruchtbarkeit* bezieht sich auf die Wahrscheinlichkeit, mit der die betreffende Person unmittelbar zur Fortpflanzung beitragen kann. Ein Beispiel macht den Unterschied zwischen diesen beiden Konzepten klar: Im Gegensatz zu einer 12-jährigen könnte eine 30-jährige unmittelbar schwanger werden und ein Kind bekommen; die Fertilität einer 30-jährigen ist daher in der Regel höher als die Fertilität einer 12-jährigen. Umgekehrt kann zu einem gegebenen Zeitpunkt eine 12-Jährige potentiell mehr Nachkommen haben als eine 30-Jährige; das heißt, der reproduktive Wert einer 12-Jährigen ist im allgemeinen höher als der einer 30-Jährigen. (Henss, 1992, S. 218f)

Inbesondere in Kulturen, in denen längerfristige (oder lebenslange) Partnerschaften die Norm waren, war es daher nach dieser Sichtweise für Männer eine evolutionär erfolgreichere Strategie, möglichst junge Partnerinnen zu bevorzugen. Durch diesen Zusammenhang wurde in der Evolution des Menschen durch natürliche Selektion bei Männern eine Präferenz für sehr junge Frauen begünstigt. Kulturvergleichende Untersuchungen wie beispielsweise die Studie von Buss (1989) mit 37 verschiedenen Kulturen zeigen, dass in allen untersuchten Kulturen Männer Partnerinnen bevorzugen, die deutlich jünger sind als sie selbst (Buss, 1989; Cunningham et al., 1995), was von evolutionär orientierten Attraktivitätsforschern als Bestätigung für den Einfluss der Gene auf Partnerwahl- und Attraktivitätspräferenzen gewertet wird.

Die evolutionspsychologische Sichtweise liefert also prinzipiell durchaus eine Interpretation dafür, warum Männer Frauengesichter bevorzugen, die aussehen wie Teenager, ohne ihnen deswegen gleich pädophile Motive zu unterstellen oder zu behaupten, Männer bevorzugten sehr junge Partnerinnen, weil diese leichter zu unterdrücken und beherrschen seien, wie dies klassisch feministische Autorinnen tun (Wolf, 1991).

Die evolutionspsychologische Interpretation hat allerdings zwei Haken: Erstens kann sie nicht erklären, warum auch männliche und nicht nur weibliche Gesichter in der „verkind-

lichten“ Variante insgesamt etwas positiver bewertet werden. Denn die Argumentation hinsichtlich Fruchtbarkeit und reproduktivem Wert gilt nur für Frauen und nicht für Männer. Und zweitens sollten sich seitens der Beurteiler Geschlechtsunterschiede nachweisen lassen, wenn das Attraktivitätsurteil tatsächlich aus Partnerwahlpräferenzen resultiert. Das ist jedoch nicht der Fall. Nicht nur Männer finden kindlich aussehende Frauen attraktiver, sondern ebenso Frauen – die sogar tendenziell noch etwas mehr (vgl. Kapitel 6.3.2). Verwendet man für Attraktivitätsuntersuchungen grundsätzlich nur Versuchspersonen mit dem entgegengesetzten Geschlecht der Stimuluspersonen und findet dabei einen Effekt, neigen Autoren (Cunningham et al., 1990; Johnston et al., 2001; Little et al., 2001; Neave et al., 2003; Scheib et al., 1999; Swaddle & Reiersen, 2002) gerne dazu, die Ursache auf biologische Gründe zurückzuführen, da sie bereits mit der Grundüberzeugung ihre Untersuchung konzipiert haben, dass Attraktivität selbstverständlich etwas mit Partnerwahl zu tun habe. Doch dies bleibt trotzdem immer eine Vorannahme über die zugrunde liegenden Mechanismen, die nicht dadurch zum Ergebnis werden, weil man irgendeinen signifikanten Effekt (z. B. hinsichtlich sexuell dimorpher Merkmale) nachweisen konnte.

Umgekehrt jedoch erscheint die Interpretation von Attraktivität als Partnerwahlkriterium in einem Fortpflanzungskontext sofort weniger überzeugend, wenn man *beide* Geschlechter als Versuchspersonen untersucht und feststellt, dass beide dieselben Präferenzen haben. Dasselbe gilt im Übrigen auch für alle Attraktivitätsstudien, in denen Kinder oder Säuglinge als Versuchspersonen verwendet wurden (de-Haan, Johnson, Maurer & Perrett, 2001; Hoss & Langlois, 2003; Langlois, Ritter, Roggman & Vaughn, 1991; Langlois et al., 1987; Langlois, Roggman & Rieser-Danner, 1990; Ramsey-Rennels & Langlois, 2006; Ramsey, Langlois, Hoss, Rubenstein & Griffin, 2004; Rhodes, Geddes, Jeffery, Dziurawiec & Clark, 2002; Rubenstein et al., 1999; Samuels et al., 1994; Samuels & Ewy, 1985; Slater et al., 1998). Haben kleine Kinder bei der Beurteilung von Gesichtern denselben Geschmack wie Erwachsene, dann erscheint die Interpretation, dass die kindlichen Attraktivitätsurteile irgendetwas mit Partnerwahl zu tun haben, doch recht weit hergeholt.

(3) Der Effekt des Kindchenschemas auf eine positivere soziale Stereotypisierung: Die dritte mögliche Interpretation der gefundenen Effekte ist, dass „verkindlichte“ Gesichter deswegen attraktiver bewertet werden, weil ihnen positivere Persönlichkeitseigenschaften zugeschrieben werden. Etliche Untersuchungen zeigen, dass kindlich (bzw. feminin) aussehenden Gesichtern auch ebensolche Persönlichkeitseigenschaften zugeschrieben werden, die mit Kindlichkeit oder Femininität assoziiert werden (Berry & Brownlow, 1989; Berry & McArthur, 1985; Berry & McArthur, 1986, 1988; Berscheid & Walster, 1974;

Brownlow & Zebrowitz, 1990; Bull & Rumsey, 1988; Eagly et al., 1991; McArthur & Apatow, 1984; Penton-Voak, Perrett & Peirce, 1999; Zebrowitz-McArthur & Montepare, 1989; Zebrowitz, 1997; Zebrowitz, Tenenbaum & Goldstein, 1991).

Es wurde auch gezeigt, dass diese Stereotype altersunabhängig sind, sowohl für Beurteiler als auch für beurteilte Gesichter (Adams & Huston, 1975; Johnson & Pittenger, 1984; Langlois, 1986; Montepare & Zebrowitz-McArthur, 1989; Zebrowitz & Montepare, 1992). Diese Effekte beziehen sich nicht nur auf die soziale Wahrnehmung von Personen, sondern ziehen auch tatsächlich Konsequenzen für die soziale Interaktion nach sich (Berkowitz & Frodi, 1979; Dion, 1972, 1974; Felson, 1980; Lerner & Lerner, 1977; Salvia, Algozzine & Sheare, 1977; Zebrowitz, Kendall-Tackett & Fafel, 1991; Zebrowitz & McDonald, 1991). Für einen Überblick siehe Zebrowitz et al. (1993) und Zebrowitz (1997). Es könnte daher sein, dass die Art der sozialen Stereotypisierung Auswirkungen auf das Attraktivitätsurteil hat. Wenn also ein „verkindlichtes“ Gesicht als netter, freundlicher oder niedlicher wahrgenommen wird, resultiert daraus ein positiverer, sympathischerer Gesamteindruck, der sich auch positiv auf das Attraktivitätsurteil auswirken könnte.

Bei Frauengesichtern könnte sich zusätzlich positiv auswirken, dass die den verkindlichten Frauengesichtern zugeschriebenen Charaktereigenschaften auch in stärkerem Maße mit den klassischen Rollenerwartungen an Frauen übereinstimmen, wie sie z. B. mit dem *Sex Role Stereotype Questionnaire* (Rosenkrantz, Vogel, Bee, Broverman & Broverman, 1968), was zusätzlich zu einem positiveren globalen Gesamteindruck führen könnte.

Bei Männergesichtern wirkt das Kindchenschema zwar gegen das typisch männliche Geschlechtsstereotyp, hier wird jedoch von anderen Autoren wie Perrett et al. (1998) argumentiert, dass das weniger maskuline Männergesicht zu einer Wahrnehmung von positiveren, angenehmeren Persönlichkeitseigenschaften führt (weniger dominant, wärmer, anständiger, kooperativer und mit einer höheren Wahrscheinlichkeit ein guter Vater zu sein). „Feminization of male face shape may increase attractiveness because it ‘softens’ particular features that are perceived to be associated with negative personality traits” (Perrett et al., 1998, S. 886). Nach dieser Sichtweise wäre demnach die Ursache für die Attraktivitätssteigerung ein Halo-Effekt des kindlich-femininen Gesichts, das zu einer positiveren Einschätzung des Charakters der betreffenden Person führt und über diesen Umweg auch zu einem höheren Attraktivitätsurteil. Eine interessante Sichtweise, die durchaus Beachtung verdient, auch wenn – oder gerade *weil* – traditionell beim klassischen Attraktivitätsstereotyp („Was schön ist, ist auch gut“) stets implizit angenommen wird, dass die Attraktivität sich kausal auf die Charakterzuschreibungen auswirkt. Doch

die bekannten Befunde sind selbstverständlich erstens nur korrelative Zusammenhänge und zweitens gibt es auch über 40 Jahre nach der Beschreibung des Phänomens des Attraktivitätsstereotyps immer noch keine überzeugende Erklärung, *warum* dieser Zusammenhang besteht.

Für eine genauere Untersuchung dieser Hypothese wäre es notwendig, bei Attraktivitätsuntersuchungen zum sexuellen Dimorphismus grundsätzlich nicht nur Attraktivitätsurteile zu messen, sondern auch Zuschreibungen von Persönlichkeitseigenschaften. Falsifizierbar ist die Hypothese, wenn sich eine Manipulation in Richtung Kindchenschema in unterschiedliche Richtungen auf Charakterzuschreibung und Attraktivität auswirkt, also positiv auf zugeschriebene Persönlichkeitseigenschaften, aber nicht positiv auf das Aussehen, bzw. umgekehrt, positiv auf das Aussehen, aber nicht auf zugeschriebene Persönlichkeitseigenschaften. Hierbei wäre es zudem sinnvoll, neben positiven (zugeschriebenen) Charaktereigenschaften, die vermutlich positiv mit dem Kindchenschema in Zusammenhang stehen, auch solche Eigenschaften zu testen, die damit vermutlich weniger kompatibel sind, jedoch ebenfalls positiv gesehen werden (z. B. erfahren, unabhängig, durchsetzungsfähig, kompetent, stark). Denn es ist unplausibel, dass kindliches Aussehen *nur* positive Konsequenzen für Charakter-Stereotypisierungen haben soll und keine negativen. Für den Attraktivitätsforscher kommt es daher auch darauf an, die relevanten Variablen zu untersuchen und damit überhaupt die richtigen Fragen zu stellen.

7.4.2 Mögliche neutrale oder attraktivitätssenkende Merkmale des Kindchenschemas

Selbst wenn die drei Untersuchungen klar zeigen könnten, dass sich die Attraktivität von Frauengesichtern und sogar von Männergesichtern durch kindliche Gesichtszüge steigern lässt, so sollte man dennoch nicht zu unkritisch sein und annehmen, dass *alle* für das Kindchenschema charakteristischen Merkmale attraktivitätserhöhend wirken. Das Kindchenschema ist schließlich kein Einzelmerkmal, sondern eine komplexe Merkmalskombination. Es wäre also gut denkbar, dass nur *manche* kindlichen Merkmale für die Attraktivitätserhöhung kausal verantwortlich sind, während andere möglicherweise irrelevant und wiederum andere vielleicht sogar attraktivitätssenkend sind.

Hierzu zwei Beispiele: Kindliche Gesichter weisen keine ausgeprägten Wangenknochen auf, dafür jedoch rundliche Pausbacken. Studien zeigen jedoch, dass das Gegenteil, nämlich ausgeprägte Wangenknochen und konkave (nach innen gehende Wangen) ein Gesicht attraktiv machen (Cunningham, 1986; Cunningham et al., 1990). Nach Cunning-

ham sei es bei Frauengesichtern gerade die Kombination aus Jugendmerkmalen und Reifekennzeichen, die für Attraktivität entscheidend sei.

Bei den Wangenknochen sei an dieser Stelle auf eine häufige Quelle von Missverständnissen hingewiesen. Hohe, ausgeprägte Wangenknochen sind ein Reifekennzeichen, das erst durch den Einfluss von Testosteron während der Pubertät entsteht (Farkas, 1987; Farkas & Posnick, 1992). Gleichzeitig wird jedoch gelegentlich behauptet, dass Frauen (im Durchschnitt) stärker ausgeprägte Wangenknochen hätten als Männer. Dieser Eindruck beruht jedoch auf einer optischen Täuschung. Weibliche Wangenknochen wirken v. a. im Halbprofil stärker ausgeprägt als männliche. Dies liegt jedoch daran, dass bei Frauen der Nasensteg flacher verläuft und die Nase und die Stirn weniger protrusiv sind als bei Männern. Dadurch wirken aus einem Blickwinkel von 45° die Wangenknochen bei Frauen optisch ausgeprägter, obwohl sie es in Wirklichkeit nicht sind. Diesen Zusammenhang erläutert Enlow (1989, S. 21ff) sehr gut und illustriert ihn mit anschaulichen Abbildungen. Es ist also keineswegs ein Widerspruch, wenn ausgeprägte, hohe Wangenknochen als testosteronbedingtes Reifekennzeichen eingeordnet werden. Dieses Beispiel macht aber deutlich, dass die Beurteilung von Gesichtsmerkmalen oft komplizierter ist, als es auf den ersten Blick scheint, weil es nicht nur um das Merkmal an sich (z. B. die Merkmalsgröße) geht, sondern bei der Gesichtswahrnehmung jedes Merkmal immer im Kontext des ganzen Gesichts gesehen wird.

Bei den Pausbacken gibt es ebenfalls eine wichtige methodische Einschränkung, die ebenfalls leicht zu Missverständnissen führen kann. Die für die Studie fotografierten Kindergartenkinder, aus denen das Kindchenschema erzeugt wurde, besaßen alle Pausbacken. Dementsprechend sind bei der verwendeten Grafik des Kindchenschemas auch entsprechende „Ausbuchtungen“ zu sehen. Dennoch bedeutet dies nicht, dass bei Anwendung dieses Schemas auf ein Erwachsenengesicht dieses Gesicht auch „pausbäckiger“ wird. Das Problem ist, dass der Wahrnehmungseindruck von Pausbacken nicht nur durch die frontalen Gesichtskonturen entsteht, sondern v. a. durch eine spezielle Verteilung von Farbwerten, die durch den natürlichen Lichteinfall auf einem Foto entstehen. Das Licht, das normalerweise von oben kommt, verursacht bei allen dreidimensionalen Strukturen einen charakteristischen Schattenwurf (Metzger, 1975) – so auch bei Pausbacken. Mit anderen Worten: Auch die Textur eines Gesichts, d. h. die Verteilung von Farbwerten, liefert dem Gehirn Informationen über dessen Form. Da jedoch bei den Manipulationen mit Morphing-Software die Farbwerte der Originalgesichter nicht verändert, sondern lediglich Proportionen verzerrt wurden, entsteht bei den auf diese Weise verkindlichten Gesichtern nicht der Eindruck von Pausbacken. Die „Ausbeulung“ am Unterkiefer,

die beim Kindchenschema ursächlich durch Pausbacken entstanden ist, wird bei Anwendung auf ein Erwachsenengesicht optisch eher als Kieferknochen (Gonion) interpretiert. Den Effekt, den die Farbwertverteilung (aufgrund des Schattenwurfs) auf die Wahrnehmung von Formen hat, kann man veranschaulichen, wenn man sich vorstellt, wie verschiedene Objekte wirken, wenn man sie mit einem großen einheitlich weißen Tuch verdeckt. Auch wenn sämtliche Konturen des Objekts verschwunden sind und das Auge nur noch unterschiedlich dunkle Abstufungen von Weiß sieht, erkennt das Gehirn dennoch problemlos, ob unter dem Tuch ein Würfel, eine Kugel, ein Teller, ein Becher oder eine Flasche steht.

Die Anwendung des Kindchenschemas auf erwachsene Gesichter unter Beibehaltung der Originaltextur imitiert also nur einen Teil des kindlichen Aussehens. Der optische Eindruck von Pausbacken entsteht dadurch *nicht*. Prinzipiell wäre es natürlich möglich, Manipulationen in Richtung Kindchenschema durchzuführen, in dem man Erwachsenengesichter mit Durchschnittsbildern von Kindern kreuzt und dabei auch Farbwerte mittelt. Dies würde jedoch zu vielen anderen Bildverarbeitungsartefakten führen (wie z. B. makelloosere Haut, Diskussion siehe Kapitel 3.1.2), die einen weitaus größeren Störfaktor darstellen würden als das Fehlen von Pausbacken. Festzuhalten bleibt: Da die Pausbäckigkeit nicht manipuliert wurde, gibt es auf der Grundlage der durchgeführten Experimente keinen Grund zu der Annahme, dass Pausbacken ein Gesicht attraktiv machen, obwohl Pausbacken (neben großen, rundlichen Augen) das wohl auffälligste Einzelmerkmal des Kindchenschemas sind.

Eine kritische Reflexion ist auch bei der kindlichen Augenpartie angebracht. Es wird gerne gesagt, dass Kinder „große Augen“ haben, doch diese Aussage ist sehr pauschal (Berry & Brownlow, 1989; Berry & McArthur, 1985; Zebrowitz, 1997). Präziser formuliert, ist das kindliche Auge im Vergleich zum erwachsenen Auge dadurch gekennzeichnet, dass

- der sichtbare Teil des Augapfels in Relation zum Schädel größer ist,
- die Iris im Verhältnis zur Sklera eine größere Fläche einnimmt,
- die Iris durch das Oberlid weniger stark bedeckt ist,
- das Auge rundlicher und weniger „mandelförmig“ ist,
- der Winkel zwischen Unterlid und Oberlid am inneren Augenwinkel (Endocanthion) größer ist.

Diese Merkmale wurden auch durch den Morphing-Prozess verändert. Hinzu kommen noch weitere Merkmale, die in diesem Experiment durch das Morphen *nicht* verändert wurden, wie der Unterschied, dass

- die Augenwimpern länger sind,
- die Pupillen bei gleicher Lichtintensität größer sind (Korczyn, Laor & Nemet, 1976),
- die Sklera heller ist.

Welche dieser kindlichen Merkmale tatsächlich eine erwachsene Augenpartie attraktiver machen, ob es alle diese Merkmale sind oder nur manche, ist schlichtweg völlig unbekannt. Auf Grundlage der durchgeführten Experimente kann ja noch nicht einmal gesagt werden, dass die Augen *überhaupt* einen positiven Einfluss auf das Attraktivitätsurteil hatten, denn vielleicht ist die Attraktivitätssteigerung auf ganz andere Gesichtsmerkmale zurück zu führen. Eine Aussage wie „große Augen machen attraktiv“ ist deswegen gleich in zweifacher Hinsicht nicht sinnvoll, da die Aussage viel zu pauschal und die Kausalität empirisch nicht belegt ist.

Beim hier verwendeten Kindchenschema (vgl. Abbildung 59, S. 254) fällt auch auf, dass der Mund des Kindchenschemas relativ klein ist. Volle Lippen gelten jedoch erstens als Attraktivitätsmerkmal und zweitens als typisch feminin. Wie ist dies zu bewerten?

Besonders gut erkennt man an der Abbildung 58 (S. 251), dass bei der Bildmanipulation in Richtung Kindchenschema tatsächlich der Mund der erwachsenen Frau stark verkleinert wird. Die durchgeführten Experimente zur Durchschnittshypothese haben jedoch gezeigt, dass die attraktivsten Gesichter (vgl. auch die Prototypen für attraktive und unattraktive Gesichter, Abbildung 5, S. 20 und Abbildung 21, S. 48) deutlich vollere Lippen besitzen als der Durchschnitt (vgl. Durchschnittsgesichter, Abbildung 6, S. 21 und Abbildung 22, S. 49). Dass diese und andere Frauen durch die Manipulation in Richtung Kindchenschema dennoch an Attraktivität gewinnen, dürfte also eher nicht *wegen*, sondern *trotz* der weniger vollen Lippen der Fall sein.

Dass volle Lippen als feminines Merkmal gelten, jedoch gleichzeitig im Kindchenschema nicht vorhanden sind, ist dadurch zu erklären, dass sich bei Frauen die Lippenfülle als geschlechtsdimorphes Merkmal erst in der Pubertät voll ausbildet. Denn verantwortlich für die Ausprägung voller Lippen ist das Östrogen (Johnston & Franklin, 1993; Symons, 1995; Thornhill & Moller, 1997). Insofern sind volle Lippen beim sexuellen Dimorphismus eine Ausnahme: Sie sind ein Reifekennzeichen, jedoch eines, das nicht durch Testosteron, sondern durch Östrogen ausgebildet wird. Die Lippengröße ist also in gewisser Wei-

se ein Problemfall, da hier durch die Anwendung des Kindchenschemas ein Gesicht nur kindlicher, nicht jedoch femininer gemacht wird.

Insgesamt bleibt festzuhalten, dass die Deutlichkeit, mit der das Kindchenschema sich in den drei durchgeführten Experimenten als attraktivitätsförderlich erwies, keinesweg dazu verleiten sollte, pauschal die Neotenie als Attraktivitätsideal zu postulieren (Eibl-Eibesfeldt, 1997). Bei einer kritischen Betrachtung lässt sich begründen, dass im Kindchenschema auch solche Merkmale enthalten sind, bei denen *nicht* davon ausgegangen werden kann, dass sie zur Attraktivitätssteigerung beitragen. Reifekennzeichen wie ausgeprägte Wangenknochen und konkave Wangen tragen vermutlich zur Attraktivität bei (Cunningham, 1986; Cunningham et al., 1990), stehen jedoch im Widerspruch zum Kindchenschema. Welche Merkmale im Einzelnen tatsächlich kausal für die Attraktivität neotener Gesichtszüge verantwortlich sind, muss durch zukünftige Untersuchungen geklärt werden.

7.5 Schlussfolgerungen und Ausblick: Von globalen Theorien und vom Schema-Denken zur Analyse von konkreten Einzelmerkmalen

Die Ausführungen des vorherigen Kapitels machen deutlich, dass eine pauschale Postulierung von neotenen Gesichtszügen als Attraktivitätsideal problematisch ist, da es zu viele Einschränkungen dieser allgemeinen Regel gibt. Zwar haben kindliche Gesichtszüge insgesamt einen positiven Effekt auf die Gesichtsattraktivität, aber es bleibt ungeklärt, welche Merkmale konkret für diesen Gesamteffekt verantwortlich sind. Die Aufgabe der Attraktivitätsforschung muss es also künftig sein, die Rolle dieser einzelnen kindlichen Gesichtsmerkmale für die Attraktivität eines Gesichts zu untersuchen.

Bei künftigen Untersuchungen sollten v. a. die in Kapitel 7.4.1 erläuterten möglichen Erklärungen beachtet werden, *warum* kindliche Gesichtszüge attraktiv wirken könnten. Dies ist in erster Linie der verjüngende Effekt des Kindchenschemas. Daraus ergeben sich einige konkrete Vorhersagen, die empirisch überprüfbar sind. Denn beim verjüngenden Effekt des Kindchenschemas ist davon auszugehen, dass eigentlich kein Erwachsener wie ein Kind aussehen möchte, sondern vielmehr, dass er jugendlich oder nicht alt aussehen will. Es sollten demnach solche Kindchenmerkmale weniger wirksam sein, die ein

Merkmal von sehr kleinen Kindern sind, und umgekehrt solche wirksamer, die charakteristisch für Jugendliche sind.

Konkret lassen sich daraus für Gesichter beider Geschlechter folgende Hypothesen ableiten:

- Pausbacken (= Kindchenschema-Merkmal) sollten *nicht* attraktiv sein. Denn Pausbacken sind ein Merkmal von sehr kleinen Kindern, und bereits bei Jugendlichen sind sie verschwunden. Stattdessen sollten konkave Wangen (= Reifekennzeichen) attraktiv sein, zumal diese auch ein Merkmal von Schlankheit sind (vgl. auch die Prototypen, Abbildung 5, S. 20 und Abbildung 21, S. 48).
- Das Fehlen von hohen, ausgeprägten Wangenknochen (= Kindchenschema-Merkmal) sollte *nicht* attraktiv sein. Denn auch dies ist ein Merkmal von kleinen Kindern, nicht jedoch von Jugendlichen.
- Ein breiter Nasenrücken (= Kindchenschema-Merkmal) sollte *nicht* attraktiv sein. Denn auch dieser ist ein Merkmal von kleinen Kindern, nicht jedoch von Jugendlichen. Nach der Pubertät verändert sich der Nasenrücken praktisch nicht mehr, er enthält also keine Informationen über das Alter einer erwachsenen Person (Coleman & Grover, 2006). Der Nasenrücken einer 50jährigen Person hat immer noch die gleiche Form wie der einer 20jährigen. Um jugendlicher zu wirken, macht es daher keinen Sinn, die breite Nasenrückenform eines Kindes zu haben. Zudem zeigen auch die Ergebnisse aus Kapitel 3.3, dass schmale Nasen attraktiv sind (vgl. die Abbildungen mit den Prototypen, Abbildung 5, S. 20 und Abbildung 21, S. 48).
- Bei der Augenpartie sollte es für eine Attraktivitätssteigerung *nicht* darauf ankommen, dass ein Auge in der Form rundlicher ist (d. h. weiter Winkel am inneren Augenwinkel) und dass die Iris eine große Fläche im Vergleich zur Sklera einnimmt. Denn auch dies sind Merkmale von kleinen Kindern, nicht jedoch von Jugendlichen. Stattdessen sollte entscheidend sein, dass die Iris durch das Oberlid wenig verdeckt ist, denn der Augenöffnungsgrad ist ein Merkmal, das sich mit zunehmendem Alter verändert: Bei älteren Personen verliert das Oberlid an Spannkraft und sinkt nach unten ab (Fagien, 2002; Rohrich et al., 2004).

Ein kleiner Mund (nicht volle Lippen, = Kindchenschema-Merkmal) sollte nicht attraktiv machen, denn er ist ein Merkmal von kleinen Kindern. Bei jungen Frauen sind die Lippen aufgrund des Östrogen-Einflusses in der Pubertät voller als bei Kindern und daher sowohl ein Reifekennzeichen als auch ein Merkmal für junges Alter, da die Lippenfülle (sowohl das sichtbare Lippenrot als auch die Lippen im anatomischen Sinn) bei Erwachsenen mit

höherem Alter aufgrund des Abbaus von perioralem Fettgewebe abnimmt (Coleman & Grover, 2006). Da die Ergebnisse in Kapitel 3.3 gezeigt haben, dass auch attraktive Männer sich von unattraktiven ebenfalls durch vollere Lippen unterscheiden (vgl. Prototypen, Abbildung 5, S. 20) sollte dieser Zusammenhang auch für Männer gelten.

Während es bei Frauengesichtern bei der Untersuchung des Effekts der Neotenie um Nuancen geht, d. h. um die Frage, ob das Ideal ein kindliches oder jugendliches Merkmal ist, liegt der Fall bei männlichen Gesichtern völlig anders. Denn hier widerspricht der positive Effekt der Neotenie ja der Theorie des sexuellen Dimorphismus, derzufolge kindliche Merkmale bei erwachsenen Männergesichtern ja eigentlich gar nicht attraktiv sein sollten, obwohl zahlreiche Studien bereits zu diesem widersprüchlichen Resultat gekommen sind (Ishi, Gyoba, Kamachi et al., 2004; Little et al., 2001; Little & Hancock, 2002; Perrett et al., 1998; Rhodes et al., 2000; Swaddle & Reiersen, 2002).

Die allgemeine Aussage: „Je männlicher ein Männergesicht, desto attraktiver“ ist also sicher nicht länger haltbar. Auf der anderen Seite ist jedoch auch die gegenteilige Aussage „je weiblicher (oder kindlicher) ein Männergesicht, desto attraktiver“ offenkundig wenig sinnvoll. Ein anschauliches Gegenbeispiel ist der attraktive männliche Prototyp (Gesicht K in Untersuchung *Dimorph 3*, vgl. Abbildung 67, S. 271), der durch die Verkindlichung weitaus unattraktiver bewertet wurde. Hinzu kommen korrelative Untersuchungen (Cunningham et al., 1990; Grammer & Thornhill, 1994; Koehler et al., 2004; Penton-Voak, Jones et al., 2001; Scheib et al., 1999), die durch Vermessen von *Facial Landmarks* einen positiven Effekt von typisch männlichen Merkmalen fanden, insbesondere bei der Größe des Unterkiefers und des Kinns.

Die heterogenen Befunde bei männlicher Gesichtsattraktivität legen den Schluss nahe, dass es bei attraktiven männlichen Gesichtern offensichtlich nicht auf ein Extrem ankommt, sondern auf die richtige Mischung aus maskulinen Merkmalen einerseits und kindlich-femininen Merkmalen andererseits. Zu diesem Ergebnis kommt auch die relativ neue Untersuchung von Rennels et al. (2008, S. 891), deren Ergebnisse aufgrund der verwendeten Methodik nicht ganz leicht zu interpretieren sind. Will man männliche Attraktivität wirklich verstehen, wird man sich von einem starren Schema (z. B. Kindchenschema) ebenso lösen müssen wie von einer simplen Theorie (z. B. dass *alle* Merkmale, deren Ausprägung durch Testosteron gefördert wird, attraktiv machen). Notwendig ist eine detailliertere Analyse von einzelnen Merkmalen, z. B. in Form von experimentellen Untersuchungen, bei denen einzelne Merkmale manipuliert, aber auch mögliche Interaktionen zwischen diesen Einzelmerkmalen berücksichtigt werden.

Konkret überprüfen sollte man z. B. folgende Hypothesen:

- Ein markanter kräftiger Unterkiefer und ein markantes Kinn (= Testosteron-Marker) machen ein Männergesicht attraktiv.
- Eine schmale und wenig protrusive Nase macht ein Männergesicht attraktiv, *obwohl* auch eine große Nase ein Testosteron-Marker ist.

Gerade für die Kombination aus markantem Unterkiefer und kleiner Nase lassen sich viele prominente Beispiele von Männern finden, die in der Bevölkerung im Allgemeinen als attraktiv angesehen werden (z. B. George Clooney, Pierce Brosnan, Til Schweiger, Brad Pitt), was selbstverständlich keinerlei wissenschaftliche Beweiskraft hat, aber zum Generieren von wissenschaftlich überprüfbaren Hypothesen doch allemal taugt.

Weitere Hypothesen sind:

- Schmale Lippen machen ein Männergesicht unattraktiv (*obwohl* sie Testosteron-Marker sind), da sie ein Alterskennzeichen sind. Attraktiv dürften eher mittelvolle Lippen machen. Wenn man sich am Aussehen von männlichen Fashion-Models orientiert, dann ist sogar die Hypothese naheliegend, dass volle Lippen Männer attraktiv machen (= typisch feminines Merkmal).
- Sehr schmale, optisch zu tief liegende Augen machen ein Männergesicht unattraktiv (*obwohl* sie Testosteron-Marker sind), da sie sich negativ auf die soziale Wahrnehmung auswirken (Perrett et al., 1998). Das Ideal dürften eher durchschnittlich aussehende Augen sein.
- Stark ausgeprägte Augenbrauenwülste, die den Eindruck einer deutlich fliehenden Stirn verursachen, machen ein Männergesicht unattraktiv (*obwohl* sie Testosteron-Marker sind), da sie sich negativ auf die soziale Wahrnehmung auswirken (Perrett et al., 1998). Dieses Merkmal ist jedoch besser in der Profil- oder Halbprofil-Ansicht zu beurteilen.
- Ein größerer Kiefer-Hals-Winkel (der Winkel zwischen Unterkiefer und Hals, sichtbar in der Profilansicht) sollte Männergesichter unattraktiver machen, obwohl es ein sexuell dimorphes Merkmal ist (vgl. dazu die männlichen und weiblichen Durchschnittsgesichter in der Profilansicht, Abbildung 11, S. 29 und Abbildung 27, S. 58). Attraktiver sollte ein kleiner Kiefer-Hals-Winkel sein, da er ein Merkmal für Jugendlichkeit und Schlankheit ist (Coleman & Grover, 2006).
- Geheimratsecken oder (Halb)glatzen machen ein Männergesicht unattraktiv, obwohl es (abgesehen vom Bartwuchs) kein Gesichtsmerkmal gibt, das derart exklusiv den

Männern vorbehalten ist und daher ein besonders gutes Beispiel für sexuellen Dimorphismus darstellt. Volles Haar hingegen sollte attraktiv wirken, da es ein Jugendmerkmal ist (vgl. auch die männlichen Prototypen, Abbildung 5, S. 20).

- Eine unreine und großporige Haut sollte ein Männergesicht unattraktiver machen (obwohl auch sie ein Testosteron-Marker ist). Stattdessen sollte eine reine und feinporige Haut attraktiv wirken, weil es ein Merkmal für Gesundheit und Jugendlichkeit ist. Ein deutliches Indiz für diese Hypothese liefern die eindeutigen Ergebnisse mit den gemorphten Männergesichtern aus den Experimenten *Durch 1* und *Durch 2*, bei denen gerade die Composites höherer Morph-Generationen mit einer sehr makellosen Haut besonders attraktiv bewertet wurden, und das, obwohl die extrem glatte und feine Haut dieser Gesichter völlig unmännlich aussieht.

Eine Überprüfung dieser Hypothesen könnte endlich die bis heute kaum verstandenen Gesetzmäßigkeiten für männliche Gesichtsattraktivität erhellen. Man muss sich dafür allerdings von einer beliebten Theorie (oder besser gesagt Hypothese) verabschieden – der Theorie des sexuellen Dimorphismus – die den Anspruch erhebt, männliche Attraktivität allgemein und umfassend zu erklären, diesen Anspruch jedoch nicht erfüllen kann. Zwar ist dem Kurt Lewin zugeschriebenen Satz „Nichts ist so praktisch wie eine gute Theorie“ prinzipiell zuzustimmen, allerdings liegt die Betonung auf „gute“. Eine Theorie, die in so vielen Punkten mit der Realität nicht übereinstimmt, ist für einen wissenschaftlichen Erkenntnisgewinn eher hinderlich, weil sie den Blick auf Alternativerklärungen verstellt.

Möglicherweise zeigt sich im Festhalten an der Theorie des sexuellen Dimorphismus bei männlicher Gesichtsattraktivität aber auch ein Grundproblem der aktuellen akademischen Attraktivitätsforschung. Viele Forscher scheinen weniger daran interessiert zu sein herauszufinden, was ein Gesicht tatsächlich attraktiv macht, sondern sie interessieren sich stärker für Fragestellungen und Probleme der evolutionären Psychologie. Wer an allgemeinen biologischen Gesetzmäßigkeiten menschlichen Verhaltens interessiert ist, ist bereits zufrieden, wenn er zeigen kann, dass bestimmte allgemeine Zusammenhänge (wie es sie im Tierreich gibt) auch beim Menschen eine Rolle spielen. Selbst ein äußerst geringer Effekt kann nach dieser Sichtweise interessant sein, da auch winzige Unterschiede (sofern sie adaptiv sind) sich über die langen Zeiträume, in denen die Evolutionsbiologie denkt, durchsetzen können und damit das Aussehen einer Art oder deren Partnerwahlverhalten beeinflussen können. Effektgröße ist ein Aspekt, der für die Angewandte Psychologie wichtig ist, nicht jedoch für die Evolutionäre Psychologie.

Beim Überprüfen von Hypothesen, die vom biologischen Standpunkt gesehen wichtig und bedeutsam sind, weil hier am Beispiel von menschlicher Attraktivität und Partnerwahl untersucht werden kann, wie Evolution generell funktioniert, kann man jedoch leicht übersehen, dass auf diese Art von menschlichem Empfinden und Verhalten möglicherweise ganz andere Faktoren (Sozialisierungseffekte durch Gesellschaft, Peer-Group, Massenmedien, Mode usw.) einen viel größeren Einfluss auf das aktuelle Schönheitsideal ausüben. Dies schließt nicht aus, dass es möglicherweise in den letzten 200.000 Jahren Menschheitsgeschichte tatsächlich biologisch „angeborene“ Präferenzen waren, die in der Vergangenheit das Attraktivitätsideal geprägt haben und möglicherweise auch durch sexuelle Selektion dazu geführt haben, dass der Mensch bzw. die beiden Geschlechter heute so aussehen wie sie aussehen. Dennoch ist es denkbar, dass heutzutage diese biologisch determinierten Präferenzen durch kulturell vermittelte Faktoren (stark) überlagert sind.

Ein simples Beispiel soll dies deutlich machen: Dass Männer überall auf der Welt eine stärkere Körperbehaarung (inklusive Bartwuchs) aufweisen als Frauen, mag durch sexuelle Selektion entstanden sein, d. h. dass in der Geschichte der Menschheit Frauen solche Männer als Fortpflanzungspartner bevorzugten, die eine stärkere Körperbehaarung aufwiesen (sofern Körperbehaarung die Folge von *natürlicher* Selektion ist, z. B. Schutz vor Witterungseinflüssen bei der Jagd o. Ä., hätte sie mit Attraktivität sowieso nichts zu tun gehabt). Dennoch ist es eine Tatsache, dass gegenwärtig in den meisten Ländern der Welt Vollbärte als unattraktiv gelten und sogar (v. a. in westlichen Industrienationen) vielfach ein unbehaarter Oberkörper bei Männern als attraktiv angesehen wird (Martins, Tiggemann & Churchett, 2008), während dies noch bis in die 70er Jahre anders war. Die typisch maskulinen sexuell dimorphen Merkmale Bartwuchs und starke Körperbehaarung gelten aktuell nicht als männliches Attraktivitätsideal, ganz zu schweigen von der verallgemeinernden Regel: „Je behaarter, desto attraktiver“. Dies ist ein Grunddilemma der evolutionären Psychologie, dass sie einerseits psychologische Mechanismen untersucht, die vor Jahrtausenden unter den damaligen Umweltbedingungen relevant waren, als empirische Wissenschaft sich jedoch andererseits sozialwissenschaftlicher Methoden (mit all ihren Einschränkungen) bedienen muss und dabei stets nur die gegenwärtigen in einer bestimmten modernen Kultur lebenden Menschen untersuchen kann.

Durch die Dominanz von evolutionsbiologisch orientierten Forschern in der Attraktivitätsforschung liegt der Focus aktueller Fragestellungen gegenwärtig auf den *ultimaten* Ursachen für Attraktivität, also auf den dahinter liegenden Gründen, auf dem (biologischen) Sinn und Zweck von bestimmten Merkmalen. Für das Verständnis der Frage, welche

Merkmale genau einen Menschen attraktiv machen, wäre jedoch eine Fokussierung auf die *proximaten* Ursachen notwendig, d. h. eine intensivere Beschäftigung mit den Reizgrundlagen. Welche Merkmale in welcher genauen Ausprägung machen attraktiv und welche Interaktionen zwischen den einzelnen Merkmalsausprägungen gibt es? Eine Hinwendung der Attraktivitätsforscher zu den konkreten Reizgrundlagen würde auch eine andere Forschungsstrategie mit sich bringen, nämlich weg von Top-down-Fragestellungen (z. B. „Bei Männern macht *alles* attraktiv, was typisch maskulin ist“ oder „Attraktiv ist ein Gesicht, das *in allen Merkmalen* durchschnittlich ist“) und hin zu Bottom-up-Fragestellungen (z. B. „Welche Kombination von Merkmalen muss ein attraktives Gesicht haben?“, „Was ist ein attraktiver Unterkiefer?“ oder „Welche Merkmale hat eine attraktive Nase?“).

Vor allem unter dem Aspekt der Anwendungsmöglichkeiten dieser Forschungsergebnisse wäre diese Umorientierung wichtig und wünschenswert. Denn zu dem weiten Anwendungsfeld der ästhetischen Medizin kann die klassische akademische Attraktivitätsforschung bislang nur wenig beitragen, da sie nach wie vor andere Fragestellungen und Probleme erforscht als solche, die für diejenigen von Bedeutung sind, deren Aufgabe es ist, durch Eingriffe am menschlichen Körper Attraktivität herzustellen. Angesichts der enormen Konsequenzen für das eigene Aussehen, die ein Eingriff durch einen plastischen Chirurgen oder einen Mund-Kiefer-Gesichtschirurgen für einen Patienten haben kann, wäre es jedem Patienten zu wünschen, dass er sich hinsichtlich des ästhetischen Ergebnisses nicht nur auf das subjektive Empfinden und das Bauchgefühl seines Operateurs verlassen braucht, sondern dass die ästhetischen Kriterien, nach denen sein Arzt operiert, auf einem soliden, empirisch-wissenschaftlichen Fundament basieren.

8 Experimente zum direkten experimentellen Vergleich der Hypothesen „Durchschnitt“, „Symmetrie“ und „Sexueller Dimorphismus“

Die meisten Studien beschränken sich darauf, mit einer bestimmten Methodik und einem bestimmten Stimulusmaterial eine einzige Hypothese zu überprüfen. Die Untersuchungen, die publiziert werden, haben in der Regel auch signifikante Effekte gefunden. So gibt es beispielsweise sowohl positive Belege für die Durchschnittshypothese als auch für die Hypothese des sexuellen Dimorphismus, obwohl sich die beiden Hypothesen gegenseitig

widersprechen. Gleichzeitig gibt es Studien, die gefundene Effekte nicht eindeutig auf eine bestimmte Hypothese zurückführen können, da bei der gewählten Methode z. B. Durchschnittlichkeit und Symmetrie konfundiert sind (Composites sind nicht nur durchschnittlich, sondern zugleich auch symmetrisch).

Aus diesem Grund ist es interessant, alle drei Hypothesen gleichzeitig mit derselben Methodik und demselben Stimulusmaterial gegeneinander zu testen, um die erzielten Effektgrößen direkt miteinander vergleichen und ihre relative Bedeutung abschätzen zu können. Zu diesem Zweck wurden sechs Männergesichter und sechs Frauengesichter mit Hilfe von Morphing-Software auf drei Arten manipuliert, nämlich durch

- Annäherung zu 90 % an ein männliches bzw. weibliches Durchschnittsgesicht,
- Symmetrisierung durch Vermorphen mit dem eigenen Spiegelbild und
- Annäherung zu 40 % an die Proportionen eines Kindchenschemas.

Die Messung der Attraktivität erfolgte in zwei unabhängigen Experimenten, nämlich,

- im *Rankingverfahren* (= Untersuchung *Vergleich 1*), bei dem jede Versuchsperson alle vier Varianten eines Gesichts gleichzeitig sah und sie nach ihrer Attraktivität in eine Rangreihenfolge bringen musste, sowie
- im *Ratingverfahren* (= Untersuchung *Vergleich 2*) und *Between-Subjects Design*, bei dem jede Versuchsperson nur stets eine Variante eines Stimulusgesichts sah und deren Attraktivität auf einer Ratingskala bewerten musste.

Das *Rankingverfahren* hat den Vorteil der höheren Sensitivität, jedoch andererseits den Nachteil der größeren Reaktivität. Das *Ratingverfahren* hingegen ist weniger sensitiv, hat dafür aber den Vorteil, dass es weniger reaktiv ist und dass für die einzelnen Gesichtervarianten auch absolute (und nicht nur relative) Attraktivitätswerte angegeben werden können.

8.1 Experiment zum direkten experimentellen Hypothesen-Vergleich mit Männer- und Frauengesichtern im Ranking-Verfahren (Untersuchung *Vergleich 1*)

8.1.1 Methode

8.1.1.1 Versuchspersonen

An der Untersuchung *Vergleich 1* nahmen insgesamt 324 Versuchspersonen teil. 161 Personen waren männlich (49.7 %), 163 weiblich (50.3 %). 134 Personen (41.4 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 16 Jahre, die älteste 79. Der Altersdurchschnitt betrug 33.1 Jahre ($SD = 15.0$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, fanden die Datenerhebungen im öffentlichen Raum statt, z. B. an Bahnhöfen, Fußgängerzonen und Einkaufszentren. Die Teilnahme am Experiment erfolgte freiwillig und wurde nicht entlohnt.

8.1.1.2 Stimulusmaterial

Grundlage für das Stimulusmaterial waren Fotos von sechs Männergesichtern und sechs Frauengesichtern. Alle fotografierten Personen waren zwischen 21 und 30 Jahre alt und wurden in einem professionellen Fotostudio unter standardisierten Bedingungen fotografiert.

Mit Hilfe von Morphing-Software (*MorphMan 4.0*) wurden sie auf drei Arten manipuliert, nämlich durch

- Annäherung zu 90 % an ein Durchschnittsgesicht (Männergesichter an das männliche Durchschnittsgesicht und Frauengesichter an das weibliche Durchschnittsgesicht),
- Symmetrisierung durch Vermorphen mit dem eigenen Spiegelbild und
- Annäherung zu 40 % an die Proportionen eines Kindchenschemas.

Alle Gesichtervarianten wurden durch die „Warp“-Technik hergestellt, d. h. nur die Gesichtsproportionen wurden verändert, während die Farbwerte des Bildes (und damit z. B. die Textur der Haut oder der Haare) unverändert blieben. Ein Beispiel für eine „Verdurschnittlichung“ eines Ausgangsgesichts zeigt Abbildung 69.

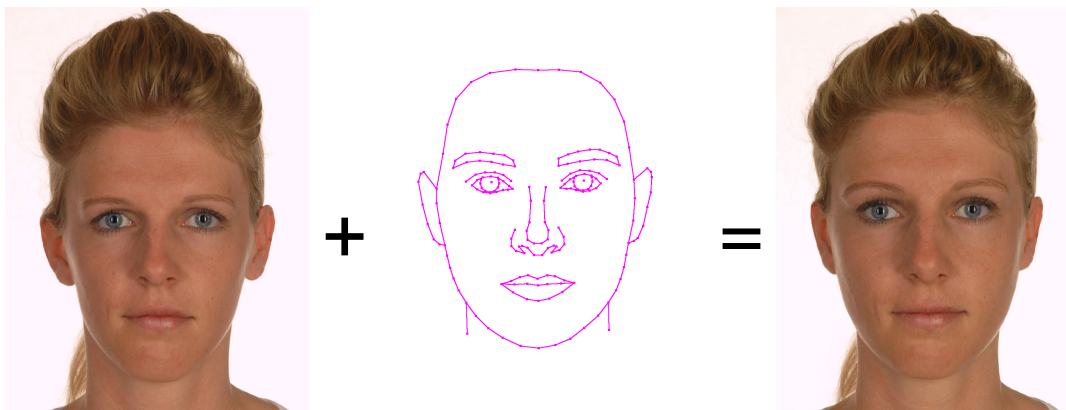


Abbildung 69: Ein Original-Frauengesicht („w60“, linkes Bild) wird zu 90 % an die Gesichtsproportionen des weiblichen Durchschnittsgesichts (berechnet aus 64 Frauengesichtern angenähert, mittleres Bild) angenähert. Es resultiert daraus das Gesicht „w6d“ (rechtes Bild), das in seinen Proportionen nahezu vollständig mit den Gesichtsproportionen des weiblichen Durchschnittsgesichts übereinstimmt.

Mit dieser Art der Verdurchschnittlichung werden die typischen Morphing-Artefakte vermieden, die für einen Großteil der Attraktivität von Durchschnittsgesichtern, die durch „Blend“-Technik (also mit Neuberechnung der Farbwerte) hergestellt wurden, verantwortlich sind. Denn wie in den Untersuchungen *Durch 1* bis *Durch 6* nachgewiesen wurde, ist es vor allem die Textur der Composites (d. h. die makellose Haut), die für die hohen Attraktivitätsbewertungen verantwortlich ist (vgl. dazu Kapitel 3.1.2). Mit dieser Untersuchung wird daher nur gezielt der Einfluss durchschnittlicher Gesichtsproportionen auf die Attraktivität untersucht.

Dass die Annäherung nur zu 90 % und nicht zu 100 % erfolgte, hat einen technischen Grund, denn bei einer vollständigen Annäherung kann es bei einigen Gesichtern (v. a. solche, die stark vom Durchschnitt abweichen) zu Bildverarbeitungsartefakten in Form von lokalen Verzerrungen an einzelnen Stellen im Gesicht kommen, die bei der Stimuluspräsentation einen Störfaktor darstellen würden. Der Unterschied zwischen einer Annäherung zu 90 % oder zu 100 % ist jedoch (abgesehen von möglichen Artefakten) ohnehin kaum erkennbar, so dass gesagt werden kann, dass die derart manipulierten Gesichtsvarianten nahezu vollständig die Proportionen eines männlichen bzw. weiblichen Durchschnittsgesichts besitzen.

Das weibliche Durchschnittsgesicht stammt aus der Untersuchung *Durch 4* (Abbildung 22, S. 49) und wurde aus 64 Frauengesichtern berechnet. Das männliche Durchschnittsgesicht ist *nicht* identisch mit dem Durchschnittsgesicht aus der Untersuchung *Durch 1* (Abbildung 6, S. 21), aber ähnlich. Es wurde nämlich noch durch 16 weitere Männergesichter einer zusätzlichen Stichprobe ergänzt. Das neue *16-face-composite* wurde mit dem alten *32-face-composite* in der Gewichtung 1:2 vermorpht, um ein neues Durchschnittsgesicht zu erzeugen, das aus 48 Gesichtern besteht und damit durch die größere Stichprobe näher an den „wahren Wert“ durchschnittlicher männlicher Gesichtsproportionen herankommt.

Auch das verwendete Kindchenschema wurde auf eine breitere empirische Basis gestellt. Das in den Untersuchungen *Dimorph 1* bis *Dimorph 3* verwendete Durchschnittskindergesicht wurde um weitere 8 Kindergesichter (nachträglich fotografiert im selben Kindergarten) erweitert, so dass das neue verwendete Kindchenschema aus insgesamt 12 Kindergesichtern bestand (siehe Anhang G, S. 401). Unter den 8 zusätzlich fotografierten Kindern waren 4 Jungen und 4 Mädchen. Alle Kinder waren 5 Jahre alt (mit Ausnahme eines 6jährigen Jungen).

Die Annäherung zu 40 % an das Kindchenschema mit der „Warp“-Technik erfolgte analog zur Erzeugung der verkindlichten Stimuli im Experiment *Dimorph 2* und *3* (Kapitel 6.3.1.2 und Kapitel 6.4.1.2). Die Symmetrisierung durch Vermorphen mit dem eigenen Spiegelbild im „Warp“-Verfahren erfolgte analog zum Vorgehen in den Experimenten *Sym 5* und *Sym 6*.

Es resultierten folglich von jedem der 12 Ausgangsgesichter jeweils 4 Varianten, nämlich eine verdurchschnittlichte, eine symmetrisch optimierte, eine verkindlichte, sowie die unveränderte Originalversion als Kontrollbedingung. Die folgenden Abbildungen illustrieren dies anhand eines Männergesichts und eines Frauengesichts.

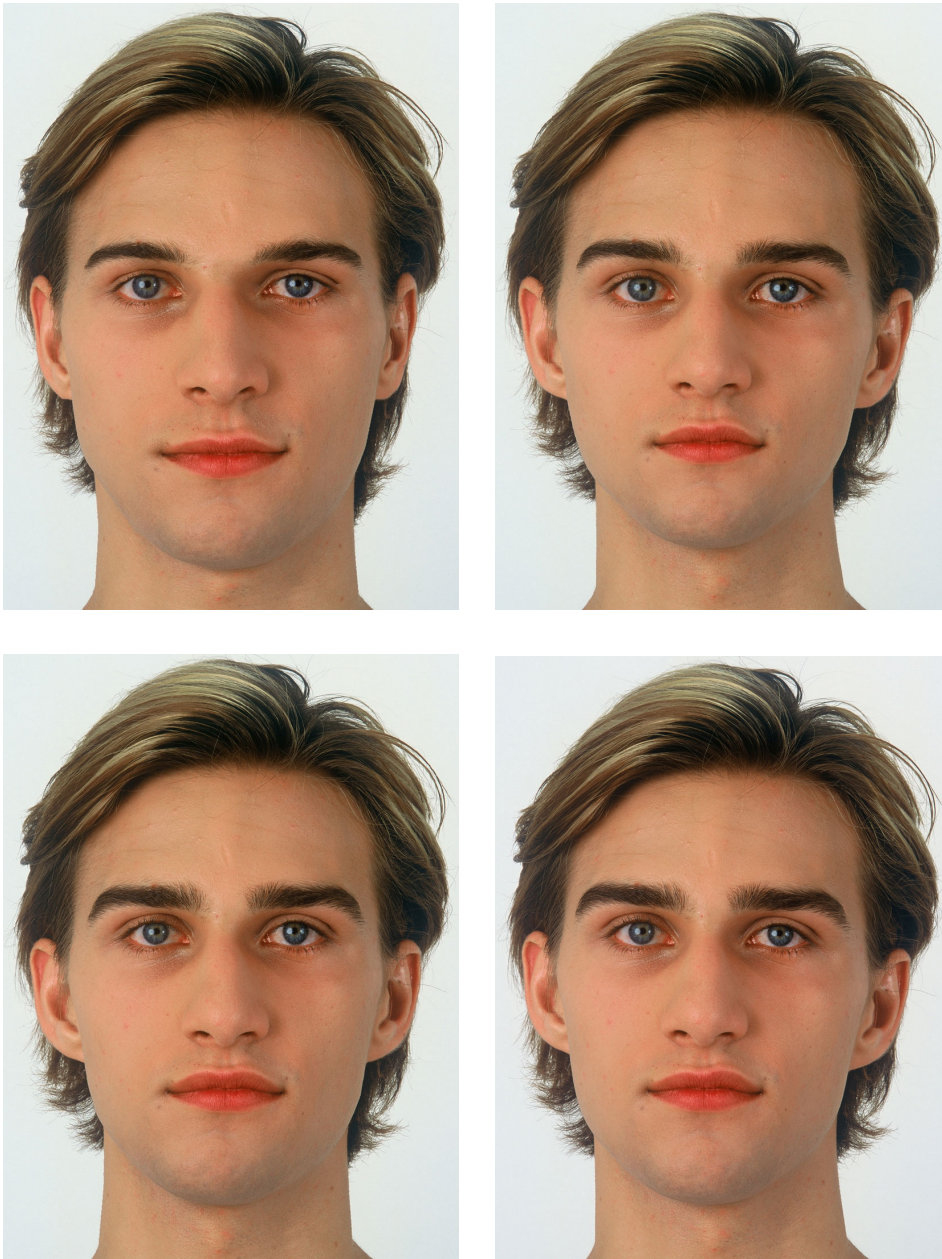


Abbildung 70: Männergesicht („m5“), das durch Morphing an die Proportionen eines Durchschnittsgesichts angepasst wurde („m5d“, oben links), zu 40 % an die Proportionen des Kindchemas angenähert wurde („m5k“, oben rechts) und symmetrisch optimiert wurde („m5s“, unten links). Unten rechts das Original („m5o“) als Kontrollbedingung.



Abbildung 71: Frauengesicht („w3“), das durch Morphing an die Proportionen eines Durchschnittsgesichts angepasst wurde („w3d“, oben links), zu 40 % an die Proportionen des Kindchemas angenähert wurde („w3k“, oben rechts) und symmetrisch optimiert wurde („w3s“, unten links). Unten rechts das Original („w3o“) als Kontrollbedingung.

Alle 48 Stimuli wurden von einem Fotodienst auf Fotopapier in einer Größe von 10 cm x 13 cm bei einer Auflösung von 1200 x 1600 Bildpunkten belichtet. Dies entspricht einer sehr hohen Auflösung von 300 dpi, so dass selbst einzelne Haare auf den Bildern erkennbar waren.

8.1.1.3 Versuchsdurchführung

Bei der Datenerhebung wurden jeder Versuchsperson alle 12 Stimulussets (12 x 4 Varianten) präsentiert (= *Within-Subjects Design*). Die 4er-Sets wurden nacheinander in zufälliger Reihenfolge gezeigt, jedoch geblockt nach Geschlecht, d. h. entweder zuerst alle männlichen Bildersets und danach die weiblichen oder umgekehrt. Die Versuchspersonen wurden instruiert, die jeweils 4 Varianten eines Gesichts nach ihrer Attraktivität in eine Rangreihenfolge zu bringen, und zwar beginnend mit dem attraktivsten Gesicht (= Platz 1) bis hin zum unattraktivsten Gesicht (= Platz 4). Gleiche Ränge durften nicht vergeben werden. Wenn die Versuchspersonen bestimmte Varianten als gleich attraktiv empfanden oder gar keinen Unterschied im Aussehen wahrnehmen konnten, sollten sie einfach raten und die Bilder per Zufall einzelnen Plätzen zuordnen.

Der Versuchsleiter notierte die Rangreihenfolge und kodierte Rang 1 mit dem Wert 1, Rang 2 mit 2, Rang 3 mit 3 und Rang 4 mit 4. Für die spätere Datenauswertung wurde die Kodierung jedoch umgekehrt, so dass hohe Zahlen hohe Attraktivität kodierten und niedrige Zahlen niedrige Attraktivität. Dies wurde so gemacht, weil dies die Interpretierbarkeit der Ergebnisse erleichtert, da auch bei allen anderen Untersuchungen dieser Arbeit hohe Attraktivität durch hohe Zahlen kodiert ist. Für die Datenerhebung wurde jedoch die nutzerfreundlichste Kodierung gewählt, da bei einer Bildung von Rangreihenfolgen es leichter zu merken ist, dass die attraktivste Person auf "Platz 1" ist, die zweitattraktivste auf „Platz 2“ usw., ähnlich wie bei einem Schönheitswettbewerb. Bei der Beurteilung der einzelnen Varianten gab es keinerlei Zeitbeschränkung.

8.1.2 Ergebnisse

Für die statistische Analyse wurden die Daten von Männergesichtern und Frauengesichtern getrennt analysiert. Bei den Männergesichtern wurde für jede der sechs männlichen Stimuluspersonen überprüft, ob sich die mittleren Ränge der vier Varianten signifikant voneinander unterschieden. Dazu wurden sechs Friedman-Tests gerechnet. Tabelle 21 listet die Ergebnisse auf.

Tabelle 21: Die mittleren Ränge für die einzelnen Varianten (Durchschnitt, Kindchenschema, Symmetrie und Original) bei den einzelnen männlichen Stimulusgesichtern („m1“ bis „m6“). Tests nach Friedman. In allen Fällen $N = 324$ und $df = 3$.

	Durchschnitt	Kindchenschema	Symmetrie	Original	χ^2	p
m1	3.35	3.17	1.77	1.71	455.09	<.001
m2	3.63	2.89	1.79	1.69	505.577	<.001
m3	3.23	1.99	2.58	2.20	172.622	<.001
m4	2.31	2.15	2.90	2.64	65.167	<.001
m5	2.87	2.59	2.39	2.15	53.937	<.001
m6	3.07	2.45	2.35	2.14	93.245	<.001
Median	3.50	2.59	2.12	1.79		

Anmerkung: Der hier angegebene Median basiert *nicht* auf den jeweils nur sechs mittleren Rängen für „m1“ bis „m6“. Stattdessen wurde für jede Versuchsperson und für jede Variante der Median über alle sechs Stimuluspersonen berechnet und anschließend für jede Variante der Median über alle 324 Versuchspersonen.

Bei allen sechs männlichen Stimuluspersonen ergaben die Tests ein hoch signifikantes Ergebnis. In fast allen Fällen (Ausnahme: „m4“) schnitten die verdurchschnittlichten Varianten attraktiver ab als sämtliche andere Varianten.

Im Anschluss an die Friedman-Tests wurden als Post-hoc-Tests für jede Stimulusperson paarweise Vergleiche in Form von Wilcoxon-Tests gerechnet. Es zeigte sich, dass sich nahezu alle Varianten in ihren mittleren Attraktivitätsrängen signifikant voneinander unterscheiden (in den allermeisten Fällen $p < .001$). Unter den insgesamt 36 Post-hoc-Tests (6 mögliche Paarvergleiche x 6 Stimuluspersonen) gab es lediglich 4 Ausnahmen: Bei Gesicht „m1“ und „m2“ unterschieden sich Originalvariante und Symmetrie-Variante nicht voneinander, bei Gesicht „m4“ die Durchschnittsvariante nicht signifikant von der Kindchenschema-Variante und bei Gesicht „m6“ die Kindchenschema-Variante nicht von der symmetrisierten Variante.

Für einen besseren Überblick über die Ergebnisse wurden die Daten der sechs männlichen Stimuluspersonen zusammengefasst, indem jeweils der Median über Stimuli einer Variante (z. B. alle verdurchschnittlichten Gesichter) berechnet wurde. Die folgende Grafik illustriert die mittleren Attraktivitätsränge für die zusammengefassten Ergebnisse aller sechs männlichen Stimulusgesichter.

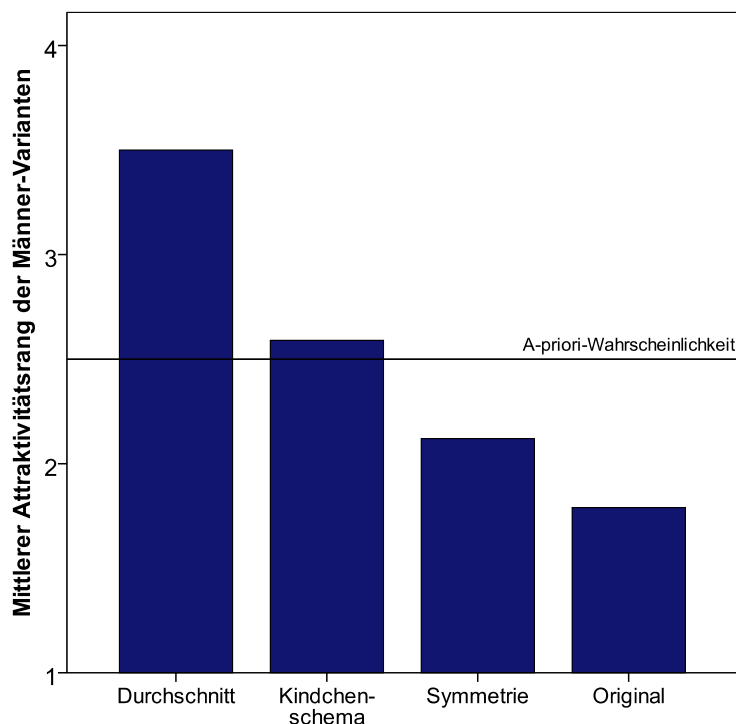


Abbildung 72: Der mittlere Attraktivitätsrang für die einzelnen Varianten der Männergesichter. Für die Grafik wurden die Ergebnisse zuvor durch Bildung von Medianen pro Versuchsperson und pro Variante über alle sechs Stimuluspersonen zusammengefasst. $N = 324$. Ein hoher Wert bedeutet hohe Attraktivität. Die verdurchschnittlichten Männergesichter werden am besten und die Originalgesichter am schlechtesten bewertet.

An den Ergebnissen fällt auf, dass die verdurchschnittlichten Männergesichter im direkten Vergleich zu den anderen Varianten am besten abschnitten. Die verkindlichten Männer-Varianten, die nach der Theorie des sexuellen Dimorphismus eigentlich am unattraktivsten (auch unattraktiver als die Originalgesichter) abschnitten sollten, wurden insgesamt weder besser noch schlechter bewertet, sondern so, wie nach der A-priori-Wahrscheinlichkeit (mittlerer Attraktivitätsrang von 2.5) zu erwarten. Kindlich-feminine

Männergesichter wurden jedoch immer noch besser bewertet als Gesichter, die „nur“ symmetrisch optimiert worden waren.



Abbildung 73: Varianten von links nach rechts: Durchschnitt, Kindchenschema, Symmetrie und Original. Stimulusperson „m4“ stellte eine Ausnahme dar. Hier wurden die symmetrische Variante und das Original am besten bewertet, die verdurchschnittlichte und die verkindlichte Variante signifikant schlechter.

Nachdem für alle sechs Stimuluspersonen separate Friedman-Tests gerechnet waren und die Ergebnisse gezeigt hatten, dass ein klarer Trend erkennbar war – nämlich die Durchschnittsvarianten als die attraktivsten Bilder, die Originalgesichter als die unattraktivsten Bilder und die symmetrisierten Gesichter als die zweitunattraktivsten – wurde für die zusammengefassten Daten (deren mittlere Ränge in den beiden Säulendiagrammen dargestellt sind) noch einmal ein Friedman-Test gerechnet. Die Ergebnisse sind in Tabelle 24 (S. 326) zusammengefasst.

In anschließenden Wilcoxon-Tests für die zusammengefassten Daten zeigte sich, dass alle möglichen paarweisen Vergleiche (insgesamt sechs mögliche paarweise Kombinationen, z. B. verdurchschnittlicht versus verkindlicht, verdurchschnittlicht versus symmetriert usw.) sich voneinander statistisch hochsignifikant unterschieden. Die sechs Testwerte lagen zwischen $Z = -5.025$ und $Z = -13.587$ (jeweils $p < .001$, vgl. Tabelle 22).

Tabelle 22: Paarweise Tests nach Wilcoxon für Attraktivitätsunterschiede zwischen den jeweiligen Gesichtervarianten der Männergesichter. Z-Werte und p -Werte. $N = 324$.

Männervarianten	Durchschnitt	Kindchen- schema	Symmetrie	Original
Durchschnitt	–	–	–	–
Kindchen- schema	-9.847 <.001	–	–	–
Symmetrie	-13.195 <.001	-5.324 <.001	–	–
Original	-13.587 <.001	-8.188 <.001	-5.025 <.001	–

Anmerkung: Die Daten wurden für jede Variante durch die Bildung von Medianen über alle sechs Stimulusgesichter zusammengefasst.

Die statistische Analyse der Frauengesichter-Varianten erfolgte analog zur Analyse der Männergesichter. Zuerst wurde für jede der sechs weiblichen Stimuluspersonen überprüft, ob sich die mittleren Ränge der vier Varianten signifikant voneinander unterscheiden. Dazu wurden sechs Friedman-Tests gerechnet. Tabelle 23 listet die Ergebnisse auf.

Tabelle 23: Die mittleren Ränge für die einzelnen Varianten (Durchschnitt, Kindchenschema, Symmetrie und Original) bei den einzelnen weiblichen Stimulusgesichtern („w1“ bis „w6“). Tests nach Friedman. In allen Fällen $N = 324$ und $df = 3$.

	Durchschnitt	Kindchenschema	Symmetrie	Original	χ^2	p
w1	3.33	2.74	2.11	1.82	236.36	<.001
w2	2.62	2.95	2.35	2.08	80.102	<.001
w3	3.72	2.81	1.79	1.68	534.374	<.001
w4	3.52	2.40	2.09	1.99	286.878	<.001
w5	2.89	2.71	2.24	2.06	105.187	<.001
w6	2.81	2.90	2.20	2.09	101.600	<.001
Median	3.52	2.88	1.91	1.69		

Anmerkung: Der hier angegebene Median basiert *nicht* auf den jeweils nur sechs mittleren Rängen für „w1“ bis „w6“. Stattdessen wurde für jede Versuchsperson und für jede Variante der Median über alle sechs Stimuluspersonen berechnet und anschließend für jede Variante der Median über alle 324 Versuchspersonen.

Bei allen sechs weiblichen Stimuluspersonen ergaben die Tests ein hoch signifikantes Ergebnis. In der einen Hälfte der Fälle („w1“, „w4“ und „w5“) schnitten die verdurchschnittlichten Varianten attraktiver ab als sämtliche anderen Varianten, in der anderen Hälfte der Fälle („w2“, „w3“ und „w6“) waren die verkindlichten Varianten die attraktivsten. Die Originalvarianten wurden in allen Fällen am schlechtesten beurteilt.

Im Anschluss an die Friedman-Tests wurden als Post-hoc-Tests für jede Stimulusperson paarweise Vergleiche in Form von Wilcoxon-Tests gerechnet. Es zeigte sich, dass sich nahezu alle Varianten in ihren mittleren Attraktivitätsrängen signifikant voneinander unterscheiden (in den allermeisten Fällen $p < .001$). Lediglich bei drei weiblichen Stimuluspersonen („w3“, „w4“ und „w6“) gab es keine signifikanten Unterschiede zwischen der symmetrisierten Version und dem Originalgesicht (bei „w6“ zusätzlich keinen Unterschied zwischen Durchschnitt und Kindchenschema).

Die folgende Grafik illustriert die mittleren Attraktivitätsränge für die zusammengefassten Ergebnisse aller sechs weiblichen Stimulusgesichter.

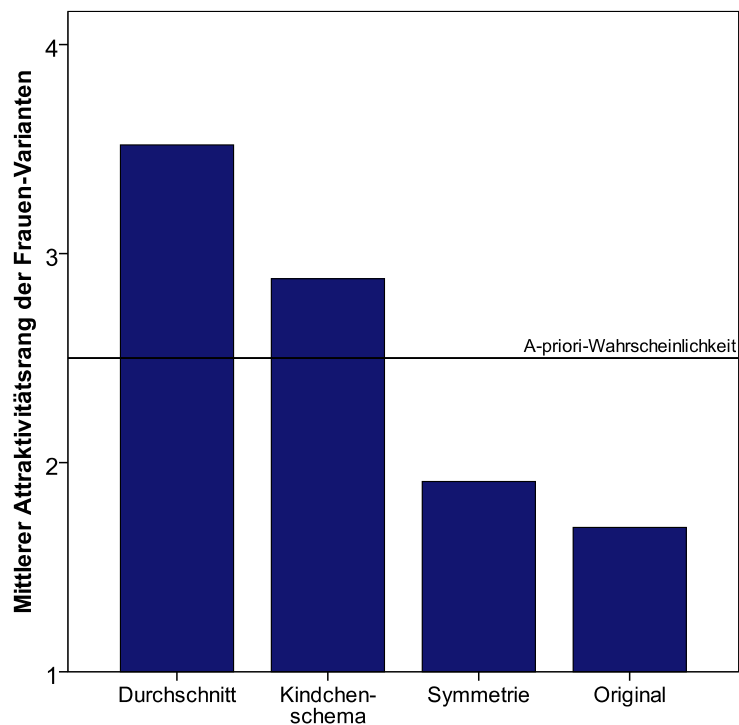


Abbildung 74: Der mittlere Attraktivitätsrang für die einzelnen Varianten der Frauengesichter. Für die Grafik wurden die Ergebnisse zuvor durch Bildung von Medianen pro Versuchsperson und pro Variante über alle sechs Stimuluspersonen zusammengefasst. $N = 324$. Ein hoher Wert bedeutet hohe Attraktivität. Die verdurchschnittlichten Frauengesichter werden am besten und die Originalgesichter am schlechtesten bewertet.



Abbildung 75: Varianten der Stimulusperson „w2“ von links nach rechts: Durchschnitt, Kindchenschema, Symmetrie und Original. Bei der Hälfte der weiblichen Stimuluspersonen wurde – so wie hier – die kindliche Variante am besten bewertet und die verdurchschnittlichte nur am zweitbesten.

Ein Blick auf das Säulendiagramm mit den zusammengefassten Daten zeigt, dass verdurchschnittlichte Frauengesichter im direkten Vergleich mit den anderen Varianten insgesamt am besten bewertet wurden. Bei der Bevorzugung der Durchschnittlichkeit ist das Ergebnis mit den Männergesichtern vergleichbar. Die verkindlichten Varianten werden ebenfalls besser bewertet als nach der A-priori-Wahrscheinlichkeit zu erwarten. Da im Ranking-Experiment jeder Rang genau einmal vergeben werden musste und es keine gleichen Ränge gab, geht die höhere Bevorzugung der kindlichen Varianten zu Lasten der Bewertungen von Symmetrie- und Originalvarianten.

Auch für die zusammengefassten Daten ist der Friedman-Test statistisch hoch signifikant (vgl. Tabelle 24). Dies gilt auch für sämtliche paarweisen direkten Vergleiche in Form von Wilcoxon-Tests (vgl. Tabelle 25). Obwohl also bei drei Frauengesichtern die verkindlichten Gesichter bessere Rankings bekamen als die verdurchschnittlichten, schnitten insgesamt dennoch die durchschnittlichen Frauengesichter signifikant besser ab.

Tabelle 24: Die mittleren Ränge für die Varianten (Durchschnitt, Kindchenschema, Symmetrie und Original) bei weiblichen sowie männlichen Stimulusgesichtern. Tests nach Friedman. In beiden Fällen $N = 324$ und $df = 3$.

	Durchschnitt	Kindchenschema	Symmetrie	Original	χ^2	p
Männergesichter	3.50	2.59	2.12	1.79	350.788	<.001
Frauengesichter	3.52	2.88	1.91	1.69	453.666	<.001

Anmerkung: Die Daten wurden zusammengefasst, indem für jede Versuchsperson und jede Variante über alle sechs Stimuluspersonen eines Geschlechts der Median gebildet wurde.

Tabelle 25: Paarweise Tests nach Wilcoxon für Attraktivitätsunterschiede zwischen den jeweiligen Gesichtervarianten der Frauengesichter. Z-Werte und p -Werte. $N = 324$.

Frauenvarianten	Durchschnitt	Kindchenschema	Symmetrie	Original
Durchschnitt	–	–	–	–
Kindchenschema	-8.025 <.001	–	–	–
Symmetrie	-13.843 <.001	-11.242 <.001	–	–
Original	-14.218 <.001	-12.542 <.001	-4.216 <.001	–

Anmerkung: Die Daten wurden für jede Variante durch die Bildung von Medianen über alle sechs Stimulusgesichter zusammengefasst.

In einem weiteren Schritt wurde überprüft, ob die Bewertungen für die einzelnen Gesichtervarianten möglicherweise vom Geschlecht der Beurteiler beeinflusst waren. Hierzu

wurden die zusammengefassten Daten analysiert. Die folgenden beiden Tabellen listen die Ergebnisse auf.

Tabelle 26: Ergebnisse der vier Tests nach Mann-Whitney, mit denen für die zusammengefassten Daten der sechs *männlichen* Stimulusgesichter überprüft wurde, ob ein Zusammenhang zwischen dem mittleren Attraktivitätsrang einer Gesichtsvariante und dem Geschlecht der Beurteiler bestand.

Variante des Stimulusgesichts	Geschlecht der Urteiler	<i>N</i>	Mittlerer Rang der Männergesichter	<i>Z</i>	<i>p</i>
Durchschnitt	männlich	161	165.66	-0.633	.526
	weiblich	163	159.38		
Kindchenschema	männlich	161	147.80	-2.881	.004
	weiblich	163	177.02		
Symmetrie	männlich	161	171.98	-1.910	.056
	weiblich	163	153.14		
Original	männlich	161	170.89	-1.662	.097
	weiblich	163	154.21		

Anmerkung: **Fett** gedruckte Ergebnisse sind statistisch signifikant.

Tabelle 27: Ergebnisse der vier Tests nach Mann-Whitney, mit denen für die zusammengefassten Daten der sechs weiblichen Stimulusgesichter überprüft wurde, ob ein Zusammenhang zwischen dem mittleren Attraktivitätsrang einer Gesichtsvariante und dem Geschlecht der Beurteiler bestand.

Variante des Stimulusgesichts	Geschlecht der Urteiler	N	Mittlerer Rang der Frauengesichter	Z	p
Durchschnitt	männlich	161	155.53	-1.432	.152
	weiblich	163	169.38		
Kindchenschema	männlich	161	157.50	-0.997	.319
	weiblich	163	167.44		
Symmetrie	männlich	161	170.80	-1.669	.095
	weiblich	163	154.30		
Original	männlich	161	165.61	-0.624	.532
	weiblich	163	159.42		

Wie aus Tabelle 26 und Tabelle 27 zu entnehmen ist, gibt es bei fast allen Gesichtsvarianten keine Unterschiede in Abhängigkeit vom Geschlecht der Versuchspersonen. Die einzige Ausnahme bilden die Männergesichter, die an die Proportionen des Kindchenschemas angenähert wurden. Verkindlichte (femininisierte) Männergesichter wurden von Frauen attraktiver bewertet als von Männern. Dieser Befund ist v. a. vor dem Hintergrund solcher Studien interessant, die sexuellen Dimorphismus bei Männergesichtern untersucht haben, hierfür jedoch nur Frauen als Versuchspersonen (und keine Männer als Kontrollgruppe) verwendet haben (z. B. Scheib et al., 1999, Johnston et al., 2001, Little et al., 2001, Swaddle & Reiersen, 2002, Neave et al., 2003). Denn die Entscheidung, nur das entgegengesetzte Geschlecht zur Attraktivität von Gesichtern zu untersuchen, erfolgt ja in der Regel mit der Grundannahme, dass Attraktivitätsurteile etwas mit Partnerwahl und Fortpflanzung zu tun haben. Wenn sich dann herausstellt, dass gerade dieses Geschlecht (Frauen) offensichtlich nicht so empfindet, wie die Theorie vorhersagt, das andere Geschlecht (Männer) jedoch tendenziell noch eher (obwohl es keinen Zusammenhang mit Partnerwahl und Fortpflanzung gibt), dann stellt dies die biologische Sichtweise auf die Attraktivitätswahrnehmung zweifach in Frage.

Zuletzt wurde noch überprüft, ob das Alter der Versuchspersonen einen Einfluss auf die Präferenz für einzelne Gesichtsvarianten hatte. Dazu wurde die untersuchte Versuchspersonenstichprobe in drei Alterskategorien unterteilt: 16 bis 25 Jahre ($N = 160$), 26 bis 40 Jahre ($N = 64$) und 41 Jahre und älter ($N = 100$). Es wurden für die männlichen Stimulusgesichter mit den zusammengefassten Daten vier Kruskal-Wallis-Tests gerechnet, um zu überprüfen, ob die mittleren Attraktivitätsränge für die Gesichtsvarianten in Abhängigkeit von der Altersklasse der Beurteiler signifikant verschieden waren. Dasselbe wurde für die weiblichen Stimulusgesichter durchgeführt. Es zeigten sich jedoch keinerlei signifikante Unterschiede; die p -Werte lagen zwischen .100 und .741. Das Alter der Versuchspersonen hat also keinen Einfluss auf die Präferenzen für einzelne Gesichtsvarianten.

Zusammenfassend kann festgehalten werden, dass sich sowohl bei Männergesichtern als auch bei Frauengesichtern klare Attraktivitätspräferenzen nachweisen ließen. Wurden den Versuchspersonen alle Varianten eines Stimulusgesichts gleichzeitig präsentiert, dann erhielten die Gesichter, deren Proportionen mit Morphing-Software durchschnittlicher gemacht waren, die höchsten Attraktivitätsränge. Es folgten die verkindlichten Gesichter als die zweitattraktivsten Varianten. Erst an dritter Stelle standen die symmetrisch optimierten Gesichter. Die unveränderten Originalgesichter hingegen wurden im direkten Vergleich am schlechtesten bewertet. Diese Rangreihenfolge war für männliche und weibliche Stimulusgesichter gleich.

8.2 Experiment zum direkten experimentellen Hypothesen-Vergleich mit Männer- und Frauengesichtern im Rating-Verfahren und Between-Subjects Design (Untersuchung *Vergleich 2*)

8.2.1 Methode

8.2.1.1 Versuchspersonen

An der Untersuchung *Vergleich 2* nahmen insgesamt 468 Versuchspersonen teil. 238 Personen waren männlich (50.9 %), 230 weiblich (49.1 %). 207 Personen (44.2 %) waren Studierende, jedoch befanden sich keine Psychologiestudierenden in der Stichprobe. Die jüngste Person war 13 Jahre, die älteste 90. Der Altersdurchschnitt betrug 31.7 Jahre ($SD = 15.6$). Um eine für die Bevölkerung möglichst repräsentative Stichprobe zu erhalten, fanden die Datenerhebungen im öffentlichen Raum statt, z. B. an Bahnhöfen, Fußgän-

gerzonen und Einkaufszentren. Die Teilnahme am Experiment erfolgte freiwillig und wurde nicht entlohnt.

8.2.1.2 Stimulusmaterial und Versuchsdurchführung

Das Stimulusmaterial war identisch mit dem Stimulusmaterial von Untersuchung *Vergleich 1*. Der Unterschied zum vorhergehenden Experiment liegt in der Art und Weise der Erhebung der Attraktivitätsurteile. Im *vorangegangenen* Experiment gaben die Versuchspersonen Rangurteile ab, verglichen dabei stets alle vier Varianten einer Stimulusperson direkt miteinander und wurden zu Präferenzurteilen „gezwungen“ (da gleiche Rangplätze nicht erlaubt waren). Daher lieferte die Untersuchung *Vergleich 1* lediglich *relative* Attraktivitätsurteile.

Die Untersuchung *Vergleich 2* hingegen liefert *absolute* Attraktivitätsurteile, denn hier wurde Attraktivität über Ratingskalen (von 1 = „sehr unattraktiv“ bis 7 = „sehr attraktiv“, analog zu den Untersuchungen *Durch 1* bis *6* sowie *Sym 2*, *Sym 4* und *Sym 6*) gemessen. Zusätzlich wurde ein *Between-Subjects Design* gewählt, d. h. dass jeder Versuchsperson stets nur *eine einzige* Variante einer Stimulusperson präsentiert wurde. Von der Existenz weiterer Varianten derselben Stimulusperson ahnten die Probanden nichts.

Insgesamt gab es vier unabhängige Gruppen von Versuchspersonen – so viele wie Versuchsbedingungen. Da in jeder Versuchspersonengruppe von jeder Stimulusperson nur eine Variante gezeigt und bewertet wurde, sah jede Versuchsperson nur insgesamt 12 Bilder (so viele, wie es Stimuluspersonen gab). Ein Beispiel: Die verdurchschnittlichte Variante des Männergesichts „m1“ wurde nur der Versuchspersonengruppe 1 gezeigt, die verkindlichte Variante wurde nur der Gruppe 2 gezeigt, die symmetrisierte nur Gruppe 3 und die Originalvariante dieses Männergesichts nur Gruppe 4. Beim Männergesicht „m2“ sahen die Gruppen dann jeweils eine andere Variante usw.

Das Versuchsdesign war vollständig ausbalanciert: Jede Versuchspersonengruppe sah jede Art der Manipulation gleich oft (nämlich 3 Mal) und jedes Bild wurde von gleich vielen Versuchspersonen beurteilt (nämlich von 117 Personen). Tabelle 28 zeigt einen Ausschnitt aus dem Versuchsplan und illustriert das Prinzip.

Tabelle 28: Ausschnitt aus dem Versuchsplan des Experiments *Vergleich 2*. Insgesamt gab es vier unabhängige Gruppen von Versuchspersonen – so viele wie Versuchsbedingungen. Jede Gruppe sah von einer Stimuluspersion stets nur eine Variante. Das Versuchsdesign war vollständig ausbalanciert.

	Stimuluspersion m1				Stimuluspersion m2				usw.
	m1d	m1k	m1s	m1o	m2d	m2k	m2s	m2o	...
Vpn-Gruppe 1	X					X			...
Vpn-Gruppe 2		X					X		...
Vpn-Gruppe 3			X					X	...
Vpn-Gruppe 4				X	X				...

Anmerkung: „X“ bedeutet, dass dieser Versuchspersonengruppe ein bestimmter Stimulus präsentiert wurde. Der Zusatz „d“, „k“, „s“ oder „o“ zeigt an, auf welche Art ein Stimulus-gesicht manipuliert wurde (d = durchschnittlich, k = verkindlicht, s = symmetrisch optimiert und o = original). Weitere Erläuterungen im Text.

Wozu ein so kompliziertes Versuchsdesign? Gewiss wäre es einfacher gewesen, ein *Between-Subjects Design* so zu entwerfen, dass die vier Experimentalgruppen nach der Art der manipulierten Bilder unterteilt sind, die sie zu sehen bekommen. So hätte man einer Gruppe „Durchschnitt“ nur verdurchschnittlichte Gesichter präsentieren können, einer Gruppe „Kindchenschema“ nur verkindlichte Gesichter usw. Dies hätte jedoch die Gefahr von Adaptationseffekten mit sich gebracht. Angenommen, eine bestimmte Variante wäre besonders attraktiv (z. B. die durchschnittliche), dann bestünde das Risiko, dass die Versuchspersonen an das höhere Attraktivitätsniveau adaptieren und insgesamt strengere Maßstäbe anlegen, insgesamt also schlechtere Ratings vergeben (= Kontext-Effekt), als wenn man dieselben Gesichter in einem Kontext von normalen, unveränderten Gesichtern präsentieren würde (Haubensak, 1985; Henss, 1993; Higgins & Stangor, 1988; Kenrick & Gutierrez, 1980; Kenrick et al., 1989). Zum anderen bestünde jedoch auch das Risiko, dass die Versuchspersonen an die Gesichtsproportionen dieser gezeigten Variante adaptieren, dass Ihnen also z. B. verkindlichte Männergesichter ganz normal erscheinen. Und drittens bestünde das Risiko, dass die Versuchspersonen vor lauter Gesichtern mit stets denselben Gesichtsproportionen das, was in diesem Experiment eigentlich von Interesse ist – nämlich die Proportionen – mehr oder weniger stark ignorie-

ren, da sie ja eine konstante Größe sind, und stattdessen auf die Unterschiede zwischen den Stimuluspersonen stärker fokussieren. Merkmale wie die Beschaffenheit der Haut oder die Farbe von Augen oder Haaren würden dann viel stärker ins Gewicht fallen. Um diese Probleme auszuschließen, wurde der kompliziertere ausbalancierte Versuchsplan gewählt.

Jede Versuchsperson beurteilte die 12 präsentierten Gesichter auf der siebenstufigen Ratingskala (von 1 = „sehr unattraktiv“ bis 7 = „sehr attraktiv“). Während der Beurteilung lag vor der Versuchsperson ein Ausdruck dieser Skala, so dass die Bedeutung der einzelnen Skalenstufen nicht im Gedächtnis behalten werden musste, sondern stets im Blickfeld war. Die Gesichter wurden wie im vorangegangenen Experiment auf Fotopapier belichtet präsentiert (in einer Größe von 10 cm x 13 cm bei einer Auflösung von 1200 x 1600 Bildpunkten). Vor jedem Versuch mischte der Versuchsleiter den Stapel mit den 12 Fotos, so dass jeder Versuchsperson die Gesichter in einer zufälligen Reihenfolge gezeigt wurden. Die Antworten der Probanden wurden vom Versuchsleiter notiert. Bei der Beurteilung der Gesichtsattraktivität gab es keinerlei zeitliche Beschränkungen.

8.2.2 Ergebnisse

Für die statistische Analyse wurden die Daten von Männergesichtern und Frauengesichtern getrennt analysiert. Bei den Männergesichtern wurde für jede der sechs männlichen Stimuluspersonen überprüft, ob sich die Mittelwerte der Attraktivitäts-Ratings für die vier Varianten signifikant voneinander unterschieden. Dazu wurden sechs einfaktorielle ANOVAs gerechnet. Tabelle 29 listet die Ergebnisse auf.

Tabelle 29: Die mittleren Attraktivitätsbewertungen für die Varianten der *Männern*gesichter mit den Testergebnissen der einfaktoriellen ANOVAs. Die drittletzte Spalte listet als Effektgröße Cohen's *d* auf und vergleicht jeweils die Attraktivität einer manipulierten Variante mit der Attraktivität der unveränderten Originalvariante.

Stimulusperson	Variante	<i>N</i>	<i>M</i>	<i>SD</i>	Cohen's <i>d</i> zu Originalversion	<i>F</i>	<i>p</i>
m1	Durchschnitt	117	2.08	1.10	0.17	2.401	.067
	Kindchenschema	117	1.74	0.77	-0.17		
	Symmetrie	117	1.85	1.01	-0.05		
	Original	117	1.90	1.07			
	Gesamt	468	1.89	1.00			
m2	Durchschnitt	117	2.94	1.11	0.88	17.820	<.001
	Kindchenschema	117	2.59	1.04	0.55		
	Symmetrie	117	2.21	1.01	0.16		
	Original	117	2.06	0.87			
	Gesamt	468	2.45	1.07			
m3	Durchschnitt	117	3.79	1.17	0.75	16.354	<.001
	Kindchenschema	117	2.97	1.19	0.02		
	Symmetrie	117	2.97	1.02	0.02		
	Original	117	2.95	1.07			
	Gesamt	468	3.17	1.17			

Fortsetzung

Fortsetzung von Tabelle 29

Stimulus-person	Variante	<i>N</i>	<i>M</i>	<i>SD</i>	Cohen's <i>d</i> zu Original- version	<i>F</i>	<i>p</i>
m4	Durchschnitt	117	4.03	1.16	-0.26	3.351	.019
	Kindchenschema	117	4.23	1.12	-0.08		
	Symmetrie	117	4.48	1.12	0.15		
	Original	117	4.32	1.05			
	Gesamt	468	4.26	1.12			
m5	Durchschnitt	117	4.08	1.38	0.11	0.747	.524
	Kindchenschema	117	3.81	1.25	-0.09		
	Symmetrie	117	3.92	1.43	-0.01		
	Original	117	3.93	1.38			
	Gesamt	468	3.94	1.36			
m6	Durchschnitt	117	4.44	1.12	0.22	3.797	.010
	Kindchenschema	117	3.94	1.15	-0.21		
	Symmetrie	117	4.09	1.17	-0.08		
	Original	117	4.19	1.16			
	Gesamt	468	4.16	1.16			

Anmerkung: **Fett** gedruckte Effektgrößen (Cohen's *d*) sind statistisch signifikant.

Bei vier von sechs männlichen Stimuluspersonen ergaben sich signifikante Mittelwertsunterschiede ($p < .05$). Bei Gesicht „m1“ waren die Mittelwertsunterschiede nur knapp nicht signifikant ($p = .067$), bei Gesicht „m5“ gab es keinerlei Unterschiede ($p = .524$). Im An-

schluss an die ANOVAs wurden Post-hoc-Tests nach Scheffé gerechnet, um zu prüfen, welche Varianten sich genau von welchen anderen signifikant unterschieden. Die allermeisten paarweisen Tests erwiesen sich als nicht signifikant. Die häufigsten signifikanten Unterschiede bestanden zwischen der verdurchschnittlichten Version und der Originalversion bzw. der meist sehr ähnlich bewerteten symmetrisierten Version. Auf einen vollständigen Abdruck aller Post-hoc-Tests wurde daher verzichtet. Stattdessen sind in Tabelle 29 die Effektgrößen (Cohen's d) durch Fettdruck hervorgehoben, wenn der Mittelwertsunterschied zwischen einer manipulierten Variante und dem Originalgesicht als Kontrollbedingung statistisch signifikant war ($p < .05$). Als Faustregel lässt sich sagen, dass ein Mittelwertsunterschied bei einem Cohen's d größer als 0.42 statistisch signifikant wurde. Dies entspricht einem Mittelwertsunterschied von ungefähr 0.6 Punkten auf der Ratingskala von 1 bis 7.

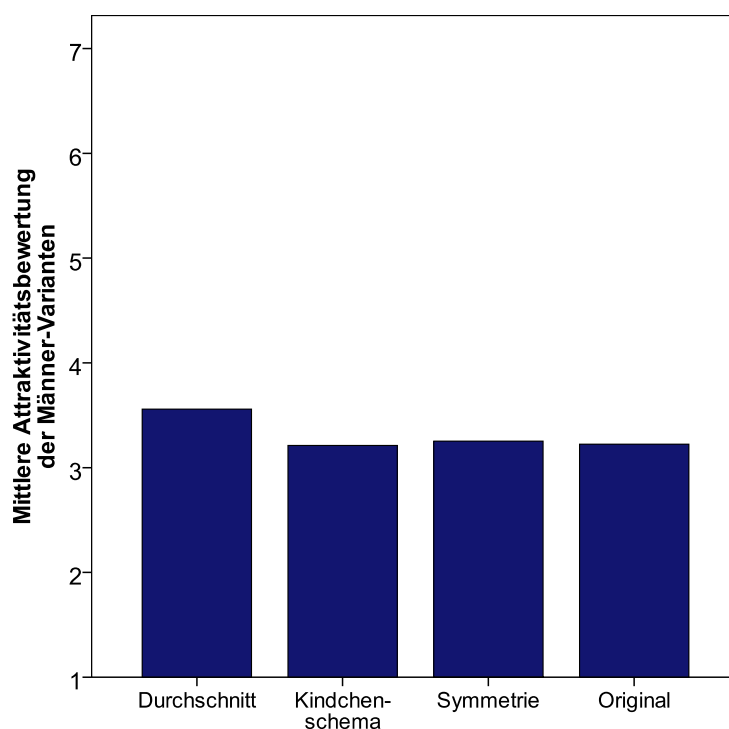


Abbildung 76: Die durchschnittlichen Attraktivitätsratings für die einzelnen Varianten der Männergesichter. Für die Grafik wurden die Attraktivitätsbewertungen pro Variante über alle sechs Stimuluspersonen zusammengefasst. $N = 468$ Versuchspersonen, n pro Variante = 117 Versuchspersonen. Die verdurchschnittlichten Männergesichter wurden am besten bewertet, zwischen den anderen drei Varianten bestand kaum ein Attraktivitätsunterschied. Keine Variante erreichte hohe durchschnittliche Attraktivitätsbewertungen.

In Tabelle 29 sind auch Effektgrößen in Form eines Cohen's d angegeben. Dieser Wert bezeichnet den Attraktivitätsunterschied zwischen einer durch Morphing-Software manipulierten Variante und dem jeweiligen Originalgesicht als Kontrollbedingung. Die stärksten Effekte finden sich in der Regel zwischen der verdurchschnittlichten Variante und dem Originalgesicht; sie reichen dort von $d = 0.11$ und $d = 0.88$ (bei Gesicht „m4“ ist dieser Effekt jedoch negativ, d. h. das verdurchschnittlichte Gesicht war unattraktiver als das Original, $d = -0.26$).

Die statistische Analyse der Frauengesichter-Varianten erfolgte analog zur Analyse der Männergesichter. Auch hier wurden für alle weiblichen Stimuluspersonen sechs ANOVAs gerechnet, um zu überprüfen, ob sich die Attraktivitätsratings für die einzelnen Varianten signifikant voneinander unterscheiden. Tabelle 30 listet die Ergebnisse auf.

Tabelle 30: Die mittleren Attraktivitätsbewertungen für die Varianten der *Frauengesichter* mit den Testergebnissen der einfaktoriellen ANOVAs. Die drittletzte Spalte listet als Effektgröße Cohen's *d* auf und vergleicht jeweils die Attraktivität einer manipulierten Variante mit der Attraktivität der unveränderten Originalvariante.

Stimulus-person	Variante	<i>N</i>	<i>M</i>	<i>SD</i>	Cohen's <i>d</i> zu Original- version	<i>F</i>	<i>p</i>
w1	Durchschnitt	117	4.21	1.24	0.63	9.685	<.001
	Kindchenschema	117	3.99	1.24	0.46		
	Symmetrie	117	3.69	1.11	0.24		
	Original	117	3.40	1.31			
	Gesamt	468	3.82	1.26			
w2	Durchschnitt	117	4.38	1.21	-0.31	10.390	<.001
	Kindchenschema	117	4.93	1.19	0.17		
	Symmetrie	117	4.16	1.17	-0.51		
	Original	117	4.74	1.09			
	Gesamt	468	4.55	1.20			
w3	Durchschnitt	117	4.54	1.04	1.03	19.019	<.001
	Kindchenschema	117	3.80	1.28	0.30		
	Symmetrie	117	3.68	1.24	0.20		
	Original	117	3.45	1.08			
	Gesamt	468	3.87	1.23			

Fortsetzung

Fortsetzung von Tabelle 30

Stimulus-person	Variante	<i>N</i>	<i>M</i>	<i>SD</i>	Cohen's <i>d</i> zu Original- version	<i>F</i>	<i>p</i>
w4	Durchschnitt	117	4.77	1.12	0.21	4.714	.003
	Kindchenschema	117	4.21	1.28	-0.27		
	Symmetrie	117	4.36	1.25	-0.14		
	Original	117	4.53	1.16			
	Gesamt	468	4.47	1.22			
w5	Durchschnitt	117	4.86	1.16	0.91	17.503	<.001
	Kindchenschema	117	4.45	1.15	0.58		
	Symmetrie	117	4.43	1.18	0.55		
	Original	117	3.74	1.30			
	Gesamt	468	4.37	1.26			
w6	Durchschnitt	117	4.51	1.25	0.31	18.115	<.001
	Kindchenschema	117	4.74	1.27	0.50		
	Symmetrie	117	3.63	1.28	-0.43		
	Original	117	4.15	1.09			
	Gesamt	468	4.26	1.29			

Anmerkung: **Fett** gedruckte Effektgrößen (Cohen's *d*) sind statistisch signifikant.

Bei allen sechs weiblichen Stimuluspersonen ergaben sich signifikante Mittelwertsunterschiede in der Attraktivität der einzelnen Gesichtervarianten (jeweils $p < .01$). Anschließend wurden Post-hoc-Tests nach Scheffé gerechnet. Wie die fett hervorgehobenen Effektgrößen in Tabelle 30 verdeutlichen, zeigte sich dabei, dass

- bei drei Stimuluspersonen die Durchschnittsvariante attraktiver als das jeweilige Originalgesicht war,
- bei drei Stimuluspersonen die verkindlichte (femininisierte) Variante attraktiver als das Originalgesicht war und
- die symmetrisierte Variante in einem Fall attraktiver und in zwei Fällen unattraktiver als das Originalgesicht war.

Beim direkten Vergleich „Durchschnitt“ versus „Kindchenschema“ schnitt zweimal die verdurchschnittlichte und einmal die verkindlichte Variante signifikant besser ab – in den übrigen Fällen bestand kein Unterschied. Auch bei den Frauengesichtern gilt als Faustregel, dass ein Mittelwertsunterschied bei einem Cohen's d größer als 0.42 statistisch signifikant wurde. Dies entspricht einem Mittelwertsunterschied von ungefähr 0.6 Punkten auf der Ratingskala von 1 bis 7. Abbildung 77 veranschaulicht die Attraktivitätsratings der einzelnen Gesichtervarianten. Die Daten aller Stimuluspersonen wurden dazu zu je einem Mittelwert pro Variante zusammengefasst.

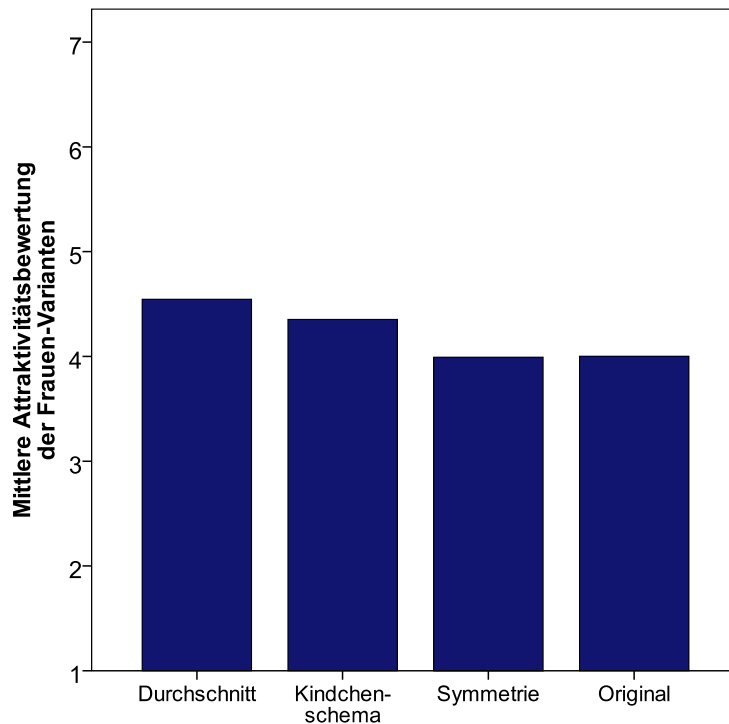


Abbildung 77: Die durchschnittlichen Attraktivitätsratings für die einzelnen Varianten der Frauengesichter. Für die Grafik wurden die Attraktivitätsbewertungen pro Variante über alle sechs Stimuluspersonen zusammengefasst. $N = 468$ Versuchspersonen, n pro Variante = 117 Versuchspersonen. Die verdurchschnittlichten Frauengesichter wurden am besten bewertet, zwischen den symmetrisch optimierten Varianten und den Originalgesichtern bestand kein Attraktivitätsunterschied. Keine Variante erreichte hohe durchschnittliche Attraktivitätsbewertungen.

Man erkennt an diesem Säulendiagramm, dass (wie auch bei den Männergesichtern) die verdurchschnittlichten Frauengesichter insgesamt am besten bewertet wurden, während zwischen Symmetrie-Version und Original kein Attraktivitätsunterschied besteht. Dazwischen liegen die verkindlichten Gesichter, die bei den Frauengesichtern – konform mit der Theorie des sexuellen Dimorphismus – besser bewertet wurden als die Originalgesichter, während sie bei den Männergesichtern praktisch gleich bewertet wurden.

Frauengesichter wurden insgesamt attraktiver bewertet als Männergesichter, d. h. alle Attraktivitätsratings sind bei den Frauengesichtern (auch bei den Originalen) auf der y-Achse ein Stück nach oben verschoben (ca. 0.8 Skalenpunkte). Dieser Befund, der die Floskel von Frauen als dem schöneren Geschlecht zu bestätigen scheint, ist in der Attrak-

tivitätsforschung ein bekanntes Phänomen und zeigte sich hier bereits in den Experimenten zur Durchschnittshypothese *Durch 1* und *Durch 4* (vgl. Kapitel 2.3 und 2.7). Zumindest gilt dies, sofern man mit jungen Gesichtern und nicht mit älteren experimentiert, da sich bei älteren Gesichtern der Effekt umkehrt (= *double standard of aging*), wie zahlreiche Untersuchungen zeigen (Alley, 1988; Deutsch et al., 1986; England & McClintock, 2009; Falser et al., 2008; Giesen, 1989; Henss, 1992; Jones, 1995; Teuscher & Teuscher, 2007; Wilcox, 1997).

Trotz der im Vergleich zu den Männergesichtern höheren Bewertungen der Frauengesichter sind dennoch alle Attraktivitätsratings nur mäßig gut, v. a. wenn man bedenkt, dass darunter ja Gesichtsvarianten sind, die den drei getesteten Attraktivitätstheorien zufolge ja hoch attraktive Gesichtsproportionen besitzen sollten, da sie z. B. extrem durchschnittlich oder perfekt symmetrisch sind. Trotz teilweise signifikanter Mittelwertsunterschiede zwischen den Varianten ist dennoch keine Variante dabei, die insgesamt von den Versuchspersonen als sehr attraktiv angesehen würde. Selbst wenn man einzelne Gesichtstimuli betrachtet, so muss man konstatieren, dass selbst die attraktivsten Gesichtsvarianten lediglich durchschnittliche Attraktivitätswerte zwischen 4 (= mittelmäßig attraktiv) und 5 (= eher attraktiv) erhielten. Wirklich attraktive Gesichter (z. B. 6 = ziemlich attraktiv, 7 = sehr attraktiv) konnten bei den Manipulationen durch Morphing-Software in Richtung Durchschnitt, Kindchenschema oder perfekter Symmetrie offensichtlich nicht erzeugt werden.

Dieses Fazit gilt ausdrücklich auch für die verdurchschnittlichten Gesichtsvarianten, die im Vergleich zu den anderen Varianten am besten abgeschnitten haben. Trotz sehr durchschnittlicher Gesichtsproportionen erhielten sie auch nur mäßige Attraktivitätsbewertungen. Dies widerlegt die immer wieder von den Befürwortern der Durchschnittshypothese verbreiteten Behauptungen, durchschnittliche Gesichtsproportionen seien extrem attraktiv. So fassen z. B. Rubenstein, Langlois und Roggman im Standardwerk Attraktivitätsforschung von Rhodes und Zebrowitz (2002) im Kapitel über die Durchschnittshypothese den Stand der Forschung so zusammen: „These averaged faces and any individual faces that are close in configuration to the mathematical average are extreme, not average, in attractiveness“ (S. 11).

Zusammenfassend lässt sich feststellen, dass sich sowohl bei Männergesichtern als auch bei Frauengesichtern eine Präferenz für verdurchschnittlichte Gesichter nachweisen ließ. Bei Frauengesichtern schnitten insgesamt zudem verkindlichte Varianten besser ab als Originalgesichter, jedoch nicht ganz so gut wie verdurchschnittlichte Gesichter. Bei den

Männern Gesichtern hatte die Verkindlichung insgesamt keinen Effekt auf die Attraktivität, obwohl die Theorie des sexuellen Dimorphismus einen Attraktivitätsverlust vorhersagt. Eine symmetrische Optimierung des Gesichts hatte bei Männer- und Frauengesichtern gar keinen Einfluss auf die Attraktivitätsbewertung. Die Präsentation dieser Stimuli im nicht-reaktiven *Between-Subjects Design* und die Messung der Attraktivität durch Ratingskalen führten im Vergleich zu den Ergebnissen aus dem Ranking-Experiment (Untersuchung *Vergleich 1*) zu einem massiven Schrumpfen der beobachteten Effekte. Während im *Ranking-Experiment* fast alle Attraktivitätsunterschiede zwischen zwei beliebigen Gesichtervarianten statistisch signifikant waren, gab es (trotz der großen Stichprobe von 468 Probanden) beim *Rating-Experiment* – selbst im direkten Vergleich zwischen manipulierter Variante und Originalgesicht – nur noch sehr wenige signifikante Mittelwertsunterschiede. Ein weiteres Ergebnis ist, dass durch das Anpassen der Gesichtsproportionen in Richtung auf ein bestimmtes Schema offensichtlich gar keine hochattraktiven Gesichter erzeugt werden konnten, da selbst die attraktivsten Gesichtsvarianten bestenfalls durchschnittliche Bewertungen im Mittelfeld der verwendeten Attraktivitätsskala bekamen.

9 Zusammenfassende Diskussion aller Experimente zum direkten experimentellen Vergleich der Hypothesen „Durchschnitt“, „Symmetrie“ und „Sexueller Dimorphismus“

Das Ziel der beiden letzten von insgesamt 20 Untersuchungen dieser Arbeit war es, abschließend alle drei zuvor bereits ausführlich einzeln untersuchten Attraktivitätstheorien in einer Art Gesamtschau zu überprüfen. Denn was die ersten 18 Untersuchungen nicht leisten konnten, war der direkte Vergleich der einzelnen Hypothesen untereinander. Welche Hypothese erklärt im Vergleich zu den anderen am meisten? Macht die Durchschnittlichkeit, die Symmetrie oder der sexuelle Dimorphismus den größten Effekt für die Attraktivität eines Gesichts aus?

Die meisten früheren Studien beschränkten sich darauf, lediglich eine oder bestenfalls zwei Hypothesen auf einmal zu überprüfen. Da die Ergebnisse naturgemäß auch immer abhängig sind von der konkret verwendeten Methodik und dem konkreten Stimulusmaterial, ist es daher kaum möglich, die Effekte der einzelnen Hypothesen miteinander zu

vergleichen. Daher wurden in den letzten zwei Untersuchungen (*Vergleich 1* und *2*) alle drei Hypothesen *gleichzeitig* mit derselben Methodik und demselben Stimulusmaterial gegeneinander getestet, um die erzielten Effektgrößen direkt miteinander vergleichen und ihre relative Bedeutung abschätzen zu können.

Zusätzlich wurden auch zwei verschiedene Versuchsdesigns verwendet, nämlich einerseits das *Rankingverfahren* mit seinen direkten Vergleichsmöglichkeiten der einzelnen Gesichtervarianten untereinander und andererseits das weniger reaktive *Ratingverfahren* im *Between-Subjects Design*, bei dem absolute Attraktivitätsurteile erhoben wurden. Die Untersuchung der Effekte auf die Attraktivität durch zwei unabhängige Messverfahren (in getrennten Datenerhebungen mit unabhängigen Versuchspersonen-Kollektiven) erscheint angesichts der Erfahrungen aus den neun durchgeführten Untersuchungen zur Symmetriehypothese nicht nur sinnvoll, sondern sogar notwendig, da dort schließlich das zentrale Ergebnis war, dass die Untersuchungsmethodik (z. B. Versuchsdesign, Art der Datenerhebung, Paarvergleich versus Rating) einen immensen Einfluss auf die erzielten Ergebnisse hatte (vgl. Kapitel 5.2).

Bei der Diskussion der Ergebnisse dieser letzten beiden Experimente soll es daher *nicht* um den Vergleich dieser Ergebnisse mit Studien *anderer* Autoren gehen, denn all diese Vergleiche wurden bereits an früherer Stelle in dieser Arbeit gezogen, nämlich dort, wo die eigenen Ergebnisse zu einer einzelnen bestimmten Hypothese vor dem Hintergrund des aktuellen Forschungsstandes diskutiert wurden. Denn neu an den letzten beiden Experimenten ist lediglich die *gleichzeitige* Untersuchung der drei Attraktivitätstheorien. Durch sie besteht daher die Möglichkeit, nochmals ein neues Licht auf die bisherigen 18 durchgeführten Experimente zu werfen, ihre Ergebnisse miteinander zu vergleichen und die in den „Einzel-Hypothesen-Untersuchungen“ gezogenen Schlussfolgerungen gegebenenfalls zu relativieren. Anstatt des üblichen Vergleichs mit Studien anderer Forscher erfolgt daher hier der Vergleich mit den Ergebnissen der in dieser Arbeit durchgeführten 18 eigenen Untersuchungen.

Bei diesem Vergleich fällt auf, dass grundsätzlich sämtliche Ergebnisse der bisherigen „Einzeluntersuchungen“ bestätigt und untermauert werden. Im Folgenden soll dies näher erläutert werden.

Schlussfolgerungen zur Durchschnittshypothese

Hinsichtlich der Durchschnittlichkeit von Gesichtern bestätigen beide Untersuchungen (*Vergleich 1* und *Vergleich 2*) die Ergebnisse der sechs Untersuchungen zur Durch-

schnittshypothese (*Durch 1* bis *6*, Kapitel 2). Verdurchschnittlichte Gesichter schnitten insgesamt attraktiver ab als unveränderte Originalgesichter. Dies gilt sowohl für Männer als auch für Frauengesichter. Am besten vergleichbar sind die Ergebnisse mit den Experimenten *Durch 3* und *Durch 6*. Denn hier wurden Profilgesichter ohne Textur (= Schattenprofile) als Stimuli verwendet, bei denen nur die Gesichtsproportionen schrittweise durchschnittlicher gemacht wurden, während die Textur beibehalten wurde. Den Experimenten *Durch 3*, *Durch 6*, *Vergleich 1* und *Vergleich 2* ist damit gemeinsam, dass tatsächlich nur Gesichtsproportionen verdurchschnittlicht wurden, ohne dass es dabei zu den Morphing-typischen Artefakten der glatten, makellosen Haut kam, die das zentrale methodische Problem bei fast allen experimentellen Untersuchungen zur Durchschnittshypothese sind (vgl. Kapitel 3.1.2), und die auch in den Experimenten *Durch 1* und *2* sowie *Durch 4* und *5* nachgewiesen wurden (vgl. Kapitel 2.6 und 2.10).

Bei diesen Experimenten mit verdurchschnittlichten Schattenprofilen zeigten sich geringe Effektgrößen, nämlich bei Männerprofilen $r_s = .38$ ($p < .038$) und bei Frauen $r_s = .30$ ($p = .101$), wobei der Effekt für Frauengesichter nicht signifikant war. Beim Experiment *Vergleich 2* war die mittlere Effektstärke der Durchschnittlichkeit vergleichbar groß. Betrachtet man jeweils das d nach Cohen beim Vergleich „Durchschnitt versus Original“, dann lag für Männergesichter der Median über alle sechs Cohen's d bei 0.20 und für Frauengesichter bei 0.47 (wobei bei Männergesichtern nur 2 von 6 und bei Frauengesichtern nur 3 von 6 Post-hoc-Tests überhaupt statistisch signifikant waren). Fazit: Durchschnittlichkeit hat insgesamt einen positiven Effekt auf viele Gesichter (aber eben nicht auf alle); jedoch ist dieser Effekt klein und gerade an der Grenze zur statistischen Nachweisbarkeit. Klarer erscheinen zwar die Ergebnisse des Ranking-Experiments (*Vergleich 1*), wo die verdurchschnittlichten Gesichter beider Geschlechter insgesamt eindeutig höhere Attraktivitätsränge bekamen als die Originalversionen ($p < .001$), doch handelte es sich hier eben nur um einen *relativen* Attraktivitätsvergleich, wo bereits kleinste Attraktivitätsunterschiede im direkten Vergleich und bei 324 Versuchspersonen leicht zu einem statistisch signifikanten Ergebnis führen können. Bei den anderen vier genannten Untersuchungen bildeten hingegen Ratings, also absolute Attraktivitätsurteile, die Datengrundlage. Beurteilten Versuchspersonen also die absolute Attraktivität eines Gesichts und berücksichtigten damit auch andere Gesichtsmerkmale (wie zum Beispiel Merkmale der Haut), dann hatten durchschnittliche Gesichtsproportionen nur einen sehr geringen Einfluss auf die Attraktivitätsbewertung.

Sexueller Dimorphismus

Wie aus beiden Untersuchungen (*Vergleich 1* und *2*) übereinstimmend hervorgeht, ist Durchschnittlichkeit im Vergleich zu den anderen beiden überprüften Hypothesen das stärkste Attraktivitätskriterium. Die Theorie des sexuellen Dimorphismus hat folglich einen noch geringeren Einfluss.

Am besten lassen sich die Ergebnisse aus dem Experiment *Vergleich 1* mit den Ergebnissen aus *Dimorph 2* und *Dimorph 3* vergleichen, denn in allen Untersuchungen betrug die Stärke der Annäherung an das Kindchenschema 40 % und war den Versuchspersonen ein direkter Vergleich zwischen verkindlichter Version und Originalversion möglich. Die Ergebnisse bestätigten (mit neuen Stimuluspersonen und einer neuen Probandenstichprobe) vollauf die Befunde der Experimente zum sexuellen Dimorphismus. Entgegen den Vorhersagen der Theorie wurden die verkindlichten Männergesichter keineswegs schlechter, sondern tendenziell eher besser als die Originalgesichter bewertet (in 4 von 6 Fällen signifikant besser als die jeweilige Originalversion). Genauso war es auch beim Experiment *Dimorph 3*: Hier wurden 6 von 10 Stimulusgesichter in der verkindlichten Version signifikant besser bewertet, also bei beiden Experimenten bei rund zwei Drittel der Stimuluspersonen. Für Männergesichter wurde die Theorie des sexuellen Dimorphismus also erneut widerlegt.

Bei den Frauengesichtern fiel der positive Effekt der Verkindlichung (= Femininisierung) erwartungsgemäß und hypothesenkonform noch etwas positiver aus: Alle auf diese Weise manipulierten Versionen erhielten signifikant höhere Attraktivitätsränge als die Originalgesichter; im Experiment *Dimorph 2* konnte zuvor bereits ein positiver signifikanter Effekt des Kindchenschemas ebenfalls bei 9 von 10 Stimuluspersonen nachgewiesen werden.

Dies sind jedoch nur die Ergebnisse aus den *relativen* Vergleichen, wenn Probanden also im Paarvergleich oder im Ranking-Experiment die direkte Vergleichsmöglichkeit zwischen Kindchenschema-Version und Originalversion hatten. Neu sind hier nun die Ergebnisse aus dem Rating-Experiment (*Vergleich 2*), bei dem die Versuchspersonen absolute Attraktivitätsurteile abgaben und aufgrund des *Between-Subjects Designs* nicht wussten, dass es die Stimuluspersonen, die sie zu bewerten hatten, auch noch in anderen Versionen gab.

Die Ergebnisse fielen hier deutlich anders aus: Von 6 männlichen Stimulusgesichtern war im Post-hoc-Test in nur einem einzigen Fall ein verkindlichtes Gesicht signifikant attrakti-

ver als die Originalversion, und der Median über alle 6 Cohen's d betrug -0.09 , der Effekt ist also praktisch gleich Null. Bei den Frauengesichtern immerhin erwiesen sich in 3 von 6 Post-hoc-Tests die kindlichen Varianten als signifikant attraktiver. Der Median über alle 6 Cohen's d betrug hier 0.38 , d. h. der attraktivitätserhöhende Effekt des Kindchenschemas ließ sich reproduzieren, selbst dann, wenn Versuchspersonen absolute statt relativer Attraktivitätsurteile abgaben. Mit einem mittleren Cohen's d von 0.38 ist der Effekt kindlicher Gesichtsproportionen jedoch schwächer ausgeprägt als der Effekt durchschnittlicher Gesichtsproportionen ($d = 0.47$, siehe oben!).

Bei der Interpretation aller in dieser Arbeit erzielten Ergebnisse zum sexuellen Dimorphismus ist stets im Hinterkopf zu behalten, dass das Treatment, d. h. die Annäherung von Originalgesichtern an die Proportionen eines durchschnittlichen Kindergesichts zu 40% sehr stark war. Die Gesichtsproportionen wurden dadurch massiv femininisiert und sind in keiner Weise mit den kaum wahrnehmbaren marginalen Feminisierungseffekten vergleichbar, mit denen beispielsweise in den Studien von Perrett et al. (1998) oder Rhodes et al. (2000) experimentiert wurde (vgl. dazu die ausführliche Diskussion in Kapitel 7.3). Aber obwohl das Treatment so massiv war, wirkte sich dennoch der Effekt der Verkindlichung / Femininisierung nur schwach auf die absoluten Attraktivitätsbewertungen von Frauengesichtern und praktisch gar nicht auf die von Männergesichtern aus. Auch dies ist (wie schon im Fall durchschnittlicher Proportionen) wieder ein Indiz dafür, dass für das absolute Attraktivitätsurteil offensichtlich zusätzlich noch viele andere Kriterien eine Rolle spielen, die nichts mit geschlechtstypischen Gesichtsproportionen zu tun haben.

Symmetrie

Auch beim Einfluss der Symmetrie auf die Attraktivität bestätigten die Experimente *Vergleich 1* und *Vergleich 2* die Ergebnisse der zuvor durchgeführten Symmetrie-Untersuchungen. Dieselbe Methode der Symmetrie-Manipulation (*Warp*-Technik) und der Datenerhebung (Paarvergleich bzw. Rating) wurden bereits in den Experimenten *Sym 5* und *Sym 6* eingesetzt. Beim Paarvergleichsexperiment *Sym 5* (bei dem ja nur Frauengesichter untersucht worden waren) gab es eine leichte Präferenz für symmetrisierte Gesichter (in 25 von 30 Fällen), aber nur in 7 von 30 Fällen war dieser Unterschied statistisch signifikant. Im Experiment *Vergleich 1* gab es bei den Frauengesichtern bei allen 6 Stimuluspersonen eine schwache Präferenz für die symmetrisierten Varianten, die jedoch nur bei 3 von 6 Fällen signifikant attraktiver als die Originalversionen bewertet wurden. Dasselbe gilt für die Männergesichter im Experiment *Vergleich 1*, bei denen der Unter-

schied in 4 von 6 Fällen signifikant war. Hatten die Versuchspersonen eine direkte Vergleichsmöglichkeit zwischen symmetrisierter Version und Originalversion, zeigten sie in beiden Fällen eine leichte Präferenz für symmetrisch optimierte Gesichter.

Anders hingegen, wenn diese Vergleichsmöglichkeit aufgrund des *Between-Subjects Designs* fehlte und absolute Attraktivitätsurteile verlangt waren. Im Rating-Experiment Sym 6 gab es praktisch keine Attraktivitätsunterschiede. Lediglich bei 6 von 30 Fällen waren die Ergebnisse signifikant, davon jedoch war in 5 Fällen die symmetrische Version schlechter (!) und nur in einem einzigen Fall besser bewertet worden. Das Experiment *Vergleich 2* kam zum gleichen Ergebnis: Bei den 6 weiblichen Stimulusgesichtern wurde einmal die symmetrische Version signifikant besser und zweimal signifikant schlechter bewertet als die Originalversion; bei den anderen Frauengesichtern gab es keine Unterschiede. Bei den Männergesichtern gab es keinen einzigen signifikanten Unterschied. Die nicht nachweisbaren Unterschiede wiegen angesichts der großen Stichprobe von 469 Versuchspersonen (117 Urteiler pro Variante) umso schwerer. Doch auch unabhängig von Aspekten der statistischen Signifikanz sagen eigentlich bereits die Effektgrößen alles: Der Median über alle Cohen's d beträgt bei den Frauengesichtern $d = 0.03$ und bei den Männergesichtern $d = 0.01$. Es dürfte ziemlich schwierig sein, irgendein anderes Merkmal zu finden, das nachweislich noch irrelevanter für die Attraktivität eines Gesichts ist als die Symmetrie.

Damit hat sich auch die seit 20 Jahren geführte akademische Diskussion (Alley & Cunningham, 1991), ob der Attraktivitätsgewinn gemorphter Gesichter nicht durch die perfekte Symmetrie aufgrund der Mittelwertbildungen zu erklären sei, erledigt. Man braucht dafür nur die Attraktivitätsbewertungen der verdurchschnittlichten Varianten (die natürlich zugleich auch hoch symmetrisch sind) mit den Bewertungen der „nur“ symmetrisch optimierten Varianten zu vergleichen. Da der Effekt der Symmetrie Null ist, kann die kleine Attraktivitätssteigerung der Durchschnittsgesichter logischerweise nicht auf die Symmetrie zurückzuführen sein.

Relative versus absolute Attraktivitätsurteile

Die beiden letzten Experimente *Vergleich 1* und *Vergleich 2* liefern hinsichtlich der Wichtigkeit der drei Hypothesen eine klare Aussage. Sie kamen übereinstimmend zu dem Ergebnis, dass die Durchschnittlichkeit den größten attraktivitätserhöhenden Effekt auf die Gesichter hatte, danach folgten kindliche Gesichtsproportionen und an letzter Stelle die Symmetrie. Im *Ranking*-Experiment waren diese Unterschiede in der Rangreihenfolge auch signifikant. Dieses erfreulich klare Ergebnis wird jedoch durch die Ergebnisse des

Rating-Experiments getrübt, denn hier zeigte sich, dass die absoluten Effekte auf die Attraktivitätssteigerung nur schwach bis (im Fall von Symmetrie) praktisch nicht vorhanden waren. Während also das *Ranking*-Experiment mit einer hoch sensitiven Methode (dem intensiven Vergleichen von Varianten derselben Stimulusperson) nachweisen konnte, dass tatsächlich winzige signifikante Attraktivitätsunterschiede zwischen der Versionen bestanden, wies auf der anderen Seite das *Rating*-Experiment nach, dass dieselben Unterschiede für hohe Attraktivität von Gesichtern keine große Bedeutung haben.

Welchem Ergebnis man mehr Beachtung schenkt, hängt in erster Linie vom Interesse und vom eigenen Selbstverständnis eines Forschers ab. Sie lassen sich auf zwei verschiedene Typen reduzieren, nämlich den (1) klassischen Grundlagenforscher und (2) den anwendungsorientiert denkenden Forscher:

(1) Ein typischer Grundlagenforscher ist jemand, der sich für diese drei untersuchten Theorien der Attraktivitätsforschung begeistern kann, oder eine dieser Theorien besonders favorisiert und selbst daran arbeitet zu beweisen, dass Theorie X den Theorien Y und Z (die von seinen konkurrierenden Attraktivitätsforscher-Kollegen favorisiert werden) überlegen ist. Er wird mit Interesse die Ergebnisse des *Ranking*-Experiments (*Vergleich 1*) zur Kenntnis nehmen und wird darauf verweisen, dass das *Rating*-Experiment (*Vergleich 2*) gar nicht sensitiv genug sei, um die Unterschiede zwischen den einzelnen Gesichtervarianten zu messen. Zudem sei es ja ohnehin trivial, dass auch noch viele weitere Merkmale zur Attraktivität beitragen, so dass es nicht erstaunlich sei, dass bei absoluten Attraktivitätsbewertungen sich die Mittelwerte der Ratings kaum veränderten. Denn schließlich seien nur die drei Kriterien von Interesse (Durchschnitt, sexueller Dimorphismus und Symmetrie) experimentell verändert worden, andere Kriterien jenseits dieser Theorien aber nicht. Gerade aus diesem Grund sei es nicht nur vertretbar, sondern notwendig, ein sehr messempfindliches Verfahren wie die Methode des Rankings zu verwenden, um vorhandene Attraktivitätsunterschiede auch nachweisen zu können. Für einen evolutionsbiologisch denkenden Grundlagenforscher ist zudem der Hinweis auf sehr kleine Effektstärken überhaupt kein Argument, da auch winzige Unterschiede (sofern sie einen Fitnessvorteil bieten) sich evolutionär auswirken können – aber eben in den Zeiträumen, in denen Evolutionsbiologen denken, nämlich Zehntausenden oder Millionen von Jahren.

(2) Der anwendungsorientiert denkende Forscher hingegen wird sich für die Ergebnisse des *Rating*-Experiments interessieren. Er will ganz pragmatisch wissen, welche Merkmale ein Gesicht attraktiv machen und wie groß dieser Effekt ist. Daher ist für ihn die zentra-

le Erkenntnis, dass keine der drei untersuchten Theorien einen nennenswerten Effekt hat. Auch nach Anpassung der einzelnen Stimulusgesichter an die Proportionen bestimmter Schemata erhalten diese Gesichter nur Attraktivitätswerte im mittleren Bereich und kommen bestenfalls knapp über das Label „mittelmäßig attraktiv“ hinaus. Nach seiner Ansicht hätten daher die Ergebnisse des *Ranking*-Experiments kaum Aussagekraft, da sie nur eine Rangreihenfolge zwischen drei Theorien aufstellten, die allesamt nur ein geringes Erklärungspotenzial besäßen. Da es ihm auf die praktische Verwendbarkeit der erzielten Ergebnisse ankommt, beispielsweise im Kontext der ästhetischen Medizin, nützen ihm die Ergebnisse des *Ranking*-Experiments nichts; er braucht stattdessen Angaben über Effektgrößen und absolute Attraktivitätsurteile. Zudem sind für ihn absolute Attraktivitätsurteile realitätsnäher, da dies der Attraktivitätswahrnehmung im Alltag näher komme, während das direkte Vergleichen verschiedener Varianten derselben Person wie im *Ranking*-Experiment eine reaktive und per se ökologisch nicht valide Methode sei, für die es im realen Leben keine Entsprechung gebe und die daher leicht Methodenartefakte produzieren könne.

Es erscheint kaum möglich, diese beiden konträren Sichtweisen miteinander zu vereinen und gleichsam zwischen diesen Standpunkten einen Kompromiss zu finden – zu unterschiedlich sind dafür die zugrunde liegenden Interessen und Verständnisse von Wissenschaft. Doch dies muss nicht unbedingt ein Nachteil sein – im Gegenteil. In den unterschiedlichen Methoden, Ergebnissen und Ergebnisinterpretationen der beiden letzten Experimente spiegelt sich keineswegs eine Eigenheit der Attraktivitätsforschung, sondern vielmehr ein Grundkonflikt der Wissenschaft, der letztlich in allen Disziplinen anzutreffen ist: Der Gegensatz von Grundlagenforschung und anwendungsorientierter Forschung.

Dass dieser Gegensatz bei den beiden abschließenden Experimenten (und in abgeschwächter Form generell in der vorliegenden Arbeit) so deutlich zu Tage tritt, liegt allein daran, dass sich die vorliegende Arbeit den Luxus erlaubt, das Thema Attraktivität umfassend von diesen beiden unterschiedlichen Standpunkten aus zu beleuchten. Die allermeisten anderen Arbeiten zur Attraktivitätsforschung hingegen sind entweder nur aus der einen oder nur aus der anderen Perspektive verfasst: In biologischen Journals publizierte Studien zu Attraktivität sind *immer* und Studien in psychologischen Journals (mittlerweile) *meistens* aus Sicht der Grundlagenforschung konzipiert; Attraktivitätsstudien in medizinischen Journals hingegen meistens aus Sicht der angewandten Forschung. Die Autoren erwecken dabei jedoch in der Regel den Eindruck, als gäbe es auf den Forschungsgegenstand nur die jeweils selbst eingenommene Perspektive und keine andere.

Die vorliegende Arbeit hingegen unterscheidet sich bewusst von dieser Haltung. Die Untersuchungen identischer Fragestellung mit unterschiedlichen Forschungsmethoden (am deutlichsten sichtbar bei den neun Untersuchungen zur Symmetrie) erfolgten im vollen Bewusstsein ihrer unterschiedlichen Vor- und Nachteile. Die teilweise unterschiedlichen und widersprüchlichen Ergebnisse, sind daher selbstverständlich keine Pannen, sondern wichtige Erkenntnisse, die neben den Rückschlüssen zu Attraktivität auch viel über die Abhängigkeit von Forschungsergebnissen von der jeweils verwendeten Methodik aussagen (im Fall der Symmetrie dürften diese Ergebnisse zur Methodik sogar spannender sein als die Ergebnisse zur Bedeutung der Symmetrie für Attraktivität).

Die erzielten Ergebnisse sind zwar teilweise unterschiedlich – verwirrend sind sie jedoch keineswegs. Denn sie sind ja eben nicht zufällig oder beliebig, sondern werden auf systematische Weise von bestimmten Methoden begünstigt und sie fügen sich stimmig in ein Gesamtbild ein. So wurde in diesem Kapitel aufgezeigt, dass die Ergebnisse der letzten beiden Experimente (*Vergleich 1* und *2*) widerspruchsfrei frühere Ergebnisse dieser Arbeit replizierten – sofern diese auch mit den gleichen Methoden untersucht worden waren.

Damit weisen die teils unterschiedlichen Ergebnisse dieser Arbeit auf einen Umstand hin, der für wissenschaftliche Laien irritierend, für reflektierte und selbstkritische Wissenschaftler jedoch eine Selbstverständlichkeit ist: Wissenschaftliche Erkenntnisse sind immer abhängig von den Methoden, mit denen sie gewonnen wurden. Welche Methoden man jedoch auswählt, hängt von der jeweiligen Fragestellung ab, vom Ziel und Zweck der Forschung und damit letztlich von den Interessen und dem wissenschaftlichen Weltbild des Forschenden.

Auch in anderen Bereichen der Psychologie hängen bestimmte Erkenntnisse oft an einer einzigen oder an wenigen Methoden oder Paradigmen, mit denen sie gewonnen wurden. Innerhalb eines Forscherkreises von Gleichgesinnten ist ein solches Paradigma oft relativ unumstritten – ähnlich wie es unter akademischen Attraktivitätsforschern (aber auch nur dort) unumstritten ist, dass man die Durchschnittlichkeit von Gesichtern erforschen kann, indem man mit Bildverarbeitungssoftware (Morphing) aus Fotos von Gesichtern virtuelle Gesichter erzeugt, die es so in der Realität gar nicht geben kann. (In diesem Fall steht und fällt die Durchschnittshypothese – eine Hypothese, die sogar Aussagen über ca. zwei Millionen Jahre menschlicher Evolution macht! – also mit dem Paradigma des Morphens, d. h. einer speziellen Computersoftware.) Einem reflektierten Wissenschaftler sollte dies zu denken geben und es sollte seinen Ehrgeiz wecken, bekannte Befunde auch mit an-

deren Paradigmen zu replizieren als mit den gewohnheitsmäßig verwendeten, auch wenn dies viel mühsamer ist. Eine gute Methodik bedeutet dabei selbstverständlich auch weit- aus mehr als lediglich die Auswahl eines korrekten statistischen Tests zur Datenanalyse (worüber gerade Psychologen oft so leidenschaftlich streiten können). Denn es nützt der korrekteste Test nichts, wenn man sich nicht bewusst macht, dass die zu analysierenden Daten bereits zuvor auf unterschiedlichste Weise beeinflusst wurden: Durch das Versuchsdesign, die Art des Stimulusmaterials, die Darbietung des Stimulusmaterials, die Art der Messung der abhängigen Variablen und vieles mehr.

Am Beispiel der Attraktivitätsforschung lässt sich hervorragend zeigen, wie durch das wissenschaftliche Weltbild, durch Motive und Ziele von Forschern bestimmte Methoden ausgewählt werden; diese Methoden beeinflussen wiederum die Art der gewonnenen Ergebnisse. Sich dies zu vergegenwärtigen, könnte zu einem etwas selbstkritischeren Umgang mit verwendeten Forschungsmethoden führen. Wenn diese Arbeit ein wenig dazu beitragen kann, dann würde dies nicht nur einen Erkenntnisfortschritt für die vergleichsweise wenigen Attraktivitätsforscher darstellen, sondern könnte darüber hinaus auch in anderen Disziplinen innerhalb der Psychologie von Nutzen sein.

10 Fazit

In der vorliegenden Arbeit wurden die drei wichtigsten Theorien der Attraktivitätsforschung ausführlich untersucht: Die Durchschnittshypothese, die Symmetrie-Hypothese und die Theorie des sexuellen Dimorphismus. Als gemeinsames Ergebnis muss festgestellt werden, dass alle drei Theorien die Attraktivität von Gesichtern bei Weitem nicht so überzeugend erklären können, wie die Mehrzahl der publizierten Forschungsarbeiten suggeriert.

Durchschnittshypothese: Die Durchschnittshypothese wurde mit insgesamt sechs Experimenten mit einem qualitativ hochwertigen Stimulusmaterial und einer sehr systematischen und gründlichen Methodik getestet. Es wurden Männer- und Frauengesichter untersucht, Gesichter in der Frontal- und Profilansicht, Gesichter mit den charakteristischen Morphing-Artefakten der glatten, makellosen Haut und Gesichter ohne jegliche Texturunterschiede (Schattenprofile). Durchschnittlichkeit wurde systematisch manipuliert anstatt wie bei der Mehrzahl früherer Untersuchungen über wenig valide *distinctiveness*-Ratings erfasst.

Die Ergebnisse zeigten übereinstimmend, dass der Attraktivitätszugewinn bei gemorpheten Composites zum größten Teil auf die Bildverarbeitungsartefakte der makelloseren Haut und nicht auf durchschnittlichere Gesichtsproportionen zurückzuführen war. Die attraktivsten Gesichter waren Composites aus wenigen attraktiven Originalgesichtern (= „Raffael-Effekt“). Diese Composites zeigten Gesichtsmerkmale, die sich klar vom Durchschnitt unterschieden. Zusätzlich illustriert wurde dieser Effekt durch ein Composite aus den acht attraktivsten Finalistinnen der Miss-Germany-Wahl 2007, das sich vom Durchschnittsgesicht aus 64 „normalen“ Frauen deutlich unterschied und gerade deswegen deutlich attraktiver bewertet wurde (Kapitel 3.2). Die Ergebnisse der Experimente *Durch 1* bis *Durch 6* wurden zusätzlich durch die beiden hypothesenvergleichenden Experimente am Schluss (Kapitel 8) noch einmal bestätigt, bei dem nur Gesichtsproportionen, jedoch keine Gesichtstexturen verändert wurden. Zwar wurde im Experiment *Vergleich 1* ein attraktivitätssteigernder Effekt durchschnittlicher Proportionen bestätigt, jedoch nur im direkten *relativen* Vergleich mit unveränderten Originalgesichtern oder Gesichtern, die in Richtung Symmetrie oder sexuellem Dimorphismus manipuliert waren. Der *absolute* Effekt auf die Attraktivität wurde im Experiment *Vergleich 2* im Rating-Verfahren gemessen: Dort wurde das Resultat der sechs vorigen Experimente zur Durchschnittshypothese bestätigt, dass Gesichter mit durchschnittlichen Gesichtsproportionen nur mittelmäßige Attraktivitätsbewertungen erhielten.

Insgesamt machten diese Experimente deutlich, dass die Durchschnittshypothese innerhalb der Attraktivitätsforschung maßlos überschätzt wird. Durchschnittliche Gesichtsproportionen bewahren in erster Linie vor Hässlichkeit, stellen jedoch nicht das Schönheitsideal der Bevölkerung dar. Für hohe Attraktivität sind zahlreiche vom Durchschnitt abweichende Gesichtsmerkmale verantwortlich, die in der deskriptiven Analyse herausgearbeitet wurden (Kapitel 3.3). Die Gründe für ein so starkes Überschätzen der Durchschnittlichkeit als Attraktivitätskriterium wurden in Kapitel 3.4 ausführlich dargelegt.

Symmetrie-Hypothese: Der Einfluss von Symmetrie auf die Attraktivität von Gesichtern wurde ebenfalls äußerst gründlich und systematisch mit verschiedenen Methoden untersucht. Insgesamt wurden neun Untersuchungen durchgeführt. Hinzu kommen einerseits eine Kontroll-Untersuchung zum Nachweis der fehlenden Validität von direkten Symmetrie-Ratings durch Beurteiler und andererseits die zwei vergleichenden Untersuchungen (Kapitel 8), bei denen der Einfluss von Symmetrie im direkten Vergleich mit Originalgesichtern getestet wurde sowie im Vergleich mit Gesichtern, die im Sinne der anderen beiden Hypothesen verändert worden waren.

Dabei kamen sämtliche Methoden (mit Ausnahme von anthropometrischen Gesichtsvermessungen) zum Einsatz, mit denen auch in bisherigen Studien ein Einfluss von Symmetrie auf Attraktivität untersucht worden war: Experimentelle Verfahren mit symmetrischer Optimierung von Gesichtern durch Chimärengesichter, Morphing-Software (mit und ohne Neuberechnung der Textur), jeweils in Paarvergleichsexperimenten oder Rating-Experimenten im *Between-Subjects Design* sowie korrelative Verfahren mit Quantifizierungen des Ausmaßes an Symmetrie durch direkte Symmetrie-Ratings oder Ähnlichkeitsratings von separaten Gesichtshälften oder Chimärengesichtern. Zur Anwendung kamen ebenfalls verschiedene Methoden der Datenauswertung, nämlich Verfahren mit und ohne Reduktion von Versuchspersonen-Urteilsvarianz.

Die durchgeführten Untersuchungen *Sym 1* bis *Sym 9* haben klar gezeigt, dass die Effekte der Symmetrie auf die Attraktivität von Gesichtern in Richtung, Größe und Signifikanz stark von der jeweils verwendeten Untersuchungsmethode abhängen. Dies ist das wichtigste Ergebnis der Untersuchungsreihe. Tabelle 15 (S. 205) fasst die Effektstärken der einzelnen Untersuchungen auf einen Blick zusammen. Die Unterschiede bei Richtung und Größe der Effektstärke waren teilweise so erheblich, dass der Schluss gezogen werden muss, dass vermeintliche Einflüsse von Symmetrie auf die Attraktivität von Gesichtern in erster Linie auf Methodenartefakten basieren.

Insgesamt führt dies zur der Schlussfolgerung, dass die Symmetrie-Hypothese als widerlegt angesehen werden muss. In Kapitel 5 wurde ausführlich dargelegt, dass alle Methoden, die zu mittelstarken oder starken Effekten führten, mit schweren methodischen Mängeln behaftet sind. Die Mehrzahl der Methoden führte zudem ohnehin zu *negativen* Effekten der Symmetrie, die jedoch ebenfalls teilweise methodische Schwächen aufweisen (z. B. statistische Tricks mit der Eliminierung von Urteilsvarianz), und die Effektstärken der validesten Methoden sind marginal. Die valideste Methode (Optimierung der Symmetrie durch Morphing-Software unter Beibehaltung der Hauttextur) und Attraktivitätsbewertung im Rating-Verfahren wurde zudem ein weiteres Mal im letzten Experiment (*Vergleich 2*) angewandt. Mit einer neuen Stichprobe an männlichen und weiblichen Stimulusgesichtern wurde der vorherige Befund repliziert, dass bei normalen Gesichtern mit natürlichen Asymmetrien (keine pathologischen Asymmetrien) Symmetrie als Attraktivitätskriterium völlig irrelevant ist. Die Attraktivität solcher Gesichter kann auch durch eine perfekt symmetrische Optimierung ihrer Gesichtsproportionen nicht gesteigert werden (Kapitel 8.2).

Durch das systematische Variieren der verwendeten Methoden konnte klar belegt werden, dass die vermeintlich attraktivitätssteigernde Wirkung von Symmetrie in Wirklichkeit auf Methodenartefakten beruht. Dieser Schluss ist besonders zwingend, weil bei allen Studien Gesichter derselben 30 Stimuluspersonen verwendet wurden und alle Versuchspersonenstichproben hinsichtlich Größe und Zusammensetzung gleich waren. Die erzielten Unterschiede können daher nur auf die verschiedenen Methoden zurückzuführen sein. Mit dem Nachweis mangelnder Validität der zur Untersuchung von Symmetrie angewandten Methoden ist auch die Gültigkeit früherer Studien, die mit denselben Methoden arbeiteten, radikal in Frage gestellt.

Sexueller Dimorphismus: Die Theorie des sexuellen Dimorphismus wurde mit drei Experimenten überprüft. Bei den beiden durchgeführten Experimenten *Dimorph 1* und *Dimorph 2* ließ sich die Theorie des sexuellen Dimorphismus für *Frauengesichter* klar bestätigen. Frauengesichter, die an das Kindchenschema angenähert und damit femininer gemacht worden waren, wurden klar attraktiver beurteilt. Männergesichter hingegen, die auf dieselbe Weise femininisiert worden waren, wurden entgegen der Voraussage der Theorie kaum schlechter bewertet, sondern (abhängig von der Attraktivität der Originalgesichter) gleich attraktiv oder attraktiver. Sowohl bei Frauen- als auch bei Männergesichtern stimmten die Ergebnisse gut mit Befunden bisheriger Studien überein, die ebenfalls für Frauengesichter einen attraktivitätssteigernden Effekt typisch femininer Gesichtszüge fanden, für Männergesichter hingegen ähnlich heterogene Ergebnisse wie beim hier durchgeführten Experiment *Dimorph 3*.

Die Ergebnisse zum sexuellen Dimorphismus stimmen gut überein mit den Ergebnissen der deskriptiven Analyse der attraktiven Prototypen in Kapitel 3.3, in der sich mit einer ganz anderen Methodik ebenfalls zeigte, dass hochattraktive Frauengesichter sich von durchschnittlichen bzw. unattraktiven durch typisch feminine Merkmale unterscheiden. Und auch hochattraktive Männergesichter zeichneten sich teilweise ebenfalls durch Merkmale aus, die eigentlich typisch weiblich sind (klare Ausnahmen sind jedoch der markante, kräftige Unterkiefer). Auch die beiden letzten hypothesenvergleichenden Untersuchungen *Vergleich 1* und *Vergleich 2* bestätigen die Ergebnisse der Experimente zum sexuellen Dimorphismus: Frauengesichter gewannen durch eine Annäherung ihrer Gesichtsproportionen an Attraktivität, während dieselbe Veränderung bei Männergesichtern (entgegen der Vorhersagen der Theorie) keinen negativen Effekt brachte.

Insgesamt scheint also die Theorie des sexuellen Dimorphismus teilweise korrekt zu sein. Frauengesichter profitierten tatsächlich von kindlich-femininen Gesichtszügen. Dass je-

doch ein Männergesicht attraktiv ist, wenn es typisch maskulin aussieht, ließ sich empirisch nicht bestätigen. Die Daten sprechen eher für einen allgemeinen attraktivitätserhöhenden Effekt von neotenen und femininen Gesichtszügen. Was speziell am Kindchenschema attraktiv wirken könnte und was nicht, bzw. inwiefern dahinter das Attraktivitätskriterium der Jugendlichkeit stecken könnte, wurde in Kapitel 7.4 ausführlich dargelegt.

In der gesamten vorliegenden Arbeit wurde großer Wert auf eine solide, gründliche Methodik gelegt. Dies beginnt beim verwendeten Stimulusmaterial, bei dem die gemorphten Gesichter eine sehr hohe Bildqualität mit einem Minimum an Artefakten aufweisen, bei denen (außer bei Composites, die aus sehr vielen Gesichtern berechnet wurden) meist noch nicht einmal für Versuchspersonen erkennbar ist, dass es sich um virtuelle Gesichter handelt. Bei zahlreichen Studien sehr renommierter Attraktivitätsforscher hingegen ist die Bildqualität von gemorphten Gesichtern (sofern sie Beispiele dafür überhaupt in einem Artikel abdrucken) häufig so schlecht und mit Bildverarbeitungsartefakten konfundiert, dass damit kaum extern valide Ergebnisse zu erzielen sind. Desweiteren wurde durch die Verwendung von Fotos von Finalistinnen einer Miss-Germany-Wahl sichergestellt, dass Attraktivität auch tatsächlich mit attraktiven Gesichtern untersucht wurde und nicht nur mit bestenfalls mittelmäßig aussehenden.

Bei den Datenerhebungen aller 20 Untersuchungen wurde Wert auf eine für die Bevölkerung möglichst repräsentative Stichprobe gelegt: Es wurden stets *beide* Geschlechter befragt (und nicht nur das den Stimuluspersonen entgegengesetzte Geschlecht), es wurden Versuchspersonen aller Altersklassen befragt, Versuchspersonen wurden größtenteils im öffentlichen Bereich (z. B. in Fußgängerzonen und Einkaufszentren) rekrutiert, während Psychologie-Studierende (aufgrund möglichen Vorwissens) grundsätzlich ausgeschlossen waren. Auch darin unterscheiden sich die hier durchgeführten Untersuchungen von den meisten anderen Studien, die ihr Versuchspersonen-Klientel meist ausschließlich unter Studierenden der Psychologie in den ersten Semestern rekrutieren, was in Hinblick auf die Verallgemeinerbarkeit der Ergebnisse grundsätzlich Fragen aufwirft.

Auf häufig verwendete Methoden, die zwar für den Versuchsleiter bequem sind, aber zu Ergebnissen von geringer Validität führen, wurde grundsätzlich verzichtet. Dazu zählen v. a. der Karikatur-Algorithmus von Morphing-Software (vgl. Kapitel 3.4 und 6.1.3) und die Quantifizierung von Durchschnittlichkeit, Symmetrie oder sexuellem Dimorphismus durch Versuchspersonen-Ratings („*How distinctive / symmetric / feminine / masculine is this*

face?"). Die einzige Ausnahme stellte die zusätzliche Datenerhebung *Sym 7 Kontrolle* dar, die lediglich den Zweck hatte, nachzuweisen, dass derartige Ratings unsinnig sind.

Ebenso wurde auf statistische Kunstgriffe verzichtet, die lediglich dem Ziel dienen, marginale Effekte statistisch signifikant zu bekommen (z. B. durch die künstliche Reduzierung von Versuchspersonen-Urteilsvarianz), wie dies bei etlichen früheren Studien (v. a. bei Symmetrie) angewandt wurde (vgl. Kapitel 4 und 5.2.3). Wo sie dennoch gerechnet wurden, wurden sie gezielt den ehrlicheren, konservativeren Auswertungsverfahren gegenübergestellt, um auf die Effekte dieser Art von Ergebnis-Kosmetik hinzuweisen. Statt statistischer Signifikanzen wurden bei allen Untersuchungen immer wieder die Effektstärken in den Fokus des Interesses gerückt. Ebenso wurde systematisch getestet, wie groß der Einfluss des Versuchsdesigns auf die Ergebnisse eines Attraktivitätsexperiments ist – zu nennen ist hier insbesondere der direkte Paarvergleich im *forced-choice*-Verfahren gegenüber dem Rating-Experiment im *Between-Subjects Design* (Experimente *Sym 1*, *Sym 3* und *Sym 5* versus *Sym 2*, *Sym 4* und *Sym 6* sowie ferner *Vergleich 1* versus *Vergleich 2*).

Die Ergebnisse zur Bedeutung der drei untersuchten Hypothesen zur Attraktivität sind daher nicht das einzige Ergebnis der vorliegenden Arbeit. Ähnlich wichtig sind die Ergebnisse zu den verschiedenen Methoden zur Untersuchung von Attraktivität. Denn sie beweisen, dass zahlreiche frühere Studien unter teilweise starken Mängeln leiden und stellen dadurch viele häufig zitierte Studien bekannter Attraktivitätsforscher in Frage. In den jeweiligen Diskussionsteilen zu den durchgeführten eigenen Untersuchungen wurde ausführlich auf methodisch fragwürdige Praktiken früherer Studien hingewiesen und durch Bildmaterial oder eigene Daten belegt (vgl. insbesondere Kapitel 3.1, 3.4, 3.5, 4.10.4, 5.2, 5.3 und 7.3).

Im vorangegangenen Kapitel 9 wurde dargelegt, wie die Verwendung bestimmter Untersuchungsmethoden für Fragestellungen zur Attraktivität von Gesichtern vom jeweiligen wissenschaftlichen Weltbild und von den jeweiligen Interessen von Forschern abhängt. Die eine typische Gruppe sind Grundlagenforscher, denen es in erster Linie darum geht, zu zeigen, dass eine bestimmte Theorie (meist mit evolutionspsychologischem Hintergrund) Daten besser erklärt als eine andere (konkurrierende) Theorie. Sie bevorzugen hoch sensitive Messverfahren, während ihnen die Effektgröße oder gar der absolute Effekt auf die Attraktivitätswahrnehmung ziemlich gleichgültig ist. Die andere typische Gruppe sind anwendungsorientiert denkende Forscher, die meist im Kontext der ästhetischen Medizin forschen und arbeiten. Bei ihnen ist es genau umgekehrt: Sie interessie-

ren sich vor allem für die Effektgröße eines Attraktivitätsmerkmals (während ihnen Attraktivitätstheorien eher gleichgültig sind) und legen eher Wert auf eine hohe ökologische Validität ihrer Methoden, da sie die Ergebnisse ihrer Forschung hinterher auch beim Patienten anwenden möchten. Entsprechend ihrer Forschungsmotivation bevorzugen diese beiden Forschergruppen unterschiedliche Methoden (z. B. hoch sensitive Ranking-Verfahren oder Paarvergleiche anstatt ökologisch validerer und weniger reaktiver Rating-Verfahren im *Between-Subjects Design*).

In Kapitel 3.5 (im Kontext der Durchschnittshypothese) wurde dargelegt, wie die klassische, akademische Attraktivitätsforschung und die anwendungsorientierte, medizinisch-ästhetische Attraktivitätsforschung isoliert und nahezu ohne gegenseitigen Austausch nebeneinander existieren. Während letztere durch den medizinisch-technischen Fortschritt in der ästhetischen Medizin immer mehr an Bedeutung und Einfluss gewinnt, scheint sich die akademische Attraktivitätsforschung an den Universitäten selbst immer stärker ins Abseits zu manövrieren. Da die meisten von ihr bearbeiteten Fragestellungen und die meisten ihrer Forschungsergebnisse aufgrund der in der Regel kleinen Effektgrößen für eine praktische Verwendbarkeit ohne Bedeutung sind, wird die akademische Attraktivitätsforschung von keiner anderen Disziplin gebraucht. Das wäre an sich nicht weiter schlimm, schließlich ist dies häufig die Situation von Grundlagenforschung. Doch damit hängt die Existenz der akademischen Attraktivitätsforschung allein am Wahrheitsgehalt ihrer Theorien. Seit über 20 Jahren erforscht sie intensiv „ihre“ drei großen, wichtigen Hypothesen der Gesichtsattraktivität: Durchschnittlichkeit, Symmetrie und sexuellen Dimorphismus. Stellen sich diese drei Hypothesen als falsch heraus, dann steht diese Disziplin mit praktisch nichts in der Hand da. Wenn bewiesen wird, dass diese Kriterien Gesichter gar nicht attraktiv machen, dann sind auch alle in den letzten Jahren vor evolutionspsychologischem Hintergrund entstandenen Studien, *warum* diese Kriterien attraktiv machen (z. B. Signale für Fruchtbarkeit, Gesundheit, Parasitenresistenz, „gute Gene“) das Papier nicht mehr wert, auf dem sie gedruckt sind.

Durch die eindeutigen Ergebnisse der 20 Untersuchungen dieser vorliegenden Arbeit wird sich der Druck auf die klassische, akademische Attraktivitätsforschung verstärken. Denn beispielsweise bei Symmetrie wurde mit insgesamt 10 Experimenten klar nachgewiesen, dass die normalen, nicht-pathologischen, natürlichen Asymmetrien, die jedes Gesicht mehr oder weniger aufweist, für die Attraktivitätsbewertung ohne Bedeutung sind. Und auch perfekte Symmetrie macht eben *nicht* attraktiv, auch wenn zahlreiche frühere Untersuchungen zu anderen Schlussfolgerungen kamen. Hingegen konnte eindeutig

nachgewiesen werden, dass die verwendeten Methoden anderer Studien zahlreiche und starke Artefakte produzieren (vgl. Kapitel 5).

Ähnlich ist die Situation bei der Durchschnittshypothese, bei der hier mit sechs Experimenten übereinstimmend gezeigt wurde, dass es eben nicht die durchschnittlichen Gesichtsproportionen sind, die ein Gesicht attraktiv machen, sondern vielmehr die typischen Morphing-Artefakte der glatteren und makelloseren Haut von gemorphten Gesichtern. Das Problem wird dadurch verschärft, weil die Durchschnittshypothese untrennbar mit einem ganz bestimmten Paradigma verbunden ist, nämlich der experimentellen Manipulation von Gesichtern durch Verwendung von Morphing-Software. Durchschnittliche Gesichtsproportionen also führen eben doch nur zu einem mittelmäßigen Aussehen, während die attraktivsten Gesichter Merkmale aufweisen, die systematisch vom Durchschnitt abweichen – also ganz so, wie der naive Laie es mit seinem gesunden Menschenverstand annehmen würde, auch wenn unter akademischen Attraktivitätsforschern nach wie vor ein anderer Konsens herrscht.

Beim sexuellen Dimorphismus hingegen ist bereits schon länger bekannt, dass die Theorie ein Problem mit der Erklärung männlicher Gesichtsattraktivität hat; die hier durchgeführten Experimente haben dies nochmals bestätigt. Für die Hälfte der Weltbevölkerung – für Männer – scheint die Vorhersage „je männlicher ein Männergesicht, desto attraktiver“ also ohnehin nicht zu gelten. Lediglich für Frauengesichter finden sich genügend empirische Belege.

Insgesamt ist das also keine besonders schmeichelhafte Bilanz für die akademische Attraktivitätsforschung, wenn sich von den drei wichtigsten und meistuntersuchten Theorien zweieinhalb als empirisch nicht haltbar herausstellen. Für die kommenden Jahre wird es interessant zu beobachten sein, wie lange es in der Szene der akademischen Attraktivitätsforscher dauern wird, bis sich dort die Einsicht durchgesetzt hat, dass insbesondere Durchschnittlichkeit und Symmetrie für die Attraktivität von Gesichtern in Wahrheit bedeutungslos sind. Und es bleibt eine spannende Frage, ob und wenn ja mit welchem Ansehensverlust die akademische Attraktivitätsforschung aus dieser Sackgasse wieder herausfindet. Für die Zukunft weitaus besser aufgestellt ist hingegen die anwendungsorientiert arbeitende Attraktivitätsforschung, denn sie profitiert vom Boom der ästhetischen Medizin und wird zweifellos an Einfluss gewinnen. Denn während einerseits immer bessere und schonendere Operationsmethoden und medizinische Produkte (z. B. Filler, Implantate) entwickelt werden, die bei der Bevölkerung auf immer größere Akzeptanz und Nachfrage stoßen, klafft auf der anderen Seite eine Wissenslücke bei der Frage,

was genau ein hochattraktives Gesicht ausmacht. Das optimale Aussehen dieser Merkmale zu erforschen, wird reichlich Stoff für viele künftige Untersuchungen zur Attraktivität von Gesichtern bieten.

11 Zusammenfassung

Hintergrund: In der vorliegenden Arbeit wurden die drei wichtigsten Theorien der Attraktivitätsforschung zur Attraktivität von Gesichtern umfassend untersucht: (1) Die Durchschnittshypothese, nach der ein Gesicht dann besonders attraktiv sein soll, wenn seine Gesichtsproportionen dem mathematischen Durchschnitt einer Population entsprechen, (2) die Symmetrie-Hypothese, nach der attraktive Gesichter sich durch hohe Symmetrie auszeichnen sollen, sowie (3) die Theorie des sexuellen Dimorphismus, nach der *Frauengesichter* besonders attraktiv sein sollen, wenn sie typisch feminin aussehen und *Männern*gesichter, wenn sie typisch maskulin aussehen. Insgesamt wurden 20 verschiedene Untersuchungen durchgeführt.

Methoden: Die Durchschnittshypothese wurde mit insgesamt sechs Experimenten mit einem qualitativ hochwertigen Stimulusmaterial und einer sehr systematischen und gründlichen Methodik getestet. Mit Hilfe von Morphing-Software wurde die Durchschnittlichkeit von Gesichtern experimentell manipuliert. Es wurden Männer- und Frauengesichter untersucht, Gesichter in der Frontal- und Profilsicht, Gesichter mit den charakteristischen Morphing-Artefakten einer glatten, makellosen Haut und Gesichter ohne jegliche Texturunterschiede (Schattenprofile). Durchschnittlichkeit wurde systematisch manipuliert anstatt wie bei der Mehrzahl früherer Untersuchungen über wenig valide *distinctiveness*-Ratings erfasst.

Der Einfluss von Symmetrie auf die Attraktivität von Gesichtern wurde ebenfalls äußerst gründlich und systematisch mit verschiedenen Methoden untersucht. Insgesamt wurden dazu neun Untersuchungen durchgeführt. Dabei kamen sämtliche Methoden (mit Ausnahme von anthropometrischen Gesichtsvermessungen) zum Einsatz, mit denen auch in bisherigen Studien ein Einfluss von Symmetrie auf Attraktivität untersucht worden war: Experimentelle Verfahren mit symmetrischer Optimierung von Gesichtern durch Chimären-gesichter, Morphing-Software (mit und ohne Neuberechnung der Textur), jeweils in Paarvergleichsexperimenten oder Rating-Experimenten im *Between-Subjects Design* sowie korrelative Verfahren mit Quantifizierungen des Ausmaßes an Symmetrie durch direkte Symmetrie-Ratings oder Ähnlichkeitsratings von separaten Gesichtshälften bzw. Chimären-gesichtern. Zur Anwendung kamen ebenfalls verschiedene Methoden der Datenauswertung, nämlich Verfahren mit und ohne Reduktion von Versuchspersonen-Urteilsvarianz.

Die Theorie des sexuellen Dimorphismus wurde mit drei Experimenten überprüft, bei denen weibliche und männliche erwachsene Gesichter mit Hilfe von Morphing-Software an die Proportionen des Kindchenschemas angenähert und dadurch kindlicher bzw. femininer gemacht wurden.

In zwei abschließenden Experimenten wurden die drei Hypothesen direkt gegeneinander getestet, indem mit Morphing-Software die Gesichter von Originalgesichtern entweder durchschnittlicher, perfekt symmetrisch oder femininer gemacht wurden und von Versuchspersonen nach Attraktivität in eine Rangreihenfolge gebracht (*Vergleich 1*) bzw. in einem *Between-Subjects Design* auf Ratingskalen nach Attraktivität bewertet wurden (*Vergleich 2*).

Ergebnisse: Die Ergebnisse zur Durchschnittshypothese zeigten übereinstimmend, dass der Attraktivitätszugewinn bei gemorphten Composites zum größten Teil auf die Bildverarbeitungsartefakte der makelloseren Haut und nicht auf durchschnittlichere Gesichtsproportionen zurückzuführen war. Die attraktivsten Gesichter waren Composites aus wenigen attraktiven Originalgesichtern (= „Raffael-Effekt“). Diese Composites zeigten Gesichtsmerkmale, die sich klar vom Durchschnitt unterschieden.

Die durchgeführten neun Untersuchungen zur Symmetrie haben klar gezeigt, dass die Effekte der Symmetrie auf die Attraktivität von Gesichtern in Richtung, Größe und Signifikanz stark von der jeweils verwendeten Untersuchungsmethode abhängen. Dies ist das wichtigste Ergebnis der Untersuchungsreihe. Die Unterschiede bei Richtung und Größe der Effektstärke waren teilweise so erheblich, dass der Schluss gezogen werden muss, dass vermeintliche Einflüsse von Symmetrie auf die Attraktivität von Gesichtern in erster Linie auf Methodenartefakten basieren. Dieser Schluss ist besonders zwingend, weil bei allen Studien Gesichter derselben 30 Stimuluspersonen verwendet wurden und alle Versuchspersonenstichproben hinsichtlich Größe und Zusammensetzung gleich waren. Insgesamt führt dies zur Schlussfolgerung, dass die Symmetrie-Hypothese als widerlegt angesehen werden muss.

Die Theorie des sexuellen Dimorphismus ließ sich für *Frauengesichter* bestätigen. Frauengesichter, die an das Kindchenschema angenähert und damit femininer gemacht worden waren, wurden klar attraktiver beurteilt. *Männergesichter* hingegen, die auf dieselbe Weise femininisiert worden waren, wurden entgegen der Voraussage der Theorie kaum schlechter bewertet, sondern (abhängig von der Attraktivität der Originalgesichter) gleich attraktiv oder attraktiver.

Diskussion: Als Gesamtergebnis muss festgestellt werden, dass alle drei Theorien die Attraktivität von Gesichtern bei Weitem nicht so überzeugend erklären können, wie die Mehrzahl der publizierten Forschungsarbeiten suggeriert. Die Theorie des sexuellen Dimorphismus ließ sich nur für Frauengesichter bestätigen, die Attraktivitätssteigerung gemorphter Durchschnittsgesichter ließ sich auf Morphing-Artefakte einer glatteren, makelloseren Haut zurückführen (während die attraktivsten Gesichter Merkmale aufwiesen, die sich systematisch vom Durchschnitt unterschieden), und die natürlichen (nicht-pathologischen) Asymmetrien von Gesichtern spielten für die Attraktivitätsbewertung keine Rolle.

Durch das systematische Variieren von Untersuchungsmethoden wurde zusätzlich demonstriert, wie sich verschiedene Methoden im Versuchsdesign (z. B. Paarvergleich versus Rating), bei der Erzeugung von Stimuli oder bei der Datenanalyse auf Signifikanz und Effektgröße von Ergebnissen zur Gesichterattraktivität auswirken. Die vorliegende Arbeit weist damit auch nach, dass zahlreiche Befunde früherer Studien zur Attraktivität von Gesichtern in Wirklichkeit auf Methodenartefakte zurückzuführen sind.

12 Literaturverzeichnis

- Adams, G. R. & Huston, T. L. (1975). Social perceptions of middle-aged varying in physical attractiveness. *Developmental Psychology, 11*, 658-675.
- Alex, J. C. (2004). Aesthetic considerations in the elevation of the eyebrow. *Facial Plastic Surgery, 20*, 193-198.
- Alexander, J. & Stimson, W. H. (1988). Sex-hormones and the course of parasitic infection. *Parasitology Today, 4*, 189-193.
- Alley, T. R. (1988). The effects of growth and aging on facial aesthetics. In T. R. Alley (Ed.), *Social and applied aspects of perceiving faces* (pp. 51-62). Hillsdale, NJ: Erlbaum.
- Alley, T. R. & Cunningham, M. R. (1991). Averaged faces are attractive, but very attractive faces are not average. *Psychological Science, 2*, 123-125.
- Anderson, M. (1994). *Sexual selection*. Princeton: Princeton University Press.
- Bartlett, S. P., Grossman, R. & Whitaker, L. A. (1992). Age-related-changes of the craniofacial skeleton - an anthropometric and histologic analysis. *Plastic and Reconstructive Surgery, 90*, 592-600.
- Baudouin, J.-Y. & Tiberghien, G. (2004). Symmetry, averageness, and feature size in the facial attractiveness of women. *Acta Psychologica, 117*, 313-332.
- Benson, P. & Perrett, D. (1992). Face to face with the perfect image. *New Scientist, 133*, 32-35.
- Berkowitz, L. & Frodi, A. (1979). Reactions to a child's mistakes as affected by her/his looks and speech. *Social Psychology Quarterly, 42*, 420-425.
- Berry, D. S. & Brownlow, S. (1989). Were the physiognomists right? Personality correlates of facial babyishness. *Personality and Social Psychology Bulletin, 15*, 266-279.
- Berry, D. S. & McArthur, L. Z. (1985). Some components and consequences of a baby-face. *Journal of Personality and Social Psychology, 48*, 312-323.
- Berry, D. S. & McArthur, L. Z. (1986). Perceiving character in faces - the impact of age-related craniofacial changes on social-perception. *Psychological Bulletin, 100*, 3-18.
- Berry, D. S. & McArthur, L. Z. (1988). What's in a face? Facial maturity and the attribution of legal responsibility. *Personality and Social Psychology Bulletin, 14*, 23-33.
- Berscheid, E. & Walster, G. W. (1974). Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 157-215). New York: Academic Press.
- Bortz, J. & Döring, N. (2009). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Brownlow, S. & Zebrowitz, L. A. (1990). Facial appearance, gender, and credibility in television commercials. *Journal of Nonverbal Behavior, 14*, 51-60.
- Bruce, V., Burton, A. M. & Dench, N. (1994). What's distinctive about a distinctive face? *Quarterly journal of experimental psychology, The A, Human experimental psychology, 47*, 119-141.
- Bull, R. & Rumsey, N. (1988). *The social psychology of facial appearance*. New York: Springer.

- Buss, D. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioural and Brain Sciences*, 12, 1-49.
- Buss, D. M. (1999). *Evolutionary psychology. The new science of the mind*. Boston: Allyn & Bacon.
- Cartwright, M. J., Kurumety, U. R., Nelson, C. C., Frueh, B. R. & Musch, D. C. (1992). Normative measurements of the lid and brow. *Investigative Ophthalmology & Visual Science*, 33, 1298-1298.
- Cartwright, M. J., Kurumety, U. R., Nelson, C. C., Frueh, B. R. & Musch, D. C. (1994). Measurements of upper eyelid and eyebrow dimensions in healthy white individuals. *American Journal of Ophthalmology*, 117, 231-234.
- Coleman, S. R. & Grover, R. (2006). The anatomy of the aging face: Volume loss and changes in 3-dimensional topography. *Aesthetic Surgery Journal / the American Society for Aesthetic Plastic Surgery*, 26, 4-9.
- Cunningham, M. R. (1986). Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *Journal of Personality and Social Psychology*, 50, 925-935.
- Cunningham, M. R., Barbee, A. P. & Pike, C. L. (1990). What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *Journal of Personality and Social Psychology*, 59, 61-72.
- Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B. & Wu, C.-H. (1995). "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, 68, 261-279.
- Daucher, H. (1967). *Künstlerisches und rationalisiertes Sehen. Gesetze des Wahrnehmens und Gestaltens*. München: Ehrenwirth.
- de Haan, M., Johnson, M. H., Maurer, D. & Perrett, D. I. (2001). Recognition of individual faces and average face prototypes by 1- and 3-month-old infants. *Cognitive Development*, 16, 659-678.
- DeBruine, L. M., Jones, B. C., Unger, L., Little, A. C. & Feinberg, D. R. (2007). Dissociating averageness and attractiveness: Attractive faces are not always average. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 1420-1430.
- Deffenbacher, K. A., Vetter, T., Johanson, J. & O'Toole, A. J. (1998). Facial aging, attractiveness, and distinctiveness. *Perception*, 27, 1233-1243.
- Denniston, A. K. O. & Murray, P. I. (2009). *Oxford Handbook of Ophthalmology*. Oxford: Oxford University Press.
- Deutsch, F. M., Zalenski, C. M. & Clark, M. E. (1986). Is there a double standard of aging? *Journal of Applied Social Psychology*, 16, 771-785.
- Dion, K. K. (1972). Physical attractiveness and evaluation of childrens transgressions. *Journal of Personality and Social Psychology*, 24, 207-213.
- Dion, K. K. (1974). Children's physical attractiveness and sex as determinants of adult punitiveness. *Developmental Psychology*, 10, 772-778.
- Dion, K. K., Berscheid, E. & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24, 285-290.
- Donofrio, L. M. (2000). Fat distribution: A morphologic study of the aging face. *Dermatologic Surgery*, 26, 1107-1112.

- Dunkle, J. H. & Francis, P. L. (1990). The role of facial masculinity/femininity in the attribution of homosexuality. *Sex Roles*, 23, 157-167.
- Eagly, A. H., Ashmore, R. D., Makhijani, M. G. & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110, 109-128.
- Eibl-Eibesfeldt, I. (1997). *Die Biologie des menschlichen Verhaltens: Grundriß der Humanethologie*. Weyarn: Seehamer.
- England, P. & McClintock, E. A. (2009). The gendered double standard of aging in us marriage markets. *Population and Development Review*, 35, 797-816.
- Enlow, D. H. (1966). A morphogenetic analysis of facial growth. *American Journal of Orthodontics*, 52, 283-299.
- Enlow, D. H. (1979). Facial growth and development. *International Journal of Oral Myology*, 5, 7-10.
- Enlow, D. H. (1989). *Handbuch des Gesichtswachstums*. Berlin: Quintessenz.
- Fagien, S. (2002). Advanced rejuvenative upper blepharoplasty: Enhancing aesthetics of the upper periorbita. *Plastic and Reconstructive Surgery*, 110, 278-291.
- Falser, G., Bendel, M., Voller, D. & Wegner, A. (2008). Women's attractiveness depends more on their age than men's: New evidence for a double standard of aging. *International Journal of Psychology*, 43, 76-76.
- Farkas, L. G. (1987). *Anthropometric facial proportions in medicine*. Springfield: Charles C. Thomas Publisher.
- Farkas, L. G. (1994). *Anthropometry of the head and face*. New York: Raven Press.
- Farkas, L. G. & Cheung, G. (1981). Facial asymmetry in healthy north american caucasians. An anthropometrical study. *Angle Orthodontist*, 51, 70-77.
- Farkas, L. G. & Posnick, J. C. (1992). Growth and development of regional units in the head and face based on anthropometric measurements. *Cleft Palate Craniofacial Journal Official Publication of the American Cleft Palate Craniofacial Association*, 29, 301-302.
- Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin*, 111, 304-341.
- Felson, R. B. (1980). Physical attractiveness, grades, and teachers' attributions of ability. *Representative Research in Social Psychology*, 11, 64-71.
- Folstad, I. & Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *American Naturalist*, 139, 603-622.
- Folstad, I., Nilssen, A. C., Halvorsen, O. & Andersen, J. (1989). Why do male reindeer (*rangifer-t tarandus*) have higher abundance of 2nd and 3rd instar larvae of hypoderma-tarandus tarandi than females. *Oikos*, 55, 87-92.
- Friedman, O. (2005). Changes associated with the aging face. *Facial Plastic Surgery Clinics of North America*, 13, 371-380.
- Galton, F. (1878). Composite portraits. *Journal of the Anthropological Institute of Great Britain & Ireland*, 8, 132-144.
- Galton, F. (1888). Personal identification and description. *Proceedings of the Royal Institution of Great Britain*, 12, 346-360.
- Gangestad, S. W. & Buss, D. M. (1993). Pathogen prevalence and human mate preferences. *Ethology & Sociobiology*, 14, 89-96.

- Gangestad, S. W. & Thornhill, R. (1997). Human sexual selection and developmental stability. In J. A. Simpson & D. T. Kenrick (Eds.), *Evolutionary social psychology* (pp. 169-196). Hillsdale: Erlbaum.
- Gangestad, S. W., Thornhill, R. & Yeo, R. A. (1994). Facial attractiveness, developmental stability, and fluctuating asymmetry. *Ethology and Sociobiology*, *15*, 73-85.
- Getty, T. (2002). Signaling health versus parasites. *American Naturalist*, *159*, 363-371.
- Giesen, C. B. (1989). Aging and attractiveness - marriage makes a difference. *International Journal of Aging & Human Development*, *29*, 83-94.
- Gilbert, C. & Bakan, P. (1973). Visual asymmetry in perception of faces. *Neuropsychologia*, *11*, 355-362.
- Glass, G. V., McGaw, B. & Smith, M. L. (1978). *Mata-analysis in social research*. London: Sage.
- Gosain, A., Klein, M. H., Sudhakar, P. V. & Prost, R. W. (2005). A volumetric analysis of soft-tissue changes in the aging midface using high-resolution mri: Implications for facial rejuvenation. *Plastic and Reconstructive Surgery*, *115*, 1143-1152.
- Graham, J. H., Emlen, J. M., Freeman, D. C., Leamy, L. J. & Kieser, J. A. (1998). Directional asymmetry and the measurement of developmental instability. *Biological Journal of the Linnean Society*, *64*, 1-16.
- Grammer, K. (2000). *Signale der Liebe*. München: Deutscher Taschenbuch Verlag.
- Grammer, K., Fink, B., Moller, A. P. & Manning, J. T. (2005). Physical attractiveness and health: Comment on Weeden and Sabini (2005). *Psychological Bulletin*, *131*, 658-661.
- Grammer, K., Fink, B., Moller, A. P. & Thornhill, R. (2003). Darwinian aesthetics: Sexual selection and the biology of beauty. *Biological reviews of the Cambridge Philosophical Society*, *78*, 385-407.
- Grammer, K. & Thornhill, R. (1994). Human (homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology*, *108*, 233-242.
- Grossman, C. J. (1985). Interactions between the gonadal-steroids and the immune system. *Science*, *227*, 257-261.
- Grumbach, M. M. (2000). Pubertal maturation in aromatase deficiency and resistance to estrogen. In J. P. Bourguignon & T. M. Plant (Eds.), *The onset of puberty in perspective* (pp. 247-267). Amsterdam: Elsevier.
- Gründl, M., Knoll, S., Eisenmann-Klein, M. & Prantl, L. (2011). The blue-eyes stereotype: Do eye color, pupil diameter, and scleral color affect attractiveness? *Aesthetic Plastic Surgery*, (in press).
- Guyuron, B. (1997). The aging nose. *Dermatologic Clinics*, *15*, 659-664.
- Halberstadt, J. & Rhodes, G. (2003). It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychonomic Bulletin and Review*, *10*, 149-156.
- Halleworden, J. (1902). Eine neue Methode experimenteller Physiognomik. *Psychiatrisch-Neurologische Wochenschrift*, *4*, 309-311.
- Halleworden, J. (1929). Rechts- und Linkshändigkeit und Gesichtsausdruck. *Zentralblatt für die gesamte Neurologie und Psychiatrie*, *53*, 560.
- Hamra, S. T. (1994). Surgery of the aging chin. *Plastic and Reconstructive Surgery*, *94*, 388-393.

- Hanae, I. & Jiro, G. (2001). Analyses of psychological aspects of attractiveness in feminized and juvenilized Japanese faces. *Tohoku Psychologica Folia*, 60, 29-34.
- Haubensak, G. (1985). *Absolutes und vergleichendes Urteil. Eine Einführung in die Theorie psychologischer Bezugssysteme*. Berlin: Springer.
- Heinz, G. W. & Kikkawa, D. O. (1997). The aging upper and middle face: An overview for the aesthetic surgeon. *Aesthetic Ophthalmic Plastic Surgery*, 37, 1-10.
- Henderson, J. J. A. & Anglin, J. M. (2003). Facial attractiveness predicts longevity. *Evolution and Human Behavior*, 24, 351-356.
- Henss, R. (1992). *"Spieglein, Spieglein an der Wand ..."* Weinheim: Psychologie Verlags Union.
- Henss, R. (1993). Kontexteffekte bei der Beurteilung der physischen Attraktivität. In M. Hassebrauck & R. Niketta (Hrsg.), *Physische Attraktivität* (S. 61-94). Göttingen: Hogrefe.
- Higgins, E. T. & Stangor, C. (1988). Context-driven social judgement and memory: When "Behavior engulfs the field" In reconstructive memory. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social memory of knowledge* (pp. 262-297). Cambridge: University Press.
- Hoening, J. A., Shorr, N. & Shorr, J. (1997). The suborbicularis oculi fat in aesthetic and reconstructive surgery. *Aesthetic Ophthalmic Plastic Surgery*, 37, 179-191.
- Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 199-209.
- Hönekopp, J., Bartholome, T. & Jansen, G. (2004). Facial attractiveness, symmetry, and physical fitness in young women. *Human Nature - an Interdisciplinary Biosocial Perspective*, 15, 147-167.
- Hönekopp, J., Becker, B. J. & Oswald, F. L. (2006). The meaning and suitability of various effect sizes for structured rater x ratee designs. *Psychological Methods*, 11, 72-86.
- Hosoda, M., Stone-Romero, E. F. & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel Psychology*, 56, 431-462.
- Hoss, R. A. & Langlois, J. H. (2003). Infants prefer attractive faces. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood: Current perspectives* (pp. 27-38). New York: Nova Science Publishers.
- Hume, D. K. & Montgomerie, R. (2001). Facial attractiveness signals different aspects of "quality" in women and men. *Evolution and Human Behavior*, 22, 93-112.
- Ishi, H., Gyoba, J. & Kamachi, M. (2004). The influence of dynamic presentation of smile on the perception of facial attributes. *Perception*, 33, 105-106.
- Ishi, H., Gyoba, J., Kamachi, M., Mukaida, S. & Akamatsu, S. (2004). Analyses of facial attractiveness on feminized and juvenilized faces. *Perception*, 33, 135-145.
- Jackson, L. A., Hunter, J. E. & Hodge, C. N. (1995). Physical attractiveness and intellectual competence - a metaanalytic review. *Social Psychology Quarterly*, 58, 108-122.
- Jelks, G. W. & Jelks, E. B. (1993). Preoperative evaluation of the blepharoplasty patient. Bypassing the pitfalls. *Clinics in Plastic Surgery*, 20, 213-223.

- Johnson, D. F. & Pittenger, J. B. (1984). Attribution, the attractiveness stereotype, and the elderly. *Developmental Psychology, 20*, 1168-1172.
- Johnston, V. S. (2000). Female facial beauty: The fertility hypothesis. *Pragmatics and Cognition, 8*, 107-122.
- Johnston, V. S. & Franklin, M. (1993). Is beauty in the eye of the beholder? *Ethology and Sociobiology, 14*, 183-199.
- Johnston, V. S., Hagel, R., Franklin, M., Fink, B. & Grammer, K. (2001). Male facial attractiveness: Evidence for hormone-mediated adaptive design. *Evolution and Human Behavior, 22*, 251-267.
- Johnston, V. S., Solomon, C. J., Gibson, S. J. & Pallares-Bejarano, A. (2003). Human facial beauty: Current theories and methodologies. *Archives of Facial Plastic Surgery, 5*, 371-377.
- Jones, B. C., DeBruine, L. M. & Little, A. C. (2007). The role of symmetry in attraction to average faces. *Perception and Psychophysics, 69*, 1273-1277.
- Jones, B. C., Little, A. C., Feinberg, D. R., Penton-Voak, I. S., Tiddeman, B. P. & Perrett, D. I. (2004). The relationship between shape symmetry and perceived skin condition in male facial attractiveness. *Evolution and Human Behavior, 25*, 24-30.
- Jones, B. C., Little, A. C., Penton-Voak, I. S., Tiddeman, B. P., Burt, D. M. & Perrett, D. I. (2001). Facial symmetry and judgements of apparent health: Support for a "good genes" explanation of the attractiveness-symmetry relationship. *Evolution and Human Behavior, 22*, 417-429.
- Jones, D. (1995). Sexual selection, physical attractiveness, and facial neoteny: Cross-cultural evidence and implications. *Current Anthropology, 36*, 723-748.
- Jones, D. & Hill, K. (1993). Criteria of facial attractiveness in five populations. *Human Nature, 4*, 271-296.
- Kalick, S. M., Zebrowitz, L.-A., Langlois, J.-H. & Johnson, R.-M. (1998). Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question. *Psychological Science, 9*, 8-13.
- Kant, I. (1796). *Anthropologie in pragmatischer Hinsicht*. Hamburg: Meiner.
- Keating, C. F. (1985). Gender and the physiognomy of dominance and attractiveness. *Social Psychology Quarterly, 48*, 61-70.
- Kenrick, D. T. & Gutierrez, S. E. (1980). Contrast effects and judgements of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology, 38*, 131-140.
- Kenrick, D. T., Gutierrez, S. E. & Goldberg, L. L. (1989). Influence of popular erotica on judgments of strangers and mates. *Journal of Experimental Social Psychology, 25*, 159-167.
- Koehler, N., Simmons, L. W., Rhodes, G. & Peters, M. (2004). The relationship between sexual dimorphism in human faces and fluctuating asymmetry. *Proceedings of the Royal Society B: Biological Sciences, 271*, 233-236.
- Koeslag, J. H. (1990). Koinophilia groups sexual creatures into species, promotes stasis, and stabilizes social behaviour. *Journal of Theoretical Biology, 144*, 15-35.
- Korczyński, A. D., Laor, N. & Nemet, P. (1976). Sympathetic pupillary tone in old-age. *Archives of Ophthalmology, 94*, 1905-1906.

- Kowner, R. (1996a). Facial asymmetry and attractiveness judgement in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 662-675.
- Kowner, R. (1996b). An improved technique for creating hemifacial chimeras. *Behavior Research Methods Instruments & Computers*, 28, 372-375.
- Krems, E.-B. (2005). *Der Fleck auf der Venus - 500 Künstleranekdoten von Apelles bis Picasso*. München: Beck.
- Langlois, J. H., Ritter, J. M., Roggman, L. A. & Vaughn, L. S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology*, 27, 79-84.
- Langlois, J. H. (1986). From the eye of the beholder to behavioral reality: The development of social behaviors and social relations as a function of physical attractiveness. In C. P. Herman, M. P. Zanna & E. T. Higgins (Eds.), *Physical appearance, stigma, and social behavior: The Ontario Symposium* (Vol. 3, pp. 23-51). Hillsdale, NJ: Erlbaum.
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M. & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126, 390-423.
- Langlois, J. H. & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, 1, 115-121.
- Langlois, J. H., Roggman, L. A., Casey, R. J., Ritter, J.-M., Rieser-Danner, L. A. & Jenkins, V. Y. (1987). Infant preferences for attractive faces: Rudiments of a stereotype? *Developmental Psychology*, 23, 363-369.
- Langlois, J. H., Roggman, L. A. & Musselman, L. (1994). What is average and what is not average about attractive faces? *Psychological Science*, 5, 214-220.
- Langlois, J. H., Roggman, L. A. & Rieser-Danner, L. A. (1990). Infants' differential social responses to attractive and unattractive faces. *Developmental Psychology*, 26, 153-159.
- Lerner, R. M. & Lerner, J. V. (1977). Effects of age, sex, and physical attractiveness on child-peer relations, academic performance, and elementary school adjustment. *Developmental Psychology*, 13, 585-590.
- Leung, A. K. (1998). Correcting for allometry in studies of fluctuating asymmetry and quality within samples. *Proceedings of the Royal Society B: Biological Sciences*, 265, 1623-1629.
- Leung, B., Forbes, M.-R. & Houle, D. (2000). Fluctuating asymmetry as a bioindicator of stress: Comparing efficacy of analyses involving multiple traits. *American Naturalist*, 155, 101-115.
- Light, L. L., Hollander, S. & Karya-Stuart, F. (1981). Why attractive people are harder to remember. *Personality and Social Psychology Bulletin*, 7, 269-276.
- Little, A. C., Burt, D. M., Penton-Voak, I. S. & Perrett, D. I. (2001). Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. *Proceedings of the Royal Society B: Biological Sciences*, 268, 39-44.
- Little, A. C. & Hancock, P. J. B. (2002). The role of masculinity and distinctiveness in judgments of human male facial attractiveness. *British Journal of Psychology*, 93, 451-464.

- Little, A. C. & Jones, B. C. (2003). Evidence against perceptual bias views for symmetry preferences in human faces. *Proceedings of the Royal Society B: Biological Sciences*, 270, 1759-1763.
- Lorenz, K. (1943). Die angeborenen Formen möglichen Verhaltens. *Zeitschrift für Tierpsychologie*, 5, 235-409.
- Martins, Y., Tiggemann, M. & Churchett, L. (2008). Hair today, gone tomorrow: A comparison of body hair removal practices in gay and heterosexual men. *Body Image*, 5, 312-316.
- Mazzella, R. & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology*, 24, 1315-1344.
- McArthur, L. Z. & Apatow, K. (1984). Impressions of baby-faced adults. *Social Cognition*, 2, 315-342.
- Mealey, L., Bridgstock, R. & Townsend, G. C. (1999). Symmetry and perceived facial attractiveness: A monozygotic co-twin comparison. *Journal of Personality and Social Psychology*, 76, 151-158.
- Mertens, I., Siegmund, H. & Grüsser, O. J. (1993). Gaze motor asymmetries in the perception of faces during a memory task. *Neuropsychologia*, 31, 989-998.
- Metzger, W. (1975). *Gesetze des Sehens*. Frankfurt a. M.: Kramer.
- Meyer, D. A. & Quong, M. W. (1999). The bio-logic of facial geometry. *Nature*, 397, 661-662.
- Moller, A. P., Christe, P. & Lux, E. (1999). Parasitism, host immune function, and sexual selection. *Quarterly Review of Biology*, 74, 3-20.
- Montepare, J. M. & Zebrowitz-McArthur, L. (1989). Childrens perceptions of baby-faced adults. *Perceptual and Motor Skills*, 69, 467-472.
- Morris, P. E. & Wickham, L. H. V. (2001). Typicality and face recognition: A critical re-evaluation of the two factor theory. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 54, 863-877.
- Neave, N., Laing, S., Fink, B. & Manning, J. T. (2003). Second to fourth digit ratio, testosterone and perceived male dominance. *Proceedings of the Royal Society B: Biological Sciences*, 270, 2167-2172.
- O'Toole, A. J., Deffenbacher, K. A., Valentin, D. & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory and Cognition*, 22, 208-224.
- O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D. & Abdi, H. (1998). The perception of face gender: The role of stimulus structure in recognition and classification. *Memory and Cognition*, 26, 146-160.
- Palmer, A. P. & Strobeck, C. (2003). Fluctuating asymmetry analyses revisited. In M. Polak (Ed.), *Developmental instability: Causes and consequences* (pp. 279-319). Oxford: Oxford University Press.
- Palmer, A. R. & Strobeck, C. (1986). Fluctuating asymmetry: Measurement, analysis, patterns. *Annual Review of Ecology and Systematics*, 17, 392-421.
- Pawlowski, B., Boothroyd, L. G., Perrett, D. I. & Kluska, S. (2008). Is female attractiveness related to final reproductive success? *Collegium Antropologicum*, 32, 457-460.
- Peck, S., Peck, L. & Kataja, M. (1991). Skeletal asymmetry in esthetically pleasing faces. *Angle Orthodontist*, 61, 43-48.

- Penton-Voak, I. S. & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior*, 25, 229-241.
- Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M. & Perrett, D. I. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society B: Biological Sciences*, 268, 1617-1623.
- Penton-Voak, I. S. & Perrett, D. I. (2001). Male facial attractiveness: Perceived personality and shifting female preferences for male traits across the menstrual cycle. *Advances in the Study of Behavior*, 30, 219-259.
- Penton-Voak, I. S., Perrett, D. I. & Peirce, J. W. (1999). Computer graphic studies of the role of facial similarity in judgements of attractiveness. *Current Psychology*, 18, 104-117.
- Perrett, D.-I., Burt, D. M., Penton-Voak, I.-S., Lee, K.-J., Rowland, D.-A. & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior*, 20, 295-307.
- Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi, S. P., Castles, D. L. & Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, 394, 884-887.
- Perrett, D. I., May, K. A. & Yoshikawa, S. (1994). Facial shape and judgments of female attractiveness. *Nature*, 368, 239-242.
- Peskin, M. & Newell, F. N. (2004). Familiarity breeds attraction: Effects of exposure on the attractiveness of typical and distinctive faces. *Perception*, 33, 147-157.
- Peters, A. (2000). Testosterone treatment is immunosuppressive in superb fairy-wrens, yet free-living males with high testosterone are more immunocompetent. *Proceedings of the Royal Society B: Biological Sciences*, 267, 883-889.
- Pittenger, J. B. (1991). On the difficulty of averaging faces - comments on Langlois and Roggman. *Psychological Science*, 2, 351-353.
- Pollard, J., Shepherd, J. & Shepherd, J. (1999). Average faces are average faces. *Current Psychology: Developmental, Learning, Personality, Social*, 18, 98-103.
- Pollard, J. S. (1995). Attractiveness of composite faces: A comparative study. *International Journal of Comparative Psychology*, 8, 77-83.
- Pospeschill, M. (2006). *Statistische Methoden - Strukturen, Grundlagen, Anwendungen in Psychologie und Sozialwissenschaften*. München: Spektrum.
- Poulin, R. (1996). Sexual inequalities in helminth infections: A cost of being a male? *American Naturalist*, 147, 287-295.
- Raffaello, S. & Camesasca, E. (Eds.). (1994). *Raffaello, gli scritti - lettere, firme, sonetti, saggi tecnici e teorici*. Milano: Rizzoli.
- Ramsey-Rennels, J. L. & Langlois, J. H. (2006). Infants' differential processing of female and male faces. *Current Directions in Psychological Science*, 15, 59-62.
- Ramsey, J. L., Langlois, J. H., Hoss, R. A., Rubenstein, A. J. & Griffin, A. M. (2004). Origins of a stereotype: Categorization of facial attractiveness by 6-month-old infants. *Developmental Science*, 7, 201-211.
- Rennels, J. L., Bronstad, P. M. & Langlois, J. H. (2008). Are attractive men's faces masculine or feminine? The importance of type of facial stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 884-893.

- Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, 199-226.
- Rhodes, G., Chan, J., Zebrowitz, L. A. & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proceedings of the Royal Society B: Biological Sciences*, 270, S93-95.
- Rhodes, G., Geddes, K., Jeffery, L., Dziurawiec, S. & Clark, A. (2002). Are average and symmetric faces attractive to infants? Discrimination and looking preferences. *Perception*, 31, 315-321.
- Rhodes, G., Harwood, K., Yoshikawa, S., Nishitani, M. & McLean, I. (2002). The attractiveness of average faces: Cross-cultural evidence and possible biological basis. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness. Evolutionary, cognitive, and social perspectives* (pp. 35-58). Westport, CT: Ablex.
- Rhodes, G., Hickford, C. & Jeffery, L. (2000). Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive? *British Journal of Psychology*, 91, 125-140.
- Rhodes, G., Proffitt, F., Grady, J. M. & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin and Review*, 5, 659-669.
- Rhodes, G., Roberts, J. & Simmons, L. W. (1999). Reflections on symmetry and attractiveness. *Psychology, Evolution, & Gender*, 1, 279-295.
- Rhodes, G., Simmons, L. W. & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success? *Evolution and Human Behavior*, 26, 186-201.
- Rhodes, G., Sumich, A. & Byatt, G. (1999). Are average facial configurations attractive only because of their symmetry? *Psychological Science*, 10, 52-58.
- Rhodes, G. & Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychological Science*, 7, 105-110.
- Rhodes, G., Yoshikawa, S., Clark, A., Lee, K., McKay, R. & Akamatsu, S. (2001). Attractiveness of facial averageness and symmetry in non-western cultures: In search of biologically based standards of beauty. *Perception*, 30, 611-625.
- Rhodes, G., Zebrowitz, L. A., Clark, A., Kalick, S. M., Hightower, A. & McKay, R. (2001). Do facial averageness and symmetry signal health? *Evolution and Human Behavior*, 22, 31-46.
- Rikowski, A. & Grammer, K. (1999). Human body odour, symmetry and attractiveness. *Proceedings of the Royal Society B: Biological Sciences*, 266, 869-874.
- Rohrich, R. J., Coberly, D. M., Fagien, S. & Stuzin, J. M. (2004). Current concepts in aesthetic upper blepharoplasty. *Plastic and Reconstructive Surgery*, 113, 32-42.
- Rosenkrantz, P. S., Vogel, S. R., Bee, H., Broverman, I. K. & Broverman, D. M. (1968). Sex role stereotypes and self-concept in college students. *Journal of Consulting and Clinical Psychology*, 32, 287-295.
- Rosenthal, R. & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.
- Rubenstein, A. J., Kalakanis, L. & Langlois, J. H. (1999). Infant preferences for attractive faces: A cognitive explanation. *Developmental Psychology*, 35, 848-855.
- Rubenstein, A. J., Langlois, J. H. & Roggman, L. A. (2002). What makes a face attractive and why: The role of averageness in defining facial beauty. In G. Rhodes & F. Zeng (Eds.), *Facial attractiveness. Evolutionary, cognitive, and social perspectives* (pp. 1-33). Westport, CT: Ablex.

- Sackeim, H. A. (1985). Morphologic asymmetries of the face: A review. *Brain and Cognition*, 4, 296-312.
- Salvia, J., Algozzine, R. & Sheare, J. B. (1977). Attractiveness and school achievement. *Journal of School Psychology*, 15, 60-66.
- Samuels, C. A., Butterworth, G., Roberts, T., Graupner, L. & Hole, G. (1994). Facial aesthetics: Babies prefer attractiveness to symmetry. *Perception*, 23, 823-831.
- Samuels, C. A. & Ewy, R. (1985). Aesthetic perception of faces during infancy. *British Journal of Developmental Psychology*, 3, 221-228.
- Sarno, J. A. & Alley, T. R. (1997). Attractiveness and the memorability of faces: Only a matter of distinctiveness? *American Journal of Psychology*, 110, 81-92.
- Scheib, J. E., Gangestad, S. W. & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society B: Biological Sciences*, 266, 1913-1917.
- Shackelford, T. K. & Larsen, R. J. (1997). Facial asymmetry as an indicator of psychological, emotional, and physiological distress. *Journal of Personality and Social Psychology*, 72, 456-466.
- Shackelford, T. K. & Larsen, R. J. (1999). Facial attractiveness and physical health. *Evolution and Human Behavior*, 20, 71-76.
- Simmons, L. W., Rhodes, G., Peters, M. & Koehler, N. (2004). Are human preferences for facial symmetry focused on signals of developmental instability? *Behavioral Ecology*, 15, 864-871.
- Slater, A., Von-der-Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S. & Samuels, C. (1998). Newborn infants prefer attractive faces. *Infant Behavior and Development*, 21, 345-354.
- Swaddle, J. P. & Cuthill, I. C. (1995). Asymmetry and human facial attractiveness: Symmetry may not always be beautiful. *Proceedings of the Royal Society B: Biological Sciences*, 261, 111-116.
- Swaddle, J. P. & Reiersen, G. W. (2002). Testosterone increases perceived dominance but not attractiveness in human males. *Proceedings of the Royal Society B: Biological Sciences*, 269, 2285-2289.
- Symons, D. (1979). *The evolution of human sexuality*. Oxford: Oxford University Press.
- Symons, D. (1995). Beauty is in the adaptations of the beholder: The evolutionary psychology of human female sexual attractiveness. In *Chicago series on sexuality, history, and society; sexual nature sexual culture* (pp. 80-118). Chicago: University of Chicago Press.
- Symons, D. P. (1987). An evolutionary approach. Can Darwin's view of life shed light on human sexuality? In J. H. Geer & W. T. O'Donohue (Eds.), *Theories of human sexuality* (pp. 91-125). New York: Plenum Press.
- Tanner, J. M. (1978). *Foetus into man: Physical growth from conception to maturity*. London: Open Books.
- Teuscher, U. & Teuscher, C. (2007). Reconsidering the double standard of aging: Effects of gender and sexual orientation on facial attractiveness ratings. *Personality and Individual Differences*, 42, 631-639.
- Thornhill, R. & Gangestad, S.-W. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature*, 4, 237-269.

- Thornhill, R. & Gangestad, S.-W. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, 3, 452-460.
- Thornhill, R. & Gangestad, S. W. (1996). The evolution of human sexuality. *Trends in Ecology and Evolution*, 11, 98-102.
- Thornhill, R. & Moller, A. P. (1997). Developmental stability, disease and medicine. *Biological Reviews of the Cambridge Philosophical Society*, 72, 497-548.
- Tipples, J., Atkinson, A. P. & Young, A. W. (2002). The eyebrow frown: A salient social signal. *Emotion*, 2, 288-296.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 43, 161-204.
- Valentine, T., Darling, S. & Donnelly, M. (2004). Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin and Review*, 11, 482-487.
- Vokey, J. R. & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, 20, 291-302.
- Volpe, C. R. & Ramirez, O. M. (2005). The beautiful eye. *Facial Plastic Surgery clinics of North America*, 13, 493-504.
- Wedekind, C. (1992). Detailed information about parasites revealed by sexual ornamentation. *Proceedings of the Royal Society B: Biological Sciences*, 247, 169-174.
- Weeden, J. & Sabini, J. (2005). Physical attractiveness and health in Western societies: A review. *Psychological Bulletin*, 131, 635-653.
- Westmore, M. G. (1974). *Facial cosmetics in conjunction with surgery*. Paper presented at the Aesthetic Plastic Surgical Society Meeting. Vancouver, British Columbia, May 7, 1974.
- Wickham, L. H. V. & Morris, P. E. (2003). Attractiveness, distinctiveness, and recognition of faces: Attractive faces can be typical or distinctive but are not better recognized. *American Journal of Psychology*, 116, 455-468.
- Wilcox, S. (1997). Age and gender in relation to body attitudes - is there a double standard of aging? *Psychology of Women Quarterly*, 21, 549-565.
- Williams, G. C. (1975). *Sex and evolution*. Princeton, NJ: Princeton University Press.
- Winkielman, P., Halberstadt, J., Fazendeiro, T. & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science*, 17, 799-806.
- Winkielman, P., Schwarz, N., Reber, R. & Fazendeiro, T. A. (2003). Cognitive and affective consequences of visual fluency: When seeing is easy on the mind. In L. M. Scott & R. Batra (Eds.), *Persuasive imagery: A consumer response perspective* (pp. 75-89). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Wolf, N. (1991). *The beauty myth: How images of beauty are used against women*. New York: W. Morrow.
- Wolff, W. (1933). The experimental study of forms of expression. *Character and Personality*, 2, 169-176.
- Yaremchuk, M. J. (2004). Improving periorbital appearance in the "morphologically prone". *Plastic and Reconstructive Surgery*, 114, 980-987.
- Zahavi, A. (1975). Mate selection: A selection for a handicap. *Journal of Theoretical Biology*, 53, 205-214.

- Zaidel, D. W. & Deblieck, C. (2007). Attractiveness of natural faces compared to computer constructed perfectly symmetrical faces. *International Journal of Neuroscience*, 117, 423-431.
- Zebrowitz-McArthur, L. & Montepare, J. M. (1989). Contributions of a babyface and a childlike voice to impressions of moving and talking faces. *Journal of Nonverbal Behavior*, 13, 189-203.
- Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Boulder, Colorado: Westview Press.
- Zebrowitz, L. A., Kendall-Tackett, K. & Fafel, J. (1991). The influence of children facial maturity on parental expectations and punishments. *Journal of Experimental Child Psychology*, 52, 221-238.
- Zebrowitz, L. A. & McDonald, S. M. (1991). The impact of litigants baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, 15, 603-623.
- Zebrowitz, L. A. & Montepare, J. M. (1992). Impressions of babyfaced individuals across the life-span. *Developmental Psychology*, 28, 1143-1152.
- Zebrowitz, L. A., Olson, K. & Hoffman, K. (1993). Stability of babyfacedness and attractiveness across the life-span. *Journal of Personality and Social Psychology*, 64, 453-466.
- Zebrowitz, L. A., Tenenbaum, D. R. & Goldstein, L. H. (1991). The impact of job applicants facial maturity, gender, and academic-achievement on hiring recommendations. *Journal of Applied Social Psychology*, 21, 525-548.
- Zebrowitz, L. A., Voinescu, L. & Collins, M.-A. (1996). "Wide-eyed" and "crooked-faced": Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin*, 22, 1258-1269.
- Zimble, M. S., Kokoska, M. S. & Thomas, J. R. (2001). Anatomy and pathophysiology of facial aging. *Facial Plastic Surgery Clinics of North America*, 9, 179-187.
- Zuk, M., Johnsen, T. S. & Maclarty, T. (1995). Endocrine-immune interactions, ornaments and mate choice in red jungle fowl. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 260, 205-210.

13 Tabellen- und Abbildungsverzeichnis

Tabellen

- Tabelle 1: Die Ergebnisse der drei ANOVAs mit Tests der Haupteffekte und Interaktionseffekte bei den drei Experimenten mit männlichen Durchschnittsgesichtern. 36
- Tabelle 2: Ergebnisse der dreifaktoriellen ANOVA. Einfluss von Durchschnittlichkeit, Attraktivität der Originalgesichter und Stimulusart auf die Attraktivität eines männlichen Composites..... 40
- Tabelle 3: Die Ergebnisse der drei ANOVAs mit Tests der Haupteffekte und Interaktionseffekte bei den drei Experimenten mit weiblichen Durchschnittsgesichtern .. 65
- Tabelle 4: Ergebnisse der dreifaktoriellen ANOVA. Einfluss von Durchschnittlichkeit, Attraktivität der Originalgesichter und Stimulusart auf die Attraktivität eines weiblichen Composites..... 70
- Tabelle 5: Übersicht der Composites, die in der qualitativen Analyse zur Identifikation von Attraktivitätsmerkmalen miteinander verglichen wurden..... 92
- Tabelle 6: Ergebnisse der deskriptiven Analyse von Composites in der *Frontalansicht*. Unterschiede zwischen Composites aus hochattraktiven Gesichtern und durchschnittlichen Composites bzw. Composites aus unattraktiven Gesichtern, getrennt nach Frauengesichtern und Männergesichtern. Aufgelistet sind die Merkmale der attraktiven Composites – für die unattraktiven gelten entgegengesetzte Merkmale 93
- Tabelle 7: Ergebnisse der deskriptiven Analyse von Composites in der *Profilansicht*. Unterschiede zwischen Composites aus hochattraktiven Gesichtern und durchschnittlichen Composites bzw. Composites aus unattraktiven Gesichtern, getrennt nach Frauengesichtern und Männergesichtern. Aufgelistet sind die Merkmale der attraktiven Composites – für die unattraktiven gelten entgegengesetzte Merkmale 96
- Tabelle 8: Ergebnisse der beiden Paarvergleichs-Experimente mit Chimärengesichtern aus linken und rechten Gesichtshälften im direkten Vergleich mit dem jeweiligen Originalgesicht. Aufgelistet sind absolute und relative Häufigkeiten sowie die in r umgerechneten Effektgrößen. Bei beiden Experimenten war jeweils $N = 64$ 138

Tabelle 9: Vergleich der demografischen Daten der Versuchspersonen in den drei Versuchsbedingungen.....	143
Tabelle 10: Ergebnisse der beiden Rating-Experimente mit Chimärengesichtern aus linken und rechten Gesichtshälften sowie den jeweiligen Originalgesichtern (= Kontrollgruppe). Aufgelistet sind die Attraktivitätsmittelwerte der jeweiligen Stimuli. Die letzten beiden Spalten listen die Effekte der Symmetrisierung in Form von Mittelwertsdifferenzen auf. In allen drei Experimentalgruppen war jeweils $N = 64$	147
Tabelle 11: Ergebnisse des Paarvergleichs-Experiments mit symmetrisierten gemorphten Gesichtern mit Farbwertmittelung („Blends“) im direkten Vergleich mit dem jeweiligen Originalgesicht. Aufgelistet sind absolute und relative Häufigkeiten sowie die in r umgerechneten Effektgrößen. $N = 64$	155
Tabelle 12: Ergebnisse des Rating-Experiments mit durch das Morphing-Verfahren mit Mittelwertbildung symmetrisch optimierten Gesichtern sowie den jeweiligen Originalgesichtern (= Kontrollgruppe). Aufgelistet sind die Attraktivitätsmittelwerte der jeweiligen Stimuli. Die drittletzte Spalte listet den Effekt der Symmetrisierung in Form von Mittelwertsdifferenzen sowie in Klammern Cohen's d auf. Die beiden letzten Spalten listen F - und p -Werte des Tests für Zwischensubjekteffekte der MANOVA auf. In beiden Versuchsgruppen war jeweils $N = 64$	160
Tabelle 13: Ergebnisse des Paarvergleichs-Experiments mit symmetrisierten gemorphten Gesichtern <i>ohne</i> Farbwertmittelung („Warps“) im direkten Vergleich mit dem jeweiligen Originalgesicht. Aufgelistet sind absolute und relative Häufigkeiten sowie die in r umgerechneten Effektgrößen. $N = 64$	170
Tabelle 14: Ergebnisse des Rating-Experiments mit durch das Morphing-Verfahren <i>ohne</i> Mittelwertbildung symmetrisch optimierten Gesichtern („Warps“) sowie den jeweiligen Originalgesichtern (= Kontrollgruppe). Aufgelistet sind die Attraktivitätsmittelwerte der jeweiligen Stimuli. Die drittletzte Spalte listet den Effekt der Symmetrisierung in Form von Mittelwertsdifferenzen sowie in Klammern Cohen's d auf. Die beiden letzten Spalten listen F - und p -Werte des Tests für Zwischensubjekteffekte der MANOVA auf. In beiden Versuchsgruppen war jeweils $N = 64$	174
Tabelle 15: Überblick über die Effektgrößen (standardisiert nach r) aller neun Untersuchungen zur Symmetrie-Hypothese in Abhängigkeit von der Berücksichtigung der Varianz von Versuchspersonen-Urteilen bei der Datenanalyse.....	205

Tabelle 16: Korrelationsmatrix. Der Zusammenhang (r) zwischen der Attraktivität der Originalgesichter und der Attraktivität symmetrisierter Varianten (Chimärengesichter, Blends und Warps). Zusätzlich die Korrelationen zwischen verschiedenen Symmetrie-Maßen (Symmetrie-Rating Originale, Ähnlichkeitsrating Gesichtshälften und Ähnlichkeitsrating Chimärengesichter) und der Attraktivität symmetrisierter Gesichter..221

Tabelle 17: Simulation der Korrelationskoeffizienten (= Effektstärken) für *Sym 7* unter der Annahme unterschiedlich großer Versuchspersonen-Stichproben.....231

Tabelle 18: Absolute und relative Häufigkeiten der Bevorzugung der einzelnen Gesichtervarianten für die Stimuluspersonen (A bis F) in Abhängigkeit vom Kindchenschema-Anteil des Gesichts. Die beiden letzten Zeilen listen die Testwerte und p -Werte der einzelnen χ^2 -Anpassungstests auf.257

Tabelle 19: Ergebnisse des Paarvergleichs-Experiments mit „verkindlichten“ Frauengesichtern (40 % Kindchenschema) im direkten Vergleich mit dem jeweiligen unveränderten erwachsenen Gesicht. Aufgelistet sind absolute und relative Häufigkeiten sowie die in r umgerechneten Effektgrößen. $N = 104$265

Tabelle 20: Ergebnisse des Paarvergleichs-Experiments mit „verkindlichten“ Männergesichtern (40 % Kindchenschema) im direkten Vergleich mit dem jeweiligen unveränderten erwachsenen Gesicht. Aufgelistet sind absolute und relative Häufigkeiten sowie die in r umgerechneten Effektgrößen. $N = 104$272

Tabelle 21: Die mittleren Ränge für die einzelnen Varianten (Durchschnitt, Kindchenschema, Symmetrie und Original) bei den einzelnen männlichen Stimulusgesichtern („m1“ bis „m6“). Tests nach Friedman. In allen Fällen $N = 324$ und $df = 3$319

Tabelle 22: Paarweise Tests nach Wilcoxon für Attraktivitätsunterschiede zwischen den jeweiligen Gesichtervarianten der Männergesichter. Z -Werte und p -Werte. $N = 324$322

Tabelle 23: Die mittleren Ränge für die einzelnen Varianten (Durchschnitt, Kindchenschema, Symmetrie und Original) bei den einzelnen weiblichen Stimulusgesichtern („w1“ bis „w6“). Tests nach Friedman. In allen Fällen $N = 324$ und $df = 3$323

Tabelle 24: Die mittleren Ränge für die Varianten (Durchschnitt, Kindchenschema, Symmetrie und Original) bei weiblichen sowie männlichen Stimulusgesichtern. Tests nach Friedman. In beiden Fällen $N = 324$ und $df = 3$	326
Tabelle 25: Paarweise Tests nach Wilcoxon für Attraktivitätsunterschiede zwischen den jeweiligen Gesichtervarianten der Frauengesichter. Z-Werte und p -Werte. $N = 324$	326
Tabelle 26: Ergebnisse der vier Tests nach Mann-Whitney, mit denen für die zusammengefassten Daten der sechs <i>männlichen</i> Stimulusgesichter überprüft wurde, ob ein Zusammenhang zwischen dem mittleren Attraktivitätsrang einer Gesichtsvariante und dem Geschlecht der Beurteiler bestand.....	327
Tabelle 27: Ergebnisse der vier Tests nach Mann-Whitney, mit denen für die zusammengefassten Daten der sechs <i>weiblichen</i> Stimulusgesichter überprüft wurde, ob ein Zusammenhang zwischen dem mittleren Attraktivitätsrang einer Gesichtsvariante und dem Geschlecht der Beurteiler bestand.....	328
Tabelle 28: Ausschnitt aus dem Versuchsplan des Experiments <i>Vergleich 2</i> . Insgesamt gab es vier unabhängige Gruppen von Versuchspersonen – so viele wie Versuchsbedingungen. Jede Gruppe sah von einer Stimulusperson stets nur eine Variante. Das Versuchsdesign war vollständig ausbalanciert.....	331
Tabelle 29: Die mittleren Attraktivitätsbewertungen für die Varianten der <i>Männern</i> gesichter mit den Testergebnissen der einfaktoriellen ANOVAs. Die drittletzte Spalte listet als Effektgröße Cohen's d auf und vergleicht jeweils die Attraktivität einer manipulierten Variante mit der Attraktivität der unveränderten Originalvariante.....	333
Tabelle 30: Die mittleren Attraktivitätsbewertungen für die Varianten der <i>Frau</i> engesichter mit den Testergebnissen der einfaktoriellen ANOVAs. Die drittletzte Spalte listet als Effektgröße Cohen's d auf und vergleicht jeweils die Attraktivität einer manipulierten Variante mit der Attraktivität der unveränderten Originalvariante.....	337

Abbildungen

Abbildung 1: Screenshot aus einer Morphing-Software (*MorphMan 4.0*), hier mit Frauengesichtern. Die weißen Punkte sind die definierten Referenzpunkte, die blauen Linien definieren die Konturen der einzelnen Gesichtsmerkmale. Beim Verrechnen der beiden Gesichter werden die korrespondierenden Referenzpunkte mit den entsprechenden Konturen jeweils aufeinander abgebildet. 14

Abbildung 2: Ein Beispiel für das Morphen von Gesichtern, hier mit den Frauengesichtern aus Abbildung 1. Aus den Gesichtern a und b ergibt sich bei einer Gewichtung von 50:50 das Gesicht c, das die beiden Ausgangsgesichter zu gleichen Anteilen enthält. 15

Abbildung 3: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Männergesichter in der Frontalansicht. Je durchschnittlicher ein Gesicht ist, desto attraktiver wird es tendenziell auch bewertet. Allerdings gibt es etliche Gesichter, die weniger durchschnittlich als das Gesamtdurchschnittsgesicht aus allen 32 Gesichtern (Generation 5) sind und trotzdem attraktiver bewertet werden. 18

Abbildung 4: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Männergesichter in der Frontalansicht. Je attraktiver die Originalgesichter waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver wurden auch die daraus resultierenden gemorphten Gesichter beurteilt. 19

Abbildung 5: „m(1-4)“, ein Composite aus den vier unattraktivsten männlichen Originalgesichtern (links) und „m(29-32)“, ein Composite aus den vier attraktivsten männlichen Originalgesichtern. Nach der Durchschnittshypothese müssten sie gleich attraktiv sein, da sie gleich durchschnittlich sind, tatsächlich ist jedoch der Durchschnitt aus den Schönsten weit attraktiver (mittlerer Attraktivitätswert von 5.55 ($SD = 1.01$) vs. 3.94 (1.05), $p < .001$). 20

Abbildung 6: „m(1-32)“, das Composite aus allen 32 männlichen Originalgesichtern (links) und „m(25-32)“, ein Composite aus den acht attraktivsten männlichen Originalgesichtern (rechts). Nach der Durchschnittshypothese müsste das Gesamtdurchschnittsgesicht „m(1-32)“ attraktiver sein, da es durchschnittlicher ist als „m(25-32)“, tatsächlich ist es jedoch umgekehrt (mittlerer Attraktivitätswert von „m(1-32)“ = 5.19 ($SD = 1.24$) vs. „m(25-32)“ = 5.58 ($SD = 1.12$), $p = .004$). 21

Abbildung 7: Ein Schattenprofil eines männlichen Originalgesichts, wie es beim Vortest als Stimulusmaterial verwendet wurde.....	23
Abbildung 8: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Männerprofile mit Texturinformation. Je durchschnittlicher ein Profil ist, desto attraktiver wird es tendenziell auch bewertet. Allerdings gibt es etliche Gesichter, die weniger durchschnittlich als das Gesamtdurchschnittsgesicht aus 32 Gesichtern (Generation 5) sind und trotzdem attraktiver bewertet werden.....	26
Abbildung 9: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Männerprofile mit Texturinformation. Je attraktiver die Originalgesichter waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver wurden auch die daraus resultierenden gemorphten Gesichter beurteilt.	27
Abbildung 10: „m(1-4)“, ein Composite aus den vier unattraktivsten Originalprofilen (links) und „m(29-32)“, ein Composite aus den vier attraktivsten Originalprofilen. Nach der Durchschnittshypothese müssten sie gleich attraktiv sein, da sie gleich durchschnittlich sind, tatsächlich ist jedoch der Durchschnitt aus den Schönsten weit attraktiver (mittlerer Attraktivitätswert von 4.56 ($SD = 1.05$) vs. 2.67 ($SD = 1.02$), $p < .001$).....	28
Abbildung 11: „m(1-32)“, das Composite aus allen 32 Originalprofilen (links) und „m(25-32)“, ein Composite aus den acht attraktivsten Originalprofilen (rechts). Nach der Durchschnittshypothese müsste das absolute Durchschnittsgesicht „m(1-32)“ attraktiver sein, da es durchschnittlicher ist als „m(25-32)“, tatsächlich ist es jedoch umgekehrt (mittlerer Attraktivitätswert von „m(1-32)“ = 4.48 ($SD = 1.09$) vs. „m(25-32)“ = 5.02 ($SD = 0.89$), $p < .001$).....	29
Abbildung 12: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Männerprofile ohne Texturinformation (= Schattenprofile). Je durchschnittlicher ein Profil ist, desto attraktiver wird es tendenziell auch bewertet. Allerdings gibt es etliche Schattenprofile, die weniger durchschnittlich als das Gesamt-Durchschnittsgesicht aus 32 Schattenprofilen (Generation 5) sind und trotzdem attraktiver bewertet werden. Der Zusammenhang ist jedoch eher gering ($r_s = .38$, $p = .037$).....	32

Abbildung 13: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Männerprofile ohne Texturinformation (= Schattenprofile). Je attraktiver die originalen Schattenprofile waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver wurden auch die daraus resultierenden gemorphten Schattenprofile beurteilt. Der Zusammenhang ist sehr hoch ($r_s = .86, p < .001$). 33

Abbildung 14: „m(1-32)“, das Composite aus allen 32 Originalschattenprofilen (links) und „m(25-32)“, ein Composite aus den acht attraktivsten Originalschattenprofilen (rechts). Nach der Durchschnittshypothese müsste das absolute Durchschnittsgesicht „m(1-32)“ attraktiver sein, da es durchschnittlicher ist als „m(25-32)“, tatsächlich ist es jedoch auch hier wieder umgekehrt (mittlerer Attraktivitätswert von „m(1-32)“ = 4.61 ($SD = 1.02$) vs. „m(25-32)“ = 4.96 ($SD = 1.11$), ($t(197) = 4.02, p < .001$). 34

Abbildung 15: Die mittlere Attraktivitätsbewertung von gemorphten Männergesichtern in der Frontalansicht in Abhängigkeit von der ursprünglichen Attraktivität der Originalgesichter (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). Die beiden Haupteffekte sind klar erkennbar. 37

Abbildung 16: Die mittlere Attraktivitätsbewertung von gemorphten Männerprofilen mit Textur in Abhängigkeit von der ursprünglichen Attraktivität der Originalprofile (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). 38

Abbildung 17: Die mittlere Attraktivitätsbewertung von gemorphten Männerprofilen ohne Textur (= Schattenprofile) in Abhängigkeit von der ursprünglichen Attraktivität der Originalprofile (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). Hier besteht nur noch ein Haupteffekt der Attraktivität der Originalgesichter. Der Effekt der Durchschnittlichkeit wird nicht mehr signifikant. 39

Abbildung 18: Die mittlere Attraktivität der gemorphten Männergesichter in Abhängigkeit von ihrer Durchschnittlichkeit und der Art der Stimuli. Der Interaktionseffekt zwischen Durchschnittlichkeit und Stimulusart wird durch die Schattenprofile ohne Textur verursacht, da bei ihnen die Attraktivität fast unabhängig von der Durchschnittlichkeit ist. 41

Abbildung 19: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Frauengesichter in der Frontalansicht. Je durchschnittlicher ein Gesicht ist, desto attraktiver wird es tendenziell auch bewertet.

Allerdings gibt es etliche Gesichter, die weniger durchschnittlich als das Gesamtdurchschnittsgesicht aus allen 64 Gesichtern (Generation 6) sind und trotzdem attraktiver bewertet werden. 46

Abbildung 20: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Frauengesichter in der Frontalansicht. Je attraktiver die Originalgesichter waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver werden auch die daraus resultierenden gemorphten Gesichter beurteilt. 47

Abbildung 21: „w(1-4)“, ein Composite aus den vier unattraktivsten weiblichen Originalgesichtern (links) und „w(61-64)“, ein Composite aus den vier attraktivsten weiblichen Originalgesichtern. Nach der Durchschnittshypothese müssten sie gleich attraktiv sein, da sie gleich durchschnittlich sind, tatsächlich ist jedoch der Durchschnitt aus den Schönsten weit attraktiver (mittlerer Attraktivitätswert von 6.19 ($SD = 0.80$) vs. 3.29 ($SD = 0.96$), $p < .001$). 48

Abbildung 22: „w(1-64)“, das Composite aus allen 64 weiblichen Originalgesichtern (links) und „w(57-64)“, ein Composite aus den acht attraktivsten Originalgesichtern (rechts). Nach der Durchschnittshypothese müsste das Gesamtdurchschnittsgesicht „w(1-64)“ attraktiver sein, da es durchschnittlicher ist als „w(57-64)“, tatsächlich ist es jedoch umgekehrt (mittlerer Attraktivitätswert von 6.33 ($SD = 0.72$) vs. 5.56 ($SD = 1.10$), $p < .001$). 49

Abbildung 23: Beispiel für Bildretusche. Zwei Frauenprofile (a und b) werden miteinander vermorpht und ergeben ein Composite (c), das an den Haaren starke Artefakte enthält. Diese Bildstörungen werden wegretuschiert und ergeben ein natürlich aussehendes Frauenprofil (d), dem man die Künstlichkeit praktisch nicht ansieht. Bereits die Bilder a und b sind gemorphte Profile (ebenfalls mit nachträglichen Bildretuschen an der Frisur), die aus je zwei Originalfotos erstellt wurden. 53

Abbildung 24: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Frauenprofile mit Texturinformation. Je durchschnittlicher ein Profil ist, desto attraktiver wird es tendenziell auch bewertet. Allerdings gibt es etliche Gesichter, die weniger durchschnittlich als das Gesamtdurchschnittsgesicht aus 32 Gesichtern (Generation 5) sind und trotzdem attraktiver bewertet werden. 55

Abbildung 25: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Frauenprofile mit Texturinformation. Je attraktiver die Originalgesichter waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver wurden auch die daraus resultierenden gemorphten Gesichter beurteilt. Die Lücke in der Punkteverteilung resultiert aus dem Umstand, dass bei der Auswahl der 32 zu vermorphenden Gesichter (aus 78 Originalen) weniger Gesichter mit mittlerer Attraktivität verwendet wurden. 56

Abbildung 26: „w(1-4)“, ein Composite aus den vier unattraktivsten weiblichen Originalprofilen (links) und „w(29-32)“, ein Composite aus den vier attraktivsten weiblichen Originalprofilen. Nach der Durchschnittshypothese müssten sie gleich attraktiv sein, da sie gleich durchschnittlich sind, tatsächlich ist jedoch der Durchschnitt aus den Schönsten weit attraktiver (mittlerer Attraktivitätswert von 5.20 ($SD = 1.11$) vs. 2.13 ($SD = 1.06$), $p < .001$). 57

Abbildung 27: „w(1-32)“, das Composite aus 32 Originalprofilen (links) und „w(25-32)“, ein Composite aus den acht attraktivsten Originalprofilen (rechts). Nach der Durchschnittshypothese müsste das Gesamt-Durchschnittsgesicht „w(1-32)“ attraktiver sein, da es durchschnittlicher ist als „w(25-32)“, tatsächlich ist jedoch umgekehrt (mittlerer Attraktivitätswert von „w(1-32)“ = 5.02 ($SD = 1.18$) vs. „w(25-32)“ = 5.76 ($SD = 0.99$), $p < .001$). 58

Abbildung 28: Die Korrelation zwischen Generation (d. h. Durchschnittlichkeit) und der mittleren Attraktivitätsbewertung der gemorphten Frauenprofile ohne Texturinformation (= Schattenprofile). Je durchschnittlicher ein Profil ist, desto attraktiver wird es tendenziell auch bewertet. Der Zusammenhang ist jedoch nur sehr schwach ($r_s = .30$) und nicht signifikant. Zudem gibt es etliche Schattenprofile, die weniger durchschnittlich als das Durchschnittsgesicht aus 32 Schattenprofilen (Generation 5) sind und trotzdem attraktiver bewertet werden. 61

Abbildung 29: Die Korrelation zwischen dem theoretischen Attraktivitätswert und der mittleren (empirischen) Attraktivitätsbewertung der gemorphten Frauenprofile ohne Texturinformation (= Schattenprofile). Je attraktiver die originalen Schattenprofile waren, aus denen die Composites erstellt wurden (je höher der theoretische Attraktivitätswert), desto attraktiver werden auch die daraus resultierenden gemorphten Schattenprofile beurteilt. Der Zusammenhang ist sehr stark ($r_s = .87$, $p < .001$). 62

Abbildung 30: „w(1-32)“, das Composite aus 32 Originalschattenprofilen (links) und „w(25-32)“, ein Composite aus den acht attraktivsten Originalschattenprofilen (rechts). Nach der Durchschnittshypothese müsste das absolute Durchschnittsgesicht „w(1-32)“ attraktiver sein, da es durchschnittlicher ist als „w(25-32)“, tatsächlich ist es jedoch auch hier wieder umgekehrt (mittlerer Attraktivitätswert von „w(1-32)“ = 4.85 (*SD* = 1.00) vs. „w(25-32)“ = 5.23 (*SD* = 1.10), (*t*(39) = 2.02, *p* = .050)..... 63

Abbildung 31: Die mittlere Attraktivitätsbewertung von gemorphten Frauengesichtern in der Frontalansicht in Abhängigkeit von der ursprünglichen Attraktivität der Originalgesichter (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). Die beiden Haupteffekte sind klar erkennbar. 67

Abbildung 32: Die mittlere Attraktivitätsbewertung von gemorphten Frauenprofilen mit Textur in Abhängigkeit von der ursprünglichen Attraktivität der Originalprofile (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). Der „Sprung“ zwischen dem 2. und 3. Quartil resultiert aus dem Umstand, dass in der Auswahl der 32 (von den ursprünglich 78 Originalgesichtern) zu vermorphenden Profile solche Gesichter mit mittlerer Attraktivität unterrepräsentiert waren. 68

Abbildung 33: Die mittlere Attraktivitätsbewertung von gemorphten Frauenprofilen ohne Textur (= Schattenprofile) in Abhängigkeit von der ursprünglichen Attraktivität der Originalprofile (unterteilt nach Quartilen) und der Durchschnittlichkeit (= Generation). Hier besteht nur noch ein Haupteffekt der Attraktivität der Originalgesichter. Der Effekt der Durchschnittlichkeit wird nicht mehr signifikant..... 69

Abbildung 34: Die mittlere Attraktivität der gemorphten Frauengesichter in Abhängigkeit von ihrer Durchschnittlichkeit und der Art der Stimuli. Der Interaktionseffekt zwischen Durchschnittlichkeit und Stimulusart wird durch die Schattenprofile ohne Textur verursacht, da bei ihnen die Attraktivität fast unabhängig von der Durchschnittlichkeit ist. 71

Abbildung 35: Die Münder von vier Originalgesichtern (A, B, C and D, von links nach rechts), die in der Studie zu einem Durchschnittsgesicht gemorpht wurden. 78

Abbildung 36: Bildausschnitte (240 x 170 Pixel) einzelner Gesichter im Vergleich: Bild ganz links: Das Composite aus den Gesichtern A und B; zweites Bild von links: Das Composite aus den Gesichtern A, B, C und D; drittes Bild von links: Das Gesicht A weichgezeichnet (Gaußscher Weichzeichnungsfiler mit 4 Pixeln Radius); Bild ganz

rechts: Das Composite aus den Gesichtern A, B, C und D weichgezeichnet (Gaußscher Weichzeichnungsfiter mit 4 Pixeln Radius). 79

Abbildung 37: Drei Ausschnitte (200 x 430 Pixel) aus einem Originalgesicht (die ersten drei Bilder von links) im Vergleich mit einem gemorphten Gesicht (Bild ganz rechts). Eine gerötete Hautstelle bei einem Originalgesicht unterhalb des Ohres ohne Filter (linkes Bild), derselbe Bildausschnitt mit einem Gaußschen Weichzeichnungsfiter von 4 Pixeln Radius (zweites Bild von links) und mit 8 Pixeln Radius (drittes Bild von links). Egal wie stark man weichzeichnet, die Rötung bleibt immer erkennbar. Zum Vergleich (ganz rechts) das *2-face-composite*, das dieses Original zu 50 % enthält und zu 50 % ein anderes Originalgesicht, das an dieser Stelle eine gesunde Haut hatte. Obwohl es sich *nur* um ein *2-face-composite* handelt, ist die Rötung nur noch schwach zu erkennen..... 80

Abbildung 38: Die „Virtuelle Miss Germany“, ein aus den acht attraktivsten Finalistinnen der Miss-Germany-Wahl 2007 berechnetes Composite (links). Zum Vergleich daneben das Durchschnittsgesicht „w(1-64)“ aus 64 „normalen“ Frauengesichtern (rechts). Abgesehen von den üblichen Morphing-Artefakten wie der makellosen Haut und dem Strukturverlust an den Haaren visualisiert das linke Bild, was den schönsten Frauen gemeinsam ist, während das rechte Bild lediglich zeigt, was für junge Frauen im Alter zwischen 17 und 30 Jahren normal ist..... 85

Abbildung 39: “Der Triumph der Galatea” von Raffael (Deckenfresko der Villa Farnesina in Rom, um 1512). Die dargestellte Seegöttin Galatea (Bildmitte) entsprach offenbar dem Schönheitsideal des 16. Jahrhunderts, zumindest wurde die Frauenfigur wegen ihrer außerordentlichen Schönheit bewundert. Raffael selbst erklärte in einem Brief auf die Frage, wie es ihm gelungen sei, eine derart schöne Frau zu malen, er habe in dieser Figur die Essenz und die Idee von vielen einzelnen schönen Frauen vereint..... 87

Abbildung 40: Übersicht über die verschiedenen Methoden zur Untersuchung des Einflusses der Symmetrie auf die Attraktivität. 125

Abbildung 41: Chimärengesicht aus zwei rechten Gesichtshälften (links), Originalgesicht (Mitte) und Chimärengesicht aus zwei linken Gesichtshälften (rechts)..... 126

Abbildung 42: Das asymmetrische Originalgesicht (links), ein symmetrisch optimiertes Gesicht nach dem herkömmlichen Morphing-Verfahren mit Neuberechnung der Farbwerte, wodurch die Haut glatter und makelloser erscheint (Mitte), sowie ein

symmetrisch optimiertes Gesicht (rechts), bei dem nur die Proportionen mit dem eigenen Spiegelbild vermorpht wurden, die Original-Farbwerte jedoch beibehalten wurden..... 128

Abbildung 43: Beispiel für das verwendete Stimulusmaterial. Das mittlere Bild zeigt ein unverändertes Originalgesicht. Das linke Bild zeigt ein Chimärengesicht, das aus der rechten Gesichtshälfte des Originalgesichts erzeugt wurde. Das rechte Bild zeigt ein Chimärengesicht aus der linken Gesichtshälfte. 135

Abbildung 44: Der Zusammenhang zwischen der Attraktivität eines Originalgesichts und der Attraktivitätsveränderung durch das Symmetrisieren mit der Chimärengesichter-Methode. Das linke Streudiagramm zeigt die Ergebnisse für Chimärengesichter aus linken Gesichtshälften, das rechte Diagramm für Chimärengesichter aus rechten Hälften. In beiden Fällen erkennt man, dass diese Art von Symmetrisierung sich auf fast alle Gesichter negativ auswirkt, und zwar tendenziell umso negativer, je attraktiver das Originalgesicht ist. 150

Abbildung 45: Illustrierung des Morph-Prozesses. Das Originalgesicht (links) wird mit seinem eigenen Spiegelbild (Mitte) zu einem neuen Durchschnittsgesicht („Blend“, rechts) verrechnet. Durch die Mittelwertbildung der Farbwerte verwischt die Struktur der Haare und die Haut erscheint feiner. Eventuelle Makel, wie beispielsweise das Muttermal auf der Stirn werden zwar verdoppelt, scheinen jedoch nur noch jeweils zu 50 % durch. Kleine Hautunreinheiten wie Mitesser oder vergrößerte Poren sind dadurch nicht mehr wahrnehmbar..... 153

Abbildung 46: Der Zusammenhang zwischen der Attraktivität eines Originalgesichts und der Attraktivitätsveränderung durch das Symmetrisieren mit der *Blend*-Methode. Bei den normalen Frauengesichtern hat die Symmetrisierung keinen Effekt, bei den attraktiven Model-Gesichtern mindert sie die Attraktivität tendenziell. 164

Abbildung 47: Illustrierung des Morph-Prozesses. Das Originalgesicht (links) wird mit seinem eigenen Spiegelbild (Mitte) zu einem neuen Durchschnittsgesicht („Warp“, rechts) verrechnet. Lediglich die Koordinatenwerte der einzelnen *Facial Landmarks* werden gemittelt, nicht jedoch die Farbwerte der Textur. Dadurch ist das Ergebnis in seinen Proportionen symmetrisch, die Textur des Originalgesichts bleibt jedoch erhalten. Es entstehen keine Artefakte an Haut oder Haaren. 168

Abbildung 48: Der Zusammenhang zwischen der Attraktivität eines Originalgesichts und der Attraktivitätsveränderung durch das Symmetrisieren mit der *Warp*-Methode. Bei den

attraktiven Model-Gesichtern hat die Symmetrisierung keinen Effekt, bei den normalen Frauengesichtern mindert sie die Attraktivität tendenziell..... 178

Abbildung 49: Die Korrelation zwischen dem mittleren Symmetrie-Rating der Originalgesichter durch Versuchspersonen und der mittleren Attraktivitätsbewertung durch (eine andere Gruppe von) Versuchspersonen. Je höher die beurteilte Symmetrie, desto höher die Attraktivität ($r = .78, p < .001$). Die Richtung der Kausalität ist jedoch unklar. Denkbar ist auch: Je höher die Attraktivität, desto höher die Symmetrie-Beurteilung. 184

Abbildung 50: Die Korrelation zwischen dem mittleren Symmetrie-Rating der perfekt symmetrischen Warp-Gesichter durch Versuchspersonen und der mittleren Attraktivitätsbewertung der Originalgesichter. Die Warp-Gesichter weisen objektiv keine Asymmetrien auf und sind alle gleich perfekt symmetrisch. Dennoch wird deren Symmetrie von den Versuchspersonen unterschiedlich beurteilt. Je attraktiver ein Gesicht, desto höher wird seine Symmetrie eingeschätzt ($r = .87, p < .001$). In diesem Fall ist die Richtung der Kausalität eindeutig: Attraktivität bedingt Symmetrie-Urteil..... 188

Abbildung 51: Beispiel für das verwendete Stimulusmaterial. Ähnlich wie bei der Erzeugung der Chimärengesichter wurde jedes Stimulusgesicht exakt in der Mitte geteilt. Die linke Gesichtshälfte wurde horizontal gespiegelt und wie in der Abbildung präsentiert, um zu verhindern, dass die Probanden die beiden Gesichtshälften als ein einziges Gesicht ganzheitlich wahrnahmen. 191

Abbildung 52: Die Korrelation zwischen dem mittleren Ähnlichkeits-Rating der Gesichtshälften als Maß für Symmetrie und der Attraktivität der Originalgesichter. Der Zusammenhang ist positiv und mittelstark ausgeprägt ($r = .47, p = .009$). Je ähnlicher die Gesichtshälften eines Originalgesichts (und damit je symmetrischer das Originalgesicht), desto attraktiver wird dieses Originalgesicht (von einer anderen Stichprobe von Versuchspersonen) bewertet..... 193

Abbildung 53: Beispiel für das verwendete Stimulusmaterial. Links ein Chimärengesicht aus zwei linken Gesichtshälften, rechts ein Chimärengesicht aus zwei rechten Gesichtshälften. Die Frage, die den Probanden gestellt wurde lautete: „Wie ähnlich sind die ‚Zwillinge‘?“ 200

Abbildung 54: Die Korrelation zwischen dem mittleren Ähnlichkeits-Rating der Chimärengesichter als Maß für Symmetrie und der Attraktivität der Originalgesichter. Der

Zusammenhang ist positiv und mittelstark ausgeprägt ($r = .48, p = .007$). Je ähnlicher die Chimärengesichter (d. h. je ähnlicher die Gesichtshälften und damit je symmetrischer das Originalgesicht), desto attraktiver wird ein Originalgesicht (von einer anderen Stichprobe von Versuchspersonen) bewertet.....202

Abbildung 55: Beispiel für Morphing-Artefakte, die aus einem gedrehten Gesicht bei der Aufnahme des Fotos resultieren (Bild links). Die linke Gesichtshälfte ist dadurch auf dem Foto kleiner abgebildet, obwohl sie es in Wirklichkeit nicht ist. Bei Verdopplung dieser Hälfte führt dies zu einem unnatürlich schmal wirkenden Gesicht. Bei der rechten Gesichtshälfte ist der Effekt umgekehrt. Unnatürlich wirken auch die symmetrischen Muttermale an der Oberlippe.....209

Abbildung 56: Ein eigentlich sehr symmetrisches Originalgesicht (Bild links), bei dem der Hals zum Aufnahmezeitpunkt jedoch in einer leicht schiefen Position war. Dies führt zu Chimärengesichtern mit einem entweder unnatürlich dünnen Hals oder einem breiten Hals, der die gesamte Person übergewichtig wirken lässt.....209

Abbildung 57: Das fotografierte Originalgesicht (Bild links) weist tatsächlich eine Asymmetrie auf. Der Mund ist zur rechten Gesichtshälfte hin verschoben. Dies macht es unmöglich, das Bild beim Setzen der Symmetrieachse fehlerfrei zu teilen. Ginge die Achse exakt durch die Nase, würde der Mund nicht mittig geteilt und es entstünde ein seltsames Artefakt an der Oberlippe. Lässt man hingegen die Symmetrieachse durch die Mitte der Oberlippe gehen, wird der Nasenrücken nicht mittig geteilt und ist bei den Chimärengesichtern einmal zu dünn (Bild Mitte) und einmal zu dick (Bild rechts). Zusätzlich entsteht beim rechten Bild ein Artefakt aufgrund der dunklen Linie auf dem Nasenrücken und den zwei Glanzpunkten auf der Nasenspitze.210

Abbildung 58: Gesicht einer erwachsenen Frau (Bild 6), das mit Morphing-Software schrittweise an die Gesichtsproportionen des Kindchenschemas angenähert wurde. Bei Bild 1 wurde das Kindchenschema mit 50 % gewichtet, bei Bild 2 mit 40 %, bei Bild 3 mit 30 %, bei Bild 4 mit 20 % und bei Bild 5 mit 10%. Es wird die Hypothese aufgestellt, dass Varianten mit einem hohen Kindchenschema-Anteil als attraktiver empfunden werden als das erwachsene Originalgesicht.251

Abbildung 59: Das für das Experiment verwendete Kindchenschema, das erzeugt wurde, indem aus vier Kindern im Kindergartenalter mit Morphing-Software ein Durchschnittsgesicht berechnet wurde.254

Abbildung 60: Bild 6 zeigt den attraktiven weibliche Prototyp („w(61-64)“, ein Composite aus den vier attraktivsten weiblichen Originalgesichtern aus dem Experiment *Durch 4*), der in der früheren Untersuchung als hochattraktiv bewertet worden war. Dieses Gesicht wurde mit Morphing-Software schrittweise an die Gesichtsproportionen des Kindchenschemas angenähert wurde. Bei Bild 1 wurde das Kindchenschema mit 50 % gewichtet, bei Bild 2 mit 40 %, bei Bild 3 mit 30 %, usw. 255

Abbildung 61: Häufigkeitsverteilung in Prozent, wie oft Frauengesichter mit einem bestimmten Kindchenschema-Anteil als attraktivstes Gesicht ausgewählt wurden. Die horizontale Linie markiert die A-priori-Wahrscheinlichkeit von 16.7 %. Am seltensten werden die Varianten des erwachsenen Originalgesichts als am attraktivsten ausgewählt – am häufigsten die Varianten mit einem hohen Kindchenschema-Anteil von 40 und 50 %. 259

Abbildung 62: Unattraktives erwachsenes Frauengesicht (Gesicht D, links) und mit Morphing-Software verkindlichte Variante, bei der das Kindchenschema mit 40 % gewichtet wurde (rechts). 262

Abbildung 63: Attraktives erwachsenes Frauengesicht (Gesicht F, links) und mit Morphing-Software verkindlichte Variante, bei der das Kindchenschema mit 40 % gewichtet wurde (rechts). 263

Abbildung 64: Obwohl bei allen Stimulusgesichtern die Annäherung an das Kindchenschema immer 40 % betrug, wirkt sich diese Manipulation nicht auf jedes Gesicht gleich aus. Beim linken Gesichterpaar (Gesicht C) ist die Veränderung zum Kindchenschema aufgrund des länglicheren Gesichts und des (in der Vertikalen) langen Kinns deutlicher sichtbar und der Attraktivitätsgewinn deutlich (74 % der Beurteiler bevorzugten die kindliche Version). Beim rechten Bildpaar (Gesicht G) ist der Unterschied viel schwächer ausgeprägt und es gab keinen signifikanten Attraktivitätsunterschied... 266

Abbildung 65: Linkes Bildpaar: Der unattraktive weibliche Prototyp (E). In der „verkindlichten“ Variante ist der Attraktivitätsgewinn deutlich (80.8 % der Versuchspersonen bevorzugten diese). Rechtes Bildpaar: Der attraktive weibliche Prototyp (K). In der kindlichen Version wird er leicht bevorzugt (62.5 % der Versuchspersonen). 267

Abbildung 66: Unattraktives erwachsenes Männergesicht (D, links) und mit Morphing-Software verkindlichte Variante, bei der das Kindchenschema mit 40 % gewichtet wurde (rechts).	270
Abbildung 67: Attraktives erwachsenes Männergesicht (K, links) und mit Morphing-Software verkindlichte Variante, bei der das Kindchenschema mit 40 % gewichtet wurde (rechts).	271
Abbildung 68: Die Stimuluspersonen B (linkes Paar) und C (rechtes Paar). Bei unattraktiven Männergesichtern wirkte sich das Verzerren der Gesichtsproportionen in Richtung Kindchenschema fast immer attraktivitätserhöhend aus – ein Widerspruch zur Theorie des sexuellen Dimorphismus.	274
Abbildung 69: Ein Original-Frauengesicht („w6o“, linkes Bild) wird zu 90 % an die Gesichtsproportionen des weiblichen Durchschnittsgesichts (berechnet aus 64 Frauengesichtern angenähert, mittleres Bild) angenähert. Es resultiert daraus das Gesicht „w6d“ (rechtes Bild), das in seinen Proportionen nahezu vollständig mit den Gesichtsproportionen des weiblichen Durchschnittsgesichts übereinstimmt.....	314
Abbildung 70: Männergesicht („m5“), das durch Morphing an die Proportionen eines Durchschnittsgesichts angepasst wurde („m5d“, oben links), zu 40 % an die Proportionen des Kindchenschemas angenähert wurde („m5k“, oben rechts) und symmetrisch optimiert wurde („m5s“, unten links). Unten rechts das Original („m5o“) als Kontrollbedingung. ...	316
Abbildung 71: Frauengesicht („w3“), das durch Morphing an die Proportionen eines Durchschnittsgesichts angepasst wurde („w3d“, oben links), zu 40 % an die Proportionen des Kindchenschemas angenähert wurde („w3k“, oben rechts) und symmetrisch optimiert wurde („w3s“, unten links). Unten rechts das Original („w3o“) als Kontrollbedingung. ...	317
Abbildung 72: Der mittlere Attraktivitätsrang für die einzelnen Varianten der Männergesichter. Für die Grafik wurden die Ergebnisse zuvor durch Bildung von Medianen pro Versuchsperson und pro Variante über alle sechs Stimuluspersonen zusammengefasst. $N = 324$. Ein hoher Wert bedeutet hohe Attraktivität. Die verdurchschnittlichten Männergesichter werden am besten und die Originalgesichter am schlechtesten bewertet.....	320
Abbildung 73: Varianten von links nach rechts: Durchschnitt, Kindchenschema, Symmetrie und Original. Stimulusperson „m4“ stellte eine Ausnahme dar. Hier wurden die	

symmetrische Variante und das Original am besten bewertet, die verdurchschnittlichte und die verkindlichte Variante signifikant schlechter. 321

Abbildung 74: Der mittlere Attraktivitätsrang für die einzelnen Varianten der Frauengesichter. Für die Grafik wurden die Ergebnisse zuvor durch Bildung von Medianen pro Versuchsperson und pro Variante über alle sechs Stimuluspersonen zusammengefasst. $N = 324$. Ein hoher Wert bedeutet hohe Attraktivität. Die verdurchschnittlichten Frauengesichter werden am besten und die Originalgesichter am schlechtesten bewertet. 324

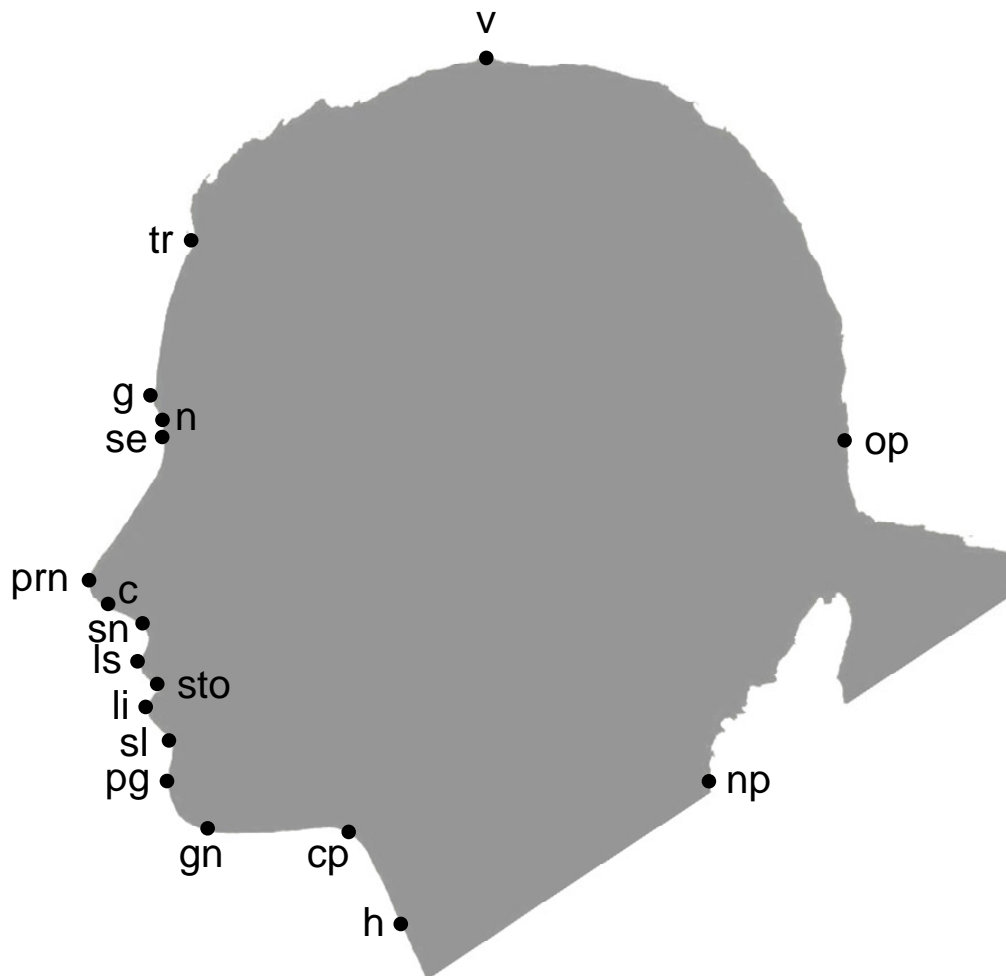
Abbildung 75: Varianten der Stimulusperson „w2“ von links nach rechts: Durchschnitt, Kindchenschema, Symmetrie und Original. Bei der Hälfte der weiblichen Stimuluspersonen wurde – so wie hier – die kindliche Variante am besten bewertet und die verdurchschnittlichte nur am zweitbesten. 325

Abbildung 76: Die durchschnittlichen Attraktivitätsratings für die einzelnen Varianten der Männergesichter. Für die Grafik wurden die Attraktivitätsbewertungen pro Variante über alle sechs Stimuluspersonen zusammengefasst. $N = 468$ Versuchspersonen, n pro Variante = 117 Versuchspersonen. Die verdurchschnittlichten Männergesichter wurden am besten bewertet, zwischen den anderen drei Varianten bestand kaum ein Attraktivitätsunterschied. Keine Variante erreichte hohe durchschnittliche Attraktivitätsbewertungen. 335

Abbildung 77: Die durchschnittlichen Attraktivitätsratings für die einzelnen Varianten der Frauengesichter. Für die Grafik wurden die Attraktivitätsbewertungen pro Variante über alle sechs Stimuluspersonen zusammengefasst. $N = 468$ Versuchspersonen, n pro Variante = 117 Versuchspersonen. Die verdurchschnittlichten Frauengesichter wurden am besten bewertet, zwischen der symmetrisch optimierten Varianten und den Originalgesichtern bestand kein Attraktivitätsunterschied. Keine Variante erreichte hohe durchschnittliche Attraktivitätsbewertungen. 340

14 Anhang

Anhang A



In der Anthropometrie standardmäßig verwendete Referenzpunkte (*Facial Landmarks*) zur Vermessung von Gesichtern. Die Abkürzungen sind aus dem Standardwerk von Farkas „*Anthropometry of the head and face*“ (1994) entnommen.

Abkürzung	Facial Landmark	Abkürzung	Facial Landmark
v	Vertex	sto	Stomion
tr	Trichion	li	Labiale inferius
g	Glabella	sl	Sublabiale
n	Nasion	pg	Pogonion
se	Sellion	gn	Gnathion
prn	Pronasale	cp	Cervical point
c	Columella	h	Halspunkt
sn	Subnasale	np	Neckpoint
ls	Labiale superius	op	Opisthocranion

Beispiele für die Schreibweise bestimmter Merkmale:

Strecken:

$sn-ls$ → Der Abstand zwischen der Unterkante der Nase und der Oberlippe

Streckenverhältnisse:

$sn-gn : tr-n$ → Das Verhältnis der vertikalen Länge des Unterkiefers zur Länge des Obergesichts

Winkel:

$gn-cp-h$ → Der Kiefer-Hals-Winkel, d. h. der Winkel der durch die Strecken $gn-cp$ und $cp-h$ aufgespannt wird

Anhang B

Ergebnisse der beiden Rating-Experimente mit Chimärengesichtern aus linken und rechten Gesichtshälften sowie den jeweiligen Originalgesichtern. Angegeben sind hier auch die Standardabweichungen zu den jeweiligen Attraktivitätsmittelwerten, die in Tabelle 10 (S. 147) aus Platzgründen weggelassen wurden. Pro Gruppe $N = 64$.

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i> Originale	<i>SD</i> Originale	<i>M</i> Chimäre links	<i>SD</i> Chimäre links	<i>M</i> Chimäre rechts	<i>SD</i> Chimäre rechts
1	Miss	3.91	0.92	3.27	1.09	3.50	1.44
2	Miss	4.59	1.03	4.22	1.13	4.31	1.13
3	Miss	4.78	1.02	4.19	1.49	3.81	1.32
4	Miss	4.38	1.08	4.47	1.31	3.83	1.16
5	Miss	4.33	1.10	3.48	1.30	4.77	1.18
6	Miss	4.59	1.19	4.31	1.08	3.80	1.24
7	Miss	5.22	0.77	4.38	1.28	4.88	1.15
8	Miss	4.81	1.17	4.50	1.15	4.28	1.25
9	Miss	4.92	0.96	4.09	1.28	4.20	1.22
10	Miss	4.70	1.00	3.78	1.34	4.19	1.60
11	Miss	4.78	1.37	4.38	1.33	4.27	1.34
12	Miss	4.50	1.14	3.16	1.31	4.73	1.31
13	Miss	5.28	1.00	4.88	1.43	3.86	1.41
14	Miss	5.09	1.08	4.42	1.31	4.81	1.19

Fortsetzung

Fortsetzung der Tabelle in Anhang B:

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i> Originale	<i>SD</i> Originale	<i>M</i> Chimäre links	<i>SD</i> Chimäre links	<i>M</i> Chimäre rechts	<i>SD</i> Chimäre rechts
15	Miss	5.58	1.08	<i>4.31</i>	1.34	<i>4.80</i>	1.53
16	Normale	2.83	1.13	2.28	1.13	<i>1.50</i>	0.71
17	Normale	2.64	1.00	2.64	1.16	<i>1.89</i>	0.82
18	Normale	1.94	0.96	1.69	0.83	1.67	0.87
19	Normale	2.34	1.16	1.66	1.00	2.31	1.05
20	Normale	2.67	1.14	2.22	1.12	<i>2.13</i>	1.09
21	Normale	2.45	0.99	<i>1.72</i>	0.79	2.59	1.19
22	Normale	2.80	1.04	<i>2.02</i>	1.02	<i>2.30</i>	1.11
23	Normale	2.81	1.08	3.05	1.30	<i>2.13</i>	0.98
24	Normale	2.53	0.93	2.13	0.98	2.61	1.15
25	Normale	2.17	0.95	<i>1.61</i>	0.85	<i>2.83</i>	1.23
26	Normale	2.06	0.94	2.31	0.99	1.84	1.09
27	Normale	2.80	1.04	2.63	0.97	2.42	1.08
28	Normale	2.44	1.04	2.47	1.30	2.30	1.26
29	Normale	1.47	0.78	1.34	0.57	1.30	0.61
30	Normale	2.95	0.92	<i>3.42</i>	1.10	<i>2.03</i>	1.05
<i>M</i>		3.61	1.03	3.17	1.14	3.20	1.16

Anmerkung: Kursive Mittelwerte zeigen einen signifikanten ($p < .05$) Unterschied im Vergleich zum Originalgesicht an (Post-Hoc-Tests nach Scheffé).

Anhang C

Die Ergebnisse der direkten Symmetrie-Ratings der Originalgesichter durch Versuchspersonen in der Untersuchung *Sym 7* (S. 179). $N = 60$.

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i>	<i>SD</i>	Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i>	<i>SD</i>
1	Miss	3.72	1.44	16	Normale	2.97	1.10
2	Miss	4.85	1.10	17	Normale	3.25	1.05
3	Miss	4.35	1.29	18	Normale	3.13	1.49
4	Miss	3.43	1.16	19	Normale	3.12	1.28
5	Miss	4.62	1.32	20	Normale	3.80	1.22
6	Miss	4.47	1.36	21	Normale	3.30	1.45
7	Miss	4.43	1.37	22	Normale	3.65	1.55
8	Miss	4.73	1.23	23	Normale	3.43	1.27
9	Miss	4.10	1.24	24	Normale	2.87	1.20
10	Miss	4.30	1.15	25	Normale	2.87	1.37
11	Miss	4.65	1.26	26	Normale	3.82	1.32
12	Miss	4.17	1.24	27	Normale	3.00	1.10
13	Miss	4.62	1.21	28	Normale	4.87	1.58
14	Miss	5.28	0.98	29	Normale	3.07	1.52
15	Miss	5.17	1.21	30	Normale	4.00	1.22

Anhang D

Die Ergebnisse der direkten Symmetrie-Ratings der symmetrisch optimierten *Warp*-Gesichter durch Versuchspersonen in der Untersuchung *Sym 7 Kontrolle* (S. 185). $N = 64$.

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i>	<i>SD</i>	Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i>	<i>SD</i>
1	Miss	4.53	1.22	16	Normale	3.97	1.21
2	Miss	4.91	1.32	17	Normale	4.11	1.37
3	Miss	5.30	1.22	18	Normale	4.16	1.79
4	Miss	5.36	1.26	19	Normale	3.81	1.60
5	Miss	4.75	1.37	20	Normale	4.16	1.63
6	Miss	4.89	0.96	21	Normale	4.33	1.49
7	Miss	5.45	1.26	22	Normale	4.27	1.71
8	Miss	5.09	1.24	23	Normale	4.59	1.33
9	Miss	5.00	1.37	24	Normale	4.02	1.43
10	Miss	4.78	1.50	25	Normale	3.67	1.48
11	Miss	5.34	1.03	26	Normale	4.31	1.49
12	Miss	4.83	1.19	27	Normale	4.45	1.46
13	Miss	4.75	1.26	28	Normale	4.45	1.60
14	Miss	5.20	1.40	29	Normale	3.69	1.70
15	Miss	5.09	1.20	30	Normale	4.86	1.62

Anhang E

Die Ergebnisse der Ähnlichkeitsbewertungen von Gesichtshälften durch Versuchspersonen als Maß für die Symmetrie eines Gesichts in der Untersuchung *Sym 8* (S. 190). $N = 64$.

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i>	<i>SD</i>	Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i>	<i>SD</i>
1	Miss	4.16	1.48	16	Normale	3.02	1.62
2	Miss	4.77	1.43	17	Normale	4.19	1.28
3	Miss	4.61	1.49	18	Normale	4.28	1.45
4	Miss	4.73	1.21	19	Normale	3.78	1.46
5	Miss	4.56	1.36	20	Normale	4.39	1.58
6	Miss	4.38	1.65	21	Normale	4.25	1.41
7	Miss	4.91	1.31	22	Normale	4.63	1.68
8	Miss	4.67	1.48	23	Normale	4.55	1.59
9	Miss	4.33	1.43	24	Normale	3.55	1.54
10	Miss	4.31	1.48	25	Normale	3.19	1.51
11	Miss	5.47	1.08	26	Normale	4.11	1.37
12	Miss	3.97	1.39	27	Normale	3.88	1.55
13	Miss	4.47	1.45	28	Normale	5.25	1.27
14	Miss	5.22	1.39	29	Normale	4.52	1.79
15	Miss	4.69	1.72	30	Normale	4.25	1.47

Anhang F

Die Ergebnisse der Ähnlichkeitsbewertungen von Chimären Gesichtern durch Versuchspersonen als Maß für die Symmetrie eines Gesichts in der Untersuchung *Sym 9* (S. 197).
N = 64.

Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i>	<i>SD</i>	Stimulus-Gesicht	Stimulus-Gruppe	<i>M</i>	<i>SD</i>
1	Miss	5.55	1.15	16	Normale	3.81	1.31
2	Miss	5.52	0.80	17	Normale	3.89	1.16
3	Miss	4.80	1.06	18	Normale	4.50	0.91
4	Miss	4.80	0.89	19	Normale	4.55	1.45
5	Miss	5.59	0.92	20	Normale	4.78	1.16
6	Miss	5.70	0.79	21	Normale	4.63	0.93
7	Miss	5.14	0.79	22	Normale	5.55	0.75
8	Miss	5.69	0.73	23	Normale	5.41	1.15
9	Miss	4.84	1.31	24	Normale	5.33	1.22
10	Miss	5.14	1.40	25	Normale	4.08	0.93
11	Miss	5.53	0.76	26	Normale	5.72	0.93
12	Miss	4.81	1.59	27	Normale	4.56	1.61
13	Miss	5.11	0.88	28	Normale	5.14	0.71
14	Miss	4.97	0.89	29	Normale	4.03	0.98
15	Miss	5.64	0.95	30	Normale	4.77	1.12

Anhang G



Ein Durchschnittsgesicht von 12 Kindern im Kindergartenalter, das als Kindchenschema für die Untersuchungen *Vergleich 1* und *Vergleich 2* (Kapitel 8, S. 311) verwendet wurde. Die an den Haaren und an der Kleidung vorhandenen Überblendungen sind typische Morphingartefakte, die immer entstehen, wenn man nicht nach jedem paarweisen Vermorphen von zwei Gesichtern die entstandenen Überblendungen mit einer Bildverarbeitungssoftware wegretuschiert. Für alle anderen gemorphen Gesichter wurden diese zeitintensiven Retuschen stets unternommen. Bei diesem Composite konnte jedoch darauf verzichtet werden, da dieses Bild nicht Versuchspersonen für eine Bewertung vorgelegt wurde, sondern lediglich als Schema verwendet wurde, an das erwachsene Originalgesichter in ihren Proportionen angenähert wurden. Für diese Manipulation sind daher die Farbwerte dieses Durchschnittsgesichts ohne Bedeutung.

Kontakt:

Dr. Martin Gründl (martin.gruendl@googlemail.com)