

Inauguraldissertation
zur Erlangung des akademischen Grades eines Dr. phil.
im Fach Psychologie
an der Fakultät für Verhaltens- und Empirische Kulturwissenschaften
der Ruprecht-Karls-Universität Heidelberg

Titel der Dissertation

Kein einheitliches Phänomen!
Der Einfluss von
Informationsvalenz, Urteilsdimension und Kontextfaktoren
auf die Qualität und Quantität von Halo-Effekten

vorgelegt von

Dipl.-Psych. Michael Gräf
aus Jena

Jahr der Einreichung

Heidelberg, Oktober 2012

Dekan:

Prof. Dr. Klaus Fiedler

Ruprecht-Karls-Universität Heidelberg

Berater:

Prof. Dr. Christian Unkelbach

Universität zu Köln

Prof. Dr. Klaus Fiedler

Ruprecht-Karls-Universität Heidelberg

„[...] denn an sich ist nichts weder gut noch böse, das Denken macht es erst dazu.“

(William Shakespeare)

Danksagung

An dieser Stelle möchte ich all denen Menschen danken, die mich bei meiner Dissertation unterstützt und mir zur Seite gestanden haben:

Zuallererst danke ich meinem Doktorvater, Prof. Dr. Christian Unkelbach, der mir stets mit seinem Fachwissen zur Seite stand und mich durch Ideen, Anregungen und konstruktive Kritik tatkräftig bei der Erstellung dieser Arbeit unterstützte.

Ebenfalls möchte ich mich herzlich bei meinem Zweitgutachter, Prof. Dr. Klaus Fiedler, für die Betreuung meiner Dissertation, den einen oder anderen Diskussionsansatz und das stets angenehme Arbeitsumfeld bedanken.

Bei meinen studentischen Hilfskräften, Laura, Jonas und Hella, bedanke ich mich besonders für die große Hilfe bei der Datenerhebung und die vielen nützlichen Anmerkungen und Hinweise bei der Konzipierung der Experimente.

Ein weiterer großer Dank geht an die Deutsche Forschungsgemeinschaft, deren finanzielle Unterstützung maßgeblich zu zügigen Fortschritten bei der Durchführung der Experimente beigetragen hat.

Ich danke meiner lieben Schwester Silvana, die mir eine großartige Hilfe bei der Gewinnung von Probanden sowie bei der Datenerhebung bei zweien der Experimente war und mir über längere Zeit ihre Räumlichkeiten zur Verfügung stellte.

Ebenso möchte ich mich bei meinen Freunden Anne-Kathrin und Stefan für das fleißige Korrekturlesen dieser Arbeit sowie auch für ihre Freundschaft und die damit verbundenen schönen Zeiten bedanken.

Meinem Freund Jürgen danke ich für seine moralische Unterstützung, die vielen gemeinsamen Unternehmungen, durch die ich oft wieder einen klaren Kopf bekommen konnte, und seine stets ehrliche und aufrichtige Meinung.

Des Weiteren möchte ich mich bei den zahlreichen Versuchsteilnehmern, insbesondere den Mitgliedern des Fitnessstudios *Elan* sowie des Tanzclubs *Kristall*, bedanken, ohne die diese Dissertation nicht hätte entstehen können.

Mein tiefster Dank gilt meinen Eltern, Michael und Erika, für die jahrelange Unterstützung, die vielen wichtigen Tipps und Hinweise, das Korrekturlesen dieser Arbeit und vor allem für das stetige Motivieren und den festen Glauben an mich.

Inhaltsverzeichnis

Zusammenfassung.....	2
1 Einleitung	4
2 Theoretischer Hintergrund	6
2.1 Stand der bisherigen Forschung	6
2.1.1 Auftrittsbereiche von Halo-Effekten	7
2.1.2 Ursachen von Halo-Effekten	14
2.1.3 Vermeidung von Halo-Effekten.....	18
2.1.4 Erfassung von Halo-Effekten	22
2.2 Einheitlicher Halo-Effekt versus vielfältige Halo-Effekte.....	25
2.2.1 Offene Fragen der bisherigen Forschung.....	26
2.2.2 Belege für die Vielfältigkeit von Halo-Effekten	29
2.3 Zentrale Hypothesen	34
2.3.1 Einfluss der Informationsvalenz	35
2.3.2 Einfluss der Urteilsdimension.....	38
2.3.3 Einfluss von Kontextfaktoren	40
3 Methodische Umsetzung	44
3.1 Konzipierung und Ablauf der Experimente	44
3.2 Überprüfung des grundlegenden Modells	46
3.3 Auswahl des abhängigen Maßes	47
3.4 Berechnung der Halo-Effekte	48
3.5 Überblick über die durchgeführten Experimente	52
4 Relevanz der postulierten Einflussfaktoren: Allgemeine Voruntersuchung	54
4.1 Methode	54
4.2 Ergebnisse.....	56
4.3 Diskussion	60
5 Einflüsse von Informationsvalenz und Urteilsdimension auf Vorannahmen	62
5.1 Quantitative Unterschiede verschiedener Halo-Effekte: Experiment 1.....	62
5.1.1 Vorerhebung	63
5.1.2 Methode.....	64
5.1.3 Ergebnisse	66

5.1.4	Diskussion.....	68
5.2	Qualitative Unterschiede verschiedener Halo-Effekte: Experiment 2.....	69
5.2.1	Methode.....	70
5.2.2	Ergebnisse.....	72
5.2.3	Diskussion.....	76
5.3	Bedeutung des Abfragemodus: Experiment 3.....	77
5.3.1	Vorerhebung.....	78
5.3.2	Methode.....	80
5.3.3	Ergebnisse.....	82
5.3.4	Diskussion.....	89
5.4	Post-Hoc-Untersuchung zur Baseline von Vorannahmen.....	91
5.4.1	Methode.....	93
5.4.2	Ergebnisse.....	94
5.4.3	Diskussion.....	95
5.5	Zusammenfassung der Experimente 1 bis 3.....	96
6	Einflüsse von Informationsvalenz und Urteilsdimension auf Verzerrungen.....	98
6.1	Robustheit von Vorannahmen: Experiment 4.....	98
6.1.1	Methode.....	99
6.1.2	Ergebnisse.....	101
6.1.3	Diskussion.....	105
6.2	Urteilsverzerrungen bei paralleler Darbietung: Experiment 5.....	106
6.2.1	Methode.....	107
6.2.2	Ergebnisse.....	109
6.2.3	Diskussion.....	113
6.3	Urteilsverzerrungen bei sequentieller Darbietung: Experiment 6.....	114
6.3.1	Methode.....	115
6.3.2	Ergebnisse.....	118
6.3.3	Diskussion.....	123
6.4	Zusammenfassung der Experimente 4 bis 6.....	125
7	Einflüsse von Kontextfaktoren auf Vorannahmen und Verzerrungen.....	127
7.1	Bedeutung des Berufs der beurteilten Person: Experiment 7.....	127
7.1.1	Methode.....	128

7.1.2	Ergebnisse	130
7.1.3	Diskussion.....	133
7.2	Bedeutung der Situation und des Beurteilenden: Experiment 8	134
7.2.1	Methode.....	135
7.2.2	Ergebnisse	137
7.2.3	Diskussion.....	142
7.3	Generalisierbarkeit der Ergebnisse: Experiment 9.....	143
7.3.1	Methode.....	144
7.3.2	Ergebnisse	146
7.3.3	Diskussion.....	151
7.4	Zusammenfassung der Experimente 7 bis 9.....	152
8	Allgemeine Diskussion	154
8.1	Zusammenfassung der Ergebnisse	154
8.2	Einordnung und Implikationen der Befunde	156
8.3	Ausblick auf zukünftige Forschungsmöglichkeiten	164
9	Fazit.....	166
	Literaturverzeichnis.....	167
	Abbildungsverzeichnis.....	191
	Tabellenverzeichnis.....	192
	Anhang.....	193
	Anhang A: Fragebogen zur Generierung gegensätzlicher Eigenschaften	194
	Anhang B: Fragebogen zur Gegenüberprüfung gegensätzlicher Eigenschaften	197
	Anhang C: Ergebnisse der Vorerhebung zur Eigenschaftsauswahl	199
	Anhang D: Verwendete Bilder und deren Attraktivitätsratings.....	201
	Anhang E: Ergebnisse der Erhebung zur Generierung von Verhaltensweisen	203
	Anhang F: Ergebnisse der Gegenüberprüfung der Verhaltensweisen	208
	Anhang G: Fragebogen zur Post-Hoc-Analyse unipolarer Urteilsdimensionen	211
	Anhang H: Fragebogen zur Post-Hoc-Analyse bipolarer Urteilsdimensionen	213
	Anhang I: Verwendete Berufsbezeichnungen und deren Häufigkeiten.....	215
	Selbständigkeitserklärung	216

Kein einheitliches Phänomen!

Der Einfluss von

Informationsvalenz, Urteilsdimension und Kontextfaktoren

auf die Qualität und Quantität von Halo-Effekten

Zusammenfassung

In den mittlerweile fast 100 Jahren seit der Prägung ihres Namens haben sich Halo-Effekte zu einem der bedeutendsten Phänomene in der Psychologie etabliert. Jedoch hat sich in dieser Zeit trotz umfassender Forschung kein festes Erklärungsmodell durchsetzen können. Im Wesentlichen existieren zwei verschiedene Ansätze zur Beschreibung und Erklärung von Halo-Effekten: Der erste Ansatz definiert einen Halo-Effekt als kognitive Verzerrung, die aus dem Einfluss einer globalen Einschätzung auf die Beurteilung individueller Attribute von Personen oder Objekten resultiert. Das zweite Modell von Halo-Effekten besagt, dass besonders saliente Merkmale die Einschätzung anderer Attribute direkt beeinflussen, ohne über einen globalen Eindruck vermittelt zu werden.

Das große Problem beider Modelle besteht in der Tatsache, dass ihre jeweiligen Vertreter von der Annahme ausgehen, dass es sich bei Halo-Effekten um ein universelles Phänomen handelt, dessen Prozesse und Stärke einem einheitlichen Mechanismus unterliegen. Da jedoch beide Modelle in einer Vielzahl von Studien bestätigt werden konnten, ist das Konzept eines einheitlichen Halo-Effekts fraglich. Ein weiterer Schwachpunkt der bisherigen Untersuchungen liegt in der Fixierung bestimmter Kontexte und damit verbundener Attribute. So beschäftigen sich beispielsweise viele Studien nur mit positiven Halo-Effekten bei der Beurteilung der Kompetenz von Personen im schulischen oder beruflichen Umfeld. Negative Halo-Effekte, auch Horn-Effekte genannt, werden in der Literatur weitgehend vernachlässigt. Das Gleiche gilt für Halo-Effekte im Rahmen persönlicher Kontexte, wie etwa der Partnerwahl.

Die vorliegende Arbeit soll, basierend auf diesen Überlegungen, die bisher verbreitete Annahme eines einheitlichen Halo-Effekts in Frage stellen und die eben genannten Einflussfaktoren untersuchen. Zu diesem Zweck wurde ein Modell entwickelt, welches die beiden Arten von Halo-Effekten zusammenführt, also sowohl Effekte des globalen Eindrucks als auch spezifische Effekte einzelner salienter Attribute berücksichtigt. Das Modell sagt vorher, dass die Art und das bloße Ausmaß von Halo-Effekten unabhängig voneinander durch verschiedene Faktoren beeinflusst werden können.

Die Art von Halo-Effekten sollte aufgrund der unterschiedlichen Charakteristika positiver und negativer Informationen durch Valenz bestimmt werden: Von *hilfsbereit* auf *fleißig* wird über einen globalen Eindruck geschlossen, von *egoistisch* auf *faul* durch einen

spezifischen Übertrag. Demgegenüber sollte die Dimension einer Eigenschaft zwar das Ausmaß, nicht aber die Art von Halo-Effekten beeinflussen: Communion-Attribute führen durchschnittlich zu stärkeren, aber nicht globaleren Halo-Effekten als Agency-Attribute, da sie in einem größeren Maße zwischenmenschliche Interaktionen und Sympathie beeinflussen. In beruflichen Kontexten, in denen Agency-Attribute von höherer Bedeutung sind, sollte sich dieser Unterschied relativieren. So wird zum Beispiel die Eigenschaft *hilfsbereit* in privaten Situationen zu stärkeren Halo-Effekten führen als in beruflichen Situationen, wogegen für das Attribut *ordentlich* das umgekehrte Prinzip gilt.

Zur Untersuchung dieser Hypothesen wurden insgesamt neun verschiedene Experimente durchgeführt. In diesen Studien wurden den Versuchsteilnehmern diverse Persönlichkeitseigenschaften oder Verhaltensweisen von Personen präsentiert. Auf Basis dieser Informationen sollten die Probanden die Personen auf verschiedenen Dimensionen einschätzen. In den unterschiedlichen Experimenten wurden unter anderem Valenz, Dimension, Kontext und Eindeutigkeit der dargebotenen Informationen variiert.

Die Ergebnisse sprechen deutlich für die oben genannten Hypothesen und das zugrunde liegende Modell: Positive Valenz führt zu globalen und allgemein größeren Halo-Effekten, wogegen negative Valenz in spezifischen und eher schwachen Halo-Effekten resultiert. Zudem gehen Communion-Attribute in rein privaten Situationen und im Kontext sozialer Berufe mit stärkeren, aber nicht globaleren Halo-Effekten einher als Agency-Attribute. Dieses Muster kehrt sich um, wenn Menschen mit naturwissenschaftlichen und technischen Berufen beurteilt werden. Die Befunde dieser Arbeit haben bedeutende Implikationen für bereits existierende Modelle und die bisher verbreitete Annahme eines einheitlichen Halo-Effekts. Gleichzeitig tragen sie zu einem besseren Erkennen und Verstehen von Halo-Effekten in verschiedenen Situationen bei, um optimale Urteile und Entscheidungen zu gewährleisten.

1 Einleitung

Halo-Effekte gehören zu den bekanntesten und am längsten erforschten Arten von Urteilsverzerrungen. Unter einem Halo-Effekt versteht man, einfach gesagt, eine verzerrte, nicht objektive Wahrnehmung von Personen oder Objekten, weil bestimmte Eindrücke einzelne Eigenschaften regelrecht überstrahlen. Beispielsweise werden attraktive Personen oft als intelligenter, dominanter und geselliger eingeschätzt als sie es in Wirklichkeit sind (Feingold, 1992). Dies führt häufig dazu, dass Urteile zu positiv oder negativ ausfallen und aus diesem Grund schlechte Entscheidungen getroffen werden. Zwar sind mittlerweile fast 100 Jahre seit dem erstmaligen Auftreten des Begriffs *Halo-Effekt* vergangen (Thorndike, 1920), jedoch hat sich in dieser Zeit kein festes Modell durchsetzen können. Vielmehr existieren verschiedene Erklärungsansätze, die alle in einem gewissen Maße gültig zu sein scheinen. Auch gibt es unterschiedliche Messmethoden, die nicht immer zu den gleichen Ergebnissen führen. Die verbreitete Annahme, dass es sich bei Halo-Effekten um ein einheitliches Phänomen handelt, kann aus diesem Grund in Frage gestellt werden. Für die Urteils- und Entscheidungsforschung ist es von großer Bedeutung, das korrekte Modell zu kennen und die zugrunde liegenden Prozesse von Halo-Effekten zu verstehen.

Die Bedeutsamkeit von Halo-Effekten erwächst aus der Tatsache, dass Urteile und Entscheidungen essentielle Bestandteile des alltäglichen Lebens sind. Was ziehe ich heute Morgen an? Wo gehe ich heute Mittag essen? Welchen Film sehe ich mir heute Abend an? All dies sind Fragen, die sich mehr oder weniger tagtäglich stellen und Entscheidungen verlangen. Um solche Entscheidungen treffen zu können, ist es notwendig, die zur Wahl stehenden Optionen zu bewerten und gegeneinander abzuwägen. Zum Teil erfolgt dies auf Basis kognitiver Daumenregeln, sogenannter Heuristiken (Tversky & Kahneman, 1974). Oft existieren auch externe Entscheidungshilfen, wie der Wetterbericht oder Filmkritiken, an denen man sich orientieren kann. Jedoch verlangen insbesondere Entscheidungen mit weitreichenden Konsequenzen nach komplexeren und sorgfältigeren Urteilen. Welchen Studiengang wähle ich? Ist der Job der richtige für mich? Soll ich jetzt schon heiraten? All dies sind Entscheidungen, die uns nur selten im Leben begegnen und häufig nicht nur uns selbst, sondern auch andere Menschen betreffen. Umso bedeutsamer ist es, Urteile zu fällen, die möglichst rational und unverfälscht sind. Nur so ist es möglich, die richtigen Handlungen zu unternehmen und persönliche Ziele optimal zu erreichen.

In der Realität sind menschliche Urteile selten vollkommen objektiv. Das liegt teilweise daran, dass oft nur ein Bruchteil aller notwendigen Informationen herangezogen werden kann (Fiedler, 2000; Juslin & Olsson, 1997). Auch spiegelt die subjektive Wahrnehmung von Häufigkeiten, welche maßgeblich für Urteile und Entscheidungen ist (Schwarz et al., 1991; Tversky & Koehler, 1994), nicht immer die objektiven Gegebenheiten wider (Kleiter, Doherty, & Brake, 2002). Urteile können zudem durch die Manipulation der dargebotenen Informationen verzerrt werden. Beispielsweise führt die Formulierung eines Problems als Gewinn oder Verlust zu Unterschieden bezüglich der favorisierten Lösung (Kahneman & Tversky, 1984). Des Weiteren können selbst offensichtliche Wiederholungen (Hintzman & Block, 1971) oder die Unterteilung von Informationen in mehrere kleine Einheiten (Fiedler & Armbruster, 1994) mit Urteilsverzerrungen einhergehen.

Halo-Effekte sind eine besonders wichtige Art der Urteilsverzerrung, da sie bereits ohne eine gezielte Manipulation der Informationen auftreten können. Schon ein schönes Gesicht oder ein besonders freundliches Wesen kann dazu führen, dass einer Person viele positive Attribute zugeschrieben und selbst negative Eigenschaften vergleichsweise positiv wahrgenommen werden. Daraus kann sich zum Beispiel ergeben, dass ein ungeeigneter Bewerber eingestellt wird oder die Wahl auf einen unpassenden Partner fällt. Da eine schlechte Entscheidung in solchen Fällen suboptimale oder sogar negative Konsequenzen haben kann, ist es von besonderer Wichtigkeit, Halo-Effekte in ihren Ursachen und Wirkungsprozessen genau zu untersuchen und zu verstehen. In diesem Zusammenhang existiert noch eine Vielzahl offener Fragen. Welche Eigenschaften führen zu stärkeren und welche zu schwächeren Halo-Effekten? Wie verhalten sich positive und negative Halo-Effekte zueinander? In welchem Maße hängen verschiedene Halo-Effekte von dem jeweiligen Urteilskontext ab? Und welches der existierenden Modelle trifft unter welchen Umständen zu? Die vorliegende Arbeit hat das Ziel, diese Fragen aufzugreifen und die Vielfältigkeit von Halo-Effekten genauer zu beleuchten.

2 Theoretischer Hintergrund

Die vorliegende Arbeit untersucht den Einfluss von Informationsvalenz, Urteilsdimension und verschiedenen Kontextfaktoren auf die Qualität und Quantität von Halo-Effekten. Im Folgenden wird zunächst ein Überblick über die bisherige Forschung zu Halo-Effekten gegeben. Anschließend werden verschiedene, bisher nicht beantwortete Fragen näher beleuchtet. Auf Basis dieser Überlegungen werden danach die zentralen Hypothesen dieser Arbeit im Detail erläutert.

2.1 Stand der bisherigen Forschung

Halo-Effekte wurden erstmals in einer Untersuchung von Wells (1907) beschrieben, jedoch noch nicht als solche bezeichnet. In dem Experiment bewerteten die Versuchsteilnehmer zehn Autoren auf verschiedenen Dimensionen sowie im Gesamten. Es zeigte sich, dass die Gesamtbewertungen die Beurteilungen der spezifischen Kategorien beeinflussten und diese damit stärker verbunden zu sein schienen, als sie es in Wirklichkeit waren. Die Prägung des Begriffs *Halo-Effekt* geht auf Thorndike (1920) zurück, der die Beurteilung von Armee-Offizieren durch deren Vorgesetzten und andere Offiziere untersuchte. Dabei stellte er fest, dass die Korrelationen zwischen den einzelnen Bewertungen höher und ähnlicher zueinander waren als sie es sein müssten. Der Autor führte dieses Ergebnis darauf zurück, dass die Beurteilung spezifischer Eigenschaften von einem generellen Halo geprägt sei, der dazu führe, Personen nur allgemein als eher positiv oder negativ zu betrachten. Ausgehend von dieser Feststellung werden Halo-Effekte heutzutage von einem Großteil der Literatur als kognitive Verzerrungen definiert, die aus dem Einfluss einer globalen Einschätzung auf die Beurteilung individueller Attribute von Personen oder Objekten resultieren (Nisbett & Wilson, 1977).

Seit den Berichten von Wells (1907) und Thorndike (1920) wurde eine große Anzahl an Untersuchungen zu Halo-Effekten durchgeführt. Neben dem bloßen Aufzeigen des Phänomens in verschiedenen Bereichen beschäftigten sich viele Studien auch mit dessen Ursachen und Einflussfaktoren. In diesem Zusammenhang wurde ebenfalls untersucht, ob und durch welche Maßnahmen sich Halo-Effekte reduzieren oder verhindern lassen. Ein

letzter Schwerpunkt der bisherigen Forschung lag in der Frage, wie Halo-Effekte am besten zu erfassen sind. Im Verlauf dieser Entwicklung wurde zudem die Problematik fokussiert, Halo-Effekte von korrekten Urteilen zu trennen. Die Vielzahl der Untersuchungen mit den verschiedenen Schwerpunkten führte dazu, dass nach dem heutigen Stand der Forschung bereits ein umfangreiches Wissen über Halo-Effekte existiert.

2.1.1 Auftrittsbereiche von Halo-Effekten

In den mittlerweile über 100 Jahren seit ihrer erstmaligen Beschreibung wurden Halo-Effekte in einer Vielzahl von Kontexten untersucht und nachgewiesen. Ausgehend von der vorhandenen Literatur können dabei fünf große Forschungsbereiche voneinander unterschieden werden.

Halo-Effekte in beruflichen Kontexten

Ein großer Teil der Forschung beschäftigt sich mit Halo-Effekten im beruflichen Leben. Dabei standen bisher vor allem Feldstudien im Zentrum der Aufmerksamkeit: Thorndike (1920) zeigte Halo-Effekte bei der Beurteilung von Armee-Offizieren hinsichtlich physischer Qualität, Intelligenz, Führungsstil, Charakter und allgemeinem Wert für den Armeedienst. Ähnliche Befunde im Kontext des Wehrdiensts ergaben sich auch in aktuelleren Studien (Frone, Adams, Rice, & Instone-Noonan, 1986; Van Scotter, Moustafa, Burnett, & Michael, 2007). Halo-Effekte konnten zudem bei der Bewertung von Polizisten durch ihre Vorgesetzten (King, Hunter, & Schmidt, 1980; Lee, Malone, & Greco, 1981) und durch sich selbst (Fox, Caspy, & Reisler, 1994) nachgewiesen werden. Auch im sportlichen Bereich existieren Belege für das Wirken von Halo-Effekten, so etwa bei der Einschätzung real existierender Baseballspieler auf verschiedenen Leistungsdimensionen (Kozlowski & Kirsch, 1987; Kozlowski, Kirsch, & Chao, 1986).

Der wohl größte Forschungsbereich innerhalb von Arbeitssituationen widmete sich der Untersuchung klassischer Angestelltenberufe. Toops (1923), der zu den ersten Wissenschaftlern auf diesem Gebiet gehörte, stellte zum Beispiel fest, dass oft eine viel zu hohe Korrelation zwischen der Beurteilung allgemeiner Fähigkeiten und der Stimme von Lehrern besteht. In späteren Untersuchungen konnten diese und vergleichbare Befunde repliziert werden (Henrikson, 1943; Phelps, Schmitz, & Boatright, 1986). Halo-Effekte

wurden in der bisherigen Forschung auch bei der Bewertung von Krankenschwestern durch Kollegen (Brown, 1968; Ostroff & Ilgen, 1992), von Verkaufspersonal durch deren Abteilungsleiter (Jackson & Furnham, 2001), von Medizinstudenten durch Patienten (Iramaneerat & Yudkowsky, 2007) und von studentischen Hilfskräften durch deren Arbeitgeber (Gardner, 1974) dokumentiert. Zudem existieren Belege für das Wirken von Halo-Effekten bei der Einschätzung von verschiedenen Dimensionen der eigenen Arbeitszufriedenheit (Grove & Kerr, 1951; Paine, Carroll, & Leete, 1966).

Viele Studien untersuchten das Wirken von Halo-Effekten im Kontext von verschiedenen Managementberufen, ohne dabei näher auf die spezifischen Branchen einzugehen. Bereits bei der Bewertung von Bewerbern durch den Arbeitgeber wurden systematische Urteilsverzerrungen beobachtet (Crissy & Regan, 1951). Die gleichen Effekte zeigten sich, wenn Bewerber sich selbst einschätzten (Jones & Fletcher, 2002) oder durch Psychologen beurteilt wurden (Zysberg & Nevo, 2004). Holzbach (1978) analysierte die Ratings von Kollegen, Vorgesetzten und durch die eigene Person von insgesamt 183 Angestellten in höheren Berufspositionen. Dabei wurden Halo-Effekte im Bereich verschiedener Leistungsdimensionen, wie etwa Produktivität, Arbeitsqualität und Eigeninitiative, deutlich. Diverse andere Autoren konnten vergleichbare Ergebnisse für die unterschiedlichsten Leistungsmaße im Rahmen noch umfangreicherer Untersuchungen nachweisen (F. J. Landy, Vance, Barnes-Farrell, & Steele, 1980; Mount, 1984; Mount & Thompson, 1987; Tsui & Barry, 1986; Tziner, 1984).

Ein eher kleiner Forschungsbereich im Rahmen beruflicher Situationen fokussierte keine Feldstudien, sondern fiktive Personen und Szenarien. Die Untersuchungen erfolgten daher hauptsächlich im Labor statt in realen Umgebungen. So beschrieb Leitner (1972) seinen Versuchsteilnehmern zuvor konstruierte Gespräche zwischen Therapeuten und Patienten, wobei die Qualität der Beratung systematisch variiert wurde. Geling es den Probanden nicht, die Kompetenz der Therapeuten adäquat einzuschätzen, sprach dies für das Wirken von Halo-Effekten. In anderen Experimenten wurden Angestellte in diversen Managementberufen mit unterschiedlichen Verhaltensweisen beschrieben. Auf dieser Grundlage schätzten die Versuchsteilnehmer verschiedene Leistungsaspekte sowie die Geeignetheit der Personen für den jeweiligen Beruf ein. Auch hier wurden nicht korrekte Urteile als Halo-Effekte interpretiert (Borman, 1975; Rizzo & Frank, 1977).

Halo-Effekte in schulischen und studentischen Kontexten

Neben der Betrachtung verschiedener Berufe standen in der bisherigen Forschung vor allem schulische und studentische Kontexte im Vordergrund. Die Abgrenzung zu beruflichen Situationen ist dabei aber nicht immer eindeutig. Im Folgenden werden unter schulischen und studentischen Szenarien all jene Untersuchungen zusammengefasst, in denen Schüler oder Studenten durch Lehrer beziehungsweise Dozenten, oder umgekehrt, beurteilt wurden. Wie auch bei beruflichen Kontexten fand ein Großteil der Analysen in realen Umgebungen, in diesem Fall in Schulen und Universitäten, statt. Eines der ersten Experimente auf diesem Gebiet stammt von Symonds (1925), der die Bewertung von 40 Schülern durch zwei Lehrer untersuchte. Als relevante Eigenschaften standen dabei Ehrlichkeit, Gehorsam, Höflichkeit, Ordentlichkeit, Sauberkeit, Sportsgeist sowie Schnelligkeit im Fokus der Aufmerksamkeit. Es zeigte sich, dass die Korrelationen zwischen den Ratings der beiden Lehrer maßgeblich durch Halo-Effekte erhöht wurden. Ähnliche Ergebnisse waren auch dann erkennbar, wenn die Bewertungen durch Lehrer mit den Resultaten objektiver Tests verglichen wurden (Caldwell & Mowry, 1934).

Einige Studien wiesen Halo-Effekte im schulischen Kontext nach, indem sie auf ungewöhnliche starke Korrelationen zwischen den Bewertungen von eigentlich unzusammenhängenden Charakteristika, wie etwa Leistungsmotivation und physischer Erscheinung, verwiesen (Jaeger & Freijo, 1975; Newcomb, 1931; Remmers & Martin, 1944; Sawyer & Raybould, 1981). Die gleichen Verzerrungen wurden auch bei der Benotung unterschiedlicher Schulfächer, beispielsweise Französisch und Mathematik, gefunden (Dompnier, Pansu, & Bressoux, 2006). In einer Untersuchung von Anastasiow (1963) ergab sich zudem, dass guten Schülern verstärkt positive Persönlichkeitseigenschaften auf einer Checkliste zugeschrieben wurden, während für wenig erfolgreiche Schüler das Gegenteil galt. Auch bei der Bewertung schriftlicher Arbeiten existieren Belege dafür, dass allgemein gute Schüler und Studenten oft bessere Noten erhalten als ihre Klassenkameraden beziehungsweise Kommilitonen (Archer & McCarthy, 1988; Dennis, 2007; Gibb, 1983). Werden Studenten dazu aufgefordert, sich selbst zu beurteilen, neigen sie oft dazu, ihr Lernverhalten und die eigene Entwicklung zu überschätzen (Pike, 1999).

Bei der Bewertung von Lehrern und Dozenten standen in der bisherigen Forschung hauptsächlich Attribute, wie Fairness, sympathische Ausstrahlung, fachliche Kompetenz, Interesse für das eigene Fachgebiet und äußerliche Erscheinung, im Mittelpunkt des

Interesses. Viele Experimente zeigten in diesem Zusammenhang ungewöhnlich hohe Korrelationen zwischen den einzelnen Bewertungen durch Schüler und Studenten auf (Coren, 1998; Feeley, 2002b; Follman, Wiley, Geiger, & Lavelly, 1974; Remmers, 1934). Andere Studien wiesen Halo-Effekte durch zu hohe oder niedrige Ratings (Davison, 1973; Edwards & Waters, 1984) oder anhand sehr geringer Varianzen bei der Einschätzung verschiedener Attribute (Bernardin, 1978; Bernardin & Walter, 1977) nach. Eine besonders umfangreiche Längsschnittuntersuchung, in der über 1000 Studenten verschiedene Fakultätsmitglieder bewerteten, stammt von Jacobs und Kozlowski (1985). Den Autoren gelang es hierbei, Halo-Effekte mit Hilfe von vier unterschiedlichen Methoden und über einen Zeitraum von zehn Wochen nachzuweisen.

Wie auch bei beruflichen Kontexten widmete sich ein gewisser Teil der Forschung keinen Feldstudien, sondern der Schaffung und Variierung fiktiver Szenarien: Nisbett und Wilson (1977) boten 118 Studenten eines von zwei zuvor gedrehten Videos eines Dozenten dar. Dieser verhielt sich in einem der Videos warm und freundlich, in dem anderen stattdessen kühl und distanziert. Alle anderen Attribute, wie etwa Auftreten, äußerliche Erscheinung und Akzent, waren in beiden Videos identisch. Die Halo-Effekte äußerten sich dahingehend, dass diese Eigenschaften als attraktiv und anziehend beurteilt wurden, wenn der Dozent ein freundliches Verhalten an den Tag legte. Diejenigen Probanden, die ein distanzierteres Verhalten beobachteten, schätzten Auftreten und Akzent des Dozenten dagegen als eher irritierend ein. In späteren Experimenten gelang es durch ähnliche Manipulationen in Form von Videosequenzen (Darley & Gross, 1983; Goffin, Jelley, & Wagner, 2003; Murphy & Anhalt, 1992; Wetzell, Wilson, & Kort, 1981) oder Verhaltensbeschreibungen (Becker & Cardy, 1986; Suter & Roberts, 1987) vergleichbare Halo-Effekte aufzuzeigen. Eine weitere Methode stellt die Simulation einer kompletten Seminarsitzung dar. Hierbei ließ eine Studie erkennen, dass Dozenten besser beurteilt wurden, wenn sie den Studenten im Vorfeld als warmherzig beschrieben wurden. Dieses Muster wurde zusätzlich verstärkt, wenn auch die Dozenten vor der Sitzung positive Beschreibungen über die Studenten erhielten (Goldman, Cowles, & Florez, 1983).

Halo-Effekte im diagnostischen und klinischen Bereich

Der dritte große Schwerpunkt bei der Erforschung von Halo-Effekten umfasst klinische und diagnostische Bereiche. Ein Teil der Studien überschneidet sich dabei mit

schulischen Kontexten, wobei im Folgenden jedoch ausschließlich die Bewertung von Intelligenz und klinischen Störungen im Vordergrund steht. Beispielsweise wurde in verschiedenen Untersuchungen deutlich, dass Kinder, die zuvor gute Leistungen gezeigt hatten, bei Intelligenztests besser bewertet wurden als Schüler mit durchschnittlichen Leistungen, weil mehrdeutige Antworten eher als korrekt interpretiert wurden (Sattler, Hillix, & Neher, 1970; Sattler & Winget, 1970).

Im klinischen Bereich boten Foster und Ysseldyke (1976) ihren Probanden ein Video eines normalen Viertklässlers dar. Einige Versuchsteilnehmer erhielten jedoch die Information, dass der Schüler geistig zurückgeblieben, emotional gestört oder aber lernbehindert sei. Die Ergebnisse machten deutlich, dass die Probanden das Kind entsprechend ihren Erwartungen durch die vorhergehende Beschreibung bewerteten, obwohl es keine objektiven Anzeichen für eine klinische Störung gab. Durch ähnliche Manipulationen wurden Halo-Effekte auch bei der Beurteilung von Angstzuständen, Depressionen und Aggression (Epkins, 1994) sowie Hyperaktivität, Unaufmerksamkeit und aufsässigen Verhaltensweisen (Abikoff, Courtney, Pelham, & Koplewicz, 1993; Schachar, Sandberg, & Rutter, 1986; Stevens, Quittner, & Abikoff, 1998) nachgewiesen.

Einige Autoren widmeten sich der Untersuchung von Halo-Effekten im Bereich der klinischen Diagnostik von Erwachsenen: McNeil und Janzen (1987) präsentierten 224 Psychologie-Studenten eine von acht Personenbeschreibungen, die sich lediglich in Bezug auf Trinkverhalten, sozioökonomischen Status und Geschlecht unterschieden. Die Befunde wiesen auf einen starken Halo-Effekt des Trinkverhaltens hin, der alle anderen Effekte überstrahlte: Fiktive Personen, die mit einem starken Trinkverhalten assoziiert wurden, wurden auf nahezu allen Dimensionen deutlich negativer bewertet als die anderen Personen. Effekte dieser Art wurden auch gefunden, wenn Patienten als unangepasst und noch nicht therapiert beschrieben wurden (Michelson, Mannarino, Marchione, Kazdin, & Costello, 1985). In einem realen Kontext entdeckte Mumma (2002), dass Ärzte häufig bestimmte Symptome ihrer Patienten, wie etwa Gewichtsverlust und Suizidgedanken, überschätzten, wenn diese depressive nonverbale Verhaltensweisen zeigten.

Halo-Effekte in der Werbung und in den Medien

Eine recht neue Strömung stellt die Erforschung von Halo-Effekten in den Bereichen von Werbung und Medien dar. In diesem Zusammenhang spielt vor allem das Image von

Personen, Marken oder Unternehmen eine besondere Rolle. Schon in einer frühen Studie zeigte sich, dass bekannten historischen Persönlichkeiten automatisch viele positive Attribute und Fähigkeiten zugeschrieben wurden (Bowden, Caldwell, & West, 1934). Ein positives Image lässt sich aber auch künstlich und durch einfache Mittel herstellen. So werden etwa Personen, die in einer Gruppe am Kopfende eines Tisches sitzen, von Außenstehenden oft am besten bewertet und tendenziell als Gruppenführer angesehen (Davenport, Brooker, & Munro, 1971; Pellegrini, 1971). Im Bereich der Politik fand Babad (2005) heraus, dass das Image von Politikern davon beeinflusst wurde, wie Interviewer sich diesen gegenüber verhielten: Zeigte ein Interviewer im Verlauf des Gesprächs feindliche oder abweisende nonverbale Verhaltensweisen, wurde auch der Politiker im Nachhinein schlechter beurteilt als bei einem Interview durch eine freundliche Person.

Im Bereich des Markenimage ergab sich in einer Studie von Min Han (1989), dass unbekannte Produkte besser beurteilt wurden, wenn ihr Herkunftsland einen guten Ruf genoss. Auch bei der Bewertung bereits etablierter Marken wurden derartige Halo-Effekte schon gefunden (Oh & Ramaprasad, 2003). Ist eine Marke einmal bekannt, kann ihr Image einen direkten Einfluss auf die Bewertung der verschiedenen Produkte und damit auch auf das Kaufverhalten ausüben (Boatwright, Kalra, & Zhang, 2008; Dillon, Mulani, & Frederick, 1984). Zum Beispiel fanden Chandon und Wansink (2007) heraus, dass Produkte von Fastfood-Ketten, die als relativ gesund galten, hinsichtlich ihrer Kalorien unterschätzt wurden und daher ungewöhnlich viel bestellt wurde. Das Image von Produkten kann auch durch gezielte Werbung, wie das Hervorheben gesundheitsfördernder Eigenschaften, nachträglich verbessert werden (Kozup, Creyer, & Burton, 2003). Als Konsequenz solcher Maßnahmen können weitreichende Halo-Effekte verursacht werden. So ergaben Studien, dass durch das Betonen bestimmter Eigenschaften auch andere Attribute desselben Produkts bessere Bewertungen erhielten (Andrews, Netemeyer, & Burton, 1998; Wansink, 2003). Ebenso wurde herausgefunden, dass die Werbung für ein Produkt den Kauf anderer Produkte beeinflussen kann (Ailawadi, Harlam, César, & Trounce, 2006).

Zu der mitunter aktuellsten Literatur im Bereich von Halo-Effekten gehören Erhebungen zum Image von Firmen und anderen Organisationen: Coombs und Holladay (2006) beschrieben ihren Probanden fiktive Szenarien, in denen allgemein bekannte und beliebte Unternehmen, wie etwa Disney, mit einer plötzlichen Krise aufgrund von technischem oder menschlichem Versagen konfrontiert waren. Anschließend füllten die

Probanden mehrere Urteile. Die Autoren fanden heraus, dass die Unternehmen nur in geringem Maße für die Krise verantwortlich gemacht wurden und die Einschätzungen vergleichsweise positiv blieben. Daraus wurde geschlussfolgert, dass der gute Ruf eines Unternehmens vor einem Reputationsverlust in Krisenzeiten schützen könne. In einer anderen Studie wurde durch inhaltliche Analysen deutlich, dass Universitäten, die in der Vergangenheit viele Erfindungen gemacht hatten, später schneller Lizenzen für ihre entwickelte Technik erhielten (Sine, Shane, & Di Gregorio, 2003). Ebenso wurden Organisationen bei einer Vielzahl von Leistungskriterien positiver beurteilt, wenn die Unternehmensführung einen guten Ruf genoss (Meindl & Ehrlich, 1987).

Halo-Effekte aufgrund von Attraktivität

Einen Sonderstatus bei der Untersuchung von Halo-Effekten nimmt die Attraktivitätsforschung ein. Während Halo-Effekte anderer Attribute und Verhaltensweisen meist nur in bestimmten Situationen fokussiert wurden, wurde die Rolle von physischer Attraktivität bereits in nahezu allen Kontexten analysiert. Bereits in frühen Experimenten konnte festgestellt werden, dass Probanden attraktive Personen verstärkt mit Werten assoziierten, die sie selbst als wichtig erachteten (Fensterheim & Tresselt, 1953). Ebenso werden attraktiven Menschen häufig eher Berufe zugeordnet, die man persönlich als sehr interessant einschätzt (DeRath & Carp, 1959). Viele Autoren zeigten Halo-Effekte dahingehend, dass attraktiven Personen automatisch viele positive Eigenschaften, wie Intelligenz, Geselligkeit, Frohsinn und Erfolg, zugeschrieben wurden (Dion, Berscheid, & Walster, 1972; Geiselman, Haight, & Kimata, 1984; Lucker, Beane, & Helmreich, 1981; A. G. Miller, 1970; Sappenfield, 1971). Dies war selbst dann der Fall, wenn nicht reale Fotos, sondern nur schematische Zeichnungen betrachtet wurden (Bradshaw, 1969). Neben der bloßen Einstellung kann Attraktivität auch das Verhalten anderer Personen beeinflussen. Beispielsweise wurden in Studien von D. Landy und Sigall (1974) und R. M. Kaplan (1978) Essays von unterschiedlicher Qualität konstruiert und zufällig mit Fotos verschiedener Frauen gepaart. Die Analysen ergaben, dass Essays von angeblich attraktiven Autorinnen besser bewertet wurden als Essays von unattraktiven Autorinnen. Dieser Unterschied war vor allem dann sehr stark, wenn die Arbeiten von schlechter Qualität waren.

Neben der bloßen physischen Attraktivität existieren weitere äußere Merkmale, die zu Halo-Effekten führen können. Lächeln ist eines dieser Attribute: Lau (1982) präsentierte

ihren Versuchsteilnehmern Bilder von Personen, die entweder lächelten oder nicht lächelten. Es wurde deutlich, dass lächelnde Personen als freundlicher, sympathischer und intelligenter eingeschätzt wurden als Personen, die nicht lächelten. In einer späteren Untersuchung konnten die gleichen Ergebnisse in Bezug auf eine Vielzahl anderer Eigenschaften, wie Optimismus und Zuverlässigkeit, gefunden werden (Otta, Barcellos Pereira Lira, Delevati, Cesar, & Salati Guirello Pires, 1993). Ein weiteres Merkmal, welches Halo-Effekte verursachen kann und mit Attraktivität assoziiert ist, ist das Gewicht: Verschiedene Erhebungen ergaben, dass normalgewichtige und dünne Menschen als enthusiastischer und erfolgreicher eingeschätzt wurden als übergewichtige Menschen. Auf der anderen Seite wurden dickeren Männern verstärkt Attribute, wie Freundlichkeit und Vertrauenswürdigkeit, zugeschrieben (Wade, Fuller, Bresnan, Schaefer, & Mlynarski, 2007; Wade, Loyden, Renninger, & Tobey, 2003). Auch die Stimme scheint eine wichtige Rolle bei der Eindrucksbildung zu spielen: Surawski und Ossoff (2006) zeigten, dass Politiker hinsichtlich Kompetenz, Vertrauenswürdigkeit und Führungsqualitäten besser bewertet wurden, wenn sie eine attraktive Stimme hatten. Auch die negativen Effekte eines unattraktiven Äußeren konnten damit zum Teil aufgehoben werden.

Aufgrund der Vielzahl der Befunde scheint Attraktivität die mitunter stärksten und weitreichendsten Halo-Effekte mit sich zu bringen, die allgemein bekannt sind. Dies wird auch in verschiedenen Zusammenfassungen und Metaanalysen deutlich (Feingold, 1992; Katz, 2003). Des Weiteren konnte belegt werden, dass nahezu alle Personengruppen dazu neigen, sich von physischer Erscheinung beeinflussen zu lassen. So zeigten Larose und Standing (1998) auf, dass nicht nur junge, sondern auch ältere und erfahrene Menschen attraktive Personen besser beurteilten als sie es objektiv betrachtet waren. Ebenfalls scheinen die positiven Halo-Effekte eines Lächelns auf nahezu alle Menschen, unabhängig von ihrer kulturellen Herkunft, in gleichem Maße zu wirken (Godoy et al., 2005).

2.1.2 Ursachen von Halo-Effekten

Ein großer Bereich der bisherigen Forschung widmete sich der Aufdeckung potentieller Ursachen beziehungsweise Einflussfaktoren von Halo-Effekten. Viele Studien versuchten, Halo-Effekte möglichst auf eine einzige Ursache zurückzuführen. Andere Autoren verwiesen dagegen auf eine ganze Reihe unterschiedlicher Gründe, die alle in

gleichem Maße für Halo-Effekte verantwortlich gemacht wurden (Cooper, 1981b; Guilford, 1954; Symonds, 1925). Im Wesentlichen lassen sich drei Gruppen von Ursachen und Einflussfaktoren unterscheiden.

Halo-Effekte aufgrund variabler Aspekte der Person

In einem Teil der Literatur werden Halo-Effekte auf unterschiedliche menschliche Aspekte zurückgeführt. Einige Untersuchungen fanden Hinweise darauf, dass es sich bei Halo-Effekten um eine Form der Diskriminierung handeln könnte. In einer Studie von Caldwell und Mowry (1934) an einer amerikanischen Schule wurde etwa deutlich, dass Schüler spanischer Herkunft tendenziell schlechter bewertet wurden als sie es objektiv betrachtet waren. Bei angloamerikanischen Schülern war eine solche Verzerrung dagegen nicht zu finden. In einem anderen Experiment zeigte sich, dass schwarze Menschen für die positiven Halo-Effekte von Attraktivität nur dann anfällig waren, wenn es sich bei der zu beurteilenden Person ebenfalls um einen Schwarzen handelte. Bei weißen Probanden war ein solcher Unterschied nicht zu finden (Zebrowitz, Montepare, & Lee, 1993).

Verschiedene Autoren versuchten Halo-Effekte schlichtweg mit mangelnder Erfahrung und Expertise der beurteilenden Personen zu erklären. Bei der Bewertung der Leistung von Baseballspielern traten Halo-Effekte vor allem dann auf, wenn die Probanden nur geringe Kenntnisse über die Sportler und den Sport an sich hatten (Kozlowski & Kirsch, 1987; Kozlowski et al., 1986). In einer Feldstudie zeigten sich Halo-Effekte besonders dann, wenn sich die beurteilende Person und der Beurteilte zuvor nicht gekannt hatten (Van Scotter et al., 2007). Eine Reihe von Experimenten ließ jedoch das genaue Gegenteil erkennen: Halo-Effekte wurden vor allem dann beobachtet, wenn sich die beurteilende Person und der Beurteilte bereits gut kannten (Jacobs & Kozlowski, 1985; Lachman & Bass, 1985; Lance, Fiscaro, & LaPointe, 1990). Auch mangelnde Erfahrung scheint keine hinreichende Begründung zu sein: Verschiedene Autoren wiesen nach, dass Halo-Effekte auch bei Experten, wie etwa Ärzten (Mumma, 2002), Psychologen (Zysberg & Nevo, 2004) oder Ernährungsfachleuten (Chandon & Wansink, 2007), auftreten können.

Einige Untersuchungen ergaben, dass die Ursache von Halo-Effekten in der Wahrnehmung des Beurteilungsobjekts durch den Beurteilenden liegt. Zum Beispiel ergab sich in einer Studie von Tsui und Barry (1986), dass Probanden vor allem dann für Halo-Effekte anfällig waren, wenn sie einen positiven oder negativen Affekt gegenüber der zu

beurteilenden Person hatten. Eine künstlich induzierte Erregung, wie etwa durch Koffein, führte zu ähnlichen Verzerrungen, sofern die Teilnehmer nicht über die wahre Quelle der Erregung informiert wurden (O'Neal, 1971). Einen weiteren Einflussfaktor stellt die soziale Distanz zwischen zwei Personen dar: Davison (1973) fand heraus, dass Studenten ihre Dozenten vor allem dann verfälscht einschätzen, wenn sie diese als höhergestellt wahrnehmen. Auch bei Schülern und Lehrern treten Halo-Effekte häufig dann am stärksten auf, wenn die soziale Distanz sehr hoch ist (Jaeger & Freijo, 1975). Durch beispielweise unterschiedliche Einstellungen zu bestimmten Themen kann sich die wahrgenommene sozialen Distanz zusätzlich vergrößern (Wicklund & Eckert-Nowack, 1989).

Halo-Effekte aufgrund stabiler Eigenschaften der Person

Die zweite bedeutende Erklärung für das Auftreten von Halo-Effekten liegt in dem Wirken bestimmter stabiler Persönlichkeitsmerkmale. Im Gegensatz zu den im letzten Abschnitt geschilderten Aspekten sind diese Eigenschaften kaum manipulierbar. Mehrere Autoren legten dar, dass Halo-Effekte vor allem dadurch verursacht werden, dass die beurteilenden Individuen keine ausreichenden kognitiven Fähigkeiten haben, um unterschiedliche Dimensionen getrennt voneinander wahrzunehmen, zu verarbeiten und auch zu erinnern (Banks & White, 1982; Bernardin, Cardy, & Carlyle, 1982; Ostroff & Ilgen, 1992). In diesem Zusammenhang scheint auch das Ausblenden irrelevanter Informationen eine große Rolle zu spielen (Gruenfeld & Arbuthnot, 1969). Andere Studien brachten Halo-Effekte unter anderem mit Schizophrenie (Glenn A. Miller & Chapman, 1968), geringer Dominanz und emotionaler Instabilität (Buffardi & Gibson, 1980), mangelnder Fähigkeit zu logischem Denken (Dickinson & Glebocki, 1990), schlechter Informationsverarbeitung und Gedächtnisleistung (Jennings, Palmer, & Thomas, 2004), geringem Kognitionsbedürfnis (Palmer & Feldman, 2005) sowie, im Falle von Selbsteinschätzungen, sehr hohem Selbstbewusstsein (Anusic, Schimmack, Pinkus, & Lockwood, 2009) in Verbindung.

Ein bestimmter Bereich der Literatur führt Halo-Effekte auf sogenannte implizite Persönlichkeitstheorien zurück. Dieser Begriff wurde von Bruner und Taguiri (1954) geprägt und beschreibt naive Annahmen über den Zusammenhang zwischen verschiedenen Eigenschaften von Personen. Gemäß dieser Erklärung werden Halo-Effekte dadurch verursacht, dass automatisch von einem Merkmal auf ein anderes geschlossen wird (Wolfradt, 2008). Beispielsweise ergaben mehrere Untersuchungen von Asch (1946), dass

fiktive Personen, die als warmherzig beschrieben wurden, auch in starkem Maße als großzügig eingeschätzt wurden, obwohl keinerlei objektive Informationen über dieses Merkmal vorhanden waren. Auch viele andere Autoren fanden Belege dafür, dass Halo-Effekte aus recht stabilen kognitiven Schemata und Kategorisierungsprozessen resultieren (Feeley, 2002a; Mount & Thompson, 1987; Nathan & Lord, 1983; Suter & Roberts, 1987). Werden solche Zusammenhänge vorausgesetzt, können Halo-Effekte auch als Heuristik interpretiert werden (Cooper, 1981a). In der Literatur gibt es Befunde, die diese Theorie zusätzlich stützen: Wie auch andere Heuristiken treten Halo-Effekte vor allem dann auf, wenn nur wenig objektive Informationen verfügbar sind, um das Risiko falscher Urteile und Entscheidungen möglichst gering zu halten (Boatwright et al., 2008; De Soto, 1961; Mills & O'Neal, 1971; O'Neal & Mills, 1969). Dies kann in einigen Fällen zutreffend sein, in anderen Fällen jedoch zu systematischen Fehlern führen. Des Weiteren können sowohl Halo-Effekte als auch Heuristiken besonders dann entstehen, wenn Urteile besonders schnell gefällt werden müssen (Klauer & Schmeling, 1990) und sich die beurteilenden Personen in einer positive Stimmungslage befinden (Sinclair, 1988).

Halo-Effekte aufgrund von Aspekten der Situation

Ein dritter Bereich in der Literatur versucht, Halo-Effekte ausschließlich durch verschiedene Merkmale der Urteilssituation zu erklären: Murphy und Anhalt (1992) vertraten die Auffassung, dass Halo-Effekte nicht etwa durch stabile kognitive Schemata, sondern vielmehr durch den spezifischen Kontext verursacht werden und damit variabel seien. Belege für diese Theorie konnten beispielsweise in einer Untersuchung von Ostroff (1993) gefunden werden, in der 64 Manager ihre Angestellten auf unterschiedlichen Dimensionen beurteilten. Das angebliche Ziel der Erhebung wurde dabei systematisch variiert. Waren die Manager der Auffassung, dass ihre Einschätzungen administrativen Zwecken, wie Beförderungen und Gehältern, dienten, wirkten starke Halo-Effekte. Ging es dagegen angeblich um Dokumentation oder Feedback, traten diese Verzerrungen in viel schwächerem Maße auf. Weitere Studien ergaben unter anderem, dass Halo-Effekte besonders dann auftraten, wenn die Attribute nicht klar definiert waren (Cooper, 1981b; Guilford, 1954; Symonds, 1925), Zeitdruck bestand (Klauer & Schmeling, 1990) oder vor Urteilen und Entscheidungen eine Gruppendiskussion durchgeführt wurde (Palmer & Loveland, 2008). Ebenfalls existieren Untersuchungen, die dem im letzten Abschnitt

geschilderten Einfluss von Gedächtnisleistungen widersprechen. So wurde etwa nachgewiesen, dass Methoden, die Gedächtnisleistungen verbessern, nicht automatisch auch zu geringeren Halo-Effekten führen (Cook, Marsh, & Hicks, 2003).

Neben den genannten äußeren Faktoren sind auch rein formale Aspekte der Urteilsabfrage in der Lage, Halo-Effekte zu verursachen: In einer Erhebung von Sawyer und Raybould (1981) schätzten Lehrer ihre Schüler hinsichtlich ihrer Fähigkeiten und Verhaltensweisen ein. Die Korrelationen zwischen den Beurteilungen der verschiedenen Dimensionen gaben Aufschluss über die Stärke der Halo-Effekte. Hohe Werte zeigten sich vor allem dann, wenn alle Items auf derselben Seite des Fragebogens zu finden waren. Dies wurde darauf zurückgeführt, dass alle Bewertungen immer präsent waren und sich gegenseitig beeinflussten. Ähnliche Befunde ergaben sich auch bei der Einschätzung der Attraktivität von unterschiedlichen Gesichtern (Geiselman et al., 1984). C. C. Clark und Lawless (1994) fanden heraus, dass auch das Fehlen bestimmter Dimensionen in einem Fragebogen zu Halo-Effekten bei der Bewertung anderer Attribute führen kann: In zwei Experimenten wurde der Geschmack von Getränken dann als süßer bewertet, wenn der Geruch nicht auch abgefragt wurde. Die fehlende Möglichkeit, den Geruch zu bewerten, führte somit zu einer extremeren Beurteilung des Geschmacks. Auch die Verwendung kontinuierlicher und rein numerischer Ratingskalen geht nachweislich mit stärkeren Halo-Effekten einher (Kasten & Weintraub, 1999).

2.1.3 Vermeidung von Halo-Effekten

Im Zuge der Erforschung der Ursachen von Halo-Effekten wurde eine Reihe von Methoden entwickelt, um solche Verzerrungen zu vermeiden. Grundsätzlich setzten diese Strategien entweder bei personellen oder aber bei situativen Aspekten an. Ein anderer Forschungsbereich stellte sich demgegenüber die Frage, ob es überhaupt notwendig und sinnvoll ist, Halo-Effekte zu reduzieren beziehungsweise gänzlich zu vermeiden.

Veränderung personeller Aspekte

Entsprechend der verschiedenen potentiellen Einflussfaktoren von Halo-Effekten wurden in der Vergangenheit viele Möglichkeiten vorgeschlagen, um die Urteilsbildung von Menschen zu verbessern. Dazu gehören zum Beispiel die Vermeidung kognitiver Erregung

(O'Neal, 1971) oder die Verbesserung der Erinnerungsleistungen durch schriftliche Aufzeichnungen vergangener Beobachtungen und Urteile (Palmer, Thomas, & Maurer, 2003). Die Ergebnisse der Untersuchungen von Davison (1973) sowie Jaeger und Freijo (1975) implizieren des Weiteren, dass Halo-Effekte vermindert werden können, wenn die wahrgenommene soziale Distanz zwischen dem Beurteilenden und der beurteilten Person verringert wird. Auch sollte sich das Ausmaß an Halo-Effekten dann reduzieren, wenn die Beurteilenden zuvor ausgiebige Informationen über die zu beurteilenden Personen und Dimensionen erhalten (Kozlowski & Kirsch, 1987; Kozlowski et al., 1986).

Verschiedene Autoren untersuchten den Einfluss vielfältiger Trainings auf die Urteilsfähigkeit von Personen. Laut Athey und McIntyre (1987) umfasst ein gutes Training drei Aspekte: eine Beschreibung der einzuschätzenden Dimensionen, Übung und Rückmeldung sowie Erklärungen korrekter Ratings durch Experten. Andere Studien verwendeten zudem Videosequenzen zur Veranschaulichung von Halo-Effekten und ließen diese in Gruppen diskutieren (Latham, Wexley, & Pursell, 1975). Im Gegensatz dazu beschränkten sich einige Autoren darauf, ihre Probanden nur über das Wirken von Halo-Effekten aufzuklären und sie zu bitten, diese zu vermeiden (Follman et al., 1974; Foon, 1986). Die Effektivität der unterschiedlichen Trainingsmethoden zur Reduzierung von Halo-Effekten konnte in einer großen Zahl an Untersuchungen nachgewiesen werden (Baltes & Parker, 2000; Bernardin & Walter, 1977; Brown, 1968; McIntyre, Smith, & Hassett, 1984; Pulakos, 1984). Jedoch existieren auch Studien, die dem Nutzen von Trainings widersprechen: Bernardin (1978) entdeckte, dass der Effekt von Trainings zwar vorhanden war, aber mit der Zeit verschwand. Andere Experimente konnten gar keinen Einfluss der Methoden feststellen (Edwards & Waters, 1984; Murphy & Anhalt, 1992; Wetzell et al., 1981). Zum Teil zeigte sich sogar, dass die Aufklärung der Probanden über Halo-Effekte erst recht zu Urteilsverzerrungen führte (Franke, Hoffmann, & Kamphaus, 1974).

Veränderung situativer Aspekte

Neben der Veränderung personeller Aspekte bietet die Literatur unterschiedliche Ansätze, um Halo-Effekte mit Hilfe von situativen Faktoren zu reduzieren: Klauer (1988) instruierte seine Versuchsteilnehmer, deutsche Politiker auf verschiedenen Skalen zu bewerten. In einer Experimentalbedingung sollte zuvor überlegt werden, welche Urteile wohl ein Vertreter gegenteiliger Ansichten abgeben würde. Die Befunde machten deutlich,

dass das bloße Bedenken der Gegenseite zu niedrigeren Halo-Effekten führte. In einem anderen Experiment konnte der Halo-Effekt von Attraktivität dadurch verringert werden, dass vor den eigentlichen Urteilen die Erwünschtheit bestimmter äußerer Merkmale eingeschätzt werden sollte. Diese Prozedur diene somit als Priming (Taylor, Bernardin, & Riegelhaupt, 1984). In anderen Studien traten Halo-Effekte auch dann seltener auf, wenn Urteile direkt im Anschluss an die entsprechenden Verhaltensbeobachtungen erfolgten. Diese Befunde wurden darauf zurückgeführt, dass zu jenem Zeitpunkt noch ausreichend valide Informationen im Gedächtnis repräsentiert waren, wogegen später im Wesentlichen nur noch ein globaler Eindruck vorhanden war (Dennis, 2007; Murphy & Balzer, 1986). Des Weiteren sollte sichergestellt sein, dass vor den Bewertungen genug Bedenkzeit zur Verfügung gestellt wird, da andernfalls eher globale Eindrücke anstelle von inhaltlichen Informationen abgerufen werden (Klauer & Schmeling, 1990).

Ein bestimmter Teil der Literatur zur Vermeidung von Halo-Effekten beschäftigt sich mit der Optimierung der Urteilsabfrage und der verwendeten Skalen: Johnson und Vidulich (1956) zeigten auf, dass sich Halo-Effekte dann reduzierten, wenn die verschiedenen Dimensionen nicht parallel, sondern zeitlich versetzt beurteilt wurden. Bei der Bewertung mehrerer Individuen hieße das, dass alle Personen erst hinsichtlich einer Eigenschaft beurteilt werden müssen, bevor die nächste Eigenschaft präsentiert wird. In anderen Untersuchungen konnten diese Befunde allerdings nicht repliziert werden (Johnson, 1963; Van Scotter et al., 2007). Schmitt (1992) fand heraus, dass sich Halo-Effekte im Laufe mehrerer Bewertungsdurchgänge verringerten, da die Probanden versuchten, vergangene Urteilsfehler wieder gut zu machen. Bezüglich der Skala selbst nannte Cooper (1983) zwei wichtige Kriterien zur Vermeidung von Halo-Effekten: Eine gute Skala dürfe die Probanden nicht dazu zwingen, sich auf ihren globalen Eindruck zu verlassen und müsse sich bei allen Dimensionen auf saliente Beobachtungen beziehen. Der Einfluss gut operationalisierter Skalen wurde schon in anderen Studien nachgewiesen (Stevens et al., 1998; Tziner, 1984). Halo-Effekte konnten auch bereits dadurch reduziert werden, dass die Skalen nicht kontinuierlich waren (Kasten & Weintraub, 1999) oder mehr positive als negative Optionen enthielten (Fox et al., 1994; Jones & Fletcher, 2002). Die verschiedenen Autoren erklärten diese Befunde damit, dass Skalen dieser Art nicht nur numerisch betrachtet, sondern in verbale Formulierungen übersetzt und damit spezifischer verarbeitet wurden.

Bedeutsamkeit erwünschter Halo-Effekte

Der Großteil der bisherigen Forschungsberichte betrachtete Halo-Effekte stets als Urteilsfehler, die es zu vermeiden gilt. Eine vergleichsweise kleine Anzahl an Studien beschäftigte sich jedoch mit der Frage, ob und in welchem Maße es sich bei Halo-Effekten um die Repräsentation wahrer Zusammenhänge handeln könnte. Bingham (1939) wies als einer der ersten Autoren darauf hin, dass Halo-Effekte nicht ausschließlich falsch sein müssen, sondern zu einem gewissen Anteil durchaus korrekt sein können. Cooper (1981b) bezeichnete diese korrekten Zusammenhänge später als illusorische Halo-Effekte, die eigentlichen Urteilsverzerrungen dagegen als wahre Halo-Effekte. In mehreren Untersuchungen konnte bereits nachgewiesen werden, dass Halo-Effekte positiv mit der Reliabilität und Validität von Ratings korreliert waren (Jackson, 1996; Kasten & Weintraub, 1999; Murphy & Balzer, 1986; Nathan & Tippins, 1990). Andere Autoren fanden dagegen keine (Goffin et al., 2003) oder ausschließlich negative Zusammenhänge (Fiscaro, 1988) zwischen Halo-Effekten und verschiedenen Maßen der Akkuratess.

Mit der Erkenntnis, dass Halo-Effekte sowohl wahr als auch illusorisch sein können, stellte sich die Frage, wie diese Phänomene zusammenhängen. Solomonson und Lance (1997) vermuteten, dass alle Urteile sowohl Effekte des globalen Eindrucks als auch tatsächliche Verhaltensweisen widerspiegeln. Gleichzeitig scheinen diese beiden Einflüsse völlig unabhängig voneinander zu sein. Kiker und Motowidlo (1998) argumentierten, dass das Ausmaß, in dem Halo-Effekte wahre Zusammenhänge widerspiegeln, durch die verarbeiteten Hinweisreize in der jeweiligen Situation bedingt sei. Auch die Interpretation von Halo-Effekten als Heuristiken (Cooper, 1981a) lässt darauf schließen, dass ihr Auftreten in vielen Kontexten sinnvoll ist und mit einer gewissen Wahrscheinlichkeit zu korrekten Urteilen führt (Boatwright et al., 2008; De Soto, 1961; Mills & O'Neal, 1971; O'Neal & Mills, 1969). Obwohl wahre und illusorische Halo-Effekte unabhängig voneinander zu sein scheinen, traten in der Vergangenheit einige Probleme im Zusammenhang mit den verschiedenen Vermeidungsstrategien auf: In einer Studie von Borman (1975) zeigte sich, dass Trainings zwar mit geringeren Halo-Effekten einhergingen, damit aber auch die Reliabilität der Urteile sank. Andere Untersuchungen machten zudem deutlich, dass Trainings zu einer Überkorrektur und damit auch zu einer geringeren Validität der Bewertungen führen können (Bernardin & Pence, 1980; Fiscaro, 1988).

2.1.4 Erfassung von Halo-Effekten

Ein wichtiges Thema in vielen vergangenen Untersuchungen war die Frage, welche Maße am besten dazu in der Lage sind, Halo-Effekte adäquat zu erfassen. Im Laufe der bisherigen Forschung wurden hierfür unterschiedliche Lösungen vorgeschlagen. Mehrere Autoren versuchten, die verwendeten Maße zusammenzufassen und gegenüberzustellen (Balzer & Sulsky, 1992; Cooper, 1981b; Fox, Bizman, & Herrmann, 1983). Auch die methodische Trennung von wahren und illusorischen Halo-Effekten war ein immer wiederkehrendes Thema in der Literatur.

Maße von Halo-Effekten

Für die Erfassung von Halo-Effekten wurden im Laufe der Zeit mehrere Ansätze verfolgt. Nach dem heutigen Stand der Dinge können fünf verschiedene Methoden voneinander unterschieden werden:

Inter-Rating-Korrelationen sind das älteste und am häufigsten verwendete Maß. Neben frühen Studien, zu deren Zeit noch keine alternativen Methoden erforscht waren (Bingham, 1939; Newcomb, 1931; Thorndike, 1920), ist dieses Maß auch noch in einem Teil der neueren Literatur zu finden (Dennis, 2007; Jackson & Furnham, 2001; Zysberg & Nevo, 2004). Der Ausgangspunkt der Erhebung ist die Einschätzung einer Person oder eines Objekts durch einen Beurteilenden auf einer Reihe unterschiedlicher Dimensionen. Anschließend werden alle Urteile paarweise korreliert und der Mittelwert aus allen Korrelationen berechnet. Ein hoher Wert der durchschnittlichen Inter-Rating-Korrelation impliziert einen starken globalen Einfluss und damit einen großen Halo-Effekt. Der Vorteil dieser Methode besteht darin, dass sie sich nah an die ursprüngliche Definition von Halo-Effekten (Thorndike, 1920) hält und damit anschaulich und valide ist (Oh & Ramaprasad, 2003). Kritiken wurden dagegen dahingehend laut, dass Halo-Effekte durch Korrelationen häufig unterschätzt werden und illusorische Halo-Effekte nur ausgeschlossen werden können, wenn die wahren Korrelationen der verschiedenen Dimensionen bekannt sind (Fiscaro & Lance, 1990; Fiscaro & Vance, 1994; Murphy & Jako, 1989).

Das zweite abhängige Maß zur Ermittlung von Halo-Effekten stellen Intra-Objekt-Varianzen dar. Dieses Maß wurde bisher sowohl in älteren (Bernardin, 1977; Bernardin & Walter, 1977; Borman, 1975) als auch neueren Studien (Jones & Fletcher, 2002; Palmer &

Loveland, 2008; Van Scotter et al., 2007) gleichermaßen verwendet. Wie auch bei Inter-Rating-Korrelationen geht die Erhebung davon aus, dass eine Person oder ein Objekt auf verschiedenen Dimensionen beurteilt wird. Die Varianz der Ratings über alle Dimensionen hinweg dient hierbei als Maß für Halo-Effekte: Eine niedrige Varianz bedeutet, dass der Beurteilende nur in äußerst geringem Maße zwischen den einzelnen Dimensionen diskriminiert. Dies wiederum ist ein Indiz für einen starken globalen Einfluss und folglich einen großen Halo-Effekt. Trotz ihrer Anschaulichkeit wurde die Methode dahingehend kritisiert, dass sie der ursprünglichen Definition von Thorndike (1920) nicht gerecht werde und in bestimmten Fällen deutlich andere Ergebnisse hervorbringe als Inter-Rating-Korrelationen (Pulakos, Schmitt, & Ostroff, 1986).

Eine dritte Methode zur Ermittlung von Halo-Effekten ist die Berechnung einer Subjekt x Objekt x Urteilsdimension ANOVA anstelle von Korrelationen oder Varianzen. Hierfür ist es notwendig, dass gleich mehrere Beurteilende, also Subjekte, vorhanden sind, die verschiedene Objekte oder Personen bewerten. Trifft mindestens ein Subjekt bei einem oder mehreren Objekten zu globale Urteile, äußert sich dies in Form einer signifikanten Interaktion von Subjekt und Objekt. Halo-Effekte können damit direkt anhand der ANOVA gezeigt werden. Im Gegensatz zu den bereits geschilderten Maßen geht diese Methode davon aus, dass Halo-Effekte nicht auf dem Beurteilenden selbst, sondern auf der spezifischen Wechselwirkung von beurteilender und beurteilter Person beruhen. Die Anwendung dieses Ansatzes ist jedoch dann problematisch, wenn die Ratings nur in geringem Maße variieren, weil beispielsweise alle Subjekte gleichermaßen durch Halo-Effekte beeinflusst werden. In solchen Fällen können signifikante Ergebnisse nur schwer gefunden werden (Balzer & Sulsky, 1992). Die Methode wird daher in der Praxis nur selten verwendet (Johnson & Vidulich, 1956; Lee et al., 1981; Mount, 1984).

Strukturgleichungsmodelle stellen die jüngste der vorhandenen Methoden dar und wurden vor allem in aktuelleren Studien den alternativen Maßen vorgezogen (Anusic et al., 2009; Dompnier et al., 2006; Iramaneerat & Yudkowsky, 2007). Bei dieser Methode wird durch verschiedene statistische Testverfahren die grundlegende Struktur von Urteilen ermittelt. Dabei können auch Halo-Effekte aufgedeckt werden. Der große Vorteil dieses Verfahrens liegt in seiner Vielseitigkeit: Alle möglichen Ursachen und Einflussfaktoren von Halo-Effekten können gleichzeitig berücksichtigt und modelliert werden (Holbrook, 1983). Jedoch ist die Verwendung von Strukturgleichungsmodellen im Vergleich zu den anderen

Methoden mit einem sehr großen Aufwand verbunden, da sie mit einem schrittweisen Testen und Anpassen einhergeht. Eine Modifizierung des Modells bedarf dabei stets der Überprüfung anhand einer neuen Stichprobe (Mulaik & Millsap, 2000).

Die fünfte und letzte Methode ist die gezielte Manipulation von dargebotenen Informationen und die Messung der daraus resultierenden Urteile. Über den Vergleich der verschiedenen experimentellen Bedingungen können Halo-Effekte berechnet werden. Als abhängiges Maß können dabei beispielsweise Inter-Rating-Korrelationen, Intra-Objekt-Varianzen, aber auch die bloßen Mittelwerte der Ratings verwendet werden. Aufgrund dieser Vielseitigkeit und Einfachheit wurde die Methode bisher in einer Vielzahl von älteren und neueren Untersuchungen angewandt (Abikoff et al., 1993; Coombs & Holladay, 2006; Surawski & Ossoff, 2006; Wetzell et al., 1981). Der Nachteil besteht jedoch darin, dass auf diese Weise keine Feldstudien durchgeführt werden können und sich die Erhebungen folglich auf reine Laborexperimente beschränken.

Trennung von wahren und illusorischen Halo-Effekten

Ein großes Problem bei der Berechnung von Halo-Effekten ist die Tatsache, dass wahre Zusammenhänge zwischen verschiedenen Eigenschaften auf methodische Weise ausgeschlossen werden müssen. In der Literatur wurden hierzu im Laufe der Zeit unterschiedliche Vorschläge gemacht. Wie bereits geschildert, können die verschiedenen abhängigen Maße nur dann als Halo-Effekte interpretiert werden, wenn die wahren Zusammenhänge und Strukturen zwischen den unterschiedlichen Urteilsdimensionen bekannt sind. In einzelnen Untersuchungen wurden Expertenratings (Engelhard, 1994; Murphy & Reynolds, 1988) oder auch Selbstbeurteilungen (Kammann, Smith, Martin, & McQueen, 1984) als Maß für die wahren Zusammenhänge verwendet und mit den eigentlichen Bewertungen verglichen. Dieses Vorgehen wurde jedoch häufig kritisiert, da Halo-Effekte oft auch bei Beurteilungen durch die eigene Person (Fox et al., 1994; Jones & Fletcher, 2002; Pike, 1999) und durch Experten (Chandon & Wansink, 2007; Mumma, 2002; Zysberg & Nevo, 2004) auftreten. Eine weitere Möglichkeit zur Abgrenzung von Halo-Effekten wurde mit der Generierung sogenannter Peabody-Quadrupeln geschaffen (Borkenau & Ostendorf, 1987; Peabody, 1967). Diese bestehen aus jeweils zwei Paaren gegensätzlicher Eigenschaften, die einmal positiv (zum Beispiel *großzügig* versus *sparsam*) und einmal negativ formuliert (zum Beispiel *verschwenderisch* versus *geizig*) sind. Werden

alle diese Attribute separat beurteilt, können rein deskriptive Aspekte von evaluativen Aspekten abgegrenzt werden. Ob diese Methode allerdings Halo-Effekte im eigentlichen Sinne wiedergibt, ist unklar und wurde bisher nicht überprüft.

Die meisten Methoden zur Trennung von wahren und illusorischen Halo-Effekten beziehen sich auf eine Reihe von statistischen Vorgehensweisen. Diese bestanden in der bisherigen Forschung vor allem darin, die Effekte von globalen Eindrücken unter der Verwendung von partiellen Korrelationen (Bartlett, 1983; F. J. Landy, Vance, & Barnes-Farrell, 1982; F. J. Landy et al., 1980; Myers, 1965) oder multiplen Regressionen (Feldman, 1986; Henik & Tzelgov, 1985) aus den spezifischen Urteilen herauszurechnen. Während einige Studien den Erfolg dieser Methoden belegten, führten andere Untersuchungen auf, dass es sich bei diesen Befunden lediglich um statistische Artefakte handele (Bownas & Bernardin, 1991; Harvey, 1982) und in Wahrheit keine Effekte bestünden (Holzbach, 1978). Mehrere Autoren kritisierten zudem, dass durch partielle Korrelationen beziehungsweise multiple Regressionen nicht nur Halo-Effekte, sondern auch wahre Zusammenhänge herausgerechnet werden und dadurch die Validität der Urteile sinke (Hulin, 1982; Lance & Woehr, 1986; Mossholder & Giles, 1983; Murphy, 1982; Nathan & Tippins, 1990).

2.2 Einheitlicher Halo-Effekt versus vielfältige Halo-Effekte

Wie im letzten Abschnitt deutlich wurde, existieren nach dem heutigen Stand der Forschung viele Studien, die sich mit verschiedenen Aspekten von Halo-Effekten befassen. Aufgrund dieser Tatsache scheint das Phänomen für viele Autoren bereits vollständig untersucht und geklärt zu sein. Allgemein betrachtet stagniert die Erforschung von Halo-Effekten mittlerweile, was sich in der geringen Anzahl von Publikationen in den letzten Jahren erkennen lässt. Dies ist sehr bedauerlich, da noch viel Potential in dieser Thematik liegt. Von besonderer Wichtigkeit ist die Frage, ob Halo-Effekte vielfältig und variabel sind oder stattdessen auf einen einheitlichen Prozess zurückgeführt werden können. Der überwiegende Teil der Literatur, der sich mit diesem Problem auseinandersetzt, postuliert den letzteren Ansatz: Halo-Effekte wurden in der Vergangenheit unter anderem als unvermeidlich (Feldman, 1986; Newcomb, 1931), universell (Cooper, 1981b), einheitlich (Nathan, 1986), allgegenwärtig (Kozlowski et al., 1986), konstant (Pike, 1999) und frei von

Widersprüchen (Pulakos et al., 1986; Saal, Downey, & Lahey, 1980) bezeichnet. Dagegen stellten nur wenige Autoren die Annahme eines einheitlichen Halo-Effekts in Frage und führten Belege für die Vielfältigkeit von Halo-Effekten auf (Balzer & Sulsky, 1992; Fox et al., 1983; Fox, Bizman, & Hoffman, 1989; Murphy, Jako, & Anhalt, 1993).

Die vorliegende Arbeit vertritt die Annahme, dass Halo-Effekte keinem universellen Prozess unterliegen, sondern verschiedene Formen annehmen und auf unterschiedliche Arten wirken können. Zwar ist das Phänomen in der Vergangenheit gut erforscht worden, jedoch existieren nach wie vor ungeklärte Fragen und zum Teil auch widersprüchliche Ergebnisse. Allein aus diesem Grund kann die Annahme eines einheitlichen Halo-Effekts angezweifelt werden. Neben diesen Unklarheiten lassen sich in der Literatur auch Befunde finden, die explizit für die Vielfältigkeit von Halo-Effekten sprechen.

2.2.1 Offene Fragen der bisherigen Forschung

In der Vergangenheit wurden Halo-Effekte in einer Vielzahl von Bereichen nachgewiesen und verschiedene mögliche Ursachen aufgezeigt. Ebenso beschäftigten sich mehrere Autoren mit Vermeidungsstrategien sowie der Erfassung von Halo-Effekten. Obwohl das Phänomen damit bereits umfassend erforscht wurde, zeigt eine genauere Betrachtung der Literatur, dass sich viele Befunde gegenseitig widersprechen und immer noch viele Ansätze für weitere Untersuchungen gegeben sind.

Widersprüchliche Befunde in der bisherigen Forschung

Die verschiedenen Untersuchungen zu Halo-Effekten brachten einige Ergebnisse hervor, die sich zum Teil widersprechen. Dies betrifft vor allem die Frage nach möglichen Strategien zur Vermeidung von Halo-Effekten. Zum Beispiel ergaben mehrere Studien, dass Halo-Effekte unwahrscheinlicher wurden, wenn sich die Beurteilenden gut mit den zu bewertenden Personen und Attributen auskannten (Kozlowski & Kirsch, 1987; Kozlowski et al., 1986). Andere Erhebungen machten dagegen deutlich, dass selbst Experten auf unterschiedlichen Gebieten nicht vor Halo-Effekten geschützt waren (Chandon & Wansink, 2007; Mumma, 2002; Zysberg & Nevo, 2004). Unterschiedliche Befunde zeigten sich auch bezüglich der Bekanntheit von den beurteilenden und beurteilten Personen: Abhängig von dem jeweiligen Experiment führte eine hohe Bekanntheit entweder zu geringeren (Van

Scotter et al., 2007) oder stärkeren Halo-Effekten (Jacobs & Kozlowski, 1985; Lachman & Bass, 1985; Lance et al., 1990). Des Weiteren wurde bereits eine Verbindung zwischen schwachen Gedächtnisleistungen und größeren Halo-Effekten gefunden (Jennings et al., 2004), die in einer anderen Untersuchung wiederum nicht vorhanden war (Cook et al., 2003). Besonders viele Widersprüche existieren in Bezug auf die Effektivität von Trainings zur Reduzierung von Halo-Effekten: Während zahlreiche Autoren einen positiven Effekt nachweisen konnten (Baltes & Parker, 2000; Bernardin & Walter, 1977; Brown, 1968; McIntyre et al., 1984; Pulakos, 1984), sprachen ähnlich viele Studien deutlich gegen die Wirksamkeit unterschiedlicher Trainingsmethoden (Bernardin, 1978; Edwards & Waters, 1984; Franke et al., 1974; Murphy & Anhalt, 1992; Wetzell et al., 1981).

Neben dem Bereich der Vermeidungsstrategien zeigten sich in der Literatur auch Widersprüche bei der Erfassung von Halo-Effekten. Einige Untersuchungen, in denen Halo-Effekte mit Hilfe mehrerer Maße berechnet wurden, ließen beispielsweise erkennen, dass unterschiedliche Methoden auch zu unterschiedlichen Ergebnissen führten (Alliger & Williams, 1989; Balzer & Sulsky, 1992; Fox et al., 1983; Fox et al., 1989; Pulakos et al., 1986). Andere Autoren stellten fest, dass je nach den äußeren Bedingungen bestimmte Maße für die Erhebung geeigneter waren als andere (Fisicaro & Vance, 1994; Saal et al., 1980). Weitere Unstimmigkeiten lassen sich bei eventuellen Korrelaten von Halo-Effekten finden: Viele Analysen machten einen positiven Zusammenhang zwischen Halo-Effekten und der Reliabilität und Validität von Urteilen deutlich (Jackson, 1996; Kasten & Weintraub, 1999; Murphy & Balzer, 1986; Nathan & Tippins, 1990). Andere Studien zeigten dagegen keine oder sogar eine negative Verbindung auf (Fisicaro, 1988; Goffin et al., 2003).

Weitere Forschungsansätze

Obwohl Halo-Effekte bereits in vielerlei Hinsicht untersucht wurden, existieren dennoch Bereiche, die bisher weitgehend vernachlässigt wurden. So beschäftigt sich zum Beispiel der überwiegende Teil der Forschung mit Halo-Effekten in beruflichen und schulischen Kontexten. In diesem Zusammenhang standen vor allem Attribute, wie etwa Intelligenz, Fleiß, Ordentlichkeit und Eigeninitiative, im Vordergrund, die sich alle auf die Kompetenz und Leistung von Personen bezogen. Soziale Kompetenzen, wie Ehrlichkeit, Hilfsbereitschaft und Geselligkeit, wurden dagegen nur vereinzelt und wenig detailliert betrachtet. Auch verschiedene Auftrittsbereiche von Halo-Effekten, wie beispielsweise

Strafrecht (D. Landy & Aronson, 1969; O'Sullivan, 2003), Partnerschaft (Murray & Holmes, 1997; Terman, 1939) und Kultur (Intons-Peterson & Samuels, 1978; Zebrowitz et al., 1993), wurden bisher nur in sehr wenigen Studien thematisiert. Weitere ungeklärte Fragen existieren bei der Erforschung möglicher Ursachen und Einflussfaktoren von Halo-Effekten. So ist zum Beispiel die Verbindung beziehungsweise Abgrenzung von Halo-Effekten und Diskriminierung, die in wenigen frühen Studien thematisiert wurde (Caldwell & Mowry, 1934; Zebrowitz et al., 1993), nach wie vor nicht eindeutig geklärt. Auch die Erklärung von Halo-Effekten mit Hilfe impliziter Persönlichkeitstheorien ist ein weitgehend unerforschtes Gebiet (Schneider, 2004; Wolfradt, 2008). Des Weiteren konnte bisher keine vollständig geeignete Methode zur Trennung von wahren und illusorischen Halo-Effekten aufgezeigt werden. Statt der Verwendung statistischer Methoden wäre es beispielsweise möglich, die Informationen auf experimentellem Weg so zu manipulieren, dass es keine tatsächlichen Zusammenhänge gibt und somit ausschließlich wahre Halo-Effekte in den Urteilen widergespiegelt werden. Ein solches Vorgehen wurde bisher nicht angewandt.

Ein großes und bisher nur zum Teil gelöstes Problem ist die quantitative Erfassung von Halo-Effekten. Mehrere Autoren führten schon auf, dass das Vorhandensein von Halo-Effekten zwar durch verschiedene Methoden nachgewiesen, die genaue Größe aber nicht erfasst werden könne (Cooper, 1981b; Thorndike, 1920; Viswesvaran, Schmidt, & Ones, 2005). Dies ist darauf zurückzuführen, dass Inter-Rating-Korrelationen und die anderen Maße nur dann quantitativ interpretierbar sind, wenn die wahren Daten bekannt sind. Da die statistische Trennung von Halo-Effekten und wahren Zusammenhängen problematisch ist, ist das aber nur dann der Fall, wenn keine Feldstudien durchgeführt, sondern die Informationen experimentell manipuliert werden. In diesem Zusammenhang war es bisher auch nur eingeschränkt möglich, verschiedene Halo-Effekte gezielt miteinander zu vergleichen. Beispielsweise existieren nur wenige Untersuchungen, in denen das Ausmaß von positiven Halo-Effekten der Größe von negativen Halo-Effekten, auch Horn-Effekte genannt, gegenübergestellt wurde (Nisbett & Wilson, 1977; Peeters, 1991; Rizzo & Frank, 1977; Tsui & Barry, 1986). Die meisten Erhebungen fokussierten stattdessen nur eine der beiden Valenzen. Auch Vergleiche von Halo-Effekten bei der Beurteilung verschiedener Persönlichkeitsdimensionen, wie etwa Agency und Communion (Judd, James-Hawkins, Yzerbyt, & Kashima, 2005; Suitner & Maass, 2008) oder den Big Five (Anusic et al., 2009; Srivastava, Guglielmo, & Beer, 2010), waren bisher die Ausnahme. Neben diesen Ansätzen

besteht auch noch Forschungsbedarf bei der Betrachtung von Kontextfaktoren, welche nachweislich große Einflüsse auf Halo-Effekte haben können (Murphy et al., 1993). So konnte schon vereinzelt gezeigt werden, dass Halo-Effekte davon beeinflusst werden, welche Personen (Palmer & Feldman, 2005; Palmer & Loveland, 2008; Palmer et al., 2003) oder Dimensionen (M. F. Kaplan, 1974; Tourangeau & Rasinski, 1988) im Einzelnen eingeschätzt werden. Auch verschiedene Eigenschaften der Beurteilenden stehen oft im Zusammenhang mit Halo-Effekten (R. M. Kaplan, 1978; Newcomb, 1931; Vance, Winne, & Wright, 1983). An dieser Stelle kann angesetzt werden, um zum Beispiel verschiedene Berufsgruppen, Situationen oder auch Stichproben miteinander zu vergleichen.

2.2.2 Belege für die Vielfältigkeit von Halo-Effekten

In einem großen Teil der Studien zu Halo-Effekten wurde der Standpunkt vertreten, dass die Wirkungsweise und Prozesse dieses Phänomens ein einheitliches Schema aufweisen. Jedoch verrät ein genauerer Blick in die Literatur, dass verschiedene Befunde nicht konform mit dieser Annahme sind und stattdessen für vielfältige und variable Mechanismen von Halo-Effekten sprechen.

Wirkungsweisen von Halo-Effekten

Die Tatsache, dass sich Halo-Effekte auf unterschiedliche Weise äußern können, wurde erstmals von Nisbett und Wilson (1977) festgestellt. In ihrer Untersuchung führten die beiden Autoren auf, dass Halo-Effekte eine schwache oder starke Form aufweisen können. Eine schwache Form bedeute dabei einerseits, dass Halo-Effekte im Sinne von Vorannahmen über nicht gegebene Informationen wirken können. Auf der anderen Seite können aber auch ambivalente Informationen positiv oder negativ geprägt werden. Unter der starken Form von Halo-Effekten verstanden die Autoren dagegen den Fall, dass zwar eindeutige Informationen vorhanden seien, diese jedoch verzerrt wahrgenommen und interpretiert werden. Betrachtet man alle bisher durchgeführten Untersuchungen, lassen sich in der Tat diese drei Wirkungsweisen unterscheiden:

Belege für Halo-Effekte als Vorannahmen lassen sich vor allem im Bereich der Attraktivitätsforschung finden. Eine Vielzahl von Experimenten wies nach, dass attraktive Personen unter anderem als besonders intelligent, erfolgreich, gesellig und fröhlich

eingeschätzt wurden, wenn, abgesehen von Bildern, keine weiteren Informationen präsentiert wurden (Dion et al., 1972; Geiselman et al., 1984; Larose & Standing, 1998; Lucker et al., 1981; A. G. Miller, 1970; Sappenfield, 1971). Auch lächelnde Gesichter (Lau, 1982; Otta et al., 1993) und schlanke, zum Teil auch dicke Körper (Wade et al., 2007; Wade et al., 2003) konnten ähnliche Effekte in Bezug auf verschiedene Attribute hervorrufen. Andere Autoren stellten fest, dass erfolgreichen Schülern (Anastasiow, 1963) oder bedeutenden historischen Persönlichkeiten (Bowden et al., 1934) ohne weiteres Vorwissen automatisch viele positive Eigenschaften zugeschrieben wurden.

Zu Halo-Effekten als Prägung ambivalenter Informationen führte besonders Asch (1946) viele Untersuchungen durch. In seinen Experimenten bot er den Probanden eine Reihe von Charaktereigenschaften von fiktiven Personen dar. Anschließend sollten die Teilnehmer die fiktiven Personen mit ihren eigenen Worten beschreiben und hinsichtlich weiterer Eigenschaften beurteilen. Es zeigte sich, dass bestimmte Attribute, je nachdem ob sie zusammen mit eindeutig positiven oder negativen Eigenschaften präsentiert wurden, auch positiv oder negativ interpretiert wurden. War eine fiktive Person beispielsweise intelligent, geschickt, fleißig, zielstrebig, erfahren, vorsichtig und warm, wurde sie unter anderem als entschlossen, aus Überzeugung handelnd, aber auch großzügig und freundlich eingeschätzt. War die Person dagegen intelligent, geschickt, fleißig, zielstrebig, erfahren, vorsichtig und kalt, wurde sie vielmehr als rücksichtslos und überheblich charakterisiert. Daraus wurde deutlich, dass die Hervorhebung menschlicher Wärme oder Kälte zu Halo-Effekten führen kann, die Eigenschaften wie Fleiß und Zielstrebigkeit positiv oder negativ prägen. Auch andere Autoren kamen zu dem Schluss, dass die Beurteilung einzelner Attribute in Abhängigkeit von den weiteren vorhandenen Eigenschaften variieren kann (Anderson, 1971; M. F. Kaplan, 1974; Petzold, 1992; Wyer, 1974).

Der Beweis von Halo-Effekten als Verzerrungen eindeutiger Informationen konnte bereits in vielen Kontexten erbracht werden. Foster und Ysseldyke (1976) wiesen zum Beispiel nach, dass Schüler mit angeblichen Beeinträchtigungen anders beurteilt wurden, obwohl sie genau die gleichen Verhaltensweisen zeigten wie die gesunden Kinder. Im studentischen Bereich wurde gezeigt, dass die Einschätzung des gleichen Essays besser ausfiel, wenn die angebliche Autorin physisch attraktiv war (R. M. Kaplan, 1978; D. Landy & Sigall, 1974). Neben dem Äußeren können auch die Reaktionen anderer Personen zu Urteilsverzerrungen führen. So fand etwa Babad (2005) heraus, dass das objektiv gleiche

Verhalten eines Politikers negativer beurteilt wurde, wenn er mit einem abweisenden anstelle eines freundlichen Interviewers interagierte.

Es ist anzumerken, dass in den einzelnen Untersuchungen nicht immer eindeutig geklärt werden kann, ob ein Halo-Effekt im Sinne einer Prägung ambivalenter oder einer Verzerrung eindeutiger Informationen zu verstehen ist. Dies ist darin begründet, dass die Wahrnehmung einer Eigenschaft als ambivalent oder eindeutig sehr subjektiv und von der jeweiligen Situation abhängig ist. Beispielsweise ergaben mehrere Experimente, dass gute Schüler bei offenen Fragen von Intelligenztests besser bewertet wurden als Schüler mit durchschnittlichen Leistungen, weil auch unklare Antworten eher als richtig interpretiert wurden (Sattler et al., 1970; Sattler & Winget, 1970). Ob die Antworten in diesen Fällen als ambivalent oder eindeutig zu bezeichnen sind, kann nicht genau gesagt werden. Würde etwa ein guter Schüler eine eindeutig schlechte Antwort geben, könnte selbst diese als ambivalent wahrgenommen werden, weil sie nicht zu den vergangenen Leistungen passt und somit nur eine Momentaufnahme darstellt.

Prozesse von Halo-Effekten

Ein wichtiger Beleg für die Vielfältigkeit von Halo-Effekten findet sich bei der Betrachtung der grundlegenden Prozesse. Diese sind nicht zu verwechseln mit den bereits geschilderten Ursachen und Einflussfaktoren von Halo-Effekten, zu denen bereits viele Untersuchungen existieren. Prozesse bezeichnen vielmehr diejenigen Mechanismen, die beim Auftreten von Halo-Effekten wirken. Die meisten Autoren vertraten implizit eine bestimmte Auffassung über diese Mechanismen, was sich in der jeweils formulierten oder zitierten Definition von Halo-Effekten niederschlägt. Eine genauere Betrachtung und Überprüfung der grundlegenden Prozesse blieb dabei jedoch meistens aus. Dieser Forschungsbereich ist daher noch weitgehend ungeklärt und nur Gegenstand weniger Studien (Lance, LaPointe, & Fiscaro, 1994). Die erste detaillierte Betrachtung von verschiedenen Definitionen und Prozessmodellen von Halo-Effekten stammt von Fiscaro und Lance (1990). In einem Rückblick auf die vergangene Forschung unterschieden die Autoren drei verschiedene Modelle, die auch in wenigen neueren Studien betrachtet und miteinander verglichen wurden (Dennis, 2007; Lance, LaPointe, & Fiscaro, 1994; Lance, LaPointe, & Stewart, 1994; Oh & Ramaprasad, 2003). Die drei Modelle sind konform mit verschiedenen Definitionen und Erhebungsmethoden von Halo-Effekten. Manche der

Widersprüche, die sich in der Vergangenheit bei der Erfassung von Halo-Effekten mit unterschiedlichen Maßen gezeigt haben (Alliger & Williams, 1989; Balzer & Sulsky, 1992; Fiscaro & Vance, 1994; Fox et al., 1983; Fox et al., 1989; Pulakos et al., 1986; Saal et al., 1980), können auf diese Weise aufgeklärt werden. Eine graphische Veranschaulichung der drei Prozessmodelle ist in Abbildung 1 dargestellt.

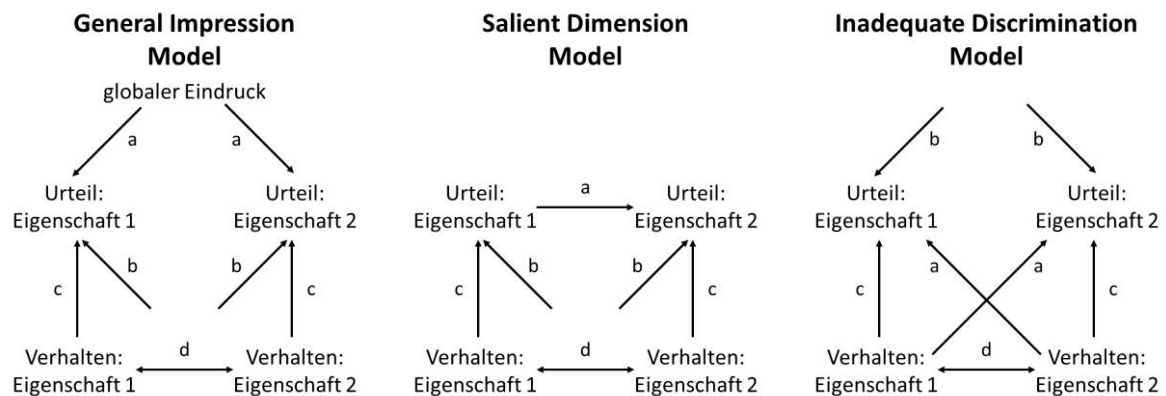


Abbildung 1. Kausale Modelle von Halo-Effekten aus der Literatur

a = Halo-Effekt; b = individueller Urteilsfehler; c = Einfluss des tatsächlichen Verhaltens auf das Urteil; d = wahre Korrelation zwischen verschiedenen Verhaltensweisen

Das erste Modell zur Beschreibung der Mechanismen von Halo-Effekten hat die Bezeichnung *General Impression Model*. Dieses Modell ist konform mit der Definition von Halo-Effekten als kognitive Verzerrungen, die aus dem Einfluss einer globalen Einschätzung auf die Beurteilung individueller Attribute von Personen oder Objekten resultieren (Nisbett & Wilson, 1977). Urteile über spezifische Eigenschaften werden gemäß dieser Auffassung nicht nur durch das tatsächliche Verhalten (und individuelle Fehler), sondern auch stets durch einen globalen Eindruck beeinflusst. Der Effekt dieses globalen Eindrucks wird mit einem Halo-Effekt gleichgesetzt. Dieses Modell spiegelt die am weitesten verbreitete Auffassung über die grundlegenden Mechanismen von Halo-Effekten wider und wurde bereits in vielen Studien indirekt bestätigt. Beckwith und Lehmann (1975) zeigten zum Beispiel, dass der allgemeine Eindruck über bestimmte Fernsehsendungen auch deren Einschätzung auf verschiedenen spezifischen Skalen beeinflusste. Bei der Beurteilung von Menschen stellten Lachman und Bass (1985) fest, dass die allgemeine Einschätzung von Personen die Beurteilung einzelner Persönlichkeitseigenschaften, wie etwa Höflichkeit, Intelligenz und Humor, prägte. Ähnliche Befunde ergaben sich in einer Vielzahl weiterer

Untersuchungen (Dennis, 2007; DeRath & Carp, 1959; Hulin, 1982; Jennings et al., 2004; Viswesvaran et al., 2005). Bei einer genauen Betrachtung der verschiedenen statistischen Methoden kann festgestellt werden, dass Inter-Rating-Korrelationen und Intra-Objekt-Varianzen ausschließlich diese globalen Halo-Effekte messen, da sie jeweils den Aspekt erfassen, der allen abgegebenen Urteilen gemein ist.

Das zweite Prozessmodell von Halo-Effekten trägt den Namen *Salient Dimension Model*. Entsprechend dieses Modells können Halo-Effekte als Einfluss der Bewertung einer einzigen salienten Eigenschaft auf die Einschätzung anderer spezifischer Attribute definiert werden (Blum & Naylor, 1968). Das bedeutet, dass einzelne Merkmale spezifische, unterschiedlich starke Halo-Effekte auf andere Attribute haben können, ohne über einen globalen Eindruck vermittelt zu werden. Ein bekanntes Beispiel für solche spezifischen Effekte findet sich in den Experimenten von Asch (1946). Der Autor wies nach, dass Menschen, die als warm beschrieben wurden, auch auf vielen anderen Dimensionen überdurchschnittlich positiv beurteilt wurden, während für als kalt bezeichnete Personen das Gegenteil galt. In einer anderen Studie zeigten Phelps und Kollegen (1986) auf, dass Lehrern, die ein kompetentes Auftreten hatten, auch ein hohes Fachwissen zugeschrieben wurde, auch wenn dies objektiv nicht der Fall war. Andere saliente Merkmale, die zu Halo-Effekten führen können, wurden unter anderem in Unsicherheit (Paine et al., 1966), Führungsqualität (Meindl & Ehrlich, 1987), Trinkverhalten (McNeil & Janzen, 1987) und depressiven Zuständen (Mumma, 2002) gefunden. Das Modell wird zusätzlich durch eine Reihe von Analysen gestützt, die ergaben, dass die Evaluation von Personen und Objekten nicht global, sondern auf spezifischen Ebenen erfolgte (Hallworth, 1961; Srivastava et al., 2010). Da die Halo-Effekte dieses Modells nicht auf alle, sondern nur bestimmte andere Urteile wirken, können sie durch Inter-Rating-Korrelationen und Intra-Objekt-Varianzen nicht beziehungsweise nur in begrenztem Maße erfasst werden.

Das dritte Modell zur Erklärung von Halo-Effekten ist unter der Bezeichnung *Inadequate Discrimination Model* bekannt. Dieses Modell entspricht der vergleichsweise wenig zitierten Definition von Halo-Effekten als Fehler, die durch die mangelnde Unterscheidung von distinkten und voneinander unabhängigen Verhaltensaspekten zustande kommen (Saal et al., 1980). Mit anderen Worten gesagt, werden bestimmte Merkmale oder Verhaltensweisen zwar korrekt wahrgenommen, aber auf eine falsche Eigenschaft zurückgeführt. Ein einfaches Beispiel wäre, dass das Stottern eines Lehrers mit

mangelnden sprachlichen Kompetenzen erklärt wird, wenn es korrekterweise durch Unsicherheit bedingt ist. Halo-Effekte dieser Art nehmen einen Sonderstatus ein, weil sie meistens dann auftreten, wenn die zu beurteilenden Attribute komplex und nicht klar definiert sind. In vielen Studien hat es den Anschein, als würden diese Halo-Effekte auftreten, auch wenn dies nicht explizit erwähnt oder diskutiert wird (Boatwright et al., 2008; Dompnier et al., 2006; Engelhard, 1994; Holzbach, 1978). Da diese Art von Fehlern durch gute Untersuchungsdesigns weitgehend ausgeschlossen werden kann, ist sie für den Großteil der Literatur nur von untergeordneter Wichtigkeit.

Von großer Bedeutung in Bezug auf Prozessmodelle ist die Frage, welche der drei Erklärungen die Natur von Halo-Effekten am besten wiedergibt. Sowohl Lance, LaPointe und Fisicaro (1994) als auch Lance, LaPointe und Stewart (1994) stellten fest, dass direkte Vergleiche am stärksten für das General Impression Model sprachen. Dieser Befund war jedoch zum Teil dadurch bedingt, dass das Modell im Vergleich zu den anderen statistisch sparsamer war, da es alle Verzerrungen auf nur einen Faktor zurückführte. Die Autoren konnten jedoch ebenfalls zeigen, dass die Gültigkeit der drei Modelle durch verschiedene äußere Faktoren beeinflusst wurde. Das heißt, dass etwa unterschiedliche Designs oder Instruktionen dazu führen können, dass einige Modelle besser und andere weniger gut zur Erklärung der jeweils wirkenden Halo-Effekte dienen. Folglich ist anzunehmen, dass alle Modelle in einem gewissen Maße gültig sind. Dies spricht wiederum dafür, dass nicht nur eine Art von Halo-Effekten existiert, sondern es mehrere, qualitativ unterschiedliche Halo-Effekte gibt, die unter verschiedenen externen Bedingungen auftreten.

2.3 Zentrale Hypothesen

Die im letzten Abschnitt geschilderten Befunde und offenen Fragen lassen den Schluss zu, dass Halo-Effekte, entgegen der weit verbreiteten Ansicht, kein einheitliches Muster aufweisen, sondern in vielen unterschiedlichen Faktoren variieren können. In der vorliegenden Arbeit soll diese Vielfältigkeit an zwei wesentlichen Aspekten verdeutlicht werden: Der erste Fokus liegt auf möglichen Unterschieden in dem Ausmaß (im Folgenden auch als Quantität bezeichnet) von Halo-Effekten, welche in der bisherigen Forschung weitgehend vernachlässigt worden sind. Den zweiten Schwerpunkt stellen potentielle

Unterschiede in den Prozessen (im Folgenden auch als Qualität bezeichnet) von Halo-Effekten, basierend auf den Modellen von Fiscaro und Lance (1990), dar.

Für die in den folgenden drei Abschnitten geschilderten Hypothesen werden Halo-Effekte vor allem als tatsächliche Fehler betrachtet. Eventuelle Nutzen des Phänomens werden dagegen nur am Rande diskutiert. Der Schwerpunkt der vorliegenden Arbeit liegt ausschließlich auf dem Bereich der Personenbeurteilung. Des Weiteren beziehen sich die folgenden Hypothesen auf alle potentiell möglichen Wirkungsweisen von Halo-Effekten. Obwohl die verschiedenen Wirkungsweisen grundsätzlich auch für die Vielfältigkeit des Phänomens sprechen, werden sie also ausschließlich als abhängige Variablen betrachtet. Halo-Effekte als Prägung ambivalenter Informationen und Halo-Effekte als Verzerrungen eindeutiger Informationen werden in allen folgenden Schilderungen als eine einzige Wirkungsweise betrachtet, da sie in der Praxis nur schwer voneinander getrennt werden können. In den folgenden Hypothesen werden beispielhaft drei Faktoren beschrieben, welche die Qualität und Quantität von Halo-Effekten beeinflussen können.

2.3.1 Einfluss der Informationsvalenz

Die erste Hypothese geht davon aus, dass Halo-Effekte durch die Valenz der dargebotenen Informationen beeinflusst werden. Das bedeutet, dass sich positive und negative Halo-Effekte maßgeblich voneinander unterscheiden sollten. In der Literatur konnten Valenzasymmetrien bei der Struktur und Verarbeitung sozialer Informationen bereits mehrfach verdeutlicht werden. Boucher und Osgood (1969) wiesen zum Beispiel nach, dass positive Wörter im alltäglichen Sprachgebrauch häufiger verwendet werden als negative. Auch werden positive Eigenschaften im Durchschnitt als Norm und negative Eigenschaften als Abweichung von der Norm angesehen (H. H. Clark & Clark, 1977). Auf der anderen Seite lassen sich negative Informationen vielfältiger beschreiben als positive. So zeigten etwa Semin und Fiedler (1992) auf, dass das englische Vokabular negativer Verben das Vokabular an positiven Verben übersteigt. Bei Adjektiven ist zusätzlich zu erwähnen, dass sie sowohl durch die Veränderung positiver Wörter als auch negativ ausgedrückt werden können. Das negative Äquivalent der Eigenschaft *ordentlich* kann beispielsweise sowohl *unordentlich* als auch *chaotisch* lauten (H. H. Clark & Clark, 1977). Eine weitere Valenzasymmetrie äußert sich darin, dass negative Verhaltensweisen von Personen

diagnostischer für die dahinterstehenden Attribute sind als positive Verhaltensweisen. So sagt zum Beispiel das Ausrauben einer Bank mehr über die Charaktereigenschaft *ehrlich* beziehungsweise *unehrlich* aus als das wahrheitsgemäße Ausfüllen der Steuererklärung (Reeder & Brewer, 1979; Skowronski & Carlston, 1989). Im Bereich der Erforschung von Gesichtern stellten Ekman und Friesen (1971) fest, dass von sechs kulturell unabhängigen Basisemotionen vier Emotionen eindeutig negativ sind, wogegen nur eine Emotion distinkt positiv ist. Weitere Befunde machten des Weiteren deutlich, dass attraktive Gesichter von anderen Personen als ähnlicher zueinander wahrgenommen werden als unattraktive Gesichter (Ekman & Friesen, 1971; Potter, Corneille, Ruys, & Rhodes, 2007). Unter dem Begriff *Density Hypothesis* wurden diese und weitere Ergebnisse zu einer allgemeinen Aussage über die Struktur positiver und negativer Informationen zusammengefasst. Die Density Hypothesis besagt, dass positive Informationen häufiger und einander ähnlich, negative Informationen dagegen grundsätzlich seltener und distinkter voneinander sind (Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008).

Im Bereich der Halo-Effekte wurde der Einfluss von Valenz bisher nur am Rande berücksichtigt. Die meisten Studien fokussierten nur eine Valenz, ohne auf die andere einzugehen. Andere Experimente untersuchten zwar positive und negative Halo-Effekte gleichermaßen, stellten diese jedoch nicht explizit gegenüber (Anastasiow, 1963; Babad, 2005; Goldman et al., 1983; A. G. Miller, 1970; Nisbett & Wilson, 1977). Ein unmittelbarer Vergleich positiver und negativer Halo-Effekte wurde bis zum heutigen Zeitpunkt nur in wenigen Untersuchungen vorgenommen. Asch (1946) stellte beispielsweise fest, dass das Vorhandensein der Eigenschaft *warm* andere Urteile stärker verzerren kann als das gegenteilige Attribut *kalt*. Auch andere Studien zu dieser Thematik ergaben, dass positive Halo-Effekte stärker sind als negative Halo-Effekte (Peeters, 1991; Rizzo & Frank, 1977; Tsui & Barry, 1986). Gänzlich unberücksichtigt blieb dagegen die Frage, ob sich positive und negative Halo-Effekte möglicherweise auch qualitativ unterscheiden. Wie in mehreren Analysen zu den drei Modelle von Fiscaro und Lance (1990) deutlich wurde, können Halo-Effekte in Abhängigkeit von diversen externen Faktoren mit unterschiedlichen Prozessen beschrieben werden (Lance, LaPointe, & Fiscaro, 1994; Lance, LaPointe, & Stewart, 1994). Von besonderer Bedeutung ist hierbei die Tatsache, dass sowohl ein globaler Eindruck (General Impression Model) als auch ein einzelnes Attribut (Salient Dimension Model) zur Entstehung von Halo-Effekten beitragen kann. Zusätzlich ist zu berücksichtigen, dass ein

globaler Eindruck nicht nur die Beurteilung einzelner Eigenschaften beeinflussen kann, sondern umgekehrt auch einzelne Eigenschaften den globalen Eindruck über eine Person oder ein Objekt prägen können (Thorndike, 1920).

Wendet man die Aussagen der Density Hypothesis auf Halo-Effekte an, ergeben sich unterschiedliche Vorhersagen in Abhängigkeit von Valenz: Wie bereits geschildert, besagt die Density Hypothesis, dass positive Informationen grundsätzlich viel ähnlicher zueinander sind als negative Informationen. Auf einen hypothetischen Raum übertragen würde das bedeuten, dass positive Attribute alle sehr nah beieinander liegen und negative Attribute weit verstreut sind. Daraus ergibt sich, dass leicht von einer positiven Eigenschaft auf viele andere positive Eigenschaften, aber nur schwer von einem negativen Attribut auf andere negative Attribute geschlossen werden kann. Positive Halo-Effekte äußern sich folglich dahingehend, dass bei der Vorgabe eines einzelnen positiven Attributs automatisch auch viele andere positive Eigenschaften als gegeben betrachtet werden, auch wenn dafür keine oder sogar nur gegenteilige objektive Indizien vorhanden sind. Durch dieses gleichzeitige Auftreten vieler spezifischer Halo-Effekte sollte darüber hinaus ein ausgeprägter globaler Eindruck entstehen, der seinerseits wieder zu Halo-Effekten auf die Beurteilung einzelner Attribute führt. Wird dagegen eine negative Eigenschaft vorgegeben, werden nur wenige andere negative Attribute automatisch als gegeben betrachtet. Das heißt, es treten nur wenige spezifische Halo-Effekte auf, so dass kein globaler Eindruck entstehen kann. Der wesentliche Unterschied zwischen positiven und negativen Halo-Effekten besteht folglich darin, dass negative Halo-Effekte ausschließlich spezifischer Natur sind und positive Halo-Effekte darüber hinaus global wirken. Ein ideales Prozessmodell sollte beide Formen von Halo-Effekten berücksichtigen. Da globale Halo-Effekte aus dem gleichzeitigen Auftreten vieler spezifischer Halo-Effekte resultieren, sind positive Halo-Effekte auch quantitativ stärker als negative Halo-Effekte. Es ist wichtig anzumerken, dass diese Valenzasymmetrie vor allem dann auftritt, wenn nur einzelne Attribute als Informationen präsentiert werden. Je mehr Eigenschaften gleichzeitig vorhanden sind, desto eher kann auch bei negativer Valenz ein globaler Eindruck entstehen. So sollten beispielsweise bei der Präsentation von Gesichtern, welche nachgewiesenermaßen sowohl global (Farah, Wilson, Drain, & Tanaka, 1998; Tanaka & Farah, 1993) als auch spezifisch (Amishav & Kimchi, 2010; Cabeza & Kato, 2000; Rotshtein, Geng, Driver, & Dolan, 2007; Schwarzer & Massaro, 2001) verarbeitet werden, nur vergleichsweise geringe Valenzasymmetrien auftreten.

Auf Basis der eben genannten Überlegungen wird folgende Valenzhypothese formuliert und in Abbildung 2 veranschaulicht: *Positive Halo-Effekte sind globaler, negative Halo-Effekte dagegen spezifischer Natur. Aus diesem Grund sind positive Halo-Effekte auch von größerem Ausmaß als negative Halo-Effekte.*

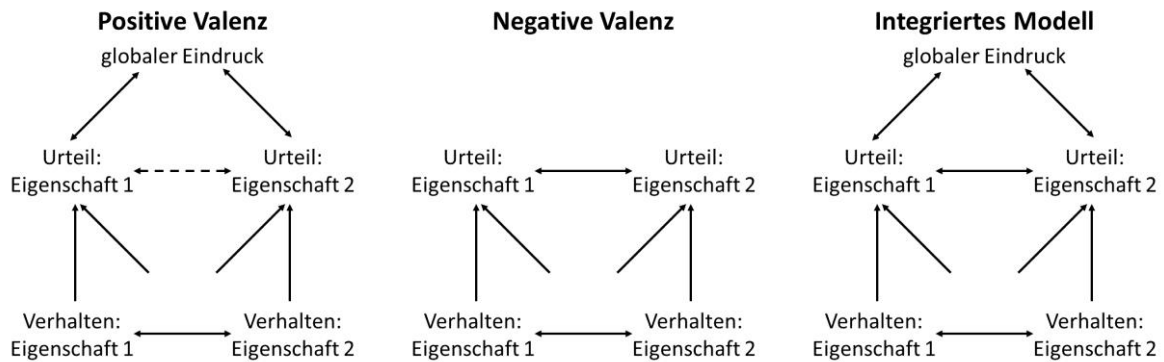


Abbildung 2. Postulierter Einfluss der Valenz auf die Qualität von Halo-Effekten

2.3.2 Einfluss der Urteilsdimension

Die zweite Hypothese postuliert einen Einfluss der Urteilsdimension auf die Stärke der auftretenden Halo-Effekte. In diesem Zusammenhang stellt sich zunächst die Frage, welche Dimensionen im Bereich der Personenbeurteilung betrachtet werden sollen. Eine Möglichkeit sind die Dimensionen des Fünf-Faktoren-Modells (Digman, 1990), welche auch unter dem Namen *Big Five* bekannt sind (Goldberg, 1990). Dabei handelt es sich um die Faktoren *Neurotizismus*, *Extraversion*, *Offenheit für Erfahrungen*, *Gewissenhaftigkeit* und *Verträglichkeit*. Mehrere Untersuchungen zeigten jedoch auf, dass diese fünf Faktoren, entgegen der ursprünglichen Ergebnisse, nicht perfekt unkorreliert sind, sondern zwei Oberkategorien ergeben (DeYoung, 2006; Digman, 1997). Da Halo-Effekte am besten dann untersucht werden können, wenn keine wahren Korrelationen zwischen den einzelnen Attributen existieren, ist das Fünf-Faktoren-Modell aus diesem Grund keine geeignete Option. Eine weitere Möglichkeit stellt die Betrachtung der zwei Dimensionen *Agency* und *Communion* dar (Bakan, 1966), die vom Prinzip her den zwei Oberkategorien der Big Five entsprechen (DeYoung, 2006; Digman, 1997). Auch andere Autoren fanden heraus, dass soziale Kognition in Form von zwei grundlegenden Dimensionen organisiert ist. Bei der Erforschung von Stereotypen wurden hierfür beispielsweise die Namen *Competence* und *Warmth* geprägt (Fiske, Cuddy, & Glick, 2007; Fiske, Cuddy, Glick, & Xu, 2002). Nach der

Ansicht von Judd und Kollegen (2005) herrscht ein beachtlicher Konsens über die Existenz zweier fundamentaler Dimensionen. Die Autoren führten auf, dass die Benennung und der Untersuchungskontext der beiden Dimensionen zwar zwischen den verschiedenen Studien variieren, aber alle Ergebnisse einen gemeinsamen Kern aufweisen. Des Weiteren gibt es Befunde, welche die Existenz von Agency und Communion in der Personenwahrnehmung auch über unterschiedliche Nationen und Sprachen hinweg bestätigten (Abele, Uchronski, Suitner, & Wojciszke, 2008). Da die beiden Faktoren aufgrund der Vielzahl ähnlicher Ergebnisse als nahezu unkorreliert betrachtet werden können, sind sie bestens dazu geeignet, quantitative Unterschiede von Halo-Effekten zu untersuchen.

Agency und Communion wurden in der Vergangenheit bereits in einer Vielzahl von Studien untersucht und miteinander verglichen. Dabei zeigte sich meistens eine höhere Bedeutung von Communion. Wojciszke und Abele (2008) wiesen beispielsweise nach, dass Communion-Attribute bei der Bewertung anderer Personen im Durchschnitt als wichtiger erachtet werden als Agency-Attribute. Dies ist vor allem dann der Fall, wenn keine enge Beziehung zwischen der beurteilenden und der beurteilten Person herrscht. Laut den Befunden einer anderen Studie spielt Communion vor allem bei der Sympathie gegenüber anderen Menschen eine größere Rolle als Agency (Wojciszke, Abele, & Baryla, 2009). Eine Untersuchung von Abele und Bruckmüller (2011), in der verschiedene Aspekte der zwei Faktoren betrachtet wurden, ergab, dass Communion insgesamt die wichtigere der beiden Dimensionen darstellt. Die Autoren fanden heraus, dass Communion-Attribute schneller hinsichtlich ihrer Valenz erkannt und auch schneller aus Verhaltensbeschreibungen erschlossen werden als Agency-Attribute. Zudem werden Communion-Eigenschaften in spontanen Beschreibungen anderer Personen eher genannt als Agency-Eigenschaften. Daraus lässt sich schlussfolgern, dass Communion im zwischenmenschlichen Bereich von höherer Bedeutsamkeit und größerem Einfluss ist als Agency.

Bei der Erforschung von Halo-Effekten wurden Agency und Communion bisher nur am Rande betrachtet. Judd und Kollegen (2005) zeigten zum Beispiel, dass Individuen mit hohen Werten auf einer Dimension auch auf der anderen Dimension extremer beurteilt werden. Des Weiteren stellten Suitner und Maass (2008) fest, dass Communion stärker mit Valenz assoziiert ist und daher die Einschätzung von Personen oder Gruppen besser vorhersagt als Agency. Somit scheinen Communion-Eigenschaften auch bei Halo-Effekten von größerer Bedeutsamkeit zu sein. Ein direkter Vergleich von Halo-Effekten, die aus

Communion- und Agency-Attributen resultieren, wurde jedoch in der Vergangenheit noch nicht vorgenommen. Auf Basis der bisherigen Befundlage kann festgestellt werden, dass Communion-Attribute bei der Wahrnehmung und Beurteilung anderer Personen von höherer Bedeutung sind als Agency-Attribute. Auch die Daten anderer Untersuchungen, welche sich nicht explizit mit den beiden Dimensionen beschäftigten, sprechen für die enorme Wichtigkeit von Communion. So führten sowohl Asch (1946) als auch Scholer und Higgins (2008) auf, dass die Attribute *warm* und *kalt* im zwischenmenschlichen Bereich von zentraler Bedeutung seien und die Wahrnehmung anderer Eigenschaften stark verzerren können. Aufgrund dieser Bedeutsamkeit kann angenommen werden, dass Communion-Attribute zu stärkeren Halo-Effekten führen als Agency-Attribute. Dabei handelt es sich jedoch lediglich um einen quantitativen Unterschied, der nicht durch die grundlegende Struktur der Informationen (wie im Falle von Valenz) bedingt ist. Agency- und Communion-Halo-Effekte unterscheiden sich somit nicht in ihren Prozessen.

Ausgehend von den geschilderten Befunden und Herleitungen wird folgende Dimensionshypothese postuliert: *Communion-Halo-Effekte sind stärker als Agency-Halo-Effekte. Bei konstanter Valenz sind Communion- und Agency-Halo-Effekte jedoch gleichermaßen global beziehungsweise spezifisch.*

2.3.3 Einfluss von Kontextfaktoren

Die dritte Hypothese besagt, dass Halo-Effekte in Abhängigkeit von verschiedenen Kontextfaktoren variieren. In der Vergangenheit wurden Kontextfaktoren zwar in vielen Studien im Bereich der Urteils- und Entscheidungspsychologie untersucht, dabei jedoch oft unterschiedlich definiert. Einige Autoren beschäftigten sich etwa mit Kontexteffekten im Sinne menschlicher Faktoren, wie beispielsweise Affekt (Tsui & Barry, 1986) oder innerer Erregung (Bagozzi, 1996; O'Neal, 1971). In anderen Experimenten standen demgegenüber verschiedene externe Variablen im Mittelpunkt des Interesses. So verstand Payne (1982) unter Kontexteffekten all diejenigen Faktoren, die mit dem Wert von Objekten innerhalb einer Entscheidungsoption assoziiert sind. In diesem Zusammenhang wurden zum Beispiel die Attraktivität und Ähnlichkeit eventuell vorhandener Alternativen (Payne, 1982), die Auftrittswahrscheinlichkeit bestimmter Ergebnisse (Payne, Bettman, & Johnson, 1988), mögliche Konflikte zwischen verschiedenen Ergebnissen (Payne, Bettman, & Luce, 1996)

sowie persönliche Erfahrungen in vergangenen Entscheidungssituationen (Vlaev & Chater, 2007) untersucht. Des Weiteren sind auch Variablen, wie die aktuell und bereits zuvor dargebotenen Informationen (Petzold, 1992; Tourangeau & Rasinski, 1988) sowie deren Formulierung (Kahneman & Tversky, 1984), externe Kontextfaktoren. Auch werden häufig Eigenschaften der Aufgabe selbst, wie etwa die Präsentationsart, der Zeitdruck und die Anzahl der Alternativen, den Kontextfaktoren zugeordnet, obwohl es sich dabei genau genommen um eine eigenständige Kategorie handelt (Payne, 1982). In der vorliegenden Arbeit werden drei Kontextfaktoren betrachtet, die bisher nur selten untersucht wurden: Dabei handelt es sich um den Beruf der zu beurteilenden Person, die Situation, in der die Urteile gefällt werden müssen, und den Beruf des Beurteilenden selbst.

Bei der Analyse von Halo-Effekten wurde in der bisherigen Forschung bereits eine Reihe von internen und externen Kontextfaktoren betrachtet. Verschiedene Erhebungen zeigten beispielsweise, dass Halo-Effekte durch positiven oder negativen Affekt (Tsui & Barry, 1986) und innere Erregung (O'Neal, 1971) verstärkt werden können. Auch stabile Personeneigenschaften, wie etwa geringe kognitive Fähigkeiten (Banks & White, 1982; Bernardin et al., 1982; Ostroff & Ilgen, 1992), emotionale Instabilität (Buffardi & Gibson, 1980), mangelnde Fähigkeit zu logischem Denken (Dickinson & Glebocki, 1990), schlechte Gedächtnisleistungen (Jennings et al., 2004) und ein sehr niedriges Kognitionsbedürfnis (Palmer & Feldman, 2005), wurden schon mit Halo-Effekten in Verbindung gebracht. Im Bereich der externen Kontextfaktoren stellte bereits Asch (1946) fest, dass die Bedeutung einiger Eigenschaften in Abhängigkeit davon variiert, welche anderen Eigenschaften zuvor präsentiert wurden beziehungsweise gleichzeitig dargeboten werden. Spätere Studien konnten vergleichbare Ergebnisse hervorbringen (M. F. Kaplan, 1974; Petzold, 1992; Tourangeau & Rasinski, 1988). Andere Autoren wiesen nach, dass Halo-Effekte bei der Beurteilung bestimmter Personen davon beeinflusst werden, welche Individuen zuvor bewertet wurden und ob diese Bewertungen positiv oder negativ ausfielen (Palmer & Loveland, 2008; Palmer et al., 2003). Auch Aufgabeneffekte standen bereits im Fokus der Forschung: Mehrere Experimente konnten zeigen, dass die Art und die Stärke von Halo-Effekten durch die Aufgabenstellung manipuliert werden kann. Je nachdem, wie die Instruktionen formuliert werden, können beispielsweise Effekte des globalen Eindrucks gezielt verstärkt oder abgeschwächt werden (Kiker & Motowidlo, 1998; Lance, LaPointe, & Fisicaro, 1994; Lance, LaPointe, & Stewart, 1994). Eine weitere, jedoch sehr unspezifische

Reihe von Ergebnissen bezieht sich auf die Eigenschaften der Beurteilenden sowie der beurteilten Individuen. Verschiedene Autoren erwähnten, dass Halo-Effekte in einem gewissen Maße von den Charakteristika der Beurteilenden selbst abhängen, dazu jedoch weitere Forschung notwendig sei (Crissy & Regan, 1951; Vance et al., 1983; Viswesvaran et al., 2005). Andere Studien belegten, dass bestimmte Halo-Effekte häufig an die Beurteilung von bestimmten Personengruppen gebunden sind (R. M. Kaplan, 1978; Murphy & Anhalt, 1992; Wade et al., 2007). So tritt etwa der Halo-Effekt von körperlicher Attraktivität besonders dann auf, wenn Frauen von Männern beurteilt werden (R. M. Kaplan, 1978). Bisher existieren keine Untersuchungen oder Befunde zu den Einflüssen des Berufs des beurteilten Individuums, der Situation, in der die Urteile abgegeben werden, oder der beruflichen Zugehörigkeit der beurteilenden Person.

Bringt man die erwähnten Kontextfaktoren mit der Dimension der dargebotenen Informationen in Verbindung, ergeben sich unterschiedliche Vorhersagen für Agency und Communion: Wie bereits im letzten Abschnitt erwähnt, ist Communion im Durchschnitt als die wichtigere der beiden Dimensionen anzusehen (Abele & Bruckmüller, 2011). Werden jedoch bestimmte Voraussetzungen variiert, kehrt sich dieses Muster zum Teil um. Abele und Wojciszke (2007) fanden beispielsweise heraus, dass Communion vor allem dann als wichtiger eingeschätzt wird, wenn es um Personen geht, zu denen man keinen Bezug hat. Stehen dagegen enge Freunde oder die eigene Person im Mittelpunkt, werden Agency-Eigenschaften als wichtiger eingestuft. In einer anderen Studie wurde ebenfalls deutlich, dass Communion dann eine größere Rolle spielt, wenn keine wechselseitige Abhängigkeit zwischen zwei Personen besteht. Wenn jedoch eine hohe Interdependenz herrscht, sind Agency-Attribute von höherer Bedeutung (Wojciszke & Abele, 2008). Des Weiteren ist Communion zwar entscheidend für die zwischenmenschliche Sympathie, aber Agency relevanter für den Respekt gegenüber anderen Personen (Wojciszke et al., 2009). In Anbetracht dieser Ergebnisse ist anzunehmen, dass verschiedene Berufsgruppen und Situationen mit unterschiedlich starken Halo-Effekten von Agency- und Communion-Attributen einhergehen. Communion-Halo-Effekte sollten vor allem in solchen Situationen verstärkt auftreten, in denen Sympathie sehr wichtig ist und eine vergleichsweise geringe Interdependenz vorliegt. Dies ist besonders in rein privaten Situationen gegeben. Auch in einigen Berufsgruppen liegt dieses Muster in höherem Maße vor als in anderen. Hier sind beispielsweise soziale und Pflegeberufe zu nennen. Somit sollten bei der Bewertung von

Personen mit einem dieser Berufe sehr starke Communion-Halo-Effekte auftreten. Zudem sollten Menschen mit sozialen und Pflegeberufen selbst sehr anfällig für Communion-Halo-Effekte sein. Agency-Halo-Effekte sollten unter den genannten Bedingungen von relativ geringer Größe sein. Jedoch ist anzunehmen, dass Agency-Halo-Effekte besonders dann recht stark sind, wenn in der Urteilsituation eine hohe Interdependenz herrscht und Respekt ein wichtiger Faktor ist. Dies ist zum Beispiel in beruflichen Situationen der Fall. Auch innerhalb bestimmter Berufsgruppen, wie etwa in naturwissenschaftlichen und technischen Berufen, sind diese Bedingungen in besonderem Maße gegeben. Es kann also angenommen werden, dass bei der Bewertung von Menschen mit naturwissenschaftlichen und technischen Berufen starke Agency-Halo-Effekte auftreten und Personen, die selbst dieser Berufsgruppe angehören, besonders anfällig für Agency-Halo-Effekte sind. Halo-Effekte von Communion-Attributen sollten unter diesen Bedingungen zwar schwächer sein als unter den weiter oben genannten Umständen, müssen aufgrund ihrer durchschnittlich hohen Bedeutsamkeit jedoch nicht unbedingt auch schwächer sein als Halo-Effekte von Agency-Attributen. Bei all den genannten Effekten handelt es sich erneut nur um rein quantitative und nicht um qualitative Unterschiede.

Auf Grundlage der vorherigen Erläuterungen wird folgende Kontexthypothese angenommen: *Erstens sind Agency-Halo-Effekte bei der Bewertung von Menschen mit naturwissenschaftlichen und technischen Berufen größer als bei der Beurteilung von Menschen mit sozialen und Pflegeberufen. Zweitens sind Agency-Halo-Effekte in rein beruflichen Situationen größer als in privaten Situationen. Drittens sind Menschen mit naturwissenschaftlichen und technischen Berufen in stärkerem Maße für Agency-Halo-Effekte anfällig als Personen mit sozialen und Pflegeberufen. Für Communion-Halo-Effekte gilt in allen drei Punkten das jeweils umgekehrte Muster.*

3 Methodische Umsetzung

Die in dieser Arbeit vorgestellten Experimente haben das Ziel, verschiedene Einflussfaktoren auf die Qualität und Quantität von Halo-Effekten zu untersuchen. Um die Vergleichbarkeit aller Ergebnisse zu gewährleisten, wurden die Messung der Halo-Effekte und der Ablauf aller Untersuchungen vereinheitlicht. Alle Experimente waren computerbasiert und wurden in der Sprache *Visual Basic* programmiert. Die Auswertung sämtlicher Daten erfolgte mit der Statistik-Software *SPSS*. Für alle statistischen Tests wurde ein Alpha-Level von .050 bei einer zweiseitig getesteten Signifikanz verwendet. Als Schätzer der Effektgröße diente das partielle Eta-Quadrat (im Folgenden mit η^2 abgekürzt).

3.1 Konzipierung und Ablauf der Experimente

Die Konzipierung und der Ablauf des Experimentalteils aller Untersuchungen erfolgten nach einem einheitlichen Schema. Zum besseren Verständnis soll dies im Folgenden anhand eines Beispiels deutlich gemacht werden:

Alle Stimuli entstammten einer von zwei vorgetesteten Datenbasen von positiven und negativen Agency- und Communion-Attributen (jeweils Gegensatzpaare). In einer Datenbasis waren die Items direkt als Eigenschaften (zum Beispiel „*ordentlich*“), in der anderen stattdessen in Form von Verhaltensweisen (zum Beispiel „*... hält die eigenen Dinge gut sortiert und lässt nichts herumliegen*“) formuliert. Die genaue Anzahl der auszuwählenden Items wurde für jedes Experiment einzeln festgelegt. Tabelle 1 zeigt ein Beispiel für ein in den Untersuchungen verwendetes Itemset.

Tabelle 1. Beispiel für ein in den Experimenten verwendetes Itemset

Agency		Communion	
positiv	negativ	positiv	negativ
ordentlich	chaotisch	ehrlich	verlogen
selbständig	abhängig	gesellig	eigenbrötlerisch
fleißig	faul	hilfsbereit	egoistisch

In jedem Experimentaldurchgang wurde eine Eigenschaft oder Verhaltensweise aus dem verwendeten Itemset vorgegeben (zum Beispiel „Eine Person hat die Eigenschaft ‚ordentlich‘.“ beziehungsweise „Eine Person hält die eigenen Dinge gut sortiert und lässt nichts herumliegen.“). In einem Teil der Experimente wurden pro Durchgang zusätzliche Stimuli, wie etwa Bilder oder weitere Verhaltensweisen, präsentiert. Diese konnten die dargebotenen Attribute entweder unterstützen (zum Beispiel ein attraktives Gesicht neben der Eigenschaft *hilfsbereit*) oder entkräften (zum Beispiel vier negative Verhaltensweisen neben der zentralen positiven Verhaltensweise ... *sagt anderen immer die Wahrheit*).

Gleichzeitig zu der Darbietung der Eigenschaft oder Verhaltensweise wurde pro Experimentaldurchgang ein anderes Attribut aus dem Itemset dargeboten. Dieses sollte anhand einer siebenstufigen unipolaren (zum Beispiel „Wie wahrscheinlich ist es, dass die Person auch folgende Eigenschaft hat: ‚fleißig‘? – keinesfalls ... ganz sicher“) oder bipolaren Skala (zum Beispiel „Wie würden Sie diese Person auf folgender Dimension einschätzen? – *faul ... fleißig*“) bewertet werden. Durch die Programmierung wurde verhindert, dass das gleiche Attribut parallel dargeboten und abgefragt werden konnte. Zwei Beispiele für unterschiedliche Experimentaldurchgänge sind in Abbildung 3 veranschaulicht.

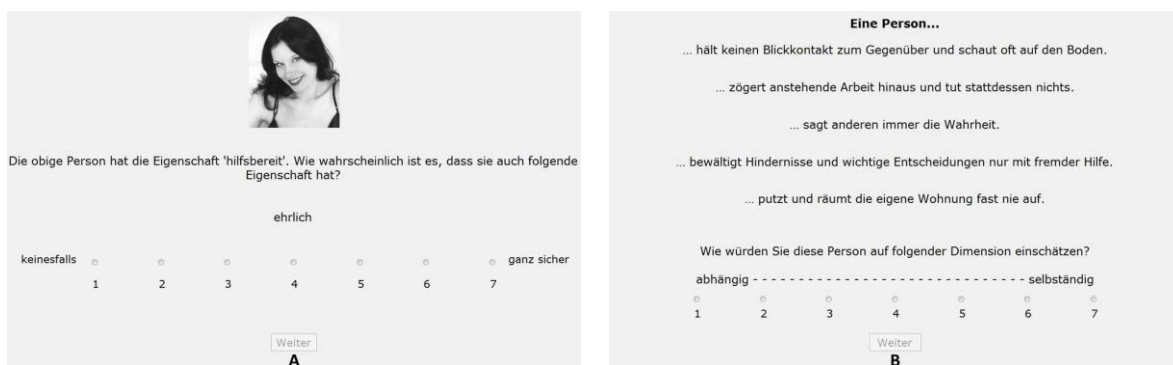


Abbildung 3. Veranschaulichung verschiedener Experimentaldurchgänge

A = Darbietung einer positiven Eigenschaft mit unterstützendem Bild und Abfrage anhand einer unipolaren Skala; B = Darbietung einer zentralen positiven Verhaltensweise (mittig) mit mehreren gegensätzlichen Verhaltensweisen und Abfrage anhand einer bipolaren Skala

Für die Zusammenstellung des gesamten Experimentalteils wurde in beiden Valenzbedingungen, getrennt voneinander, jedes Attribut mit jedem anderen kombiniert. Auf das obige Itemset bezogen würde das bedeuten, dass anhand jeder vorgegebenen Eigenschaft fünf verschiedene Urteile gefällt werden müssen. Beispielsweise müssten auf

Basis des Attributs *ordentlich* die Eigenschaften *selbständig, fleißig, ehrlich, gesellig* und *hilfsbereit* auf einer uni- oder bipolaren Skala bewertet werden. Bei der Vorgabe des Attributs *abhängig* würden Beurteilungen der Eigenschaften *chaotisch, faul, verlogen, eigenbrötlerisch* und *egoistisch* erfragt werden. Für beide Valenzbedingungen würden sich somit jeweils 30 (6 x 5) Kombinationen ergeben. Der gesamte Experimenterteil bestünde folglich aus 60 Durchgängen.

3.2 Überprüfung des grundlegenden Modells

Die wichtigste Grundlage der zentralen Hypothesen ist die Annahme einer qualitativen und quantitativen Vielfältigkeit von Halo-Effekten. Das bedeutet, dass Halo-Effekte sowohl in ihrem Ausmaß als auch ihren Prozessen beeinflusst werden können. Die bisher in der Literatur existierenden Modelle sind jedoch nicht in der Lage, qualitative Unterschiede von Halo-Effekten abzubilden. Aus diesem Grund wurde ein neues Modell konstruiert, das globale und spezifische Halo-Effekte zusammenführen und gleichermaßen berücksichtigen kann (vergleiche Abschnitt 2.3.1).

Die Überprüfung des postulierten Modells erfolgte schrittweise anhand mehrerer Experimente: Im ersten Schritt wurden lediglich die Halo-Effekte von vorgegebenen Persönlichkeitseigenschaften auf verschiedene andere Attribute betrachtet. Hierbei wurde bereits unterschieden, dass Halo-Effekte sowohl andere Eigenschaften direkt beeinflussen (spezifische Halo-Effekte) als auch indirekt über einen globalen Eindruck vermittelt (globale Halo-Effekte) wirken können. Im zweiten Entwicklungsschritt wurden nicht mehr Persönlichkeitseigenschaften, sondern einzelne konkrete Verhaltensweisen vorgegeben. Halo-Effekte einer bestimmten Eigenschaft konnten also erst wirken, wenn diese aus dem Verhalten einer Person erschlossen wurde. Der dritte Schritt stellt das vollständige Modell dar. Hierbei wurden erstmals mehrere Verhaltensweisen, die unterschiedliche Attribute beschreiben, gleichzeitig präsentiert. Während zuvor nur Halo-Effekte im Sinne von Vorannahmen betrachtet wurden, verlagerte sich der Fokus in diesem Schritt zu Halo-Effekten als Verzerrungen eindeutiger Urteile (vergleiche Abschnitt 2.2.2). Die schrittweise Überprüfung des Modells ist in Abbildung 4 graphisch dargestellt.

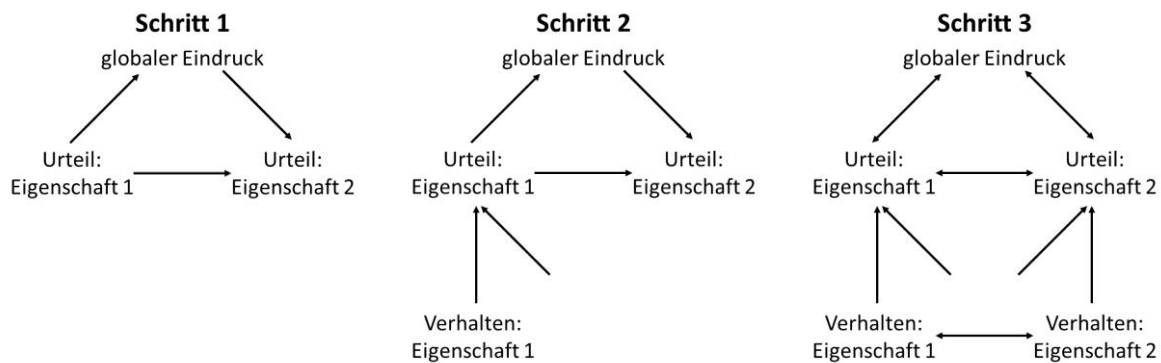


Abbildung 4. Vorgehensweise bei der Überprüfung des grundlegenden Modells

3.3 Auswahl des abhängigen Maßes

Wie in Abschnitt 2.1.3 geschildert, existiert in der Literatur eine Vielzahl von Maßen und Methoden zur Erfassung von Halo-Effekten. Unter der Annahme verschiedener, qualitativ unterschiedlicher Halo-Effekte erschwert dies die Vergleichbarkeit einzelner Befunde. Ein grundlegendes Anliegen dieser Arbeit war es daher auch, Halo-Effekte, unabhängig von der jeweiligen experimentellen Manipulation und über die verschiedenen Experimente hinweg, mit einem einheitlichen Maß zu messen.

Unter den verschiedenen Möglichkeiten fiel die Wahl auf die Methode, Halo-Effekte über den Einfluss manipulierter Informationen auf die Beurteilung nicht manipulierter Eigenschaften zu messen. Ein großer Vorteil dieser Methode bezieht sich auf die Notwendigkeit, Halo-Effekte von wahren Zusammenhängen (illusorische Halo-Effekte) abzugrenzen. Dabei wird das Problem umgangen, die wahren Daten kennen zu müssen, so wie es bei anderen Maßen der Fall ist. In der Tat können die Informationen so manipuliert und zusammengestellt werden, dass es keine realen Zusammenhänge zwischen den verschiedenen Attributen gibt und somit jedes nicht neutrale Urteil einem Halo-Effekt gleichkommt. Des Weiteren bietet die gewählte Methode die Möglichkeit, globale und spezifische Halo-Effekte voneinander abzugrenzen. Dazu ist es nur notwendig, qualitativ unterschiedliche Informationen, wie Eigenschaften und Bilder, sowohl einzeln als auch gemeinsam zu präsentieren und die resultierenden Urteile miteinander zu vergleichen. Der dritte Vorteil liegt in der einfachen methodischen Umsetzung, da Ratings als abhängige Maße einfach zu berechnen sind und unmittelbar interpretiert werden können.

Neben der ausgewählten Methode wurden auch die anderen Maße kritisch betrachtet: Sowohl Inter-Rating-Korrelationen als auch Intra-Objekt-Varianzen wurden als ungeeignet für die vorliegende Arbeit beurteilt, da sie ausschließlich globale Halo-Effekte erfassen und damit keine Überprüfung der zentralen Hypothesen zulassen. Ebenso sind diese Maße nur in der Lage, illusorische Halo-Effekte auszuschließen, wenn die wahren Korrelationen beziehungsweise Varianzen bekannt sind. Eine andere Alternative stellt die Berechnung einer Subjekt x Objekt x Urteilsdimension ANOVA dar. Eine signifikante Interaktion von Subjekt und Objekt käme hierbei einem Halo-Effekt gleich. Durch diesen Interaktionseffekt kann allerdings nicht festgestellt werden, welche Art von Halo-Effekten (global oder spezifisch) gemessen wird, so dass auch dieses Maß für diese Arbeit ausgeschlossen wurde. Zuletzt wurde die Ermittlung von Halo-Effekten anhand der Faktorstruktur der Urteile in Erwägung gezogen. Globale und spezifische Halo-Effekte können bei dieser Methode mit Hilfe von Strukturgleichungsmodellen unterschieden werden. Der Ausschluss illusorischer Halo-Effekte ist dagegen nur möglich, wenn die wahre Faktorstruktur bekannt ist und mit der ermittelten verglichen werden kann. Dies ist ein großer Nachteil gegenüber dem gewählten Maß. Da Strukturgleichungsmodelle auf der anderen Seite keinen Zugewinn gegenüber der Manipulationsmethode bieten, wurde auch dieses Maß für die vorliegende Arbeit ausgeschlossen.

3.4 Berechnung der Halo-Effekte

Die Entscheidung, Halo-Effekte durch die Variierung präsentierter Informationen zu manipulieren und anhand von Ratings zu messen, hatte wichtige Implikationen für die weitere methodische Vorgehensweise. Dies betrifft insbesondere den Anspruch, die Informationen so zu manipulieren und zusammenzustellen, dass keine realen Zusammenhänge zwischen den verschiedenen Eigenschaften existieren. Diese Bedingung wurde bereits durch die Wahl von Agency und Communion (Bakan, 1966) als grundlegende Urteilsdimensionen (vergleiche Abschnitt 2.3.2) erfüllt, da beide Dimensionen nahezu unkorreliert sind (DeYoung, 2006; Digman, 1997). Bei einem Halo-Effekt von einer Agency-Eigenschaft (zum Beispiel *fleißig*) auf ein Communion-Urteil (zum Beispiel *gesellig*) handelt es sich folglich stets um einen wahren Halo-Effekt.

Um auch Halo-Effekte innerhalb der beiden Urteilsdimensionen untersuchen zu können, wurde zusätzlich darauf geachtet, Attribute auszuwählen, die möglichst wenig wahre Rückschlüsse aufeinander zulassen. Beispielsweise ist eine fleißige Person nicht unbedingt auch selbstbewusst, obwohl es sich bei beiden Eigenschaften um Agency-Attribute handelt. Ebenso wurden die Instruktionen und Untersuchungsszenarien so gestaltet, dass sie nur die nötigsten Informationen lieferten und somit eine völlig unbekannte Urteilssituation darstellten. Die Teilnehmer erfuhren zu Beginn jedes Experiments lediglich, dass es um die Bewertung von Personen ging. Das bedeutet, dass eventuelle Kenntnisse über einzelne Zusammenhänge aufgrund von Erfahrungen in der neuen Situation keine Aussagekraft hatten und somit auch innerhalb einer Dimension alle nicht neutralen Bewertungen als Halo-Effekte interpretiert werden konnten.

In der vorliegenden Arbeit wurden zwei Wirkungsweisen von Halo-Effekten fokussiert: Zum einen wurde untersucht, inwiefern bestimmte Informationen die Vorannahmen über unbekannte Charaktereigenschaften prägen können. Zum anderen wurde die Verzerrung von Urteilen über eindeutige Sachverhalte betrachtet.

Berechnung von Halo-Effekten als Vorannahmen über unbekannte Eigenschaften

Als Basis der Ermittlung der Halo-Effekte diente die Gesamtheit der in Abschnitt 3.1 beschriebenen Experimentaldurchgänge. Die Berechnungen wurden getrennt für positive und negative Attribute vorgenommen, wobei die Vorgehensweise identisch war. Im Folgenden soll die schrittweise Berechnung der abhängigen Variablen bei positiver Valenz anhand von Beispielen dargestellt werden: Als fiktive Datenbasis werden die Agency-Eigenschaften *ordentlich*, *selbständig* und *fleißig* sowie die Communion-Eigenschaften *ehrlich*, *gesellig* und *hilfsbereit* betrachtet. In den Experimentaldurchgängen würde jedes Attribut einmal auf Basis jedes der anderen bewertet werden müssen (insgesamt 30 Durchgänge). Bei jedem nicht neutralen Urteil handelt es sich bereits um einen ganz spezifischen Halo-Effekt. Wird zum Beispiel *ordentlich* vorgegeben und *selbständig* nicht neutral bewertet, ist dies ein Halo-Effekt von *ordentlich* auf *selbständig*, da über letzteres Attribut keine aussagekräftigen Informationen existieren.

Im ersten Berechnungsschritt wurden die Halo-Effekte jeder einzelnen Eigenschaft auf Agency- sowie Communion-Attribute ermittelt. Der Halo-Effekt von *ordentlich* auf Agency-Attribute entspräche im gegebenen Beispiel dem Mittelwert aus dem Halo-Effekt

von *ordentlich* auf *selbständig* und dem Halo-Effekt von *ordentlich* auf *fleißig*. Im zweiten Schritt wurden auf dieser Basis vier Halo-Effekte auf Dimensionesebene berechnet: von Agency-Attributen auf Agency-Attribute, von Agency-Attributen auf Communion-Attribute, von Communion-Attributen auf Agency-Attribute und von Communion-Attributen auf Communion-Attribute. Der Halo-Effekt von Agency- auf die Beurteilung anderer Agency-Eigenschaften würde etwa anhand des Mittelwerts der Halo-Effekte von *ordentlich*, *selbständig* und *fleißig* auf jeweils Agency-Attribute berechnet werden. Aus den vier Variablen wurden im letzten Rechenschritt durch erneute Mittelwertberechnungen die als abhängige Variablen verwendeten Halo-Effekte ermittelt: von Agency-Eigenschaften, von Communion-Eigenschaften, auf Agency-Eigenschaften, auf Communion-Eigenschaften, bei gleicher Dimension, bei ungleicher Dimension und der Gesamt-Halo-Effekt. Beispielsweise entspricht der Halo-Effekt von Agency-Eigenschaften dem Mittelwert aus dem Halo-Effekt von Agency- auf Agency-Attribute und dem Halo-Effekt von Agency- auf Communion-Attribute. Abbildung 5 zeigt die schrittweise Berechnung der abhängigen Variablen.

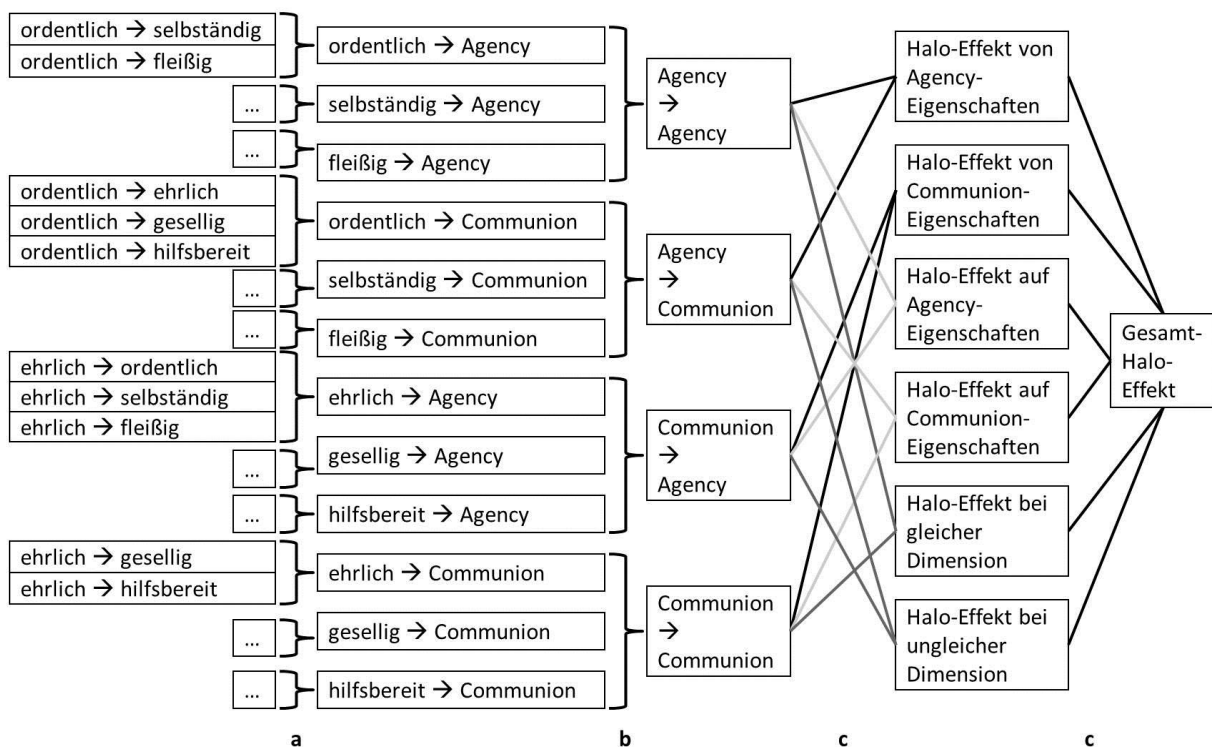


Abbildung 5. Schrittweise Berechnung der positiven abhängigen Variablen

a = Berechnung der Halo-Effekte der einzelnen Eigenschaften auf Agency- und Communion-Attribute; b = Berechnung der Halo-Effekte auf Dimensionesebene; c = Berechnung der finalen abhängigen Variablen
Die waagerechten Pfeile stellen Halo-Effekte bestimmter Eigenschaften auf andere Eigenschaften dar.

Berechnung von Halo-Effekten als Verzerrungen eindeutiger Urteile

Die Berechnung der abhängigen Variablen folgte dem gleichen schrittweisen Prinzip wie bei Halo-Effekten als Vorannahmen. So wurden nacheinander die Halo-Effekte der einzelnen Eigenschaften, die Halo-Effekte auf Dimensionsebene und schließlich die finalen abhängigen Variablen ermittelt. Sämtliche Berechnungen wurden erneut getrennt für positive und negative Valenz durchgeführt. Da sich alle Untersuchungen in diesem Bereich auf das vollständige Modell (vergleiche Abschnitt 3.2) bezogen, wurden als Stimuli ausschließlich Verhaltensweisen verwendet, die, basierend auf den Ergebnissen von Voruntersuchungen, für bestimmte Persönlichkeitseigenschaften standen.

Der grundlegende Unterschied in der Untersuchung von Verzerrungen anstelle von Vorannahmen besteht in der Ermittlung der einzelnen Halo-Effekte auf Eigenschaftsebene (zum Beispiel von *ordentlich* auf *selbständig*). Zu diesem Zweck wurden zwei verschiedene Experimentalbedingungen verwendet: In einer Baseline-Bedingung wurden eine oder mehrere Verhaltensweisen dargeboten, die bestimmte Eigenschaften beschrieben. Gleichzeitig sollten genau diese Eigenschaften beurteilt werden. Beispielsweise würde bei der Darbietung der Verhaltensweise ... *sagt anderen selten die Wahrheit* die Eigenschaft *verlogen* auf einer bipolaren Skala (*verlogen* ... *ehrlich*) abgefragt werden. Da die Verhaltensweise in einer Voruntersuchung genau aus dieser Eigenschaft generiert worden war, sollte das Urteil somit recht eindeutig ausfallen.

In der Verzerrungs-Bedingung wurden die gleichen Attribute abgefragt wie in der Baseline-Bedingung. Zusätzlich zu den schon in der Baseline-Bedingung präsentierten Verhaltensweisen wurde hierbei jedoch eine zusätzliche Verhaltensweise dargeboten, die für ein anderes Attribut stand. Bezogen auf das eben verwendete Beispiel könnte etwa zusätzlich zu ... *sagt anderen selten die Wahrheit* die Verhaltensweise ... *lernt und arbeitet viel und kontinuierlich* (stellvertretend für die Eigenschaft *fleißig*) präsentiert, aber erneut das Attribut *verlogen* abgefragt werden. Wie in der Baseline-Bedingung ist das Urteil eigentlich eindeutig, da die neu hinzugekommene Verhaltensweise keinerlei Bezug zu dem fraglichen Attribut hat und die ursprüngliche Verhaltensweise (... *sagt anderen selten die Wahrheit*) weiterhin präsent ist. Unterscheidet sich das Urteil trotzdem von dem in der Baseline-Bedingung, ist dies auf einen Halo-Effekt des Attributs *fleißig* auf die Eigenschaft *verlogen* zurückzuführen. Die Differenz der Ratings aus den beiden Bedingungen gibt Aufschluss über das genaue Ausmaß des jeweiligen Halo-Effekts.

Entsprechend dieses Prinzips konnten, je nach Datenbasis, Halo-Effekte jedes Attributs auf jedes andere Attribut ermittelt werden. Die schrittweise Berechnung der abhängigen Variablen erfolgte anschließend analog zu dem bereits für Vorannahmen geschilderten Vorgehen. Ein Beispiel für die Berechnung eines Halo-Effekts im Sinne der Verzerrung eines eindeutigen Urteils findet sich in Abbildung 6.

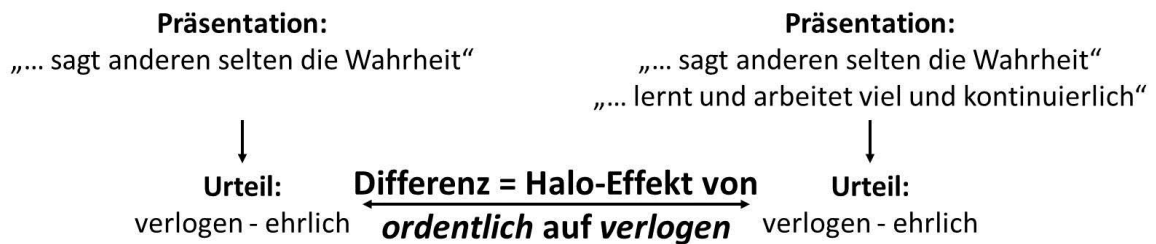


Abbildung 6. Beispiel für die Berechnung von Halo-Effekten als Verzerrung

3.5 Überblick über die durchgeführten Experimente

Aufgrund der Vielzahl potentieller Einflussfaktoren auf die jeweils wirkenden Prozesse und das Ausmaß von Halo-Effekten war eine große Anzahl an Experimenten notwendig. Mit dem Ziel einer schrittweisen Überprüfung des postulierten Halo-Effekt-Modells wurden, abgesehen von einer allgemeinen Voruntersuchung, insgesamt neun Untersuchungen, gegliedert in drei Experimentalblöcke, durchgeführt:

Der erste Experimentalblock hatte das Ziel, die Qualität und Quantität von Halo-Effekten in Abhängigkeit von Informationsvalenz und Urteilsdimension zu untersuchen. Hierbei wurden lediglich Halo-Effekte im Sinne von Vorannahmen über unbekannte Eigenschaften betrachtet. Im zweiten Block stand zuerst die Robustheit solcher Vorannahmen im Vordergrund. Anschließend verlagerte sich der Fokus auf Halo-Effekte im Sinne von Verzerrungen eindeutiger Urteile. Dabei wurden erneut Einflüsse von Valenz und Urteilsdimension betrachtet. Der dritte Experimentalblock fokussierte den Einfluss verschiedener Kontextfaktoren (Beruf des Subjekts und des Objekts sowie die allgemeine Situation) auf beide Formen von Halo-Effekten (Vorannahmen und Verzerrungen). Eine Übersicht über die durchgeführten Experimente ist in Tabelle 2 dargestellt.

Tabelle 2. Übersicht über die durchgeführten Experimente

Bl.	Exp.	Schr.	Untersuchungsziele	präsenierte Informationen
1	1	1	quantitative und qualitative	Persönlichkeitseigenschaften
	2	1	Einflüsse von Informationsvalenz und Urteilsdimension auf die Vorannahmen über unbekannte	Persönlichkeitseigenschaften und Bilder derselben Valenz
	3	2	Eigenschaften	Verhaltensweisen und Bilder derselben Valenz
2	4	2	quantitative und qualitative Einflüsse von Informationsvalenz und Urteilsdimension auf die	Verhaltensweisen und Bilder der gegensätzlichen Valenz
	5	3	Verzerrung eindeutiger Urteile und Robustheit von Halo-Effekten	zentrale Verhaltensweisen und periphere Verhaltensweisen gegensätzlicher Valenz und Dimension (gleichzeitige Darbietung)
	6	3		zentrale Verhaltensweisen und periphere Verhaltensweisen gegensätzlicher Valenz und/oder Dimension (sequentielle Darbietung)
3	7	2	quantitative und qualitative	Verhaltensweisen und Berufe
	8	2	Einflüsse verschiedener Kontextfaktoren auf Vorannahmen und Urteilsverzerrungen	Verhaltensweisen und Situationen (zusätzliche Variierung des Berufs der Probanden)
	9	3		zentrale Verhaltensweisen, Berufe und periphere Verhaltensweisen gegensätzlicher Valenz und Dimension

Anmerkung. Bl. = Experimentalblock; Exp. = Experiment; Schr. = Untersuchungsschritt im grundlegenden Modell

4 Relevanz der postulierten Einflussfaktoren: Allgemeine Voruntersuchung

Vor der Umsetzung der eigentlichen Experimente sollte zunächst sichergestellt werden, dass Informationsvalenz, Urteilsdimension und Kontextfaktoren auch relevant für Halo-Effekte sind. Um zu zeigen, dass diese Faktoren die Qualität und Quantität von Halo-Effekten potentiell beeinflussen können, wurde eine allgemeine Voruntersuchung durchgeführt. Da die Analyse von Halo-Effekten einer umfangreichen Vorauswahl und Überprüfung geeigneter Items bedarf, wurden in dieser Untersuchung stattdessen verwandte, jedoch weniger komplexe abhängige Maße betrachtet.

4.1 Methode

Teilnehmer

An der allgemeinen Voruntersuchung nahmen 60 (42 weiblich, 18 männlich; 20 in jeder Zwischensubjekt-Bedingung) Studenten der Universität Heidelberg teil, die in den Räumlichkeiten des Psychologischen Instituts geworben wurden. Das Alter der Probanden betrug im Mittel 22.15 Jahre ($SD = 3.67$, $Min = 18$, $Max = 41$). Deutsch war für insgesamt 51 Teilnehmer die Muttersprache, für fünf Teilnehmer die erste und für vier Teilnehmer die zweite Fremdsprache. Für ihre Teilnahme an der Studie erhielten die Probanden drei Euro oder eine halbe Versuchspersonenstunde.

Materialien

Der Untersuchung lag ein 3 (Beurteilungsobjekt) x 2 (Informationsvalenz) x 2 (Urteilsdimension) Design zugrunde. Das Beurteilungsobjekt wurde als Zwischensubjekt-Faktor mit den Stufen *Menschen im Allgemeinen*, *Menschen in naturwissenschaftlichen Berufen* und *Menschen in sozialen Berufen* konzipiert. Informationsvalenz (positiv versus negativ) und Urteilsdimension (Agency versus Communion) wurden innerhalb der Versuchsteilnehmer manipuliert.

Als Basis der beiden Innersubjekt-Faktoren wurden insgesamt 24 (16 positive und acht negative) Agency- und Communion-Attribute aus einer Untersuchung von Abele und

Kollegen (2008) verwendet. Die Wahl fiel auf dieses Set, da die Items auf Basis umfangreicher Tests in fünf verschiedenen Ländern, einschließlich Deutschland, getestet wurden. Zudem waren die positiven und negativen Attribute jeweils hinsichtlich ihrer Valenzstärke sowie der Worthäufigkeit im allgemeinen Sprachgebrauch vergleichbar. Das stärkere Vorkommen von positiven gegenüber negativen Attributen in diesem Set wurde nicht als problematisch betrachtet, da auch in der alltäglichen Sprache positive Wörter häufiger vorkommen als negative (Boucher & Osgood, 1969). Eine Übersicht über die verwendeten Persönlichkeitseigenschaften findet sich in Tabelle 3.

Tabelle 3. Für die Voruntersuchung verwendete Persönlichkeitseigenschaften

positive Eigenschaften		negative Eigenschaften	
Agency	Communion	Agency	Communion
klug	fürsorglich	unsicher	eingebildet
aktiv	hilfsbereit	faul	dominant
durchsetzungsfähig	treu	schüchtern	egoistisch
kreativ	höflich	verletzlich	hartherzig
selbständig	sensibel		
intelligent	einfühlsam		
rational	zuverlässig		
selbstbewusst	verständnisvoll		

Die abhängigen Maße waren drei verschiedene Variablen, die auf Basis der Literatur mit Halo-Effekten in Verbindung gebracht werden können (vergleiche Abschnitt 2.3): Valenzstärke, Generalisierbarkeit und Bedeutsamkeit für zwischenmenschliche Interaktion. Alle Variablen wurden anhand siebenstufiger Skalen gemessen. Die Erhebung der Valenzstärke erfolgte auf Grundlage der Frage *„Wie positiv oder negativ ist folgende Eigenschaft in Bezug auf Menschen [...]?“* (-3 = *sehr negativ*; 3 = *sehr positiv*). Die Generalisierbarkeit wurde mit Hilfe der Frage *„Wie viel sagt folgende Eigenschaft über Menschen [...] (im positiven wie im negativen Sinne) aus?“* (1 = *sehr wenig*; 7 = *sehr viel*) gemessen. Die Bedeutsamkeit für zwischenmenschliche Interaktion wurde durch die Frage *„Wie bedeutsam (im positiven wie im negativen Sinne) ist folgende Eigenschaft von Menschen [...] für den Umgang mit ihnen?“* (1 = *sehr unbedeutsam*; 7 = *sehr bedeutsam*) ermittelt. Je nach Zwischensubjekt-Bedingung bezogen sich die Fragen auf *„Menschen im*

Allgemeinen“, „*Menschen in naturwissenschaftlichen Berufen*“ oder „*Menschen in sozialen Berufen*“, wodurch der Faktor *Beurteilungsobjekt* manipuliert wurde. Jeder der drei abhängigen Variablen wurde anhand aller 24 Attribute erhoben, wodurch sich für jede Zwischensubjekt-Bedingung 72 verschiedene Kombinationen ergaben.

Prozedur

Die Probanden wurden über Aushänge oder persönliche Nachfrage in das Labor geladen. Dort unterschrieben sie eine Einverständniserklärung und begannen danach das Experiment. Die Zuteilung zu einer der drei Zwischensubjekt-Bedingungen erfolgte zufällig. An jedem Termin nahmen maximal sechs Personen gleichzeitig teil.

In der Instruktion wurde den Probanden mitgeteilt, dass sie im Folgenden an einer Bewertungsaufgabe teilnehmen würden, in der ihnen bestimmte Persönlichkeitsattribute präsentiert werden, die sie anhand drei unterschiedlicher Fragen bewerten sollen. Die drei Fragen wurden in der Instruktion bereits explizit genannt. Je nach Zwischensubjekt-Bedingung wurde zudem beschrieben, dass sich alle Fragen auf Menschen im Allgemeinen, in naturwissenschaftlichen Berufen oder in sozialen Berufen beziehen. Die Probanden wurden darauf hingewiesen, sich bei noch bestehenden offenen Fragen an die Versuchsleitung zu wenden und anderenfalls das Experiment zu starten.

Der anschließende Experimententeil bestand aus 72 Durchgängen, die in einer randomisierten Reihenfolge und ohne Zeitlimit dargeboten wurden. Im Durchschnitt benötigten die Probanden zehn Minuten zur vollständigen Bearbeitung aller Fragen.

Am Ende der Untersuchung wurde das Geschlecht und das Alter der Teilnehmer erfragt. Zudem gaben sie an, ob Deutsch ihre Muttersprache, ihre erste oder ihre zweite Fremdsprache sei. Nach Ende des Experiments wurden die Probanden über die Zielsetzung der Untersuchung aufgeklärt und erhielten ihre Vergütung.

4.2 Ergebnisse

Zur Analyse des Einflusses der drei Faktoren auf die abhängigen Variablen wurde jeweils eine 3 x 2 x 2 ANOVA (Beurteilungsobjekt x Urteilsdimension x Informationsvalenz) mit Messwiederholung durchgeführt. Im Falle der Valenzstärke wurden die Ratings aller

negativen Attribute vor den weiteren Analysen durch eine Multiplikation mit -1 umkodiert, um die Ergebnisse von bloßen Valenzunterschieden abgrenzen zu können.

Valenzstärke

Die Ergebnisse der ANOVA zeigten weder einen signifikanten Haupteffekt des Beurteilungsobjekts ($F(2, 57) = 1.91, p = .157, \eta^2 = .06$) noch der Urteilsdimension ($F(1, 57) = 1.40, p = .242, \eta^2 = .02$): Die durchschnittliche Valenzstärke unterschied sich nicht zwischen den verschiedenen Beurteilungsobjekten oder zwischen Agency- und Communion-Attributen. Jedoch erreichte der Haupteffekt für Valenz ein signifikantes Niveau ($F(1, 57) = 69.76, p < .001, \eta^2 = .55$): Positive Attribute ($M = 1.67, SD = 0.45$) wiesen eine höhere Valenzstärke auf als negative Attribute ($M = 1.04, SD = 0.65$).

In Bezug auf die Interaktionen gab es keinen signifikanten Effekt von Valenz x Urteilsdimension ($F(1, 57) = 0.08, p = .774, \eta^2 < .01$): Der Unterschied der Valenzstärke von positiven und negativen Attributen war für Agency- und Communion-Items gleich. Die Interaktion von Beurteilungsobjekt und Valenz wurde jedoch signifikant ($F(2, 57) = 3.91, p = .026, \eta^2 = .12$): Die Valenzstärke positiver Items ($M = 1.64, SD = 0.49$) zeigte den größten Unterschied gegenüber der Valenzstärke negativer Items ($M = 0.74, SD = 0.48$), wenn Naturwissenschaftler beurteilt wurden. Den kleinsten Unterschied zwischen positiven ($M = 1.64, SD = 0.49$) und negativen Attributen ($M = 1.26, SD = 0.78$) gab es in der Bedingung *Menschen in sozialen Berufen*. Bei der Bewertung von Menschen im Allgemeinen bewegte sich die Differenz von positiven ($M = 1.72, SD = 0.39$) und negativen Items ($M = 1.13, SD = 0.56$) zwischen denen der anderen beiden Zwischensubjekt-Bedingungen.

Es war des Weiteren eine signifikante Interaktion von Beurteilungsobjekt und Urteilsdimension erkennbar ($F(2, 57) = 12.68, p < .001, \eta^2 = .31$): Bei der Bewertung von Naturwissenschaftlern war die Valenzstärke von Agency-Attributen ($M = 1.51, SD = 0.45$) höher als die von Communion-Attributen ($M = 0.88, SD = 0.80$). Wurden Menschen mit sozialen Berufen beurteilt, zeigte sich dagegen eine geringere Valenzstärke für Agency- ($M = 1.32, SD = 0.58$) gegenüber Communion-Attributen ($M = 1.58, SD = 0.93$). Die größte Differenz zu Lasten von Agency- ($M = 1.03, SD = 0.41$) gegenüber Communion-Items ($M = 1.82, SD = 0.54$) gab es in der Bedingung *Menschen im Allgemeinen*.

Zuletzt wurde auch die Interaktion aller drei Faktoren signifikant ($F(2, 57) = 5.11, p = .009, \eta^2 = .15$): Der kleinste Unterschied in der Valenzstärke positiver ($M = 2.08, SD = 0.37$)

und negativer Communion-Attribute ($M = 1.56, SD = 0.83$) wurde bei der Beurteilung von Menschen im Allgemeinen deutlich. Für Agency-Eigenschaften ergab sich die geringste Differenz positiver ($M = 1.38, SD = 0.66$) und negativer Items ($M = 1.26, SD = 0.78$) bei der Bewertung von Menschen in sozialen Berufen.

Generalisierbarkeit

Auf Basis der ANOVA konnte kein signifikanter Haupteffekt des Faktors *Beurteilungsobjekt* gezeigt werden ($F(2, 57) = 1.21, p = .307, \eta^2 = .04$): Die mittlere Generalisierbarkeit der Persönlichkeitsattribute unterschied sich nicht zwischen den verschiedenen Beurteilungsobjekten. Im Gegensatz dazu erreichte der Haupteffekt der Urteilsdimension ein signifikantes Maß ($F(1, 57) = 10.65, p = .002, \eta^2 = .16$): Communion-Eigenschaften ($M = 5.19, SD = 0.81$) wurden stärker generalisiert als Agency-Attribute ($M = 4.88, SD = 0.69$). Des Weiteren ergab sich ein signifikanter Haupteffekt für Valenz ($F(1, 57) = 16.66, p < .001, \eta^2 = .23$): Positive Attribute ($M = 5.24, SD = 0.57$) wiesen eine höhere Generalisierbarkeit auf als negative Attribute ($M = 4.83, SD = 0.85$).

Die Interaktion von Beurteilungsobjekt und Valenz wurde nicht signifikant ($F(2, 57) = 1.34, p = .271, \eta^2 = .05$): Der Unterschied in der Generalisierbarkeit von positiven und negativen Persönlichkeitseigenschaften war für naturwissenschaftliche und soziale Berufe sowie Menschen im Allgemeinen gleich. Es konnte jedoch eine signifikante Interaktion von Valenz und Urteilsdimension deutlich gemacht werden ($F(1, 57) = 5.63, p = .021, \eta^2 = .09$): Bei negativer Valenz wurden Communion-Attribute ($M = 5.08, SD = 1.03$) stärker auf den Gesamteindruck über eine Person generalisiert als Agency-Eigenschaften ($M = 4.57, SD = 1.00$). In der positiven Valenzbedingung war die Differenz in der Generalisierbarkeit von Communion- ($M = 5.30, SD = 0.87$) und Agency-Eigenschaften ($M = 5.18, SD = 0.77$) nur von vergleichsweise geringem Maße.

Ebenfalls signifikant bestätigt wurde die Interaktion von Beurteilungsobjekt und Urteilsdimension ($F(2, 57) = 13.42, p < .001, \eta^2 = .32$): Bei der Bewertung von Personen mit naturwissenschaftlichen Berufen war die Generalisierbarkeit von Agency- ($M = 5.05, SD = 0.66$) höher als die von Communion-Items ($M = 4.67, SD = 0.81$). In der Bedingung mit Menschen in sozialen Berufen wurden Agency-Attribute ($M = 4.84, SD = 0.82$) stattdessen weniger stark generalisiert als Communion-Eigenschaften ($M = 5.40, SD = 0.68$). Bei der Beurteilung von Menschen im Allgemeinen zeigte sich sogar ein noch größerer Unterschied

in der Generalisierbarkeit zu Lasten von Agency- ($M = 4.73$, $SD = 0.57$) gegenüber Communion-Eigenschaften ($M = 5.51$, $SD = 0.70$).

Neben den schon genannten Effekten wurde auch die Dreifachinteraktion *Beurteilungsobjekt x Informationsvalenz x Urteilsdimension* signifikant ($F(2, 57) = 8.32$, $p = .001$, $\eta^2 = .23$): Die stärkste Differenz in der Generalisierbarkeit positiver ($M = 5.68$, $SD = 0.84$) und negativer Communion-Attribute ($M = 5.11$, $SD = 0.94$) gab es bei der Beurteilung von Menschen in sozialen Berufen. Gleichzeitig war in dieser Bedingung für Agency-Eigenschaften der geringste Unterschied zwischen positiven ($M = 4.90$, $SD = 0.83$) und negativen Items ($M = 4.79$, $SD = 1.13$) erkennbar.

Bedeutsamkeit für zwischenmenschliche Interaktion

Bei der ANOVA war kein signifikanter Haupteffekt des Beurteilungsobjekts zu erkennen ($F(2, 57) = 1.19$, $p = .312$, $\eta^2 = .04$): Die durchschnittliche Bedeutsamkeit der Eigenschaften unterschied sich nicht zwischen den einzelnen Beurteilungsobjekten. Signifikant wurde aber der Haupteffekt der Urteilsdimension ($F(1, 57) = 29.47$, $p < .001$, $\eta^2 = .34$): Communion-Eigenschaften ($M = 5.48$, $SD = 0.73$) wurden als bedeutsamer für die zwischenmenschliche Interaktion eingeschätzt als Agency-Attribute ($M = 4.91$, $SD = 0.76$). Des Weiteren war ein signifikanter Haupteffekt für Valenz erkennbar ($F(1, 57) = 6.52$, $p = .013$, $\eta^2 = .10$): Positive Persönlichkeitseigenschaften ($M = 5.34$, $SD = 0.52$) wiesen eine höhere Bedeutsamkeit auf als negative Eigenschaften ($M = 5.05$, $SD = 0.91$).

Bei den Interaktionen war kein signifikanter Effekt von Beurteilungsobjekt x Valenz ersichtlich ($F(2, 57) = 1.56$, $p = .219$, $\eta^2 = .05$): Der Unterschied in der Bedeutsamkeit von positiven und negativen Eigenschaften war für beide Berufsgruppen sowie Menschen im Allgemeinen gleich. Auch die Interaktion von Valenz und Urteilsdimension erreichte kein signifikantes Niveau ($F(1, 57) = 0.01$, $p = .908$, $\eta^2 < .01$): Die Differenz in der Bedeutsamkeit von positiven und negativen Persönlichkeitseigenschaften unterschied sich nicht zwischen der Agency- und der Communion-Bedingung.

Anhand der statistischen Daten wurde ein signifikanter Effekt der Interaktion *Beurteilungsobjekt x Urteilsdimension* aufgezeigt ($F(2, 57) = 6.89$, $p = .002$, $\eta^2 = .20$): Die Bedeutsamkeit von Communion-Items ($M = 5.59$, $SD = 0.75$) wies den größten Unterschied gegenüber Agency-Items ($M = 4.49$, $SD = 0.84$) auf, wenn Menschen im Allgemeinen beurteilt wurden. Den kleinsten Unterschied zwischen Communion- ($M = 5.29$, $SD = 0.85$)

und Agency-Eigenschaften ($M = 5.12$, $SD = 0.62$) gab es bei der Bewertung von Naturwissenschaftlern. In der Zwischensubjekt-Bedingung *Menschen in sozialen Berufen* bewegte sich die Differenz von Communion- ($M = 5.55$, $SD = 0.56$) und Agency-Items ($M = 5.11$, $SD = 0.65$) zwischen denen der anderen beiden Bedingungen.

Die Dreifachinteraktion von Beurteilungsobjekt, Valenz und Urteilsdimension erreichte kein signifikantes Maß ($F(2, 57) = 2.29$, $p = .110$, $\eta^2 = .07$): Positive Attribute wurden in beiden Valenzbedingungen und Urteilsdimensionen im gleichen Maße bedeutsamer eingeschätzt als negative Persönlichkeitseigenschaften.

4.3 Diskussion

Die allgemeine Voruntersuchung prüfte den Einfluss von Informationsvalenz, Urteilsdimension und Kontextfaktoren auf drei abhängige Maße, die potentiell mit Halo-Effekten in Verbindung stehen. Die Ergebnisse des Experiments sprechen klar für die Relevanz der postulierten Einflussfaktoren bezüglich Halo-Effekten.

Für alle drei betrachteten abhängigen Variablen kann ein signifikanter Einfluss der Informationsvalenz bestätigt werden: Positive Persönlichkeitseigenschaften weisen eine größere Valenzstärke, Generalisierbarkeit und Bedeutsamkeit auf als negative Attribute. Diese Befunde stützen die Hypothese, dass positive Attribute zu stärkeren Halo-Effekten führen als negative. Hinsichtlich der Globalität von Halo-Effekten lässt sich noch keine Aussage treffen, da die in den Daten gefundenen Unterschiede lediglich quantitativer und nicht qualitativer Natur sind.

Ein weiteres bedeutsames Ergebnis liegt in dem signifikanten Effekt der Urteilsdimension für zwei der drei abhängigen Maße: Communion-Eigenschaften werden stärker auf den Gesamteindruck über eine Person generalisiert und als bedeutsamer für zwischenmenschliche Interaktion eingeschätzt als Agency-Eigenschaften. Dieser Befund spricht für die Vermutung, dass Communion-Attribute mit stärkeren Halo-Effekten einhergehen als Agency-Attribute. Der fehlende signifikante Haupteffekt bezüglich der Valenzstärke lässt sich auf die Tatsache zurückführen, dass die in der Untersuchung verwendeten positiven und negativen Eigenschaften bereits jeweils hinsichtlich ihrer Valenzstärke kontrolliert waren (Abele et al., 2008).

Der dritte wichtige Befund hinsichtlich der zentralen Hypothesen liegt in der signifikanten Interaktion von Beurteilungsobjekt und Urteilsdimension für alle drei abhängigen Variablen: Bei der Bewertung von Menschen in sozialen Berufen sowie im Allgemeinen sind Communion-Items extremer valent und werden stärker generalisiert als Agency-Items. Dieses Muster kehrt sich dagegen um, wenn Naturwissenschaftler Gegenstand der Beurteilung sind. Auch hinsichtlich der Bedeutsamkeit für zwischenmenschliche Interaktion ist der Vorteil von Communion- gegenüber Agency-Eigenschaften größer, wenn Menschen im Allgemeinen oder in sozialen Berufen bewertet werden. Diese Ergebnisse lassen die Vermutung zu, dass auch Halo-Effekte durch die Interaktion von Urteilsdimension und Beurteilungsobjekt (als typisches Beispiel für Kontexteinflüsse) beeinflusst werden können.

Die Gesamtheit der Ergebnisse unterstützt die Annahme aller drei zentralen Hypothesen. Hierbei ist jedoch anzumerken, dass die verwendeten abhängigen Variablen keinesfalls mit Halo-Effekten gleichzusetzen sind, sondern lediglich auf Basis der Literatur mit ihnen in Verbindung gebracht werden können. Mögliche Zusammenhänge zwischen Halo-Effekten und den besagten drei Variablen sind bisher noch nicht explizit untersucht worden. Jedoch bieten die Befunde einen zusätzlichen Grund, die Einflüsse von Informationsvalenz, Urteilsdimension und Kontextfaktoren auf die Qualität und Quantität von Halo-Effekten genauer zu untersuchen.

5 Einflüsse von Informationsvalenz und Urteilsdimension auf Vorannahmen

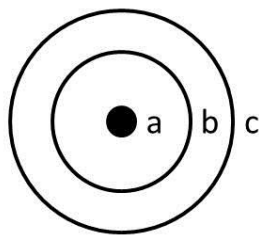
Der erste Experimentalblock untersuchte den Einfluss von Informationsvalenz und Urteilsdimension auf die Qualität und Quantität von Halo-Effekten. Hierbei standen Halo-Effekte im Sinne von Vorannahmen über unbekannte Eigenschaften im Fokus der Aufmerksamkeit. Gemäß den ersten beiden zentralen Hypothesen wurde angenommen, dass positive Halo-Effekte sowohl stärker als auch globaler sind als negative Halo-Effekte. Zudem sollten Halo-Effekte von Communion-Attributen stärker, aber nicht globaler sein als Halo-Effekte von Agency-Attributen. Zur Überprüfung dieser Postulate wurden drei Experimente durchgeführt.

5.1 Quantitative Unterschiede verschiedener Halo-Effekte: Experiment 1

Im ersten Experiment sollten zunächst nur quantitative Unterschiede verschiedener Halo-Effekte aufgezeigt werden. Da Communion-Eigenschaften bedeutsamer für die Beurteilung anderer Personen sind (Wojciszke & Abele, 2008; Wojciszke et al., 2009) und bevorzugt verarbeitet werden (Abele & Bruckmüller, 2011), sollten sie auch zu größeren Halo-Effekten führen als Agency-Eigenschaften.

Der Einfluss von Valenz ist darin begründet, dass positive Informationen einander ähnlicher sind als negative (Unkelbach et al., 2008). Folglich fällt es bei positiver Valenz leichter, von einer Eigenschaft auf eine andere zu schließen. Auf Halo-Effekte bezogen bedeutet das, dass bei der Vorgabe eines positiven Attributs (zum Beispiel das Agency-Attribut *fleißig*) auch andere positive Attribute als sehr zutreffend eingeschätzt werden sollten. Dies sollte besonders für Eigenschaften der gleichen (in diesem Beispiel etwa *selbstbewusst*), aber auch der anderen Dimension (zum Beispiel *ehrlich*) der Fall sein. Im Vergleich dazu sollten bei der Vorgabe eines negativen Attributs (zum Beispiel die Agency-Eigenschaft *faul*) andere Eigenschaften, insbesondere bei ungleicher Dimension (in diesem Beispiel etwa *verlogen*) als vergleichsweise wenig zutreffend beurteilt werden. Abbildung 7 stellt diese Valenzasymmetrie anhand eines hypothetischen Raums dar.

Positive Valenz



Negative Valenz

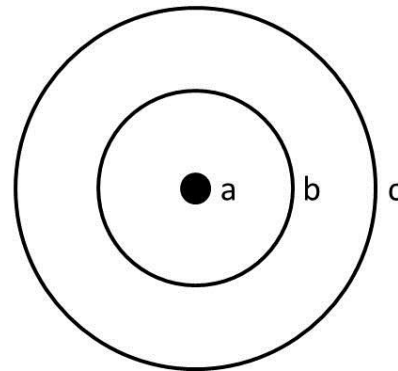


Abbildung 7. Veranschaulichung der quantitativen Valenzunterschiede

a = vorgegebene Eigenschaft; b = Eigenschaften der gleichen Dimension; c = Eigenschaften der anderen Dimension

5.1.1 Vorerhebung

Vor der Durchführung des eigentlichen Experiments war es zunächst notwendig, geeignete Items auszuwählen. Dabei stand vor allem im Vordergrund, Eigenschaftspaare gegensätzlicher Valenz zu finden, um positive und negative Halo-Effekte unmittelbar vergleichen zu können. Hierbei wurden negative von sogenannten distinkt negativen Eigenschaften unterschieden. Beispielsweise wäre das Attribut *unehrlich* zwar negativ, jedoch nicht distinkt negativ, da es einem positiven Attribut (*ehrlich*) plus einer Vorsilbe entstammt. Ein Beispiel für eine distinkt negative Eigenschaft wäre in diesem Fall das Attribut *verlogen*. Die Abgrenzung dieser beiden Arten von negativen Attributen wurde getroffen, da bei nicht distinkt negativen Persönlichkeitseigenschaften die gegensätzlichen positiven Attribute indirekt auch präsent sind und Halo-Effekte dadurch möglicherweise stärker sein könnten als bei distinkt negativen Eigenschaften.

Ziel der Vorerhebung war es, einen möglichst großen Datensatz aus gegensätzlichen positiven und distinkt negativen Eigenschaften zu generieren. Als erste Basis wurden 52 Eigenschaften (17 positive Agency-, 21 positive Communion-, sechs negative Agency- und acht negative Communion-Eigenschaften) aus einer Untersuchung von Abele und Kollegen (2008) verwendet. Die Eigenschaften wurden in einen Fragebogen eingepflegt und mit der Instruktion versehen, das jeweilige passende Gegenstück ohne die Verwendung von Vorsilben (wie zum Beispiel *un-* oder *in-*) zu finden. Insgesamt 24 Freiwillige bearbeiteten den Fragebogen. Der Fragebogen ist in Anhang A zu finden.

Anschließend wurden die aus den Items generierten Gegensätze betrachtet. Jede Eigenschaft, die von mindestens 20 Prozent der Probanden aus einem der 52 Attribute generiert worden war, wurde in die Auswahl für den nächsten Untersuchungsschritt gezogen. Generierte Gegensätze, die bereits beide im Fragebogen präsent waren (zum Beispiel die Items *faul* und *fleißig*), wurden schon in den finalen Datensatz integriert und nicht erneut untersucht. Die Auswahl bestand insgesamt aus 34 Persönlichkeitsattributen. Diese wurden in einen zweiten Fragebogen eingepflegt. Die Aufgabe lautete erneut, das jeweils passende Gegenstück zu finden. Insgesamt füllten 20 Probanden den zweiten Fragebogen aus. Der Fragebogen befindet sich in Anhang B.

Im letzten Schritt der Vorerhebung wurde analysiert, in welchem Maße die ursprünglichen Eigenschaften aus Fragebogen 1 auf Basis der 34 Eigenschaften in Fragebogen 2 generiert wurden. Dieser diente somit als Gegencheck für Fragebogen 1. In den finalen Datensatz wurden alle Eigenschaftspaare integriert, bei denen von mindestens 20 Prozent der Probanden das ursprüngliche Item generiert wurde. Es wurde auch darauf geachtet, dass jedes Attribut nur einem und nicht mehreren anderen als Gegensatz zugeordnet wurde. Die Kriterien wurden von 19 Paaren (neun Agency- und zehn Communion-Paare) erfüllt. Anschließend wurden die nicht distinkt negativen Eigenschaften für jedes positive Item formuliert. Konnte dies nicht mit Hilfe einer Vorsilbe geschehen, wurde stattdessen das Wort *nicht* verwendet (zum Beispiel *nicht fleißig*). Der finale Datensatz einschließlich der Ergebnisse der Vorerhebung findet sich in Anhang C.

5.1.2 Methode

Teilnehmer

Teilnehmer des Experiments waren 60 (44 weiblich, 16 männlich; 20 in jeder Zwischensubjekt-Bedingung) Studenten der Universität Heidelberg, die in den Räumlichkeiten des Psychologischen Instituts auf die Erhebung aufmerksam gemacht wurden. Das Alter der Probanden betrug im Mittel 22.92 Jahre ($SD = 4.25$, $Min = 18$, $Max = 41$). Deutsch war für 47 Teilnehmer die Muttersprache, für neun Teilnehmer die erste und für vier Teilnehmer die zweite Fremdsprache. Für ihre Teilnahme an dem Experiment erhielten die Probanden drei Euro oder eine halbe Versuchspersonenstunde.

Materialien

Die Basis des Experiments war ein 3 (Informationsvalenz) x 2 (Urteilsdimension) x 2 (Gleichheit von dargebotener und beurteilter Dimension) Design. Die Informationsvalenz wurde als Zwischensubjekt-Faktor mit den Ausprägungen *positiv*, *negativ* und *distinkt negativ* konzipiert. Urteilsdimension (Agency versus Communion) und Gleichheit (gleiche versus ungleiche Dimension) wurden innerhalb der Teilnehmer manipuliert.

Als Items dienten zehn Eigenschaftstripletts (jeweils fünf für Agency und Communion) aus der generierten Datenbasis. Die Wahl fiel dabei auf die Items, die in der Vorerhebung die besten Ergebnisse geliefert hatten, also am häufigsten aus ihrem entsprechenden Gegensatz generiert worden waren. Die in dieser Untersuchung verwendeten Eigenschaften sind in Anhang C markiert.

Die abhängigen Variablen der Untersuchung waren verschiedene Halo-Effekte, die aus den Mittelwerten diverser Einzelratings berechnet wurden. Die Ratings selbst wurden auf Basis folgender Frage und Skala erhoben: „*Eine Person hat die Eigenschaft [...]. Wie wahrscheinlich ist es, dass sie auch folgende Eigenschaft hat: [...]?*“ (1 = *keinesfalls*; 7 = *ganz sicher*). Die Wahrscheinlichkeit jedes Attributs wurde auf der Basis jedes anderen beurteilt, so dass sich für jede Valenzbedingung 90 (10 x 9) Durchgänge ergaben.

Prozedur

Die Teilnehmer wurden per Aushang sowie durch studentische Hilfskräfte für die Untersuchung geworben und in das Untersuchungslabor geladen. Dort unterschrieben sie eine Einverständniserklärung und begannen anschließend das Experiment. Die Zuteilung zu einer der drei Zwischensubjekt-Bedingungen erfolgte zufällig. An jedem Termin konnten maximal sechs Personen gleichzeitig teilnehmen.

In der Instruktion wurde den Probanden mitgeteilt, dass sie im Folgenden an einer Bewertungsaufgabe teilnehmen würden, in der ihnen Aussagen über verschiedene Menschen präsentiert werden und gleichzeitig bestimmte Persönlichkeitsattribute erscheinen. Dabei solle eingeschätzt werden, wie wahrscheinlich es ist, dass Menschen mit der gegebenen Eigenschaft auch die andere Eigenschaft haben. Je nach Zwischensubjekt-Bedingung wurde zudem beschrieben, dass alle Attribute positiv beziehungsweise negativ seien. Die Probanden wurden darauf hingewiesen, sich bei noch offenen Fragen an die Versuchsleitung zu wenden und anderenfalls das Experiment zu starten.

Der anschließende Experimententeil umfasste 90 Durchgänge, die in einer randomisierten Reihenfolge und ohne Zeitlimit dargeboten wurden. Die Teilnehmer benötigten etwa 15 Minuten für die Bearbeitung aller Durchgänge.

Am Ende der Untersuchung gaben die Probanden ihr Geschlecht sowie ihr Alter an. Des Weiteren wurde erfragt, ob Deutsch die Muttersprache, die erste oder die zweite Fremdsprache sei. Nach Beendigung des Experiments wurden die Teilnehmer über Inhalt und Ziel der Untersuchung aufgeklärt und erhielten ihre Vergütung.

5.1.3 Ergebnisse

Vorab-Analysen

Vor der eigentlichen Analyse der Daten wurde überprüft, ob die aus den einzelnen Ratings berechneten abhängigen Variablen auch mit Halo-Effekten gleichgesetzt werden können. Ein Urteil wurde dann als Halo-Effekt betrachtet, wenn es sich signifikant vom unteren Ende der Skala (1 = *keinesfalls*) unterschied. Zu diesem Zweck wurden für jede der drei Valenzbedingungen alle abhängigen Variablen mit Hilfe von t-Tests gegen den Wert 1 getestet. In allen drei Bedingungen war jede der abhängigen Variablen signifikant unterschiedlich zu dem Wert 1 ($ps < .050$). Die abhängigen Maße konnten folglich als Halo-Effekte betrachtet und näher untersucht werden.

In einer zweiten Vorab-Analyse wurden die Halo-Effekte von negativer und distinkt negativer Valenz miteinander verglichen. Hierzu wurden t-Tests für alle abhängigen Variablen durchgeführt. Der Gesamt-Halo-Effekt in der negativen ($M = 3.78$, $SD = 0.90$) unterschied sich nicht signifikant von dem in der distinkt negativen ($M = 3.69$, $SD = 0.67$) Bedingung ($t(38) = 0.34$, $p = .735$). Auch für alle spezifischeren Halo-Effekte (zum Beispiel der von Agency-Eigenschaften) wurden keine signifikanten Unterschiede zwischen den Bedingungen gefunden ($ps > .050$). Aus diesem Grund wurden die negative und distinkt negative Bedingung für alle folgenden Analysen zusammengefasst.

Einflüsse von Informationsvalenz und Dimensionsgleichheit

Die Überprüfung der Valenzhypothese geschah anhand einer 2 x 2 ANOVA (Informationsvalenz x Gleichheit von dargebotener und beurteilter Dimension) mit Messwiederholung. Es war ein signifikanter Haupteffekt für Valenz ($F(1, 58) = 12.45$, $p =$

.001, $\eta^2 = .18$) erkennbar: Positive Halo-Effekte ($M = 4.40$, $SD = 0.43$) waren größer als negative ($M = 3.74$, $SD = 0.78$). Auch der Haupteffekt der Dimensionsgleichheit wurde signifikant ($F(1, 58) = 61.88$, $p < .001$, $\eta^2 = .52$): Bei gleicher Dimension von dargebotenem und beurteiltem Attribut ($M = 4.28$, $SD = 0.76$) kamen stärkere Halo-Effekte zum Vorschein als bei verschiedenen Dimensionen ($M = 3.64$, $SD = 0.84$). Die Interaktion von Valenz und Gleichheit erreichte ein signifikantes Niveau ($F(1, 58) = 5.97$, $p = .018$, $\eta^2 = .09$): Negative Halo-Effekte bei gleicher ($M = 4.12$, $SD = 0.83$) und ungleicher Dimension ($M = 3.36$, $SD = 0.82$) wiesen größere Unterschiede zueinander auf als positive Halo-Effekte bei gleicher ($M = 4.60$, $SD = 0.47$) und ungleicher Dimension ($M = 4.20$, $SD = 0.56$).

Einflüsse von Informationsvalenz und Urteilsdimension

Um die Dimensionshypothese zu überprüfen, wurde eine 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung durchgeführt. Neben dem bereits im letzten Abschnitt geschilderten Haupteffekt für Valenz konnte auch ein signifikanter Effekt für den Faktor *Urteilsdimension* gefunden werden ($F(1, 58) = 9.76$, $p = .003$, $\eta^2 = .14$): Halo-Effekte von Communion-Attributen ($M = 4.03$, $SD = 0.77$) waren von stärkerem Ausmaß als Halo-Effekte von Agency-Attributen ($M = 3.89$, $SD = 0.76$). Die Interaktion von Informationsvalenz und Urteilsdimension wurde nicht signifikant ($F(1, 58) = 0.02$, $p = .903$, $\eta^2 < .01$): Positive Halo-Effekte von Agency- ($M = 4.34$, $SD = 0.45$) und Communion-Eigenschaften ($M = 4.47$, $SD = 0.48$) waren nicht unterschiedlicher zueinander als negative Halo-Effekte von Agency- ($M = 3.67$, $SD = 0.79$) und Communion-Eigenschaften ($M = 3.81$, $SD = 0.80$). In einer anschließenden Analyse wurden Agency- und Communion-Halo-Effekte noch einmal unabhängig von der Informationsvalenz miteinander verglichen. Dafür wurde jeweils ein t-Test bei gepaarten Stichproben durchgeführt: Der Unterschied zwischen positiven Agency- und Communion-Halo-Effekten wurde nicht signifikant ($t(19) = 1.63$, $p = .120$). Jedoch zeigte sich ein signifikanter Unterschied zwischen negativen Agency- und Communion-Halo-Effekten ($t(39) = 3.02$, $p = .004$).

Beeinflussbarkeit von Agency- und Communion-Urteilen

Neben der Überprüfung der Hypothesen wurde untersucht, in welchem Maße sich die Beurteilung unbekannter Agency- und Communion-Eigenschaften durch Halo-Effekte prägen lässt. Dies geschah unter der Verwendung einer 2 x 2 ANOVA (Informationsvalenz x

Dimension der beurteilten Eigenschaft) mit Messwiederholung. Neben dem Valenzeffekt wurde auch der Haupteffekt für die Dimension der beurteilten Eigenschaft signifikant ($F(1, 58) = 13.97, p < .001, \eta^2 = .19$): Vorannahmen über Communion-Attribute ($M = 4.03, SD = 0.75$) wurden stärker durch Halo-Effekte geprägt als Vorannahmen über Agency-Eigenschaften ($M = 3.89, SD = 0.77$). Bezüglich der Interaktion *Informationsvalenz x Dimension der beurteilten Eigenschaft* war kein signifikantes Ergebnis erkennbar ($F(1, 58) = 0.22, p = .643, \eta^2 < .01$): Die Vorannahmen über positive Agency- ($M = 4.34, SD = 0.45$) und Communion-Eigenschaften ($M = 4.46, SD = 0.46$) waren nicht unterschiedlicher zueinander als die Vorannahmen über negative Agency- ($M = 3.66, SD = 0.80$) und Communion-Eigenschaften ($M = 3.82, SD = 0.79$). Im nächsten Schritt wurde die Beeinflussbarkeit von Agency- und Communion-Urteilen mit Hilfe von t-Tests gegenübergestellt: Bei positiver Valenz wurde kein signifikanter Unterschied zwischen Agency- und Communion-Urteilen deutlich ($t(19) = 1.71, p = .104$). Der Unterschied zwischen negativen Agency- und Communion-Eigenschaften wurde dagegen signifikant ($t(39) = 4.03, p < .001$).

5.1.4 Diskussion

Experiment 1 widmete sich der Untersuchung quantitativer Unterschiede von Halo-Effekten in Abhängigkeit von Informationsvalenz und Urteilsdimension. Halo-Effekte von negativen und distinkt negativen Attributen konnten hierbei gleichgesetzt werden. Die Befunde bestätigen den postulierten Valenzeinfluss: Positive Halo-Effekte sind stärker als negative Halo-Effekte. Ein möglicher Hinweis auf einen qualitativen Unterschied von positiven und negativen Halo-Effekten findet sich in der signifikanten Interaktion von Informationsvalenz und Dimensionsgleichheit: Positive Halo-Effekte sind hinsichtlich ihrer Stärke untereinander ähnlicher als negative Halo-Effekte. Dieses Ergebnis ist konform mit der Density Hypothesis von Unkelbach und Kollegen (2008). Ob sich aus diesen rein quantitativen Differenzen auch qualitative Unterschiede ableiten lassen, muss in weiteren Experimenten mit Hilfe anderer Manipulationen untersucht werden.

Die Dimensionshypothese kann für die negative Valenzbedingung bestätigt werden: Negative Communion-Halo-Effekte sind stärker als negative Agency-Halo-Effekte. Dies lässt sich mit der hohen Bedeutsamkeit (Wojciszke & Abele, 2008; Wojciszke et al., 2009) sowie der bevorzugten Verarbeitung von Communion-Attributen (Abele & Bruckmüller, 2011)

begründen. Die Ergebnisse gehen auch bei positiver Valenz in die erwartete Richtung, wurden jedoch nicht signifikant. Ob sich die Tendenz auch signifikant bestätigen lässt, muss in einer weiteren Untersuchung analysiert werden.

Anhand eines weiteren Ergebnisses wird deutlich, dass Vorannahmen über negative Communion-Eigenschaften stärker durch Halo-Effekte geprägt werden als Vorannahmen über negative Agency-Eigenschaften. Die gleiche, wenn auch nicht signifikante Tendenz ist bei positiver Valenz erkennbar. Auch hier müssen weitere Experimente zeigen, ob die gezeigten Ergebnisse replizierbar sind.

5.2 Qualitative Unterschiede verschiedener Halo-Effekte: Experiment 2

Nach der Bestätigung quantitativer Unterschiede zwischen verschiedenen Halo-Effekten wurde im zweiten Experiment untersucht, ob Halo-Effekte auch hinsichtlich ihrer Qualität variieren können. Gemäß den zentralen Hypothesen sollten positive Halo-Effekte nicht nur stärker, sondern auch globaler sein als negative Halo-Effekte. Auf der anderen Seite wurde angenommen, dass Halo-Effekte von Communion-Attributen zwar stärker, jedoch nicht globaler sind als Halo-Effekte von Agency-Attributen.

Die Untersuchung der Qualität von Halo-Effekten erfolgte durch die gezielte Manipulation globaler Eindrücke. Dies geschah durch die Darbietung von attraktiven oder unattraktiven Gesichtern. Aufgrund der aktuellen Befundlage ist anzunehmen, dass Gesichter, unabhängig von ihrer Valenz, sowohl global (Farah et al., 1998; Tanaka & Farah, 1993) als auch spezifisch (Amishav & Kimchi, 2010; Cabeza & Kato, 2000; Rotshtein et al., 2007; Schwarzer & Massaro, 2001) verarbeitet werden. Folglich sollte die Darbietung eines unattraktiven Gesichts zusätzlich zu einer negativen Eigenschaft zu einem qualitativen Zugewinn führen. Der resultierende negative Halo-Effekt sollte daher bei zusätzlicher Bilddarbietung stärker sein. Im Gegensatz dazu sollte die Darbietung eines attraktiven Gesichts zusätzlich zu einer positiven Eigenschaft keinen qualitativen Zugewinn mit sich bringen, da positive Informationen alleine schon von globaler Natur sind. Positive Halo-Effekte sollten daher unabhängig von der Präsentation eines Fotos von gleicher Größe sein. Das postulierte Prinzip ist in Abbildung 8 veranschaulicht.

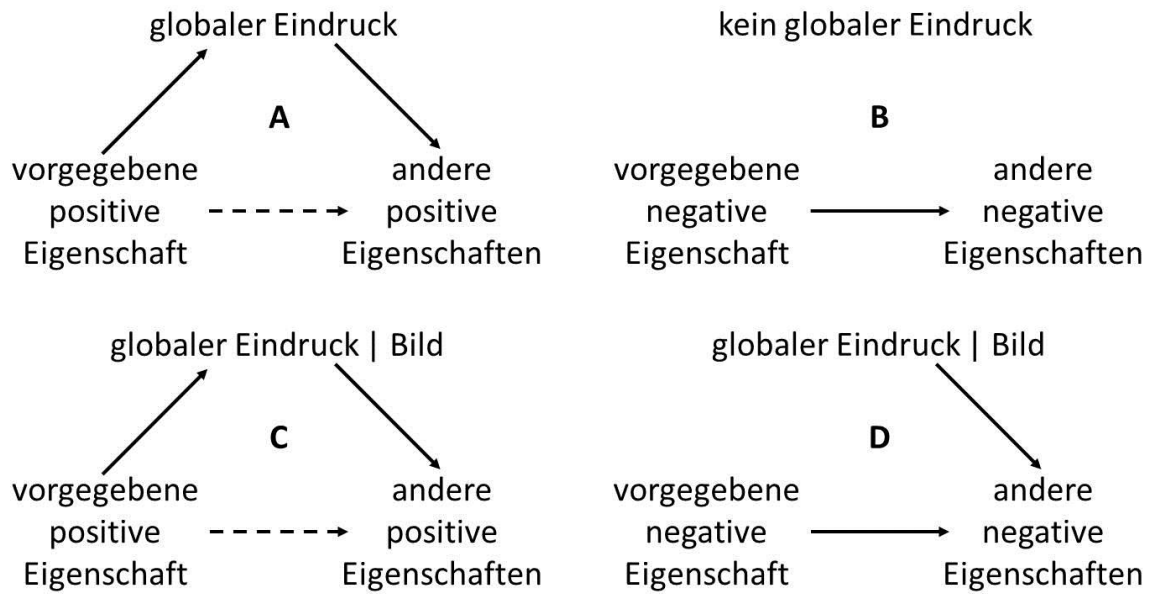


Abbildung 8. Postulierter Einfluss von Bilddarbietungen auf Halo-Effekte

A = positiver Halo-Effekt ohne Bild; B = negativer Halo-Effekt ohne Bild; C = positiver Halo-Effekt mit Bild; D = negativer Halo-Effekt mit Bild

5.2.1 Methode

Teilnehmer

Insgesamt 44 (31 weiblich, 13 männlich; 22 in jeder Zwischensubjekt-Bedingung) Studenten der Universität Heidelberg, die in der Umgebung des Psychologischen Instituts geworben wurden, nahmen an dem Experiment teil. Das Alter der Probanden betrug im Mittel 23.02 Jahre ($SD = 6.07$, $Min = 19$, $Max = 55$). Deutsch war für 41 Teilnehmer die Muttersprache, für zwei Teilnehmer die erste und für einen Teilnehmer die zweite Fremdsprache. Für ihre Teilnahme an der Untersuchung wurden die Probanden mit drei Euro oder einer halben Versuchspersonenstunde vergütet.

Materialien

Der Untersuchung lag ein 2 (Darbietungsart) x 2 (Informationsvalenz) x 2 (Urteilsdimension) x 2 (Gleichheit von dargebotener und beurteilter Dimension) Design zugrunde. Die Darbietungsart wurde zwischen den Versuchsteilnehmern manipuliert und hatte die Stufen *ohne Bild* und *mit Bild*. Informationsvalenz (positiv versus distinkt negativ), Urteilsdimension (Agency versus Communion) und Gleichheit (gleiche versus ungleiche Dimension) wurden als Innersubjekt-Faktoren konzipiert.

Als Items dienten acht Eigenschaftspaare (jeweils vier für Agency und Communion) aus der generierten Datenbasis. Die (nicht distinkt) negative Valenzbedingung wurde nicht erneut betrachtet, da sich in Experiment 1 keine signifikanten Unterschiede zur distinkt negativen Valenzbedingung gezeigt hatten. Ferner wurde im Vergleich zu Experiment 1 auf zwei Eigenschaftspaare verzichtet, um eine zu hohe Anzahl an Durchgängen (da Valenz nun ein Innersubjekt-Faktor war) zu vermeiden. Die Wahl fiel auf die Paare *klug-dumm* sowie *offen-verschlossen*, da diese am ehesten als Synonym der jeweiligen Dimensionen (Agency und Communion) statt als spezifische Attribute erschienen. Alle übrigen verwendeten Attribute waren identisch zu Experiment 1 und sind in Anhang C gekennzeichnet.

Das Bildmaterial wurde einem Datensatz von 102 Fotos aus amerikanischen High-School-Jahrbüchern entnommen. Die Fotos waren von 171 Personen hinsichtlich ihrer Attraktivität beurteilt und bereits erfolgreich in einer Studie von Fiedler und Unkelbach (2011) verwendet worden. Aus dieser Datenbasis wurden sechs attraktive und sechs unattraktive Fotos (jeweils drei Männer und drei Frauen) für die Untersuchung ausgewählt, die hinsichtlich ihres Ausmaßes an Attraktivität beziehungsweise Nichtattraktivität miteinander vergleichbar waren. Die ausgewählten Bilder mit den dazugehörigen deskriptiven Statistiken finden sich in Anhang D.

Die abhängigen Variablen der Untersuchung waren erneut verschiedene Halo-Effekte, die aus den Mittelwerten der verschiedenen Einzelratings berechnet wurden. Die Ratings wurden anhand der folgenden Frage erhoben: „*Eine Person hat die Eigenschaft [...]. Wie wahrscheinlich ist es, dass sie auch folgende Eigenschaft hat: [...]*?“ (1 = *keinesfalls*; 7 = *ganz sicher*). Die Wahrscheinlichkeit jedes Attributs wurde auf Grundlage jedes anderen Attributs der gleichen Valenz beurteilt. Es fand folglich keine Beurteilung eines negativen Attributs auf Basis einer positiven Eigenschaft, oder umgekehrt, statt. Für beide Valenzbedingungen ergaben sich damit 56 (8 x 7) Durchgänge. Insgesamt hatte jede der beiden Zwischensubjekt-Bedingungen 112 Durchgänge.

Prozedur

Die Probanden wurden durch Aushänge oder studentische Hilfskräfte in das Labor geladen. Dort unterschrieben sie eine Einverständniserklärung und begannen danach das Experiment. Die Zuteilung zu einer der beiden Zwischensubjekt-Bedingungen erfolgte zufällig. An jedem Termin nahmen maximal sechs Personen gleichzeitig teil.

In der Instruktion wurde den Probanden, abhängig von der Zwischensubjekt-Bedingung, mitgeteilt, dass sie an einer Bewertungsaufgabe teilnehmen würden, in der ihnen Aussagen beziehungsweise Aussagen und Bilder über verschiedene Menschen präsentiert werden. Gleichzeitig würden bestimmte Persönlichkeitsattribute erscheinen, deren Wahrscheinlichkeit auf Basis der dargebotenen Informationen eingeschätzt werden solle. Die Probanden wurden darauf hingewiesen, sich bei noch offenen Fragen an die Versuchsleitung zu wenden und anderenfalls das Experiment zu starten.

Der anschließende Experimententeil bestand aus 112 Durchgängen, die in einer zufälligen Reihenfolge und ohne Zeitlimit präsentiert wurden. Die Probanden benötigten etwa 20 Minuten für die vollständige Bearbeitung aller Experimentaldurchgänge. In der Zwischensubjekt-Bedingung *mit Bild* wurde in jedem Durchgang zusätzlich zu einer Eigenschaft zufällig eines der zuvor ausgewählten Fotos präsentiert. Positive Eigenschaften wurden dabei stets mit Fotos attraktiver Personen und negativen Eigenschaften immer mit Fotos unattraktiver Personen gepaart. Folglich erschien jedes Bild durchschnittlich in neun bis zehn Experimentaldurchgängen.

Am Ende des Experiments gaben die Probanden ihr Geschlecht sowie ihr Alter an. Zudem wurde erfragt, ob Deutsch die Muttersprache, die erste oder die zweite Fremdsprache sei. Nach Beendigung des Experiments wurden die Teilnehmer über Inhalt und Ziel der Untersuchung aufgeklärt und erhielten ihre Vergütung.

5.2.2 Ergebnisse

Vorab-Analysen

Wie in Experiment 1 wurde vor einer detaillierten Analyse der Daten überprüft, ob die berechneten abhängigen Variablen überhaupt mit Halo-Effekten gleichgesetzt werden können. Ein Urteil wurde dann als Halo-Effekt betrachtet, wenn es sich signifikant vom unteren Ende der Skala (1 = *keinesfalls*) unterschied. Zu diesem Zweck wurden alle abhängigen Variablen mit Hilfe von t-Tests gegen den Wert 1 getestet. In beiden Zwischensubjekt-Bedingungen waren alle abhängigen Variablen beider Valenzen signifikant unterschiedlich zu dem Wert 1 ($p < .050$). Die abhängigen Maße wurden folglich als Halo-Effekte betrachtet und weiteren Analysen unterzogen.

Replizierbarkeit der Befunde des ersten Experiments

Anfangs wurde lediglich die Zwischensubjekt-Bedingung *ohne Bild* betrachtet. Diese war konzeptuell identisch zu Experiment 1 und unterschied sich nur in der Tatsache, dass Valenz nun innerhalb der Versuchsteilnehmer manipuliert wurde. Zur Untersuchung der Replizierbarkeit der Valenzasymmetrie wurde eine 2 x 2 ANOVA (Informationsvalenz x Gleichheit von dargebotener und beurteilter Dimension) mit Messwiederholung durchgeführt. Es wurden signifikante Haupteffekte für Valenz ($F(1, 21) = 46.09, p < .001, \eta^2 = .69$) und Dimensionsgleichheit deutlich ($F(1, 21) = 91.81, p < .001, \eta^2 = .81$): Positive Halo-Effekte ($M = 4.85, SD = 0.50$) waren stärker als negative ($M = 3.94, SD = 0.57$). Auch zeigten sich bei gleicher Dimension von dargebotenem und beurteiltem Attribut ($M = 4.77, SD = 0.55$) größere Halo-Effekte als bei verschiedenen Dimensionen ($M = 4.02, SD = 0.38$). Die Interaktion der beiden Faktoren wurde ebenfalls signifikant ($F(1, 21) = 16.38, p = .001, \eta^2 = .44$): Negative Halo-Effekte bei gleicher ($M = 4.41, SD = 0.68$) und ungleicher Dimension ($M = 3.47, SD = 0.56$) hatten eine größere Differenz zueinander als positive Halo-Effekte bei gleicher ($M = 5.12, SD = 0.57$) und ungleicher Dimension ($M = 4.57, SD = 0.48$).

Zur erneuten Überprüfung der Dimensionshypothese wurde wieder eine 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung vorgenommen. Diese ließ, zusätzlich zum Valenzeffekt, einen signifikanten Haupteffekt des Faktors *Urteilsdimension* erkennen ($F(1, 21) = 19.65, p < .001, \eta^2 = .48$): Halo-Effekte von Communion-Attributen ($M = 4.51, SD = 0.40$) waren stärker als Halo-Effekte von Agency-Eigenschaften ($M = 4.28, SD = 0.50$). Auch die Interaktion von Valenz und Urteilsdimension wurde signifikant ($F(1, 21) = 8.84, p = .007, \eta^2 = .30$): Bei negativer Valenz der Items waren Halo-Effekte von Communion-Eigenschaften ($M = 4.16, SD = 0.62$) stärker als Halo-Effekte von Agency-Eigenschaften ($M = 3.72, SD = 0.59$). Im Gegensatz dazu unterschieden sich positive Halo-Effekte von Agency- ($M = 4.84, SD = 0.59$) und Communion-Eigenschaften ($M = 4.85, SD = 0.51$) nur äußerst geringfügig voneinander.

Die Beeinflussbarkeit der Vorannahmen über Agency- und Communion-Attribute wurde auf Basis einer 2 x 2 ANOVA (Informationsvalenz x Dimension der beurteilten Eigenschaft) mit Messwiederholung ermittelt. Neben dem Valenzeffekt erreichte auch der Haupteffekt für die Dimension der beurteilten Eigenschaft ein signifikantes Maß ($F(1, 21) = 15.10, p = .001, \eta^2 = .42$): Vorannahmen über Communion-Attribute ($M = 4.48, SD = 0.46$) wurden stärker durch Halo-Effekte geprägt als Vorannahmen über Agency-Attribute ($M =$

4.31, $SD = 0.43$). Auch die Interaktion *Informationsvalenz x Dimension der beurteilten Eigenschaft* wurde signifikant ($F(1, 21) = 5.19, p = .033, \eta^2 = .20$): Die Vorannahmen über positive Agency- ($M = 4.85, SD = 0.56$) und Communion-Attribute ($M = 4.84, SD = 0.50$) waren nahezu identisch. Bei negativer Valenz waren die Vorannahmen über Agency-Eigenschaften ($M = 3.77, SD = 0.53$) deutlich schwächer als die Vorannahmen über Communion-Eigenschaften ($M = 4.11, SD = 0.69$).

Einfluss der zusätzlichen Darbietung von Bildern in Bezug auf die Informationsvalenz

Die Beeinflussung positiver und negativer Halo-Effekte durch die zusätzliche Präsentation von Gesichtern wurde mit Hilfe einer $2 \times 2 \times 2$ ANOVA (Darbietungsart \times Informationsvalenz \times Gleichheit von dargebotener und beurteilter Dimension) mit Messwiederholung untersucht. Der Haupteffekt der Darbietungsart wurde nicht signifikant ($F(1, 42) = 1.75, p = .194, \eta^2 = .04$): Bei der Verwendung von Bildern ($M = 4.61, SD = 0.66$) ergaben sich im Durchschnitt keine stärkeren Halo-Effekte als ohne Bilddarbietung ($M = 4.39, SD = 0.44$). Im Gegensatz dazu konnten signifikante Haupteffekte für Valenz ($F(1, 42) = 52.29, p < .001, \eta^2 = .56$) und Dimensionsgleichheit ($F(1, 42) = 145.60, p < .001, \eta^2 = .78$) gezeigt werden: Über beide Zwischensubjekt-Bedingungen hinweg waren positive Halo-Effekte ($M = 4.85, SD = 0.57$) größer als negative ($M = 4.15, SD = 0.73$). Zudem waren Halo-Effekte bei gleicher Dimension ($M = 4.81, SD = 0.61$) von größerem Ausmaß als bei ungleicher Dimension ($M = 4.19, SD = 0.57$).

Ebenfalls signifikant wurde die Interaktion von Darbietungsart und Valenz ($F(1, 42) = 4.40, p = .042, \eta^2 = .10$): Positive Halo-Effekte wurden durch die zusätzliche Darbietung von Bildern ($M = 4.86, SD = 0.64$) im Vergleich zur Darbietung ohne Bilder ($M = 4.85, SD = 0.50$) kaum beeinflusst. Im Gegensatz dazu waren negative Halo-Effekte bei der Verwendung von Bildern ($M = 4.37, SD = 0.82$) größer als ohne Bilder ($M = 3.94, SD = 0.57$). Ein weiteres signifikantes Ergebnis zeigte sich für die Interaktion von Darbietungsart und Dimensionsgleichheit ($F(1, 42) = 6.52, p = .014, \eta^2 = .13$): Halo-Effekte bei ungleicher Dimension waren unter der Verwendung von Bildern ($M = 4.37, SD = 0.68$) stärker als ohne Bilder ($M = 4.02, SD = 0.38$). Bei gleicher Dimension waren die Halo-Effekte unter Bilddarbietung ($M = 4.86, SD = 0.67$) dagegen nur wenig größer als ohne Bilddarbietung ($M = 4.77, SD = 0.55$). Auch die Interaktion von Valenz und Dimensionsgleichheit erreichte ein signifikantes Niveau ($F(1, 42) = 4.87, p = .033, \eta^2 = .10$): Negative Halo-Effekte bei gleicher

($M = 4.50$, $SD = 0.75$) und ungleicher Dimension ($M = 3.80$, $SD = 0.80$) wiesen eine größere Differenz zueinander auf als positive Halo-Effekte bei gleicher ($M = 5.12$, $SD = 0.62$) und ungleicher Dimension ($M = 4.59$, $SD = 0.58$).

Auch die Dreifachinteraktion wurde signifikant ($F(1, 42) = 10.88$, $p = .002$, $\eta^2 = .21$): Bei negativer Valenz und ungleicher Dimension wurden Halo-Effekte am meisten durch die zusätzliche Darbietung von Bildern begünstigt. Einen schwächeren Anstieg gab es bei negativer Valenz und gleicher Dimension. So gut wie kein Unterschied war bei positiver Valenz, unabhängig von der Dimensionsgleichheit, zu verzeichnen. Die Mittelwerte und Standardabweichungen der Dreifachinteraktion sind in Tabelle 4 dargestellt.

Tabelle 4. Einfluss der zusätzlichen Darbietung von Bildern auf Halo-Effekte

	Darbietungsart		Differenz
	ohne Bild	mit Bild	
positive Valenz			
gleiche Dimension	5.12 (0.57)	5.13 (0.67)	0.01
ungleiche Dimension	4.57 (0.48)	4.60 (0.67)	0.03
negative Valenz			
gleiche Dimension	4.41 (0.68)	4.59 (0.82)	0.18
ungleiche Dimension	3.47 (0.56)	4.14 (0.87)	0.67

Einfluss der zusätzlichen Darbietung von Bildern in Bezug auf die Urteilsdimension

In einer weiteren Analyse wurde untersucht, inwiefern die Darbietung von Bildern unterschiedliche Effekte für Agency- und Communion-Attribute aufweisen kann. Hierzu wurde eine $2 \times 2 \times 2$ ANOVA (Darbietungsart \times Informationsvalenz \times Urteilsdimension) mit Messwiederholung durchgeführt. Neben den schon dargestellten Effekten für Valenz und Darbietungsart konnte auch ein signifikanter Haupteffekt der Urteilsdimension aufgezeigt werden ($F(1, 42) = 18.81$, $p < .001$, $\eta^2 = .31$): Halo-Effekte von Communion-Attributen ($M = 4.59$, $SD = 0.54$) waren über beide Zwischensubjekt-Bedingungen hinweg stärker als Halo-Effekte von Agency-Eigenschaften ($M = 4.42$, $SD = 0.62$).

Im Gegensatz dazu gab es keinen signifikanten Effekt für die Interaktion von Darbietungsart und Urteilsdimension ($F(1, 42) = 1.67$, $p = .204$, $\eta^2 = .04$): Halo-Effekte von Agency- ($M = 4.28$, $SD = 0.50$) und Communion-Attributen ($M = 4.51$, $SD = 0.40$) in der

Bedingung ohne Bilder waren nicht unterschiedlicher zueinander als Halo-Effekte von Agency- ($M = 4.55$, $SD = 0.71$) und Communion-Attributen ($M = 4.68$, $SD = 0.64$) in der Bedingung mit Bilddarbietungen. Die Interaktion von Valenz und Urteilsdimension wurde dagegen signifikant ($F(1, 42) = 14.46$, $p < .001$, $\eta^2 = .26$): Bei negativer Valenz waren Halo-Effekte von Communion-Eigenschaften ($M = 4.32$, $SD = 0.71$) größer als Halo-Effekte von Agency-Attributen ($M = 3.98$, $SD = 0.82$). Positive Halo-Effekte von Agency- ($M = 4.85$, $SD = 0.61$) und Communion-Eigenschaften ($M = 4.86$, $SD = 0.58$) unterschieden sich dagegen kaum voneinander. Eine signifikante Interaktion aller drei unabhängigen Variablen konnte nicht gefunden werden ($F(1, 42) = 1.65$, $p = .207$, $\eta^2 = .04$).

5.2.3 Diskussion

In Experiment 2 wurden sowohl qualitative als auch quantitative Unterschiede von Halo-Effekten in Abhängigkeit von Informationsvalenz und Urteilsdimension analysiert. Die Ergebnisse sprechen für das Zutreffen der Valenzhypothese: Positive Halo-Effekte sind stärker als negative Halo-Effekte. Die Befunde von Experiment 1 wurden somit repliziert. Durch den Vergleich der beiden Zwischensubjekt-Bedingungen wurde zudem deutlich, dass positive Halo-Effekte auch globaler sind als negative. Dies wird durch die Tatsache belegt, dass positive und negative Halo-Effekte in unterschiedlicher Weise durch die zusätzliche Darbietung globaler Informationen, hier in Form von Bildern (Farah et al., 1998; Tanaka & Farah, 1993), beeinflusst werden: Negative Halo-Effekte werden durch die zusätzliche Präsentation unattraktiver Gesichter stärker. Umgekehrt führt die Darbietung attraktiver Gesichter allerdings nicht zu größeren positiven Halo-Effekten.

Die Dimensionshypothese kann nur teilweise bestätigt werden: Halo-Effekte von Agency- und Communion-Eigenschaften unterscheiden sich in ihrer Qualität nicht voneinander, da sie im gleichen Maße durch die zusätzliche Darbietung von Bildern beeinflusst werden. Quantitativ sind Halo-Effekte negativer Communion-Attribute im Durchschnitt stärker als Halo-Effekte negativer Agency-Attribute. Im Gegensatz dazu sind positive Agency- und Communion-Halo-Effekte von nahezu gleicher Größe. Der letztere Befund widerspricht der Hypothese, deckt sich jedoch mit dem Ergebnis von Experiment 1, in dem ebenfalls kein signifikanter Einfluss der Urteilsdimension bei positiver Valenz gefunden werden konnte. Die Daten sind eventuell dadurch zu erklären, dass die rein

quantitativen Unterschiede von Agency- und Communion-Halo-Effekten durch die hohe Globalität bei positiver Valenz aufgehoben werden. Das heißt, dass positive Halo-Effekte aufgrund ihrer globalen Natur alle so ähnlich zueinander sind, dass Unterschiede einzelner Eigenschaften vernachlässigbar klein werden. Diese Erklärung ist konform mit den Valenzeffekten sowie der Density Hypothesis (Unkelbach et al., 2008).

Zusätzlich zu den beiden Hypothesen wurde auch die Beeinflussbarkeit von Agency- und Communion-Urteilen erneut überprüft. Auch hier wurden die Ergebnisse von Experiment 1 repliziert: Im negativen Bereich werden Vorannahmen über Communion-Eigenschaften stärker durch Halo-Effekte geprägt als Vorannahmen über Agency-Eigenschaften. Bezogen auf positive Halo-Effekte gibt es dagegen keinen signifikanten Unterschied in der Beeinflussbarkeit von Agency- und Communion-Eigenschaften. Der fehlende Effekt bei positiver Valenz geht vermutlich auf die eben geschilderte hohe Globalität positiver Informationen zurück. Da positive Halo-Effekte in ihrer Stärke alle sehr ähnlich sind, ist auch kein Unterschied in der Beeinflussbarkeit positiver Agency- und Communion-Urteile zu erwarten. Der signifikante Effekt bei negativer Valenz kann möglicherweise dadurch erklärt werden, dass Communion-Eigenschaften allgemein kritischer beurteilt werden als Agency-Eigenschaften. Diese Vermutung muss jedoch in weiteren Experimenten genauer untersucht werden.

5.3 Bedeutung des Abfragemodus: Experiment 3

In den Experimenten 1 und 2 konnte nachgewiesen werden, dass die Qualität und Quantität von Halo-Effekten durch die Valenz der dargebotenen Informationen beeinflusst werden kann. Zudem wurde belegt, dass sich Halo-Effekte von Agency- und Communion-Attributen hinsichtlich ihrer Stärke voneinander unterscheiden. Jedoch sind die beiden Experimente auch mit Einschränkungen verbunden: Erstens wurden als Stimuli lediglich Persönlichkeitseigenschaften verwendet. Dies ist nicht alltagsnah, da in der Realität keine Eigenschaften, sondern vielmehr konkrete Verhaltensweisen beobachtet werden, die auf bestimmte Eigenschaften hinweisen. Die zweite Einschränkung betrifft die randomisierte Abfrage der Urteile. In realen Kontexten erfolgt die Beurteilung unterschiedlicher Eigenschaften von verschiedenen Personen in der Regel nicht zufällig, sondern nach

bestimmten Gesichtspunkten strukturiert. Als dritte Einschränkung ist die alleinige Verwendung von Ratings als abhängige Maße zu nennen. Obwohl sich diese Methode nach eingehender Betrachtung als die beste erwiesen hat, ist es für die Generalisierbarkeit der Ergebnisse von großer Bedeutung, auch ein zusätzliches Maß zu berücksichtigen.

Die genannten Einschränkungen sollten im dritten Experiment umgangen werden: Statt Eigenschaften wurden dieses Mal konkrete Verhaltensweisen als Stimuli präsentiert, auf deren Basis die verschiedenen Urteile erfolgen sollten. Damit wurde auch der zweite Schritt im grundlegenden Modell (vergleiche Abschnitt 3.2) eingeleitet. Des Weiteren wurden Inter-Rating-Korrelationen als zusätzliches abhängiges Maß betrachtet. Da Korrelationen lediglich globale Halo-Effekte erfassen, sollte mit diesem Maß eine weitere Abgrenzung zu spezifischen Halo-Effekten vorgenommen werden. Als dritte Änderung erfolgte die Abfrage der Urteile in diesem Experiment nicht mehr randomisiert, sondern strukturiert nach Personen oder Eigenschaften. Eine Abfrage nach Personen bedeutete hierbei, dass eine Person mit einer bestimmten Verhaltensweise hinsichtlich aller Attribute beurteilt werden muss, bevor die nächste Person und damit die nächste Verhaltensweise präsentiert wird. Bei der Abfrage nach Eigenschaften wurden erst alle Personen, basierend auf ihren jeweiligen Verhaltensweisen, hinsichtlich einer Persönlichkeitseigenschaft bewertet, bevor die nächste Eigenschaft abgefragt wurde.

Für den Abfragemodus wurden unterschiedliche Effekte für positive und negative Valenz vorhergesagt: Die nach Personen gestaffelte Urteilsabfrage sollte im Vergleich zur nach Eigenschaften geordneten Abfrage zu einem Anstieg globaler Halo-Effekte führen, da damit die Bildung eines Gesamteindrucks erleichtert wird (Cooper, 1981b; Johnson & Vidulich, 1956; Sawyer & Raybould, 1981). Spezifische Halo-Effekte sollten gänzlich unabhängig vom Abfragemodus sein, da sie nicht über einen Gesamteindruck vermittelt werden. Da positive Halo-Effekte stets globaler Natur sind, sollten sie nur geringfügig vom Abfragemodus beeinflusst werden. Im Gegensatz dazu sollten negative Halo-Effekte durch eine nach Personen geordnete Abfrage globaler und somit auch stärker werden.

5.3.1 Vorerhebung

Da in der dritten Untersuchung erstmals konkrete Verhaltensweisen anstelle von Persönlichkeitseigenschaften präsentiert werden sollten, war es zunächst erforderlich,

geeignete Stimuli auszuwählen. Wichtig hierbei war, einzelne Verhaltensweisen so konkret zu formulieren, dass sie lediglich reale Rückschlüsse auf eine einzige Eigenschaft zuließen. Würde eine bestimmte Verhaltensweise für mehrere Eigenschaften gleichzeitig sprechen, könnte neben globalen und spezifischen Halo-Effekten die dritte Form von Halo-Effekten entstehen, welche im sogenannten Inadequate Discrimination Model (Lance, LaPointe, & Fisicaro, 1994; Lance, LaPointe, & Stewart, 1994) beschrieben wird. Eine solche zusätzliche Quelle für Halo-Effekte sollte in den Experimenten unbedingt vermieden werden, da sie die Interpretierbarkeit der Ergebnisse erschwert hätte.

Als Grundlage für die Generierung geeigneter Verhaltensweisen wurden die 19 Eigenschaftstripletts aus der Datenbasis der ersten beiden Experimente (siehe Anhang C) verwendet. In allen drei Bedingungen (positiv, negativ und distinkt negativ) wurden alle 19 Eigenschaften einer Stichprobe von jeweils 20 Personen vorgelegt. Die Darbietungen erfolgten computerbasiert und in einer zufälligen Reihenfolge. Aufgabe der Probanden war es, für jede Eigenschaft möglichst viele passende Verhaltensweisen zu generieren. Die Erhebung erfolgte getrennt für die drei Valenzbedingungen, um die Vielfalt der generierten Verhaltensweisen zu erhöhen. Nach der Sichtung der generierten Stimuli, wurden semantisch ähnliche Verhaltensweisen miteinander verbunden und dabei auch die beiden negativen Valenzbedingungen zusammengefasst. Durch diesen Prozess entstanden insgesamt 60 Items in jeweils einer positiv und einer negativ formulierten Version. Die 120 Verhaltensweisen sind in Anhang E ausführlich dargestellt.

Im nächsten Schritt wurden die 60 positiven Verhaltensweisen einer Stichprobe von 34 und die 60 negativen Verhaltensweisen einer Stichprobe von 32 Personen dargeboten. Erneut waren die Darbietungen computerbasiert und erfolgten in randomisierter Reihenfolge. Die Teilnehmer sollten für jede Verhaltensweise einschätzen, welche Charaktereigenschaft sie am ehesten repräsentiert. Mehrfachnennungen wurden hierbei gestattet. In die finale Datenbasis wurde jede Verhaltensweise eingeschlossen, in denen in mindestens 50 Prozent aller Fälle die Eigenschaft genannt wurde, aus der die Verhaltensweise ursprünglich generiert worden war. Gleichzeitig wurde darauf geachtet, dass jede Verhaltensweise nur mit einer einzigen Eigenschaft aus der Datenbasis in Verbindung gebracht wurde. Der endgültige Datensatz umfasste 28 Paare von Verhaltensweisen. Für acht Eigenschaftspaare wurden jeweils drei Paare von Verhaltensweisen gefunden. Vier weitere Eigenschaftspaare waren durch jeweils ein Paar

von Verhaltensweisen repräsentiert. Die vollständige Datenbasis einschließlich der Ergebnisse der Vorerhebung findet sich in Anhang F.

5.3.2 Methode

Teilnehmer

An dem Experiment nahmen 126 (97 weiblich, 29 männlich; 21 in jeder Zwischensubjekt-Bedingung) Studenten der Universität Heidelberg teil. Das Alter der Teilnehmer betrug im Mittel 22.29 Jahre ($SD = 3.16$, $Min = 18$, $Max = 36$). Deutsch war für insgesamt 104 Teilnehmer die Muttersprache, für 14 Teilnehmer die erste und für acht Teilnehmer die zweite Fremdsprache. Für ihr Mitwirken an der Untersuchung erhielten die Probanden drei Euro oder eine halbe Versuchspersonenstunde.

Materialien

Das Experiment basierte auf einem 3 (Darbietungsart) x 2 (Abfragemodus) x 2 (Informationsvalenz) x 2 (Urteilsdimension) Design. Die Darbietungsart wurde als Zwischensubjekt-Variable mit den Stufen *eine Verhaltensweise*, *drei Verhaltensweisen* und *drei Verhaltensweisen und Bilder* konzipiert. Als zweiter Zwischensubjekt-Faktor diente der Abfragemodus (nach Personen versus nach Eigenschaften geordnet). Informationsvalenz (positiv versus negativ) und Urteilsdimension (Agency versus Communion) wurden als Innersubjekt-Faktoren festgelegt.

Als Items wurden 24 Paare gegensätzlicher Verhaltensweisen (jeweils zwölf für Agency und Communion) aus der Datenbasis verwendet. Jeweils drei Verhaltensweisen konnten einer Eigenschaft zugeordnet werden, so dass insgesamt acht positive und acht negative Eigenschaften (jeweils vier für Agency und Communion) repräsentiert waren. Die verwendeten Items sind in Anhang F gekennzeichnet.

Das Bildmaterial entstammte dem gleichen Datensatz wie das Bildmaterial aus Experiment 2. Dieses Mal wurden acht attraktive (drei Männer und fünf Frauen) und acht unattraktive (fünf Männer und drei Frauen) Fotos für das Experiment ausgewählt. Erneut waren alle Bilder hinsichtlich ihres Ausmaßes an Attraktivität beziehungsweise Nichtattraktivität vergleichbar. Die ausgewählten Gesichter mit den dazugehörigen deskriptiven Statistiken finden sich in Anhang D.

Als abhängige Variablen wurden verschiedene Halo-Effekte verwendet, die aus den Mittelwerten diverser Einzelratings berechnet wurden. Die Einzelratings erfolgten auf acht verschiedenen bipolaren Skalen, die von -3 bis 3 (zum Beispiel „*abhängig ... selbständig*“) reichten. In jedem Durchgang wurde folgende Frage gestellt: „*Wie würden Sie diese Person auf folgender Dimension einschätzen?*“. Die acht bipolaren Skalen entsprachen den acht Eigenschaftspaaren, die durch die dargebotenen Verhaltensweisen repräsentiert waren (siehe Anhang F). Jede der 16 Eigenschaften (je nach Bedingung repräsentiert durch eine Verhaltensweise oder drei Verhaltensweisen) wurde auf allen acht Skalen beurteilt. Damit hatte jede Zwischensubjekt-Bedingung 128 (16 x 8) Durchgänge. Im Gegensatz zu den vorhergehenden Experimenten wurde jede Verhaltensweise dieses Mal auch auf der Skala bewertet, auf deren Basis sie in der Voruntersuchung generiert worden war. Beispielsweise musste die Verhaltensweise ... *sagt anderen immer die Wahrheit*, welche die Eigenschaft *ehrlich* repräsentierte, auch auf der Skala *verlogen ... ehrlich* beurteilt werden. Diese eindeutigen Urteile dienten jedoch nur der Kontrolle der Items und wurden nicht in die Berechnung der Halo-Effekte einbezogen.

Prozedur

Die Probanden wurden auf dem Gelände des Psychologischen Instituts sowie der Mensa geworben. Hierbei wurde in einem gesonderten Raum über der Mensa ein zusätzliches Untersuchungslabor eingerichtet. Zu Beginn jeder Sitzung unterschrieben die Teilnehmer eine Einverständniserklärung und begannen danach das Experiment. Dabei wurden sie zufällig einer der sechs Zwischensubjekt-Bedingungen zugeteilt. Gleichzeitig konnten maximal sechs Personen an einem Termin teilnehmen.

In der Instruktion wurde den Probanden mitgeteilt, dass sie im Folgenden an einer Bewertungsaufgabe teilnehmen würden, in der ihnen insgesamt 16 verschiedene Personen dargeboten werden, die jeweils eine bestimmte Verhaltensweise (in der anderen Zwischensubjekt-Bedingung drei bestimmte Verhaltensweisen) zeigen. Diese Personen würden auf acht Skalen unterschiedlicher Persönlichkeitsdimensionen bewertet werden müssen. Die Probanden wurden darauf hingewiesen, sich bei noch offenen Fragen an die Versuchsleitung zu wenden und anderenfalls das Experiment zu starten.

Der anschließende Experimentalteil umfasste 128 Durchgänge, die ohne Zeitbeschränkung präsentiert wurden. Die Teilnehmer benötigten im Durchschnitt 25

Minuten für die vollständige Bearbeitung des Experimentalteils. Jede der 16 fiktiven Personen (bezeichnet als Person 1 bis Person 16) wurde zufällig mit einer der 16 Eigenschaften kombiniert. In der Bedingung *drei Verhaltensweisen* war jede der 16 Eigenschaften durch alle drei Verhaltensweisen, in der Bedingung *eine Verhaltensweise* dagegen nur durch eine zufällige Verhaltensweise aus dem entsprechenden Triplet repräsentiert. In der Bedingung *drei Verhaltensweisen und Bilder* wurde jede Person zudem zufällig einem der 16 Fotos zugeordnet. Personen mit positiven Verhaltensweisen konnten jedoch nur mit attraktiven Bildern und Personen mit negativen Verhaltensweisen nur mit unattraktiven Bildern gekoppelt werden.

In der Abfragebedingung *nach Personen geordnet* wurde erst eine Person auf allen acht Skalen bewertet, bevor die nächste Person präsentiert wurde. Die Reihenfolge der Personen war randomisiert. Auch innerhalb jeder Person erfolgte die Abfrage der acht Skalen in einer zufälligen Reihenfolge. In der Bedingung *nach Eigenschaften geordnet* wurden zuerst alle 16 Personen auf einer Skala bewertet, bevor die nächste Skala dargeboten wurde. Die Reihenfolge der Skalen sowie die Reihenfolge der Personen innerhalb jeder Skala waren randomisiert.

Am Ende der Untersuchung gaben die Probanden ihr Geschlecht sowie ihr Alter an. Zudem wurde erfragt, ob Deutsch die Muttersprache, die erste oder die zweite Fremdsprache sei. Nach Beendigung des Experiments wurden die Teilnehmer über Inhalt und Ziel der Untersuchung aufgeklärt und erhielten ihre Vergütung.

5.3.3 Ergebnisse

Vorab-Analysen

Vor der Datenanalyse wurden zunächst die verwendeten Verhaltensweisen einer genaueren Betrachtung unterzogen. Zwar war schon vorgetestet worden, für welche Eigenschaften die Verhaltensweisen stehen, jedoch wurde dies noch einmal für alle Probanden einzeln untersucht. Zu diesem Zweck wurde für jede Verhaltensweise das Urteil auf der Dimension betrachtet, aus der die Verhaltensweise ursprünglich generiert worden war (zum Beispiel die Bewertung der Verhaltensweise ... *sagt anderen immer die Wahrheit* auf der Skala *verlogen* ... *ehrlich*). Die Bewertungen sollten bei positiven Verhaltensweisen unbedingt im positiven Bereich (Skalenwert 1 bis 3), bei negativen Verhaltensweisen im

negativen Bereich (Skalenwert -3 bis -1) sein. Nur so konnte gewährleistet werden, dass die Verhaltensweise für den Probanden auch die erwartete Eigenschaft widerspiegelte. Die Bedingung war bei allen 126 Teilnehmern für alle Verhaltensweisen erfüllt. Folglich konnte die komplette Stichprobe in alle folgenden Analysen einbezogen werden.

Als nächstes wurden die beiden Darbietungsbedingungen *eine Verhaltensweise* und *drei Verhaltensweisen*, jeweils für den Abfragemodus nach Personen beziehungsweise Eigenschaften, miteinander verglichen. Dazu wurden t-Tests für alle abhängigen Variablen durchgeführt. Für keine der abhängigen Variablen konnte ein signifikanter Unterschied zwischen den beiden Bedingungen gefunden werden ($ps > .050$). Dies galt sowohl für die getrennte Betrachtung der beiden Abfragemodi als auch über beide Abfragemodi hinweg. Aus diesem Grund wurden die beiden Darbietungsbedingungen zusammengefasst. Für die folgenden statistischen Analysen wurde die Darbietungsart daher in die Bedingungen *ohne Bild* (ehemals *eine Verhaltensweise* und *drei Verhaltensweisen*) und *mit Bild* (ehemals *drei Verhaltensweisen und Bilder*) unterteilt.

Im dritten Schritt der Vorab-Analyse wurde überprüft, ob die abhängigen Variablen mit Halo-Effekten gleichgesetzt werden können. Ein Urteil wurde dann als Halo-Effekt betrachtet, wenn es sich signifikant vom Mittelpunkt der Skala unterschied. Zu diesem Zweck wurden alle abhängigen Variablen anhand von t-Tests gegen den Wert 0 getestet. In den drei Zwischensubjekt-Kombinationen *Darbietung ohne Bild / Abfrage nach Personen*, *Darbietung ohne Bild / Abfrage nach Eigenschaften* und *Darbietung mit Bild / Abfrage nach Personen* unterschied sich die abhängige Variable *negativer Halo-Effekt bei ungleicher Dimension* nicht signifikant vom Wert 0 ($ps > .050$). Alle anderen abhängigen Variablen, insbesondere die positiven und negativen Gesamt-Halo-Effekte, zeigten in allen Zwischensubjekt-Bedingungen signifikante Unterschiede zum Wert 0. Aufgrund dieser Befunde wurde der Faktor *Gleichheit von dargebotener und beurteilter Dimension* (gleiche versus ungleiche Dimension) in diesem Experiment nicht betrachtet.

Da positive und negative Halo-Effekte auf der gleichen Skala erhoben wurden, mussten vor der Überprüfung der Hypothesen alle Ratings, die auf negativen Verhaltensweisen basierten, durch eine Multiplikation mit -1 umkodiert werden. Auf diese Weise war es möglich, die Stärke positiver und negativer Halo-Effekte, unabhängig von den offensichtlichen Valenzunterschieden, miteinander vergleichen zu können.

Einflüsse von Informationsvalenz und Urteilsdimension

Zunächst wurde überprüft, ob sich die in den letzten Experimenten aufgetretenen Effekte auch für das neue Untersuchungsdesign zeigen. Dazu wurden die beiden Innersubjekt-Faktoren einer 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung unterzogen. Es ergab sich ein signifikanter Haupteffekt für Valenz ($F(1, 125) = 88.94, p < .001, \eta^2 = .42$): Positive Halo-Effekte ($M = 0.82, SD = 0.46$) waren über alle Zwischensubjekt-Bedingungen hinweg stärker als negative Halo-Effekte ($M = 0.41, SD = 0.32$). Der Haupteffekt des Faktors *Urteilsdimension* erreichte kein signifikantes Maß ($F(1, 125) = 3.16, p = .078, \eta^2 = .03$): Halo-Effekte von Communion- ($M = 0.65, SD = 0.34$) und Agency-Verhaltensweisen ($M = 0.59, SD = 0.38$) waren im Mittel gleich stark. Auch die Interaktion von Valenz und Urteilsdimension wurde nicht signifikant ($F(1, 125) = 0.99, p = .321, \eta^2 = .01$). Da diese Ergebnisse nicht den Erwartungen entsprachen, wurden Agency- und Communion-Halo-Effekte anschließend, getrennt für beide Valenzbedingungen, mit Hilfe von t-Tests miteinander verglichen: Positive Halo-Effekte von Agency- ($M = 0.81, SD = 0.58$) und Communion-Verhaltensweisen ($M = 0.83, SD = 0.49$) unterschieden sich nicht signifikant voneinander ($t(125) = 0.51, p = .609$). In der negativen Valenzbedingung waren Halo-Effekte von Communion-Verhaltensweisen ($M = 0.46, SD = 0.38$) gegenüber Halo-Effekten von Agency-Verhaltensweisen ($M = 0.37, SD = 0.41$) von signifikant stärkerem Ausmaß ($t(125) = 2.10, p = .038$).

Einfluss der Darbietungsart

Zur Analyse des Einflusses der Darbietungsart auf die verschiedenen Halo-Effekte wurde eine 2 x 2 x 2 ANOVA (Darbietungsart x Informationsvalenz x Urteilsdimension) mit Messwiederholung durchgeführt. Die Haupteffekte von Informationsvalenz und Urteilsdimension sowie die Interaktion der beiden Faktoren waren aufgrund des neu hinzugekommenen Zwischensubjekt-Faktors nicht identisch mit den bereits geschilderten Ergebnissen. Da die Unterschiede jedoch lediglich geringfügiger Natur waren, wurden die drei Effekte nicht erneut betrachtet. Die ANOVA wies keinen signifikanten Haupteffekt des Faktors *Darbietungsart* auf ($F(1, 124) = 3.21, p = .076, \eta^2 = .03$): Unter der Verwendung von Bildern ($M = 0.69, SD = 0.35$) zeigten sich im Durchschnitt keine stärkeren Halo-Effekte als ohne Bilddarbietung ($M = 0.58, SD = 0.29$). Auch wurde keine signifikante Interaktion von Darbietungsart und Urteilsdimension deutlich ($F(1, 124) < 0.01, p = .993, \eta^2 < .01$): Die

durchschnittlichen Halo-Effekte von Agency- ($M = 0.55$, $SD = 0.37$) und Communion-Verhaltensweisen ($M = 0.61$, $SD = 0.31$) in der Bedingung ohne Bilder waren nicht unterschiedlicher zueinander als die Halo-Effekte von Agency- ($M = 0.66$, $SD = 0.40$) und Communion-Verhaltensweisen ($M = 0.72$, $SD = 0.39$) in der Bedingung mit Bildern.

Es ergab sich keine signifikante Interaktion der Faktoren *Darbietungsart* und *Informationsvalenz* ($F(1, 124) = 0.55$, $p = .459$, $\eta^2 < .01$). Dieser Befund widersprach der Vorhersage und dem bereits in Experiment 2 aufgezeigten Ergebnis. Aus diesem Grund wurden sowohl die positiven als auch die negativen Halo-Effekte zwischen den beiden Darbietungsbedingungen verglichen. Hierfür wurden t-Tests bei unabhängigen Stichproben durchgeführt: Positive Halo-Effekte waren ohne ($M = 0.80$, $SD = 0.43$) und mit der Verwendung von Bildmaterial ($M = 0.87$, $SD = 0.52$) nicht von signifikant unterschiedlicher Größe ($t(124) = 0.83$, $p = .411$). Dagegen waren negative Halo-Effekte bei zusätzlicher Bilddarbietung ($M = 0.51$, $SD = 0.36$) im Vergleich zur Darbietung ohne Bilder ($M = 0.37$, $SD = 0.29$) signifikant größer ($t(124) = 2.35$, $p = .021$). Bei der ANOVA wurde die Interaktion aller drei Faktoren nicht signifikant ($F(1, 124) < 0.01$, $p = .963$, $\eta^2 < .01$).

Einfluss des Abfragemodus

Der Einfluss des Abfragemodus wurde mit einer $2 \times 2 \times 2$ ANOVA (Abfragemodus \times Informationsvalenz \times Urteilsdimension) mit Messwiederholung untersucht. Erneut unterschieden sich die Haupteffekte von Informationsvalenz und Urteilsdimension und die Interaktion der beiden Faktoren nur geringfügig von den bereits geschilderten Ergebnissen und wurden daher nicht noch einmal betrachtet. Es konnte kein signifikanter Haupteffekt des Abfragemodus gezeigt werden ($F(1, 124) = 0.47$, $p = .496$, $\eta^2 < .01$): Die nach Personen ($M = 0.60$, $SD = 0.34$) und die nach Eigenschaften geordnete Urteilsabfrage ($M = 0.64$, $SD = 0.30$) führten nicht zu signifikant unterschiedlichen Halo-Effekten. Auch die Interaktion von Abfragemodus und Urteilsdimension wurde nicht signifikant ($F(1, 124) = 0.09$, $p = .768$, $\eta^2 < .01$): Halo-Effekte von Agency- ($M = 0.57$, $SD = 0.36$) und Communion-Verhaltensweisen ($M = 0.63$, $SD = 0.38$) bei der nach Personen gestaffelten Abfrage waren nicht unterschiedlicher zueinander als Agency- ($M = 0.61$, $SD = 0.40$) und Communion-Halo-Effekte ($M = 0.66$, $SD = 0.31$) bei der nach Eigenschaften geordneten Abfrage.

Die Interaktion von Abfragemodus und Valenz erreichte ein signifikantes Maß ($F(1, 124) = 11.52$, $p = .001$, $\eta^2 = .09$): Positive Halo-Effekte waren bei der nach Personen

geordneten Abfrage ($M = 0.87$, $SD = 0.52$) von größerem Ausmaß als bei der nach Eigenschaften geordneten Abfrage ($M = 0.77$, $SD = 0.40$). Umgekehrt waren negative Halo-Effekte bei der nach Personen geordneten Abfrage ($M = 0.33$, $SD = 0.28$) schwächer als bei der nach Eigenschaften geordneten Abfrage ($M = 0.50$, $SD = 0.34$). Der letztere Befund war das genaue Gegenteil des vorhergesagten Ergebnisses. Deswegen wurden anschließend sowohl die positiven als auch die negativen Halo-Effekten anhand von t-Tests zwischen den beiden Abfragebedingungen verglichen: Der Unterschied positiver Halo-Effekte zwischen den beiden Abfragebedingungen war, wie erwartet, nicht signifikant ($t(124) = 1.24$, $p = .217$). Allerdings zeigte sich ein signifikanter Unterschied für negative Halo-Effekte, der in dieser Richtung nicht erwartet wurde ($t(124) = 3.24$, $p = .002$). Es wurde keine signifikante Dreifachinteraktion bei der ANOVA deutlich ($F(1, 124) < 0.01$, $p = .971$, $\eta^2 < .01$).

Beeinflussbarkeit von Agency- und Communion-Urteilen

Wie in den Experimenten 1 und 2 wurde auch in dieser Untersuchung überprüft, in welchem Maße sich die Beurteilung unbekannter Agency- und Communion-Eigenschaften durch Halo-Effekte prägen lässt. Dazu wurde eine 2 x 2 ANOVA (Informationsvalenz x Dimension der beurteilten Eigenschaft) mit Messwiederholung berechnet. Der signifikante Haupteffekt für Valenz wurde bereits bei der Analyse von Informationsvalenz und Urteilsdimension genauer betrachtet. Es konnte kein signifikanter Haupteffekt für die Dimension der beurteilten Eigenschaft gezeigt werden ($F(1, 125) = 0.19$, $p = .660$, $\eta^2 < .01$): Vorannahmen über Communion- ($M = 0.61$, $SD = 0.35$) und Agency-Attribute ($M = 0.62$, $SD = 0.33$) wurden im Durchschnitt im gleichen Maße durch Halo-Effekte geprägt. Jedoch wurde die Interaktion von Informationsvalenz und Dimension der beurteilten Eigenschaft signifikant ($F(1, 125) = 6.14$, $p = .015$, $\eta^2 = .05$): Positive Halo-Effekte konnten Agency-Attribute ($M = 0.85$, $SD = 0.49$) stärker prägen als Communion-Attribute ($M = 0.79$, $SD = 0.49$). Negative Halo-Effekte hatten dagegen auf Agency-Attribute ($M = 0.39$, $SD = 0.34$) geringere Einflüsse als auf Communion-Attribute ($M = 0.44$, $SD = 0.40$).

Inter-Rating-Korrelationen in Abhängigkeit von Informationsvalenz und Urteilsdimension

In den ersten beiden Experimenten wurden lediglich Mittelwerte verschiedener Ratings als abhängige Maße verwendet. Zur zusätzlichen Untermauerung der bisher gezeigten Ergebnisse wurden in diesem Experiment zusätzlich Inter-Rating-Korrelationen

untersucht. Diese wurden, ähnlich wie auch die Ratings, in mehreren Schritten berechnet. Da die Statistik-Software *SPSS* keine geeigneten Funktionen für die notwendigen Berechnungen bot, wurden alle Daten für diesen Zweck in das Kalkulationsprogramm *Microsoft Excel* eingelesen und über eigens programmierte Funktionen analysiert:

Wie bereits geschildert, wurden in der Untersuchung 16 verschiedene Attribute (die je nach Zwischensubjekt-Bedingung jeweils in Form einer oder dreier Verhaltensweisen dargeboten wurden) auf jeweils acht Skalen beurteilt. Eines dieser Urteile diente lediglich der Kontrolle der Items und war daher nicht Teil der weiteren Analysen. Somit blieben für jedes Attribut sieben verschiedene Ratings. Bei jedem Attribut wurden die paarweisen Korrelationen aller Ratings, also insgesamt 21 Korrelationen, berechnet. Über alle Attribute hinweg wurden folglich 336 (16 x 21) Korrelationen ermittelt.

Im folgenden Schritt wurden die 21 Korrelationskoeffizienten jedes Attributs gemittelt. Zu diesem Zweck wurden die Korrelationen mit Hilfe der sogenannten Fisher Z-Transformation (Fisher, 1918) zunächst in Z-Werte umgerechnet. Diese haben gegenüber Korrelationen den Vorteil, dass sie Variablenzusammenhänge auf einer Intervallskala abbilden und ihre Mittelwerte und Varianzen folglich interpretierbar sind (Bortz, 2005). Durch die Mittelwertberechnung ergab sich ein durchschnittlicher Z-Wert für jedes der 16 Attribute. Anschließend wurden die Z-Werte der Attribute, die hinsichtlich Valenz und Dimension gleich waren, noch einmal gemittelt. Die vier daraus entstandenen Z-Werte (jeweils einer für positive Agency-, positive Communion-, negative Agency- und negative Communion-Attribute) wurden wieder in Inter-Rating-Korrelationen zurückübersetzt. Tabelle 5 zeigt eine Übersicht über die verschiedenen Korrelationen sowohl getrennt nach den Zwischensubjekt-Bedingungen als auch über alle Bedingungen hinweg.

Tabelle 5. Inter-Rating-Korrelationen in Abhängigkeit von Valenz und Dimension

	Abfrage nach Personen				Abfrage nach Eigenschaften				gesamt	
	ohne Bild		mit Bild		ohne Bild		mit Bild			
	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.
Agency	.392	.144	.462	.153	.332	.105	.418	.131	.363	.140
Communion	.388	.119	.450	.162	.319	.096	.429	.151	.367	.120
gesamt	.390	.132	.456	.158	.325	.100	.424	.141	.365	.130

Die deskriptiven Ergebnisse zeigten, dass die Inter-Rating-Korrelationen bei der Darbietung positiver Informationen ($r = .365$) durchschnittlich höher waren als bei der Darbietung negativer Informationen ($r = .130$). Dieses Muster ergab sich für beide Urteilsdimensionen sowie über alle Zwischensubjekt-Bedingungen hinweg. Für beide Valenzbedingungen wurde zudem deutlich, dass die Dimension der dargebotenen Informationen nur zu geringfügigen Unterschieden in den Inter-Rating-Korrelationen führte ($r = .363$ bei positiven Agency- und $r = .367$ bei positiven Communion-Verhaltensweisen, $r = .140$ bei negativen Agency- und $r = .120$ bei negativen Communion-Verhaltensweisen). Im Mittel ging die Darbietung mit Bild ($r = .440$ bei positiver und $r = .150$ bei negativer Valenz) mit leicht höheren Inter-Rating-Korrelationen einher als die Darbietung ohne Bild ($r = .360$ bei positiver und $r = .115$ bei negativer Valenz). Auch waren die Inter-Rating-Korrelationen bei der nach Personen geordneten Abfrage ($r = .425$ bei positiver und $r = .145$ bei negativer Valenz) im Mittel etwas höher als bei der nach Eigenschaften geordneten Abfrage ($r = .375$ bei positiver und $r = .120$ bei negativer Valenz).

Abschließend wurde überprüft, ob sich die Korrelationen auch signifikant voneinander unterscheiden. Hierfür wurde eine Formel von Steiger (1980) für den Innersubjekt-Vergleich zweier Korrelationen r_{ab} und r_{cd} verwendet, die ab einer Stichprobengröße von 20 Probanden gültig ist. Für den Signifikanztest wurden erneut Z-Werte betrachtet, da diese intervallskaliert, annähernd normalverteilt und somit interpretierbar sind (Bortz, 2005). Neben den zu vergleichenden Korrelationen r_{ab} und r_{cd} war es für den Test auch notwendig, die Korrelationen r_{ac} , r_{ad} , r_{bc} und r_{bd} zu kennen. Diese konnten ebenfalls aus den Mittelwerten der verschiedenen Einzelkorrelationen berechnet werden. Die Ermittlung der Einzelkorrelationen selbst war abhängig von der jeweiligen Hypothese. Um etwa die durchschnittlichen Inter-Rating-Korrelationen in Bezug auf die Informationsvalenz vergleichen zu können, mussten alle sechs Korrelationskoeffizienten separat für jede Urteils Paarung (zum Beispiel die Urteile auf den Skalen *faul ... fleißig* und *eigenbrötlerisch ... gesellig*) jedes gegensätzlichen Attributpaars (zum Beispiel *ehrlich-verlogen*, repräsentiert durch passende Verhaltensweisen) ermittelt werden. Entsprechend des bereits geschilderten Vorgehens wurden danach Mittelwerte für jeden der sechs Koeffizienten (getrennt nach Urteilsdimensionen sowie im Gesamten) berechnet. Ein Beispiel für die verschiedenen Einzelkorrelationen ist in Tabelle 6 gegeben.

Tabelle 6. Beispielhafte Übersicht über die berechneten Einzelkorrelationen

	präsentierte Verhaltensweise 1: ... sagt anderen immer die Wahrheit (ehrlich)		präsentierte Verhaltensweise 2: ... sagt anderen selten die Wahrheit (verlogen)	
	Urteil a: <i>faul ... fleißig</i>	Urteil b: <i>eigenbrötlerisch ... gesellig</i>	Urteil c: <i>faul ... fleißig</i>	Urteil d: <i>eigenbrötlerisch ... gesellig</i>
a	1	r_{ab}	r_{ac}	r_{ad}
b		1	r_{bc}	r_{bd}
c			1	r_{cd}
d				1

Die Hypothesentests erfolgten, gemäß der Vorgabe von Steiger (1980), unter der Verwendung der standardnormalverteilten Prüfgröße z . Zwischen positiven ($r = .363$) und negativen Agency-Informationen ($r = .140$) zeigten sich keine signifikant unterschiedlichen Inter-Rating-Korrelationen ($z(123) = 1.88, p = .060$). Der Unterschied zwischen positiven ($r = .367$) und negativen Communion-Attributen ($r = .120$) wurde dagegen signifikant ($z(123) = 2.09, p = .036$). Auch über beide Urteilsdimensionen hinweg führten positive ($r = .365$) und negative Verhaltensweisen ($r = .130$) zu signifikant unterschiedlichen Inter-Rating-Korrelationen ($z(123) = 1.99, p = .047$). Wurden statt der Gesamtstichprobe nur einzelne Zwischensubjekt-Bedingungen betrachtet, waren all die genannten Unterschiede aufgrund der geringeren Stichprobengröße nicht mehr signifikant. Auf Signifikanztests bezüglich der anderen unabhängigen Variablen (Urteilsdimension, Darbietungsart und Abfragemodus) wurde verzichtet, da bereits die deskriptiven Daten deutlich geringere Unterschiede der Inter-Rating-Korrelationen als beim Faktor *Informationsvalenz* erkennen ließen. Somit konnten signifikante Ergebnisse vorab ausgeschlossen werden.

5.3.4 Diskussion

Gegenstand von Experiment 3 war die Analyse von Halo-Effekten als Funktion von Valenz und Urteilsdimension. Im Vergleich zu den vorherigen Experimenten wurde das Untersuchungsdesign in einigen Aspekten verändert: Statt Eigenschaften wurden Verhaltensweisen dargeboten. Die Abfrage erfolgte nicht mehr randomisiert, sondern

geordnet nach zu beurteilenden Personen oder abhängigen Skalen. Zudem wurden Inter-Rating-Korrelationen als abhängiges Maß hinzugezogen. Für die Befunde spielte es keine Rolle, ob ein Attribut durch eine oder drei Verhaltensweisen repräsentiert war.

Die Valenzhypothese kann erneut belegt werden: Positive Halo-Effekte sind stärker und globaler als negative Halo-Effekte. Der qualitative Unterschied wird einerseits dadurch deutlich, dass zusätzliche Bilddarbietungen zwar negative, nicht aber positive Halo-Effekte verstärken können. Ein weiterer Beleg findet sich anhand der signifikant unterschiedlichen Inter-Rating-Korrelationen von positiven und negativen Informationen, da dieses Maß ausschließlich globale Halo-Effekte erfasst. Um signifikante Korrelationsunterschiede zu zeigen, ist jedoch eine hinreichend große Stichprobe notwendig.

Ein überraschender Befund in Bezug auf Valenzunterschiede zeigte sich für den Abfragemodus. Die nach Personen gestaffelte Urteilsabfrage sollte die Entstehung globaler Halo-Effekte begünstigen (Cooper, 1981b; Johnson & Vidulich, 1956; Sawyer & Raybould, 1981) und negative Halo-Effekte daher stärker vergrößern als positive Halo-Effekte. Wie auch erwartet hatte der Abfragemodus keinen signifikanten Einfluss auf positive Halo-Effekte. Der Einfluss auf negative Halo-Effekte war zwar signifikant, ging jedoch in eine unerwartete Richtung: Eine nach Personen geordnete Abfrage führt zu einem Nachlassen negativer Halo-Effekte gegenüber der nach Eigenschaften geordneten Abfrage. Dies ist vermutlich darauf zurückzuführen, dass auch in dieser Bedingung globale positive Halo-Effekte wirken. Eine nach Personen geordnete Abfrage scheint also generell positive Halo-Effekte zu begünstigen, wohingegen negative Halo-Effekte unter diesen Voraussetzungen nicht global wirken können. Die Befunde lassen darauf schließen, dass positive Halo-Effekte so global sind, dass sie auch bei eigentlich negativen Stimuli wirken können.

Der Einfluss der Urteilsdimension kann für die negative, aber nicht die positive Valenzbedingung belegt werden. Bei negativer Valenz sind Halo-Effekte von Communion-Verhaltensweisen stärker als Halo-Effekte von Agency-Verhaltensweisen. Positive Agency- und Communion-Halo-Effekte unterscheiden sich jedoch nicht signifikant voneinander. Die Ergebnisse sind konform mit denen der ersten beiden Experimente. Somit wird auch die Vermutung gestützt, dass Unterschiede verschiedener positiver Halo-Effekte aufgrund der hohen Globalität positiver Informationen aufgehoben werden. Die Befunde, insbesondere die ähnlich hohen Korrelationen, belegen zudem, dass sich Agency- und Communion-Halo-Effekte zwar quantitativ, nicht jedoch qualitativ voneinander unterscheiden.

Bei der Beeinflussbarkeit von Agency- und Communion-Urteilen konnte erneut repliziert werden, dass Vorannahmen über negative Communion-Eigenschaften stärker durch Halo-Effekte geprägt werden als Vorannahmen über negative Agency-Eigenschaften. Im Gegensatz zu den ersten beiden Experimenten wurde dieses Mal jedoch deutlich, dass Agency-Urteile bei positiver Valenz stärker beeinflusst werden als Communion-Urteile. Die Gesamtheit der beiden Ergebnisse lässt somit vermuten, dass Agency-Attribute allgemein positiver betrachtet werden als Communion-Attribute. Der Unterschied gegenüber den ersten beiden Studien ist möglicherweise mit der Verwendung einer bi- anstelle einer unipolaren Skala zu erklärbar. Ob Communion-Eigenschaften wirklich kritischer betrachtet werden als Agency-Eigenschaften muss in einem gesonderten Experiment und unabhängig von der Darbietung jeglicher Informationen betrachtet werden.

5.4 Post-Hoc-Untersuchung zur Baseline von Vorannahmen

Neben den Einflüssen von Informationsvalenz und Urteilsdimension wurde in allen drei Experimenten erfasst, wie stark sich die Vorannahmen über Agency- und Communion-Eigenschaften durch positive und negative Halo-Effekte prägen lassen. Hierbei zeigte sich, dass sich Communion-Urteile stärker negativ prägen lassen als Agency-Urteile. Bei positiver Valenz variierte das Ergebnis in Abhängigkeit von der verwendeten Skala: Gaben die Probanden ihre Urteile auf einer unipolaren Skala ab, gab es keine signifikanten Dimensionsunterschiede. Unter der Verwendung bipolarer Skalen wurden Agency-Urteile stärker durch positive Halo-Effekte beeinflusst als Communion-Attribute.

Die Ergebnisse lassen sich nicht nur durch unterschiedlich starke Halo-Effekte erklären. Wie die Experimente 1 bis 3 zeigen, sind positive Halo-Effekte von Agency- und Communion-Attributen in etwa gleich stark. Negative Halo-Effekte sind bei Communion-Attributen stärker als bei Agency-Attributen. Wenn sich die dargebotene und die zu beurteilende Dimension unterscheiden, sind Halo-Effekte im Mittel schwächer als bei Dimensionsgleichheit. Die bisherigen Befunde zeigen jedoch auch, dass keine qualitativen Unterschiede zwischen Agency- und Communion-Halo-Effekten existieren. Das bedeutet, dass Urteile, die auf Basis von Agency-Items getroffen werden, nicht ähnlicher zueinander sein sollten als Urteile, die auf Grundlage von Communion-Eigenschaften erfolgen. Werden

diese Prinzipien als gegeben betrachtet, sollten sich eigentlich keine Unterschiede in der Beeinflussbarkeit von Agency- und Communion-Urteilen ergeben. Diese Herleitung ist in Abbildung 9 mit Hilfe fiktiver Werte veranschaulicht.

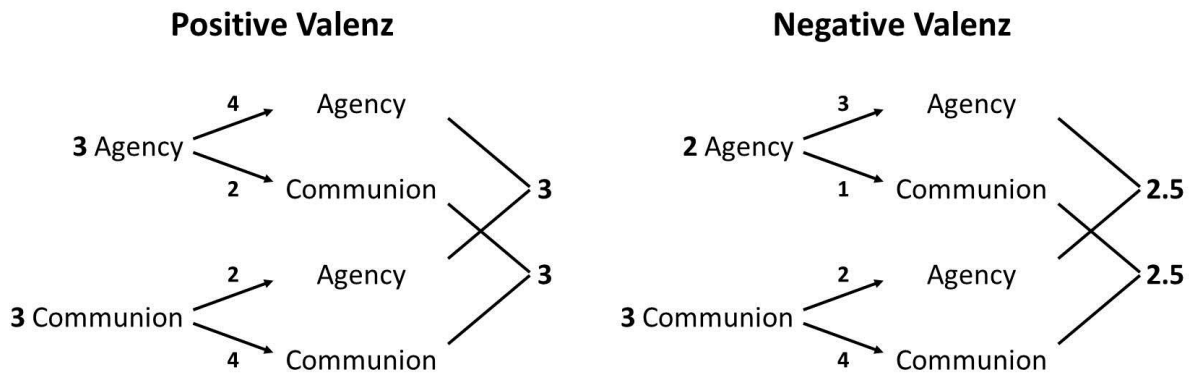


Abbildung 9. Veranschaulichung der Beeinflussbarkeit verschiedener Attribute

Da sich die empirischen Befunde von den theoretisch hergeleiteten unterscheiden, muss die Annahme gleicher Urteilsdifferenzen verletzt sein. Verändert man im obigen Beispiel etwa den negativen Halo-Effekt von Agency auf Communion von 1 auf den Wert 2, so würde die Beeinflussbarkeit von Communion-Attributen 3 statt 2.5 betragen. Die Unterschiede in der Beeinflussbarkeit können somit geklärt werden. Das bedeutet jedoch auch, dass Agency- und Communion-Urteile untereinander nicht im gleichen Maße ähnlich sein können. Im gegebenen Beispiel wären Agency-Urteile ($3 - 2 = 1$) bei negativer Valenz ähnlicher zueinander als Communion-Urteile ($4 - 2 = 2$). Eine mögliche Erklärung für diese Asymmetrie wäre die Tatsache, dass sich Agency- und Communion-Halo-Effekte doch qualitativ voneinander unterscheiden. Aufgrund der Befunde in den drei durchgeführten Experimenten kann dies jedoch ausgeschlossen werden. Die Ergebnisse können jedoch damit begründet werden, dass auch ohne die Darbietung jeglicher Informationen bereits Unterschiede zwischen Agency- und Communion-Urteilen existieren. Wie in der Literatur deutlich wird, sind Vorannahmen über bestimmte Attribute schon ohne die Vorgabe jeglicher Informationen vorhanden. Diese Vorannahmen sind jedoch keineswegs neutral, sondern eher positiv. Positive Eigenschaften werden also im Durchschnitt als Norm angesehen (H. H. Clark & Clark, 1977). Aus diesem Grund liegt der Schluss nahe, dass die Norm auch unterschiedlich für Agency- und Communion-Eigenschaften sein könnte. Beispielsweise wäre es möglich, dass positive Agency-Eigenschaften eher als gegeben und

positive Communion-Eigenschaften kritischer betrachtet werden. Um diese Vermutung zu testen, wurden zwei Post-Hoc-Untersuchungen durchgeführt.

5.4.1 Methode

Teilnehmer

An den beiden Post-Hoc-Experimenten nahmen 38 (20 weiblich, 18 männlich; erste Untersuchung) beziehungsweise 60 (34 weiblich, 26 männlich; zweite Untersuchung) Personen teil, die in der Mensa der Universität Heidelberg und im eigenen Bekanntenkreis geworben wurden. Die Teilnehmer der ersten Untersuchung waren durchschnittlich 26.89 Jahre ($SD = 5.87$, $Min = 18$, $Max = 37$), die der zweiten Untersuchung im Mittel 23.58 Jahre ($SD = 4.47$, $Min = 18$, $Max = 34$) alt. Für ihre Teilnahme an einem der beiden Experimente wurden die Probanden mit einer Tafel Schokolade vergütet.

Materialien

Die Basis der ersten Post-Hoc-Untersuchung war ein 2 (Informationsvalenz) x 2 (Urteilsdimension) Design. Sowohl Informationsvalenz (positiv versus negativ) als auch Urteilsdimension (Agency versus Communion) wurden innerhalb der Versuchsteilnehmer manipuliert. Als Items dienten 24 Eigenschaften aus einer Untersuchung von Abele und Kollegen (2008), die auch schon in der allgemeinen Voruntersuchung verwendet worden waren. Alle Attribute wurden in einer zufälligen Reihenfolge in einen Fragebogen eingepflegt. Die Teilnehmer sollten einschätzen, in welchem Maße eine durchschnittliche Person die jeweiligen Charaktereigenschaften aufweist. Die Urteile erfolgten auf derselben unipolaren Skala wie die Experimente 1 und 2 (1 = *keinesfalls*; 7 = *ganz sicher*). Der vollständige Fragebogen ist in Anhang G zu finden.

In der zweiten Post-Hoc-Untersuchung wurden, analog zu Experiment 3, bipolare Skalen von -3 bis 3 zur Urteilsabfrage verwendet (zum Beispiel „*streitsüchtig ... harmoniebedürftig*“). Aufgrund dessen fiel der Faktor *Informationsvalenz* im Untersuchungsdesign weg. Insgesamt 24 Eigenschaftspaare (elf für Agency und 13 für Communion) wurden als Items verwendet. Neunzehn dieser Paare wurden aus der Datenbasis entnommen, die in der Vorstudie von Experiment 1 erstellt worden war. Fünf weitere Paare (*kompetent-inkompetent*, *rational-irrational*, *höflich-unhöflich*, *sensibel-*

unsensibel und *zuverlässig-unzuverlässig*) wurden auf Grundlage der Daten von Abele und Kollegen (2008) generiert. Erneut wurden alle Attribute in einer zufälligen Reihenfolge in einen Fragebogen eingepflegt. Die Instruktion war identisch zur ersten Post-Hoc-Untersuchung. Der komplette Fragebogen ist in Anhang H dargestellt.

Prozedur

Ein Teil der Probanden wurde in der Mensa der Universität Heidelberg auf die Untersuchungen angesprochen, während der andere Teil dem eigenen Bekanntenkreis entstammte. Jeder Person wurde nur einer der beiden Fragebögen vorgelegt. Die Bearbeitung der Fragebögen erfolgte an Ort und Stelle und dauerte im Durchschnitt fünf Minuten. Anschließend wurden die Teilnehmer über Sinn und Zweck der Untersuchungen aufgeklärt und erhielten ihre Belohnung.

5.4.2 Ergebnisse

Baseline von Vorannahmen unter der Verwendung einer unipolaren Skala

Zur Untersuchung möglicher Valenz- und Dimensionsunterschiede wurde eine 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung durchgeführt. Es konnte ein signifikanter Haupteffekt für Valenz gezeigt werden ($F(1, 37) = 5.98, p = .019, \eta^2 = .14$): Vorannahmen über positive Eigenschaften ($M = 4.59, SD = 0.78$) waren extremer als Vorannahmen über negative Eigenschaften ($M = 4.17, SD = 0.72$). Der Haupteffekt des Faktors *Urteilsdimension* wurde nicht signifikant ($F(1, 37) = 0.42, p = .523, \eta^2 = .01$): Vorannahmen über Agency- ($M = 4.36, SD = 0.53$) und Communion-Attribute ($M = 4.40, SD = 0.60$) waren im Mittel gleich stark. Auch die Interaktion von Valenz und Urteilsdimension erreichte kein signifikantes Maß ($F(1, 37) = 0.42, p = .519, \eta^2 = .01$): Vorannahmen über positive Agency- ($M = 4.61, SD = 0.79$) und Communion-Eigenschaften ($M = 4.58, SD = 0.92$) waren nicht unterschiedlicher zueinander als Vorannahmen über negative Agency- ($M = 4.11, SD = 0.80$) und Communion-Eigenschaften ($M = 4.23, SD = 0.93$).

Baseline von Vorannahmen unter der Verwendung einer bipolaren Skala

Um die Baseline von Agency- und Communion-Urteilen zu vergleichen, wurde eine ANOVA mit Messwiederholung berechnet. Die Urteilsdimension war hierbei der einzige

Faktor. Es wurde kein signifikanter Effekt deutlich ($F(1, 59) = 1.85, p = .179, \eta^2 = .03$): Vorannahmen über Agency- ($M = 0.72, SD = 0.75$) und Communion-Attribute ($M = 0.61, SD = 0.80$) unterschieden sich nicht voneinander. Anschließend wurde getestet, ob sich die Mittelwerte der beiden Vorannahmen signifikant vom Skalenmittelpunkt (neutrales Urteil) unterscheiden. Dies geschah unter der Verwendung von t-Tests gegen den Wert 0. Sowohl Agency- ($t(59) = 7.45, p < .001$) als auch Communion-Urteile ($t(59) = 5.94, p < .001$) waren signifikant verschieden vom Skalenmittelpunkt.

5.4.3 Diskussion

Die beiden Post-Hoc-Analysen untersuchten, inwiefern sich Vorannahmen über Agency- und Communion-Attribute ohne die Darbietung jeglicher Informationen voneinander unterscheiden. Bei der Verwendung einer unipolaren Skala wurde deutlich, dass positive Urteile signifikant extremer ausfallen als negative Urteile. Wurden bipolare Skalen verwendet, so lagen die Urteile signifikant im positiven Bereich. Die Befunde bestätigen die Ansicht von H. H. Clark und Clark (1977), dass das Vorhandensein positiver Eigenschaften als Norm betrachtet wird. In diesem Zusammenhang stellt sich die Frage, ob eine Korrektur der berechneten Halo-Effekte um die jeweilige Baseline notwendig ist. Die Tatsache, dass Bewertungen von vornherein eher positiv sind, würde den starken quantitativen Unterschied positiver und negativer Halo-Effekte ein Stück weit relativieren. Auf der anderen Seite ist die Asymmetrie bezüglich der erwarteten Norm einer der Belege für die stärkere Densität positiver Informationen (Unkelbach et al., 2008) und damit auch indirekt für die Globalität positiver Halo-Effekte. Im weitesten Sinne lässt sich somit sagen, dass bereits ohne jegliche Manipulation positive Halo-Effekte wirken, die aus einer allgemeinen positiven Sicht auf die Welt (Boucher & Osgood, 1969) resultieren. Für alle durchgeführten Untersuchungen wurde aus diesem Grund die Baseline als Bestandteil von Halo-Effekten betrachtet und auf eine Korrektur der Ergebnisse verzichtet.

Bezüglich der Urteilsdimension konnten keine signifikanten Unterschiede gezeigt werden: In beiden Erhebungen waren die Vorannahmen über Agency- und Communion-Eigenschaften nicht von unterschiedlicher Stärke. Dieses Ergebnis ist dahingehend von Vorteil, dass keine Korrektur der berechneten Halo-Effekte notwendig ist. Anders als Valenzunterschiede sind die Differenzen von Agency- und Communion-Halo-Effekten nicht

durch eine unterschiedliche Density bedingt. Daher könnte die Baseline bei Einflüssen der Urteilsdimension nicht als Bestandteil der Halo-Effekte angesehen werden. Es wäre also nötig, die Baseline der Vorannahmen über Agency- und Communion-Attribute in jeder Untersuchung als Kontrollvariable einzufügen, um die wahren quantitativen Unterschiede der Halo-Effekte zu ermitteln. Aufgrund der fehlenden signifikanten Unterschiede hinsichtlich der Vorannahmen ist dies nicht notwendig. Allerdings machen beide Post-Hoc-Untersuchungen auf deskriptiver Ebene deutlich, dass Agency-Urteile im Durchschnitt positiver ausfallen als Communion-Urteile. Das ist damit begründbar, dass Communion-Eigenschaften als bedeutsamer wahrgenommen werden (Wojciszke & Abele, 2008; Wojciszke et al., 2009) und voreilige und eventuell fälschliche positive Einschätzungen stärker vermieden werden. Auch wenn der Unterschied kein signifikantes Maß erreichte, kann er doch eine Erklärung für die unterschiedliche Beeinflussbarkeit von Agency- und Communion-Urteilen in den ersten drei Experimenten bieten.

5.5 Zusammenfassung der Experimente 1 bis 3

Der erste Experimentalblock untersuchte den Einfluss von Informationsvalenz und Urteilsdimension auf die Qualität und Quantität von Halo-Effekten. Halo-Effekte wurden hierbei nur als Vorannahmen über noch unbekannte Eigenschaften betrachtet. In den Experimenten wurden fiktive Personen mit verschiedenen Eigenschaften oder Verhaltensweisen präsentiert. Auf Grundlage dieser Informationen beurteilten die Probanden die fiktiven Personen auf unterschiedlichen Dimensionen. Alle Halo-Effekte wurden mit Hilfe der unterschiedlichen Einzelratings berechnet, aber zum Teil auch mit Inter-Rating-Korrelationen untersucht. Die Verwendung von uni- oder bipolaren Skalen variierte zwischen den Experimenten. Zur Untersuchung der Hypothesen wurden neben Informationsvalenz und Urteilsdimension auch die zusätzliche Darbietung von Gesichtern sowie die Reihenfolge der Urteilsabfragen manipuliert.

Alle drei Experimente bestätigen die Valenzhypothese: Es existieren sowohl qualitative als auch quantitative Unterschiede zwischen positiven und negativen Halo-Effekten. Positive Halo-Effekte sind stark und global, während bei negativer Valenz schwächere und rein spezifische Halo-Effekte wirken. Der quantitative Unterschied wurde

wiederholt anhand von Mittelwertdifferenzen gezeigt. Der qualitative Unterschied wird dadurch deutlich, dass die Darbietung von Bildern und der Abfragemodus unterschiedlich stark auf positive und negative Halo-Effekte wirken. Zudem weisen auch die Inter-Rating-Korrelationen auf einen Qualitätsunterschied hin.

Die Ergebnisse aller drei Experimente bestätigen die Dimensionshypothese bei negativer Valenz: Negative Communion-Halo-Effekte sind stärker als negative Agency-Halo-Effekte. In der positiven Bedingung wurden dagegen keine Unterschiede bezüglich der Urteilsdimension deutlich: Positive Agency- und Communion-Halo-Effekte sind nicht unterschiedlich stark. Auch dieser Befund war über alle Untersuchungen hinweg stabil und kann dadurch erklärt werden, dass die hohe Globalität positiver Halo-Effekte die Unterschiede zwischen Agency- und Communion-Attributen aufhebt. Des Weiteren zeigen die Befunde, dass sich Agency- und Communion-Halo-Effekte zwar quantitativ, aber nicht qualitativ voneinander unterscheiden. So hatten Bilddarbietung und Abfragereihenfolge keine verschieden starke Wirkung auf Agency- und Communion-Halo-Effekte. Ebenfalls ergaben sich keine signifikant unterschiedlichen Inter-Rating-Korrelationen.

In zusätzlichen Analysen wurde deutlich, dass sich Vorannahmen über Communion-Attribute stärker durch negative Halo-Effekte prägen als Vorannahmen über Agency-Attribute. Im Falle positiver Halo-Effekte ergaben sich keine Unterschiede bei der Verwendung unipolarer Skalen. Wurden alle Urteile auf bipolaren Skalen abgegeben, wurden Agency-Attribute stärker positiv geprägt als Communion-Attribute. Die Post-Hoc-Untersuchungen konnten diese Befunde darauf zurückführen, dass Communion-Eigenschaften tendenziell, wenn auch nicht signifikant kritischer betrachtet werden als Agency-Eigenschaften: Bereits ohne das Vorhandensein jeglicher Informationen fallen Agency-Urteile im Durchschnitt positiver aus als Communion-Urteile.

6 Einflüsse von Informationsvalenz und Urteilsdimension auf Verzerrungen

Auch im zweiten Experimentalblock wurde die Qualität und Quantität von Halo-Effekten in Abhängigkeit von Informationsvalenz und Urteilsdimension untersucht. Im Gegensatz zum vorherigen Abschnitt wurden Halo-Effekte dieses Mal jedoch zum größten Teil als Verzerrungen eindeutiger Urteile betrachtet. Gemäß der Valenzhypothese sollten positive Halo-Effekte stärker und globaler sein als negative Halo-Effekte. Des Weiteren wurde vorhergesagt, dass Communion-Halo-Effekte von größerem Ausmaß, aber nicht globaler sind als Agency-Halo-Effekte. Um die beiden Annahmen zu testen, wurden drei Experimente durchgeführt.

6.1 Robustheit von Vorannahmen: Experiment 4

In den ersten drei Experimenten wurden Halo-Effekte als Vorannahmen über noch unbekannte Eigenschaften betrachtet. Die verschiedenen Urteile erfolgten dabei lediglich auf Basis kongruenter Informationen, wie zum Beispiel positiver Verhaltensweisen gepaart mit Bildern attraktiver Personen. Um Halo-Effekte im Sinne von Urteilsverzerrungen untersuchen zu können, war es jedoch notwendig, gegensätzliche Stimuli simultan darzubieten (vergleiche Abschnitt 3.4). Damit die grundlegenden Hypothesen getestet werden konnten, musste es sich bei diesen Stimuli um Verhaltensweisen unterschiedlicher Valenz und Dimension handeln.

Bevor sich der Fokus komplett auf die Verzerrung eindeutiger Urteile verlagerte, wurde im vierten Experiment ein Zwischenschritt vorgenommen. In diesem wurden Halo-Effekte noch einmal als Vorannahmen betrachtet, dieses Mal jedoch in ihrer Robustheit untersucht. Dies sollte einen ersten Hinweis darauf liefern, ob sich die für Vorannahmen gezeigten Ergebnisse auch für die stärkere Form von Halo-Effekten replizieren lassen. Die Robustheit von Vorannahmen wurde anhand der zusätzlichen Darbietung gegensätzlicher globaler Eindrücke, hier in Form von attraktiven und unattraktiven Gesichtern (Farah et al., 1998; Tanaka & Farah, 1993), geprüft. Da positive Halo-Effekte selbst globaler Natur sind, sollten sie nur geringfügig von der Darbietung unattraktiver Gesichter beeinflusst werden.

Umgekehrt sollten die rein spezifischen negativen Halo-Effekte bei der gleichzeitigen Präsentation attraktiver Personen stark geschwächt werden. Bezüglich der beiden Urteilsdimensionen wurde keine unterschiedliche Robustheit vermutet. Da Agency- und Communion-Halo-Effekte, je nach ihrer Valenz, gleichermaßen global beziehungsweise spezifisch sind, sollten sie auch nicht unterschiedlich stark durch einen gegensätzlichen globalen Stimulus beeinträchtigt werden.

6.1.1 Methode

Teilnehmer

Insgesamt 40 (32 weiblich, acht männlich; 20 in jeder Zwischensubjekt-Bedingung) Studenten der Universität Heidelberg nahmen an dem Experiment teil. Das Alter der Versuchsteilnehmer betrug im Mittel 20.98 Jahre ($SD = 1.90$, $Min = 18$, $Max = 25$). Deutsch war für 36 Teilnehmer die Muttersprache und für je zwei Teilnehmer die erste beziehungsweise zweite Fremdsprache. Für die Teilnahme an der Untersuchung erhielten die Probanden drei Euro oder eine halbe Versuchspersonenstunde.

Materialien

In dem Experiment wurde ein 2 (Darbietungsart) x 2 (Informationsvalenz) x 2 (Urteilsdimension) Design verwendet. Die Darbietungsart wurde diesbezüglich als Zwischensubjekt-Faktor mit den Stufen *ohne Bild* und *mit Bild* festgelegt. Sowohl Informationsvalenz (positiv versus negativ) als auch Urteilsdimension (Agency versus Communion) wurden innerhalb der Versuchsteilnehmer manipuliert.

Als Items wurden die gleichen 24 Verhaltenspaare wie in Experiment 3 verwendet (siehe Anhang F). Da jeweils drei Verhaltensweisen für das gleiche Attribut standen, waren insgesamt 16 Attribute repräsentiert. Auch das Bildmaterial war identisch zu dem in Experiment 3 (siehe Anhang D).

Die Einzelratings, die zur Berechnung der Halo-Effekte dienten, wurden auf acht bipolaren Skalen von -3 bis 3 (zum Beispiel „*verlogen ... ehrlich*“) erfasst. Pro Durchgang lautete die Frage wie folgt: „*Wie würden Sie diese Person auf folgender Dimension einschätzen?*“. Jedes der 16 Attribute (jeweils durch eine Verhaltensweise repräsentiert) wurde auf allen Skalen beurteilt. Damit hatte jede Zwischensubjekt-Bedingung 128 (16 x 8)

Durchgänge. Wie schon in Experiment 3 wurde jede Verhaltensweise auch auf der Skala bewertet, auf deren Basis sie ursprünglich generiert worden war. Diese Urteile dienten der Kontrolle der Items und wurden nicht in weitere Berechnungen einbezogen.

Zusätzlich wurde der Gesamteindruck auf Basis folgender Frage erfasst: „*Wie ist Ihr Gesamteindruck von dieser Person?*“. Die Bewertungen erfolgten auf einer bipolaren Skala von -3 (*sehr negativ*) bis 3 (*sehr positiv*). Da der Gesamteindruck für jedes Attribut erfragt wurde, gab es insgesamt 16 Durchgänge in jeder Zwischensubjekt-Bedingung.

Prozedur

Die Teilnehmer wurden auf dem Gelände des Psychologischen Instituts auf das Experiment angesprochen oder über Aushänge geworben. Im Untersuchungslabor unterschrieben sie zunächst eine Einverständniserklärung und begannen anschließend das Experiment. Dabei wurden sie zufällig einer der beiden Zwischensubjekt-Bedingungen zugeteilt. Es nahmen maximal sechs Personen gleichzeitig an einem Termin teil.

In der Instruktion wurde den Probanden mitgeteilt, dass sie im Folgenden an einer Bewertungsaufgabe teilnehmen würden, in der ihnen verschiedene Personen dargeboten werden, die jeweils eine bestimmte Verhaltensweise zeigen. Diese Personen würden auf acht Skalen unterschiedlicher Persönlichkeitsdimensionen bewertet werden müssen. Die Teilnehmer wurden darauf hingewiesen, sich bei noch offenen Fragen an die Versuchsleitung zu wenden und anderenfalls das Experiment zu starten.

Der erste Experimentalteil umfasste 128 Durchgänge, die randomisiert und ohne Zeitbeschränkung präsentiert wurden. Die Teilnehmer benötigten im Durchschnitt 20 Minuten für die vollständige Bearbeitung aller Durchgänge. Jedes der 16 Attribute war zufällig durch eine von drei Verhaltensweisen aus der Datenbasis repräsentiert. In der Bedingung *mit Bild* wurde jedes Attribut zudem zufällig einem der 16 Fotos zugeordnet. Positive Attribute konnten dabei nur mit unattraktiven Bildern und negative Attribute nur mit attraktiven Bildern kombiniert werden.

Im zweiten Experimentalteil wurde den Probanden mitgeteilt, dass im Folgenden erneut mehrere Personen mit unterschiedlichen Verhaltensweisen dargeboten werden würden. Dieses Mal würde jedoch ein Gesamteindruck von jeder Person gebildet werden müssen. Die 16 Durchgänge wurden randomisiert und ohne Zeitbeschränkung dargeboten. Wieder war jedes Attribut durch eine zufällig ausgewählte Verhaltensweise repräsentiert

und in der Bedingung *mit Bild* zudem an eines der 16 Fotos gekoppelt. Auch hier wurden positive Attribute stets mit unattraktiven Bildern und negative Attribute mit attraktiven Bildern kombiniert. Der zweite Experimentaltteil war unabhängig vom ersten. Die Auswahl der Verhaltensweisen und die Zuordnung der Bilder zu den Attributen waren folglich nicht zwingenderweise identisch.

Am Ende des Experiments gaben die Teilnehmer ihr Geschlecht sowie ihr Alter an. Des Weiteren wurde erfragt, ob Deutsch die Muttersprache, die erste oder die zweite Fremdsprache sei. Nach Beendigung des Experiments wurden die Teilnehmer über Inhalt und Ziel der Untersuchung aufgeklärt und erhielten ihre Vergütung.

6.1.2 Ergebnisse

Vorab-Analysen

Wie in Experiment 3 wurden vor der eigentlichen Datenanalyse die verwendeten Verhaltensweisen kontrolliert. Zu diesem Zweck wurde für jede Verhaltensweise das Urteil auf der Dimension betrachtet, aus der die Verhaltensweise ursprünglich generiert worden war. Die Bewertungen sollten bei positiven Verhaltensweisen im positiven Bereich (Skalenwert 1 bis 3) und bei negativen Verhaltensweisen dementsprechend im negativen Bereich (Skalenwert -3 bis -1) liegen. Es wurde nur die Baseline-Bedingung (ohne Bild) betrachtet, da in der Verzerrungs-Bedingung (mit Bild) auch bei eigentlich eindeutigen Urteilen gewisse Abweichungen zu erwarten waren. Die Bedingung war bei allen 20 Probanden in der Baseline-Bedingung für alle Verhaltensweisen erfüllt. Somit konnte die komplette Stichprobe in die weiteren Analysen einbezogen werden.

In einer weiteren Vorab-Analyse wurde getestet, ob die abhängigen Variablen mit Halo-Effekten gleichgesetzt werden können. Ein Urteil wurde dann als Halo-Effekt betrachtet, wenn es sich signifikant vom Mittelpunkt der Skala unterschied. Dazu wurden alle abhängigen Variablen unter der Verwendung von t-Tests gegen den Wert 0 getestet. Auch hierbei wurde nur die Baseline-Bedingung betrachtet. Die Variable *negativer Halo-Effekt bei ungleicher Dimension* wick nicht signifikant von dem Wert 0 ab ($p > .050$). Alle anderen abhängigen Variablen zeigten signifikante Unterschiede zum Skalenmittelpunkt. Folglich wurde der Faktor *Gleichheit von dargebotener und beurteilter Dimension* (gleiche versus ungleiche Dimension) in diesem Experiment nicht betrachtet.

Da positive und negative Halo-Effekte auf derselben Skala erfasst wurden, wurden vor der Überprüfung der Hypothesen alle Ratings, die auf negativen Verhaltensweisen basierten, durch eine Multiplikation mit -1 umkodiert.

Replizierbarkeit der Befunde des dritten Experiments

Zunächst wurde nur die Zwischensubjekt-Bedingung *ohne Bild* betrachtet. Da in dieser Bedingung keine widersprüchlichen Informationen dargeboten wurden, konnte untersucht werden, ob die in Experiment 3 gezeigten Valenz- und Dimensionseffekte replizierbar sind. Dies geschah unter der Verwendung einer 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung. Es zeigte sich ein signifikanter Haupteffekt für Valenz ($F(1, 19) = 23.65, p < .001, \eta^2 = .55$): Positive Halo-Effekte ($M = 0.79, SD = 0.33$) waren stärker als negative Halo-Effekte ($M = 0.44, SD = 0.28$). Der Haupteffekt des Faktors *Urteilsdimension* wurde nicht signifikant ($F(1, 19) = 1.06, p = .317, \eta^2 = .05$): Halo-Effekte von Agency- ($M = 0.58, SD = 0.33$) und Communion-Verhaltensweisen ($M = 0.65, SD = 0.27$) unterschieden sich nicht voneinander. Die Interaktion beider Faktoren erreichte ebenfalls kein signifikantes Maß ($F(1, 19) = 2.07, p = .166, \eta^2 = .10$). Zur genaueren Analyse der Daten wurden Agency- und Communion-Halo-Effekte anschließend, getrennt für beide Valenzbedingungen, mit Hilfe von t-Tests miteinander verglichen: Positive Halo-Effekte von Agency- ($M = 0.79, SD = 0.42$) und Communion-Verhaltensweisen ($M = 0.78, SD = 0.36$) unterschieden sich nicht signifikant voneinander ($t(19) = 0.14, p = .890$). Entgegen den Erwartungen waren Agency- ($M = 0.36, SD = 0.33$) und Communion-Halo-Effekte ($M = 0.52, SD = 0.36$) auch in der negativen Valenzbedingung nicht unterschiedlich zueinander ($t(19) = 1.73, p = .100$).

Die Beeinflussbarkeit der Vorannahmen über Agency- und Communion-Attribute wurde auf Basis einer 2 x 2 ANOVA (Informationsvalenz x Dimension der beurteilten Eigenschaft) mit Messwiederholung untersucht. Im Gegensatz zu dem Effekt von Valenz wurde der Haupteffekt für die Dimension der beurteilten Eigenschaft nicht signifikant ($F(1, 19) = 0.19, p = .672, \eta^2 = .01$): Vorannahmen über Communion- ($M = 0.60, SD = 0.28$) und Agency-Attribute ($M = 0.62, SD = 0.27$) wurden im Durchschnitt gleichermaßen durch Halo-Effekte geprägt. Es war jedoch eine signifikante Interaktion der beiden Faktoren erkennbar ($F(1, 19) = 7.16, p = .015, \eta^2 = .27$): Positive Halo-Effekte konnten Agency-Attribute ($M = 0.86, SD = 0.35$) stärker prägen als Communion-Attribute ($M = 0.72, SD = 0.36$). Negative

Halo-Effekte hatten dagegen auf Agency-Attribute ($M = 0.39$, $SD = 0.32$) geringere Einflüsse als auf Communion-Attribute ($M = 0.49$, $SD = 0.32$).

Einfluss der Darbietung gegensätzlicher Bilder

Der Einfluss der Darbietung gegensätzlicher Bilder wurde mit Hilfe einer $2 \times 2 \times 2$ ANOVA (Darbietungsart \times Informationsvalenz \times Urteilsdimension) mit Messwiederholung analysiert. Es war kein signifikanter Haupteffekt der Darbietungsart erkennbar ($F(1, 38) = 3.92$, $p = .055$, $\eta^2 = .09$): Durch die Darbietung gegensätzlicher Bilder ($M = 0.38$, $SD = 0.45$) veränderten sich die Halo-Effekte im Vergleich zur Darbietung ohne Bilder ($M = 0.61$, $SD = 0.26$) nicht. Der Haupteffekt der Urteilsdimension wurde auch nicht signifikant ($F(1, 38) = 2.13$, $p = .153$, $\eta^2 = .05$): Agency- ($M = 0.46$, $SD = 0.39$) und Communion-Halo-Effekte ($M = 0.53$, $SD = 0.43$) waren im Durchschnitt gleich stark. Im Gegensatz dazu ergab sich ein signifikanter Haupteffekt des Faktors *Informationsvalenz* ($F(1, 38) = 45.43$, $p < .001$, $\eta^2 = .55$): Positive Halo-Effekte ($M = 0.75$, $SD = 0.48$) waren von durchschnittlich größerem Ausmaß als negative Halo-Effekte ($M = 0.24$, $SD = 0.43$).

Es konnte keine signifikante Interaktion von Valenz und Urteilsdimension gezeigt werden ($F(1, 38) = 2.31$, $p = .137$, $\eta^2 = .06$): Positive Agency- ($M = 0.75$, $SD = 0.48$) und Communion-Halo-Effekte ($M = 0.75$, $SD = 0.55$) hatten keinen größeren Unterschied zueinander als negative Agency- ($M = 0.17$, $SD = 0.45$) und Communion-Halo-Effekte ($M = 0.31$, $SD = 0.51$). Auch die Interaktion *Darbietungsart \times Urteilsdimension* wurde bei der ANOVA nicht signifikant ($F(1, 38) < 0.01$, $p = .979$, $\eta^2 < .01$): Halo-Effekte von Agency- ($M = 0.58$, $SD = 0.33$) und Communion-Verhaltensweisen ($M = 0.65$, $SD = 0.27$) in der Bedingung ohne Bilddarbietung wiesen keine andere Differenz zueinander auf als Halo-Effekte von Agency- ($M = 0.35$, $SD = 0.41$) und Communion-Verhaltensweisen ($M = 0.42$, $SD = 0.53$) bei der Präsentation von Bildern. Die Interaktion von Darbietungsart und Valenz erreichte ein signifikantes Niveau ($F(1, 38) = 4.26$, $p = .046$, $\eta^2 = .10$): Positive Halo-Effekte wurden durch die Darbietung gegensätzlich valenter Bilder ($M = 0.71$, $SD = 0.60$) im Vergleich zur Darbietung ohne Bilder ($M = 0.79$, $SD = 0.33$) nur wenig beeinflusst. Im Gegensatz dazu waren negative Halo-Effekte bei der Präsentation gegensätzlicher Bilder ($M = 0.05$, $SD = 0.47$) geringer als ohne Bilder ($M = 0.44$, $SD = 0.28$). Eine signifikante Interaktion aller drei Faktoren konnte nicht gefunden werden ($F(1, 38) = 0.12$, $p = .729$, $\eta^2 < .01$).

Gesamteindruck in Abhängigkeit von Informationsvalenz und Urteilsdimension

Wie bei der Untersuchung der Halo-Effekte wurde zuerst nur die Zwischensubjekt-Bedingung *ohne Bild* analysiert. Hierfür wurde eine 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung verwendet. Der Haupteffekt für Valenz wurde nicht signifikant ($F(1, 19) = 0.05, p = .835, \eta^2 < .01$): Positive ($M = 1.27, SD = 0.54$) und negative Gesamteindrücke ($M = 1.24, SD = 0.53$) waren im Durchschnitt gleich stark. Es zeigte sich jedoch ein signifikanter Haupteffekt der Urteilsdimension ($F(1, 19) = 6.18, p = .022, \eta^2 = .25$): Gesamteindrücke, die auf Grundlage von Communion-Verhaltensweisen gebildet wurden ($M = 1.46, SD = 0.57$), waren extremer als Gesamteindrücke auf der Basis von Agency-Verhaltensweisen ($M = 1.06, SD = 0.60$). Es war keine signifikante Interaktion von Valenz und Urteilsdimension erkennbar ($F(1, 19) = 3.56, p = .074, \eta^2 = .16$): Positive Agency- ($M = 1.18, SD = 0.74$) und Communion-Gesamteindrücke ($M = 1.36, SD = 0.71$) waren nicht unterschiedlicher zueinander als negative Agency- ($M = 0.94, SD = 0.62$) und Communion-Gesamteindrücke ($M = 1.55, SD = 0.69$).

Gesamteindruck in Abhängigkeit von der Darbietung gegensätzlicher Bilder

Um den Einfluss der Darbietungsart untersuchen zu können, wurde eine 2 x 2 x 2 ANOVA (Darbietungsart x Informationsvalenz x Urteilsdimension) mit Messwiederholung durchgeführt. Es ergab sich kein signifikanter Haupteffekt der Darbietungsart ($F(1, 38) = 1.92, p = .174, \eta^2 = .05$): Die durchschnittlichen Gesamteindrücke mit ($M = 1.00, SD = 0.68$) und ohne Bilddarbietung ($M = 1.26, SD = 0.47$) unterschieden sich nicht voneinander. Der Haupteffekt für Valenz wurde auch nicht signifikant ($F(1, 38) = 0.39, p = .534, \eta^2 = .01$): Positive ($M = 1.16, SD = 0.69$) und negative Gesamteindrücke ($M = 1.09, SD = 0.68$) waren gleich stark. Es konnte jedoch ein signifikanter Haupteffekt des Faktors *Urteilsdimension* aufgezeigt werden ($F(1, 38) = 11.54, p = .002, \eta^2 = .23$): Gesamteindrücke, die auf Communion-Verhaltensweisen basierten ($M = 1.34, SD = 0.77$), waren im Durchschnitt extremer als Agency-Gesamteindrücke ($M = 0.91, SD = 0.65$).

Auch die Interaktion von Valenz und Urteilsdimension wurde signifikant ($F(1, 38) = 8.13, p = .007, \eta^2 = .18$): Negative Agency- ($M = 0.76, SD = 0.75$) und Communion-Gesamteindrücke ($M = 1.43, SD = 0.90$) wiesen eine höhere Differenz zueinander auf als positive Agency- ($M = 1.07, SD = 0.82$) und Communion-Gesamteindrücke ($M = 1.26, SD = 0.87$). Es ergab sich keine signifikante Interaktion von Darbietungsart und Urteilsdimension

($F(1, 38) = 0.06, p = .807, \eta^2 < .01$): Agency- ($M = 1.06, SD = 0.60$) und Communion-Gesamteindrücke ($M = 1.46, SD = 0.57$) in der Bedingung ohne Bilder waren nicht unterschiedlicher zueinander als Agency- ($M = 0.77, SD = 0.67$) und Communion-Gesamteindrücke ($M = 1.23, SD = 0.93$) in der Bedingung mit Bildern. Die Interaktion von Darbietungsart und Valenz wurde ebenfalls nicht signifikant ($F(1, 38) = 0.16, p = .692, \eta^2 < .01$): Positive Gesamteindrücke ohne ($M = 1.27, SD = 0.54$) und mit Bildern ($M = 1.06, SD = 0.81$) hatten keine andere Differenz zueinander als negative Gesamteindrücke ohne ($M = 1.24, SD = 0.53$) und mit Bilddarbietung ($M = 0.94, SD = 0.79$). Es war keine signifikante Dreifachinteraktion erkennbar ($F(1, 38) = 0.13, p = .717, \eta^2 < .01$).

6.1.3 Diskussion

In Experiment 4 wurde die Robustheit von Halo-Effekten bei der Darbietung gegensätzlicher globaler Stimuli, hier in Form von Gesichtern (Farah et al., 1998; Tanaka & Farah, 1993), untersucht. Gleichzeitig wurde versucht, die Ergebnisse von Experiment 3 zu replizieren. Erneut kann bestätigt werden, dass positive Halo-Effekte von höherem Ausmaß sind als negative Halo-Effekte. Des Weiteren können Agency-Urteile leichter positiv und Communion-Urteile leichter negativ geprägt werden. Bezüglich des Einflusses der Urteilsdimension gelang es aber nicht, die Befunde des dritten Experiments zu replizieren: Zwar waren positive Agency- und Communion-Halo-Effekte wieder durchschnittlich gleich stark, jedoch zeigten sich dieses Mal auch keine signifikanten Unterschiede zwischen negativen Agency- und Communion-Halo-Effekten. Da die deskriptiven Daten jedoch sehr ähnlich zu denen von Experiment 3 waren, ist die fehlende Signifikanz vermutlich auf die geringe Stichprobengröße zurückzuführen.

Für die Robustheit der Halo-Effekte bei der Darbietung gegensätzlich valenter Bilder kann das postulierte Muster bestätigt werden: Positive Halo-Effekte werden durch die gleichzeitige Präsentation unattraktiver Gesichter nur unmaßgeblich beeinträchtigt. Im Gegensatz dazu werden negative Halo-Effekte durch die Darbietung attraktiver Personen bedeutend schwächer. Dies bestätigt einmal mehr den qualitativen Unterschied positiver und negativer Halo-Effekte: Positive Halo-Effekte sind globaler Natur, negative Halo-Effekte dagegen spezifisch. Halo-Effekte von Agency- und Communion-Verhaltensweisen werden gleichermaßen durch die zusätzliche Präsentation gegensätzlicher Bilder beeinflusst. Auch

dieses Ergebnis wurde erwartet, da sich Agency- und Communion-Halo-Effekte zwar quantitativ, nicht jedoch qualitativ unterscheiden.

Die Analyse der Gesamteindrücke zeigt im Vergleich zu den Halo-Effekten ein verändertes Muster: Positive und negative Gesamteindrücke sind im Durchschnitt gleich stark. Jedoch sind Gesamteindrücke, die auf Basis von Communion-Verhaltensweisen gebildet werden, insgesamt stärker als Agency-Gesamteindrücke. Des Weiteren werden positive und negative Gesamteindrücke, im Gegensatz zu positiven und negativen Halo-Effekten, nicht unterschiedlich stark durch die Darbietung gegensätzlicher Bilder geschwächt. Diese Befunde implizieren einmal mehr, dass Halo-Effekte und globale Urteile bei negativer Valenz völlig unabhängig voneinander sind.

6.2 Urteilsverzerrungen bei paralleler Darbietung: Experiment 5

Nachdem Valenzunterschiede sowohl für Vorannahmen als auch deren Robustheit nachgewiesen werden konnten, wurde im fünften Experiment schließlich das komplette grundlegende Modell (vergleiche Abschnitt 3.2) untersucht. Halo-Effekte wurden folglich nicht mehr als Vorannahmen, sondern als Verzerrungen eindeutiger Urteile betrachtet. Um diese Verzerrungen untersuchen zu können, waren zwei verschiedene experimentelle Bedingungen notwendig: In der Baseline-Bedingung wurden mehrere kongruente Verhaltensweisen dargeboten, die zu eindeutigen Urteilen führen sollten (zum Beispiel die Beurteilung der Verhaltensweise ... *sagt anderen selten die Wahrheit auf der Skala verlogen ... ehrlich*). In der Verzerrungs-Bedingung wurde zusätzlich zu diesen kongruenten Verhaltensweisen eine zentrale gegensätzliche Verhaltensweise präsentiert, die keinen Bezug zu der abgefragten Skala hatte (zum Beispiel ... *lernt und arbeitet viel und kontinuierlich*). Die Urteile waren somit objektiv gesehen immer noch eindeutig. Zeigte sich jedoch eine Abweichung gegenüber den Urteilen der Baseline-Bedingung, entsprach dies einem Halo-Effekt der zentralen gegensätzlichen Verhaltensweise.

Anhand des beschriebenen Untersuchungsdesigns wurden Halo-Effekte in Abhängigkeit von Informationsvalenz und Urteilsdimension analysiert. In dieser Studie wurden ausschließlich quantitative Unterschiede betrachtet. Die Hypothesen waren identisch zu denen der vorangegangenen Experimente: Positive Halo-Effekte sollten von

größerem Ausmaß sein als negative Halo-Effekte. Des Weiteren wurde angenommen, dass Halo-Effekte von Communion-Verhaltensweisen im Durchschnitt stärker sind als Halo-Effekte von Agency-Verhaltensweisen.

6.2.1 Methode

Teilnehmer

An dem Experiment nahmen 40 (34 weiblich, sechs männlich; 20 in jeder Zwischensubjekt-Bedingung) Studenten der Universität Heidelberg teil. Das Alter der Probanden betrug im Mittel 22.13 Jahre ($SD = 2.61$, $Min = 18$, $Max = 29$). Deutsch war für insgesamt 37 Teilnehmer die Muttersprache, für zwei Teilnehmer die erste und für einen Teilnehmer die zweite Fremdsprache. Für die Teilnahme an der Untersuchung wurden die Probanden mit drei Euro oder einer halbe Versuchspersonenstunde vergütet.

Materialien

Der Untersuchung lag ein 2 (Darbietungsart) x 2 (Informationsvalenz) x 2 (Urteilsdimension) x 2 (Gleichheit von dargebotener und beurteilter Dimension) Design zugrunde. Die Darbietungsart wurde zwischen den Teilnehmern manipuliert und hatte die Ausprägungen *ohne Gegensatz* und *mit Gegensatz*. Informationsvalenz (positiv versus negativ), Urteilsdimension (Agency versus Communion) und Gleichheit (gleiche versus ungleiche Dimension) wurden als Innersubjekt-Faktoren bestimmt.

Als Items wurden die gleichen 24 Verhaltenspaare wie in den Experimenten 3 und 4 verwendet (siehe Anhang F). Da je drei Verhaltensweisen für das gleiche Attribut standen, waren insgesamt 16 Attribute repräsentiert. In der Bedingung *mit Gegensatz* wurde jedes der 16 Attribute (zentrales Attribut) mit den vier Attributen der gegensätzlichen Dimension und Valenz (periphere Attribute) kombiniert. Für das positive Agency-Attribut *ordentlich* (genauso wie für die anderen positiven Agency-Attribute) waren dies zum Beispiel die negativen Communion-Attribute *verlogen*, *eigenbrötlerisch*, *egoistisch* und *streitsüchtig*. In der Bedingung *ohne Gegensatz* waren nur vier periphere Attribute vorhanden.

Die einzelnen Ratings erfolgten auf acht bipolaren Skalen von -3 bis 3 (zum Beispiel „*schüchtern ... selbstbewusst*“). Die Frage lautet dabei stets folgendermaßen: „*Wie würden Sie diese Person auf folgender Dimension einschätzen?*“. Jedes der 16 zentralen Attribute

wurde mit allen acht Skalen kombiniert, so dass sich insgesamt 128 (16 x 8) Durchgänge ergaben. In der Baseline-Bedingung (ohne Gegensatz) waren die zentralen Attribute zwar nicht präsent, jedoch wurden die Durchgänge zur besseren Vergleichbarkeit identisch zur Verzerrungs-Bedingung (mit Gegensatz) konzipiert. Dies hatte zur Folge, dass nicht alle Urteile in der Baseline-Bedingung objektiv eindeutig waren. Wurden beispielsweise vier Agency-Attribute präsentiert, aber ein Communion-Attribut abgefragt, wäre eine nicht neutrale Bewertung eher ein Halo-Effekt statt ein korrektes Urteil. Da jedoch sowohl Agency- als auch Communion-Attribute auch auf der gegensätzlichen Dimension beurteilt werden mussten, wurde angenommen, dass die grundsätzlichen Valenz- und Dimensionsunterschiede nicht verfälscht werden. In der Verzerrungs-Bedingung gab es Durchgänge, in denen die jeweilige zentrale Verhaltensweise und die abgefragte Skala demselben Attribut (beziehungsweise Attributpaar) entsprachen und das Urteil somit eindeutig war. Analog zu den Experimenten 3 und 4 wurden diese Urteile lediglich zur Kontrolle der Items verwendet und nicht in weitere Berechnungen einbezogen.

Prozedur

Die Probanden wurden im Umfeld des Psychologischen Instituts durch Aushänge oder studentische Hilfskräfte für die Erhebung geworben und in das Labor geladen. Dort unterschrieben sie eine Einverständniserklärung und begannen danach das Experiment. Dabei wurden sie zufällig einer der beiden Zwischensubjekt-Bedingungen zugeteilt. An jedem Termin nahmen maximal sechs Personen gleichzeitig teil.

In der Instruktion wurde den Probanden, je nach Zwischensubjekt-Bedingung, mitgeteilt, dass sie im Folgenden an einer Bewertungsaufgabe teilnehmen würden, in der ihnen verschiedene Personen dargeboten werden, die jeweils vier beziehungsweise fünf bestimmte Verhaltensweisen zeigen. Diese Personen würden auf acht Skalen unterschiedlicher Persönlichkeitsdimensionen bewertet werden müssen. Die Teilnehmer wurden darauf hingewiesen, sich bei noch bestehenden offenen Fragen an die Versuchsleitung zu wenden und anderenfalls das Experiment zu starten.

Der Experimentalteil bestand aus 128 Durchgängen, die randomisiert und ohne Zeitbeschränkung präsentiert wurden. Im Durchschnitt benötigten die Probanden 25 Minuten zur vollständigen Bearbeitung aller Fragen. In jedem Durchgang waren alle dargebotenen Attribute zufällig durch eine von drei Verhaltensweisen aus der Datenbasis

repräsentiert. In der Baseline-Bedingung wurden in jedem Durchgang vier periphere Verhaltensweisen, in der Verzerrungs-Bedingung zusätzlich eine zentrale Verhaltensweise dargeboten. Die zentrale Verhaltensweise nahm stets die Position in der Mitte ein, die Positionen der vier peripheren Verhaltensweisen waren randomisiert. Tabelle 7 zeigt einen beispielhaften Durchgang in den beiden Experimentalbedingungen.

Tabelle 7. Beispiel für einen Durchgang in den beiden Experimentalbedingungen

ohne Gegensatz (Baseline-Bedingung)	mit Gegensatz (Verzerrungs-Bedingung)
<i>... hat eine aufrechte Körperhaltung (selbstbewusst)</i>	<i>... erledigt anstehende Arbeit sofort und vor dem Vergnügen (fleißig)</i>
<i>... setzt sich ohne Hilfe eigene Ziele und verfolgt diese (selbständig)</i>	<i>... hat eine aufrechte Körperhaltung (selbstbewusst)</i> <i>... redet hinter dem Rücken schlecht über andere (verlogen)</i>
<i>... hat einen guten Überblick über die eigenen Termine (ordentlich)</i>	<i>... putzt und räumt die eigene Wohnung regelmäßig auf (ordentlich)</i>
<i>... lernt und arbeitet viel und kontinuierlich (fleißig)</i>	<i>... setzt sich ohne Hilfe eigene Ziele und verfolgt diese (selbständig)</i>

Anmerkung. Die in Klammern genannten Attribute dienen nur dem besseren Verständnis und waren in den Experimentaldurchgängen selbst nie präsent.

Am Ende der Untersuchung gaben die Teilnehmer ihr Geschlecht und ihr Alter an. Zudem wurde erfragt, ob Deutsch die Muttersprache, die erste oder die zweite Fremdsprache sei. Nach Beendigung des Experiments wurden die Teilnehmer über Inhalt und Ziel der Untersuchung aufgeklärt und erhielten ihre Vergütung.

6.2.2 Ergebnisse

Vorab-Analysen

In der ersten Vorab-Analyse wurden die verwendeten Verhaltensweisen überprüft. Zu diesem Zweck wurde für jede zentrale Verhaltensweise das Urteil auf der Dimension betrachtet, aus der die Verhaltensweise ursprünglich generiert worden war. Die Bewertungen sollten bei positiven Verhaltensweisen im positiven Bereich (Skalenwert 1 bis 3) und bei negativen Verhaltensweisen im negativen Bereich (Skalenwert -3 bis -1) sein. Es

wurde nur die Verzerrungs-Bedingung betrachtet, da in der Baseline-Bedingung die zentralen Verhaltensweisen gar nicht präsent waren. Die Bedingung war bei allen 20 Teilnehmern für alle Verhaltensweisen erfüllt. Folglich konnte die komplette Stichprobe in die weiteren Analysen einbezogen werden.

Anschließend wurde überprüft, ob die Ratings in der Baseline-Bedingung nicht neutral sind. Nur dann wären Verzerrungen durch gegensätzliche Stimuli auch als Halo-Effekte interpretierbar. Ein Rating wurde dann als nicht neutral betrachtet, wenn es sich signifikant vom Skalenmittelpunkt unterschied. Zu diesem Zweck wurden alle abhängigen Variablen in der Baseline-Bedingung mit Hilfe von t-Tests gegen den Wert 0 getestet. Alle abhängigen Maße wiesen signifikante Unterschiede zum Skalenmittelpunkt auf ($p < .050$) und wurden daher in die weiteren Analysen einbezogen.

Da positive und negative Halo-Effekte auf der gleichen Skala erhoben wurden, wurden vor der Überprüfung der Hypothesen alle Ratings, die auf peripheren negativen Verhaltensweisen basierten, durch eine Multiplikation mit -1 umkodiert.

Einflüsse von Informationsvalenz und Dimensionsgleichheit ohne Gegensatz

Im ersten Schritt der Datenanalyse wurden die Effekte von Informationsvalenz und Dimensionsgleichheit in der Zwischensubjekt-Bedingung *ohne Gegensatz* betrachtet. Zu diesem Zweck wurde eine 2 x 2 ANOVA (Informationsvalenz x Gleichheit von dargebotener und beurteilter Dimension) mit Messwiederholung durchgeführt. Es war ein signifikanter Haupteffekt für Valenz erkennbar ($F(1, 19) = 17.26, p = .001, \eta^2 = .48$): Ratings, die auf der Basis von positiven Verhaltensweisen abgegeben wurden ($M = 1.43, SD = 0.49$), waren extremer als Ratings, die auf negativen Verhaltensweisen beruhten ($M = 1.16, SD = 0.41$). Auch der Haupteffekt der Dimensionsgleichheit wurde signifikant ($F(1, 19) = 251.69, p < .001, \eta^2 = .93$): Bei gleicher Dimension von präsenten Verhaltensweisen und beurteiltem Attribut ($M = 2.15, SD = 0.50$) wurden extremere Ratings abgegeben als bei unterschiedlichen Dimensionen ($M = 0.44, SD = 0.48$). Die Interaktion der beiden Faktoren erreichte kein signifikantes Niveau ($F(1, 19) = 3.67, p = .071, \eta^2 = .16$): Positive Ratings bei gleicher ($M = 2.24, SD = 0.54$) und ungleicher Dimension ($M = 0.62, SD = 0.58$) waren nicht unterschiedlicher zueinander als negative Ratings bei gleicher ($M = 2.07, SD = 0.48$) und ungleicher Dimension ($M = 0.25, SD = 0.50$).

Einflüsse von Informationsvalenz und Urteilsdimension ohne Gegensatz

Zur Analyse potentieller Einflüsse der Urteilsdimension in der Bedingung *ohne Gegensatz* wurde eine 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung durchgeführt. Neben dem bereits im letzten Abschnitt geschilderten Haupteffekt für Valenz konnte auch ein signifikanter Effekt für den Faktor *Urteilsdimension* gezeigt werden ($F(1, 19) = 5.23, p = .034, \eta^2 = .22$): Ratings, die auf der Grundlage von Communion-Verhaltensweisen erfolgten ($M = 1.34, SD = 0.44$), waren im Durchschnitt extremer als Ratings, die auf Agency-Verhaltensweisen basierten ($M = 1.25, SD = 0.44$). Die Interaktion von Informationsvalenz und Urteilsdimension wurde nicht signifikant ($F(1, 19) = 0.41, p = .530, \eta^2 = .02$): Ratings auf der Basis positiver Agency- ($M = 1.40, SD = 0.48$) und Communion-Verhaltensweisen ($M = 1.46, SD = 0.55$) waren nicht unterschiedlicher zueinander als Ratings auf der Basis negativer Agency- ($M = 1.10, SD = 0.44$) und Communion-Verhaltensweisen ($M = 1.22, SD = 0.41$).

Einfluss zentraler gegensätzlicher Verhaltensweisen in Bezug auf die Informationsvalenz

Die Stärke positiver und negativer Halo-Effekte wurde anhand des Einflusses der zentralen gegensätzlichen Verhaltensweisen auf die verschiedenen Ratings ermittelt. Dazu wurde eine 2 x 2 x 2 ANOVA (Darbietungsart x Informationsvalenz x Gleichheit von dargebotener und beurteilter Dimension) mit Messwiederholung durchgeführt. Es war ein signifikanter Haupteffekt der Darbietungsart zu erkennen ($F(1, 38) = 18.35, p < .001, \eta^2 = .33$): In der Bedingung mit einer zentralen gegensätzlichen Verhaltensweise ($M = 0.84, SD = 0.22$) zeigten sich im Durchschnitt neutralere Ratings als ohne Gegensatz ($M = 1.29, SD = 0.43$). Auch die Haupteffekte für Valenz ($F(1, 38) = 63.44, p < .001, \eta^2 = .63$) und Dimensionsgleichheit ($F(1, 38) = 383.93, p < .001, \eta^2 = .91$) wurden signifikant: Waren die vier peripheren Verhaltensweisen von positiver Valenz ($M = 1.25, SD = 0.43$), wurden extremere Ratings abgegeben als bei negativer Valenz ($M = 0.88, SD = 0.44$). Bei gleicher Dimension von peripheren Verhaltensweisen und beurteilter Eigenschaft ($M = 2.03, SD = 0.53$) wurden über beide Zwischensubjekt-Bedingungen hinweg extremere Ratings abgegeben als bei unterschiedlichen Dimensionen ($M = 0.10, SD = 0.51$).

Bezüglich der Interaktionen gab es einen signifikanten Effekt von Valenz x Dimensionsgleichheit ($F(1, 38) = 17.39, p < .001, \eta^2 = .31$): Positive Ratings bei gleicher ($M = 2.15, SD = 0.52$) und ungleicher Dimension ($M = 0.34, SD = 0.56$) wiesen eine geringere

Differenz zueinander auf als negative Ratings bei gleicher ($M = 1.91$, $SD = 0.57$) und ungleicher Dimension ($M = -0.14$, $SD = 0.55$) von peripheren Verhaltensweisen und beurteiltem Attribut. Signifikant wurde ebenfalls die Interaktion von Darbietungsart und Valenz ($F(1, 38) = 4.46$, $p = .041$, $\eta^2 = .11$): Urteile, die auf vier negativen Verhaltensweisen basierten ($M = 1.16$, $SD = 0.41$), konnten durch die zusätzliche Darbietung einer positiven Verhaltensweise ($M = 0.60$, $SD = 0.27$) stark geschwächt werden. Im Vergleich dazu wurden positive Urteile ($M = 1.43$, $SD = 0.49$) durch die zusätzliche Präsentation einer negativen Verhaltensweise ($M = 1.07$, $SD = 0.25$) nur geringfügig beeinflusst. Mit anderen Worten gesagt, waren positive Halo-Effekte stärker als negative.

Ein signifikanter Befund zeigte sich auch für die Interaktion *Darbietungsart x Dimensionsgleichheit* ($F(1, 38) = 4.67$, $p = .037$, $\eta^2 = .11$): Bei ungleicher Dimension von peripheren Verhaltensweisen und beurteiltem Attribut führte die Präsentation einer gegensätzlichen Verhaltensweise ($M = -0.24$, $SD = 0.26$) gegenüber der Baseline-Bedingung ($M = 0.44$, $SD = 0.48$) zu einer starken Verzerrung der Ratings. Im Falle einer gleichen Dimension von peripheren Verhaltensweisen und beurteilter Eigenschaft konnte die Darbietung einer gegensätzlichen Verhaltensweise ($M = 1.91$, $SD = 0.54$) die Urteile aus der Baseline-Bedingung ($M = 2.15$, $SD = 0.50$) nur in schwachem Maße beeinflussen. Das heißt im Umkehrschluss, dass Halo-Effekte größer waren, wenn die zentrale gegensätzliche Verhaltensweise dieselbe Dimension hatte wie das beurteilte Attribut. Die Interaktion aller drei Faktoren wurde nicht signifikant ($F(1, 38) = 0.67$, $p = .420$, $\eta^2 = .02$).

Einfluss zentraler gegensätzlicher Verhaltensweisen in Bezug auf die Urteilsdimension

Zur Untersuchung von Agency- und Communion-Halo-Effekten wurde eine $2 \times 2 \times 2$ ANOVA (Darbietungsart x Informationsvalenz x Urteilsdimension) mit Messwiederholung berechnet. Die Haupteffekte für Darbietungsart und Valenz sowie die Interaktion dieser Faktoren wurden bereits im letzten Abschnitt betrachtet. Es konnte ein signifikanter Haupteffekt der Urteilsdimension aufgezeigt werden ($F(1, 38) = 20.40$, $p < .001$, $\eta^2 = .35$): Ratings, die auf peripheren Communion-Verhaltensweisen basierten ($M = 1.15$, $SD = 0.41$), waren extremer als Ratings auf der Grundlage von Agency-Verhaltensweisen ($M = 0.98$, $SD = 0.44$). Die Interaktion von Valenz und Urteilsdimension erreichte dagegen kein signifikantes Niveau ($F(1, 38) = 0.02$, $p = .878$, $\eta^2 < .01$): Ratings auf der Grundlage peripherer positiver Agency- ($M = 1.17$, $SD = 0.44$) und Communion-Verhaltensweisen ($M = 1.33$, $SD = 0.49$)

zeigten keine andere Differenz zueinander als Ratings auf der Basis negativer Agency- ($M = 0.79$, $SD = 0.48$) und Communion-Verhaltensweisen ($M = 0.97$, $SD = 0.44$).

Es gab einen signifikanten Effekt für die Interaktion von Darbietungsart und Urteilsdimension ($F(1, 38) = 4.24$, $p = .047$, $\eta^2 = .10$): Ratings, die auf vier Agency-Verhaltensweisen basierten ($M = 1.25$, $SD = 0.44$), wurden durch die zusätzliche Darbietung einer gegensätzlich valenten Communion-Verhaltensweise ($M = 0.71$, $SD = 0.23$) stark beeinflusst. Demgegenüber wurden Urteile auf der Grundlage von vier Communion-Verhaltensweisen ($M = 1.34$, $SD = 0.44$) durch die Präsentation einer gegensätzlichen Agency-Verhaltensweise ($M = 0.96$, $SD = 0.28$) vergleichsweise schwach verzerrt. Communion-Halo-Effekte waren somit stärker als Agency-Halo-Effekte. Eine signifikante Dreifachinteraktion wurde nicht gefunden ($F(1, 38) = 0.60$, $p = .445$, $\eta^2 = .02$).

6.2.3 Diskussion

Experiment 5 diente der Analyse quantitativer Unterschiede von Halo-Effekten in Abhängigkeit von Informationsvalenz und Urteilsdimension. Halo-Effekte wurden hierbei erstmals als Verzerrungen eindeutiger Urteile betrachtet. Die Tatsache, dass nicht alle Urteile in der Baseline-Bedingung objektiv eindeutig waren, galt für alle Attribute in gleichem Maße. Somit wurden die grundlegenden Ergebnisse nicht verfälscht. Die Analysen bestätigen die zentralen Hypothesen: Positive Halo-Effekte sind von größerem Ausmaß als negative Halo-Effekte. Ebenso sind Halo-Effekte von Communion-Verhaltensweisen stärker als Halo-Effekte von Agency-Verhaltensweisen. Anhand eines weiteren Ergebnisses wird deutlich, dass Halo-Effekte einer Verhaltensweise größer sind, wenn diese die gleiche Dimension aufweist wie die zu beurteilende Eigenschaft. Die Befunde der vorhergehenden Experimente, in denen Halo-Effekte noch als Vorannahmen betrachtet wurden, konnten somit allesamt auf einer anderen Ebene repliziert werden.

Die separate Untersuchung der Baseline-Bedingung beweist, dass auch ohne das Wirken von Halo-Effekten positive Ratings stärker sind als negative Ratings. Des Weiteren gehen Communion-Verhaltensweisen mit extremeren Beurteilungen einher als Agency-Verhaltensweisen. Diese Befunde sind konform mit den bisherigen Ergebnissen und auch durch eine unterschiedlich hohe Density (Unkelbach et al., 2008) beziehungsweise Bedeutsamkeit (Wojciszke & Abele, 2008; Wojciszke et al., 2009) erklärbar.

Bezüglich der Halo-Effekte ist anzumerken, dass in dem Experiment nur Verhaltensweisen gegensätzlicher Valenz und Dimension miteinander gepaart wurden. Das hat zur Folge, dass das Wirken von Halo-Effekten nicht klar von der Robustheit der ursprünglichen Ratings abgegrenzt werden kann. In der Studie wurden beispielsweise Bewertungen, die auf vier negativen Agency-Verhaltensweisen basierten und durch eine zusätzliche positive Communion-Verhaltensweise stark verzerrt wurden, mit einem starken Halo-Effekt der Communion-Verhaltensweise erklärt. Alternativ wäre es jedoch ebenso möglich, dass Ratings, die auf negativen Agency-Verhaltensweisen beruhen, besonders anfällig für Verzerrungen jeder Art sind. Um Halo-Effekte komplett von der Verzerrungsanfälligkeit der Ratings abzugrenzen, ist es notwendig, Eigenschaften beider Valenzen und Dimensionen miteinander zu kombinieren.

6.3 Urteilsverzerrungen bei sequentieller Darbietung: Experiment 6

In Experiment 5 wurde gezeigt, dass die Stärke von Urteilsverzerrungen in Abhängigkeit von Informationsvalenz und Urteilsdimension variiert. Die Aussagekraft der Befunde ist jedoch in zwei Aspekten limitiert: Zum einen waren zentrale und periphere Verhaltensweisen eines Durchgangs stets von gegensätzlicher Valenz und Dimension. Dies hatte zur Folge, dass die Halo-Effekte bestimmter Attribute nicht klar von der Verzerrungsanfälligkeit der Ratings abgegrenzt werden konnten. Zum anderen konnten auf Basis des verwendeten Experimentaldesigns keine qualitativen Unterschiede positiver und negativer Halo-Effekte analysiert werden.

Das Ziel des sechsten Experiments war es, die genannten Einschränkungen zu umgehen. Zu diesem Zweck wurde ein erweitertes Design verwendet, in dem Verhaltensweisen beider Valenzen und Dimensionen miteinander kombiniert wurden. Auf diese Weise konnten Halo-Effekte und die Robustheit der ursprünglichen Ratings experimentell voneinander getrennt werden. Um auch qualitative Unterschiede verschiedener Halo-Effekte zu untersuchen, wurden die Verhaltensweisen nicht mehr gleichzeitig, sondern nacheinander dargeboten. Dabei wurde die Position der zentralen Verhaltensweise variiert. Es wurden unterschiedliche Effekte für positive und negative Valenz vorhergesagt: Da positive Halo-Effekte globaler Natur sind, sollte die Position einer

zentralen positiven Verhaltensweise keine Rolle für die Stärke der resultierenden Halo-Effekte spielen. Im Gegensatz dazu sollten die spezifischen Halo-Effekte von zentralen negativen Verhaltensweisen an bestimmten Positionen stärker sein als an anderen.

Die Umstellung von einer parallelen auf eine sequentielle Informationsdarbietung trug auch zu einer höheren Alltagsnähe des Experiments bei, da verschiedene Verhaltensweisen einer Person in der Realität meistens nur in zeitlichem Abstand beobachtet werden können. Um die bisherigen Ergebnisse zusätzlich generalisieren zu können, lag dem Experiment erstmals die komplette Datenbasis zugrunde. Die Auswahl der präsentierten Verhaltensweisen und abgefragten Eigenschaften erfolgte dabei nicht mehr nach einem festen Schema, sondern auf rein zufälliger Basis.

6.3.1 Methode

Teilnehmer

Teilnehmer der Untersuchung waren 42 (31 weiblich, elf männlich; 21 in jeder Zwischensubjekt-Bedingung) Studenten der Universität Heidelberg. Das Alter der Versuchsteilnehmer betrug im Durchschnitt 24.07 Jahre ($SD = 5.53$, $Min = 19$, $Max = 40$). Deutsch war für 39 Teilnehmer die Muttersprache, für zwei Teilnehmer die erste und für einen Teilnehmer die zweite Fremdsprache. Für die Teilnahme an der Untersuchung erhielten die Probanden drei Euro oder eine halbe Versuchspersonenstunde.

Materialien

Die Basis des Experiments war ein 4 (Darbietungsart) x 3 (Position des zentralen Attributs) x 2 (Informationsvalenz) x 2 (Urteilsdimension) x 2 (Dimension der beurteilten Eigenschaft) Design. Die Darbietungsart wurde innerhalb der Teilnehmer variiert und hatte die Baseline-Ausprägung *ohne Gegensatz* sowie die drei Verzerrungs-Bedingungen *gleiche Valenz / andere Dimension*, *andere Valenz / gleiche Dimension* und *andere Valenz / andere Dimension*. Die Position des zentralen Attributs war ein Innersubjekt-Faktor mit den Stufen *Anfang*, *Mitte* und *Ende*. Auch Informationsvalenz (positiv versus negativ) und Urteilsdimension (Agency versus Communion) wurden innerhalb der Teilnehmer manipuliert. Die Dimension der zu beurteilenden Eigenschaft (Agency versus Communion) wurde als Zwischensubjekt-Faktor konzipiert, um eine zu hohe Anzahl an experimentellen

Durchgängen und damit eine zu lange Bearbeitungsdauer des Experiments zu vermeiden. Aus diesem Grund wurde der Faktor nicht in die Datenanalyse einbezogen. Die Variable *Gleichheit von dargebotener und beurteilter Dimension* war aus Gründen der Übersichtlichkeit nicht Teil des Untersuchungsdesigns.

Als Items dienten alle 28 Verhaltenspaare aus der Datenbasis (siehe Anhang F), die insgesamt 24 Attribute repräsentierten. In der Baseline-Bedingung wurden jeweils fünf zufällig ausgewählte Attribute gleicher Valenz und Dimension (jeweils in Form einer Verhaltensweise) miteinander kombiniert, in den Verzerrungs-Bedingungen dagegen nur vier Attribute. Das fünfte Attribut in den Verzerrungs-Bedingungen hatte entweder eine andere Dimension, eine andere Valenz oder eine andere Dimension und Valenz. Bei diesem fünften handelte es sich um das zentrale Attribut. Eine komplette Übersicht über die verschiedenen Darbietungsbedingungen ist in Tabelle 8 dargestellt.

Tabelle 8. Übersicht über die verschiedenen Darbietungsbedingungen

ohne Gegensatz (Baseline-Bedingung)	mit Gegensatz (Verzerrungs-Bedingungen)		
	gleiche Valenz / andere Dimension	andere Valenz / gleiche Dimension	andere Valenz / andere Dimension
5 positive Agency- Verhaltensweisen	4 pos. Ag.-Verh. + 1 pos. Co.-Verh.	4 pos. Ag.-Verh. + 1 neg. Ag.-Verh.	4 pos. Ag.-Verh. + 1 neg. Co.-Verh.
5 positive Communion- Verhaltensweisen	4 pos. Co.-Verh. + 1 pos. Ag.-Verh.	4 pos. Co.-Verh. + 1 neg. Co.-Verh.	4 pos. Co.-Verh. + 1 neg. Ag.-Verh.
5 negative Agency- Verhaltensweisen	4 neg. Ag.-Verh. + 1 neg. Co.-Verh.	4 neg. Ag.-Verh. + 1 pos. Ag.-Verh.	4 neg. Ag.-Verh. + 1 pos. Co.-Verh.
5 negative Communion- Verhaltensweisen	4 neg. Co.-Verh. + 1 neg. Ag.-Verh.	4 neg. Co.-Verh. + 1 pos. Co.-Verh.	4 neg. Co.-Verh. + 1 pos. Ag.-Verh.

Anmerkung. pos. = positive; neg. = negative; Ag. = Agency; Co. = Communion; Verh. = Verhaltensweise(n)

Die Ratings wurden auf zufälligen bipolaren Skalen von -3 bis 3 erfasst. Die Frage lautet hierbei stets „Wie würden Sie diese Person auf folgender Dimension einschätzen?“. Jede der 16 verschiedenen Darbietungs-Kombinationen wurde mit den drei Positions-Bedingungen verknüpft. In der Bedingung *Anfang* stand die zentrale Verhaltensweise an erster, in der Bedingung *Mitte* an dritter und in der Bedingung *Ende* an fünfter und letzter Stelle. Für die Baseline-Bedingung machte dies keinen Unterschied, da es darin keine zentrale Verhaltensweise gab. Die Positionen der vier peripheren Verhaltensweisen waren

durchgehend randomisiert. Durch die verschiedenen Kombinationen kamen insgesamt 48 (16 x 3) Durchgänge in jeder Zwischensubjekt-Bedingung zustande. Die Verhaltensweisen, welche die Attribute repräsentierten, waren in jedem Durchgang randomisiert. Analog zu Experiment 5 wurde es nicht als problematisch betrachtet, dass nicht alle Urteile in der Baseline-Bedingung objektiv eindeutig waren.

Prozedur

Alle Probanden wurden in der Mensa der Universität Heidelberg durch studentische Hilfskräfte für die Untersuchung geworben. Das Labor wurde in einem gesonderten Raum über der Mensa eingerichtet. Nachdem die Teilnehmer eine Einverständniserklärung unterschrieben hatten, begannen sie mit dem Experiment. Dabei wurden sie zufällig einer der beiden Zwischensubjekt-Bedingungen zugeteilt. An einem Termin konnten maximal sechs Personen gleichzeitig teilnehmen.

In der Instruktion wurde den Probanden mitgeteilt, dass sie im Folgenden an einer Bewertungsaufgabe teilnehmen würden, in der ihnen 48 unterschiedliche Personen dargeboten werden, die jeweils fünf bestimmte Verhaltensweisen zeigen. Diese Personen würden auf verschiedenen Skalen bewertet werden müssen. In der Zwischensubjekt-Bedingung *Agency* wurde erläutert, dass sich alle Skalen auf Persönlichkeitsdimensionen beziehen, welche die Kompetenz von Personen beschreiben. In der Bedingung *Communion* wurde stattdessen mitgeteilt, dass sich alle Skalen auf menschliche Wärme beziehen. Die Versuchsteilnehmer wurden darauf hingewiesen, sich bei noch offenen Fragen an die Versuchsleitung zu wenden und anderenfalls das Experiment zu starten.

Der Experimentaltail bestand aus 48 Durchgängen, die in einer randomisierten Reihenfolge präsentiert wurden. In jedem Durchgang wurden die fünf Verhaltensweisen je vier Sekunden lang präsentiert. Anschließend sollte auf dieser Grundlage eine zufällig aus der Datenbasis ausgewählte *Agency*- beziehungsweise *Communion*-Eigenschaft beurteilt werden. Durch die Programmierung wurde verhindert, dass die zentrale Verhaltensweise und die abgefragte Skala demselben Attribut (beziehungsweise Attributpaar) entsprachen und das Urteil somit eindeutig war. Zur Bearbeitung des kompletten Experimentaltails benötigten die Teilnehmer durchschnittlich 25 Minuten.

Am Ende der Untersuchung gaben die Probanden ihr Geschlecht sowie ihr Alter an. Des Weiteren wurde erfragt, ob Deutsch die Muttersprache, die erste oder die zweite

Fremdsprache sei. Nach Beendigung des Experiments wurden die Teilnehmer über Inhalt und Ziel der Untersuchung aufgeklärt und erhielten ihre Vergütung.

6.3.2 Ergebnisse

Vorab-Analysen

In den Vorab-Analysen wurde überprüft, ob die Ratings in der Baseline-Bedingung nicht neutral sind. Nur dann wären Verzerrungen durch gegensätzliche Verhaltensweisen auch als Halo-Effekte interpretierbar. Ein Rating wurde dann als nicht neutral betrachtet, wenn es sich signifikant vom Mittelpunkt der Skala unterschied. Zu diesem Zweck wurden alle abhängigen Maße in der Baseline-Bedingung mit Hilfe von t-Tests gegen den Wert 0 getestet. Alle abhängigen Variablen wiesen signifikante Unterschiede zum Mittelpunkt der Skala auf ($p < .050$) und wurden daher in die weiteren Analysen einbezogen.

Da positive und negative Halo-Effekte auf derselben Skala erfasst wurden, wurden vor der Überprüfung der Hypothesen alle Ratings, die auf peripheren negativen Verhaltensweisen basierten, durch eine Multiplikation mit -1 umkodiert.

Einflüsse von Informationsvalenz und Urteilsdimension ohne Gegensatz

Der Fokus der Datenanalyse lag zunächst auf der Darbietungsbedingung *ohne Gegensatz*. Zur Untersuchung der Einflüsse von Informationsvalenz und Urteilsdimension wurde eine 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung vorgenommen. Es konnte kein signifikanter Haupteffekt des Faktors *Informationsvalenz* aufgezeigt werden ($F(1, 41) = 0.16, p = .693, \eta^2 < .01$): Ratings, die auf positiven ($M = 1.63, SD = 0.60$) und negativen Verhaltensweisen basierten ($M = 1.58, SD = 0.62$), unterschieden sich im Mittel nicht voneinander. Ebenfalls nicht signifikant wurde der Haupteffekt der Urteilsdimension ($F(1, 41) = 3.62, p = .064, \eta^2 = .08$): Ratings, die auf der Grundlage von Agency- ($M = 1.41, SD = 0.94$) und Communion-Verhaltensweisen erfolgten ($M = 1.80, SD = 0.67$), waren durchschnittlich gleich stark. Auch die Interaktion von Informationsvalenz und Urteilsdimension erreichte kein signifikantes Maß ($F(1, 41) = 0.43, p = .517, \eta^2 = .01$). Da diese Befunde nicht den Erwartungen entsprachen, wurden die Ratings auf der Grundlage von Agency- und Communion-Verhaltensweisen anschließend getrennt für beide Valenzbedingungen miteinander verglichen. Dafür wurde jeweils ein t-Test bei gepaarten

Stichproben durchgeführt: Ratings, die auf positiven Agency- ($M = 1.47$, $SD = 1.15$) und Communion-Verhaltensweisen basierten ($M = 1.79$, $SD = 0.75$), unterschieden sich nicht signifikant voneinander ($t(41) = 1.39$, $p = .173$). In der negativen Valenzbedingung waren Urteile auf der Grundlage von Communion-Verhaltensweisen ($M = 1.81$, $SD = 0.84$) gegenüber Ratings auf der Basis von Agency-Verhaltensweisen ($M = 1.36$, $SD = 1.02$) signifikant extremer ($t(41) = 2.08$, $p = .044$).

Einfluss von Verhaltensweisen gleicher Valenz und gegensätzlicher Dimension

Zur Analyse des Einflusses zentraler Verhaltensweisen mit gleicher Valenz und gegensätzlicher Dimension wurde eine $2 \times 2 \times 2$ ANOVA (Darbietungsart \times Informationsvalenz \times Urteilsdimension) mit Messwiederholung berechnet, wobei nur die beiden Bedingungen *ohne Gegensatz* und *gleiche Valenz / andere Dimension* betrachtet wurden. Der Haupteffekt des Faktors *Darbietungsart* erreichte ein signifikantes Maß ($F(1, 41) = 9.50$, $p = .004$, $\eta^2 = .19$): In der Verzerrungs-Bedingung ($M = 1.80$, $SD = 0.53$) zeigten sich im Durchschnitt extremere Ratings als in der Bedingung ohne Gegensatz ($M = 1.61$, $SD = 0.47$). Auch der Haupteffekt der Urteilsdimension wurde signifikant ($F(1, 41) = 6.07$, $p = .018$, $\eta^2 = .13$): Handelte es sich bei den peripheren um Communion-Verhaltensweisen ($M = 1.89$, $SD = 0.57$), wurden extremere Beurteilungen abgegeben als bei Agency-Verhaltensweisen ($M = 1.52$, $SD = 0.77$). Es konnte kein signifikanter Haupteffekt für Valenz deutlich gemacht werden ($F(1, 41) = 0.09$, $p = .772$, $\eta^2 < .01$): Ratings auf der Basis peripherer positiver ($M = 1.69$, $SD = 0.59$) und negativer Verhaltensweisen ($M = 1.72$, $SD = 0.60$) unterschieden sich im Mittel nicht voneinander.

Die Interaktion von Valenz und Urteilsdimension wurde nicht signifikant ($F(1, 41) = 0.08$, $p = .779$, $\eta^2 < .01$): Urteile auf der Basis peripherer positiver Agency- ($M = 1.51$, $SD = 0.94$) und Communion-Verhaltensweisen ($M = 1.87$, $SD = 0.62$) zeigten keine andere Differenz zueinander als Ratings auf der Basis negativer Agency- ($M = 1.52$, $SD = 0.86$) und Communion-Verhaltensweisen ($M = 1.92$, $SD = 0.80$). Ebenfalls gab es keinen signifikanten Effekt der Interaktion *Darbietungsart* \times *Informationsvalenz* ($F(1, 41) = 1.97$, $p = .168$, $\eta^2 = .05$): Positive ($M = 1.63$, $SD = 0.60$) und negative Ratings ($M = 1.58$, $SD = 0.62$) in der Baseline-Bedingung waren nicht unterschiedlicher zueinander als positive ($M = 1.75$, $SD = 0.69$) und negative Ratings ($M = 1.86$, $SD = 0.71$) in der Verzerrungs-Bedingung. Die Interaktion von Darbietungsart und Urteilsdimension erreichte kein signifikantes Maß ($F(1,$

41) = 0.04, $p = .837$, $\eta^2 < .01$): Beurteilungen auf der Grundlage peripherer Agency- ($M = 1.41$, $SD = 0.94$) und Communion-Verhaltensweisen ($M = 1.80$, $SD = 0.67$) in der Baseline-Bedingung wiesen dieselbe Differenz zueinander auf wie Bewertungen auf der Grundlage von Agency- ($M = 1.62$, $SD = 0.69$) und Communion-Verhaltensweisen ($M = 1.98$, $SD = 0.62$) in der Verzerrungs-Bedingung. Eine signifikante Interaktion aller drei Faktoren war nicht erkennbar ($F(1, 41) = 0.49$, $p = .487$, $\eta^2 = .01$).

Einfluss von Verhaltensweisen gegensätzlicher Valenz und gleicher Dimension

Der Einfluss zentraler Verhaltensweisen mit gegensätzlicher Valenz und gleicher Dimension wurde mit Hilfe einer 2 x 2 x 2 ANOVA (Darbietungsart x Informationsvalenz x Urteilsdimension) mit Messwiederholung untersucht. Es wurden nur die Darbietungsarten *ohne Gegensatz* und *andere Valenz / gleiche Dimension* betrachtet. Es konnte ein signifikanter Haupteffekt der Darbietungsart gezeigt werden ($F(1, 41) = 61.09$, $p < .001$, $\eta^2 = .60$): In der Verzerrungs-Bedingung ($M = 1.06$, $SD = 0.44$) waren die Bewertungen neutraler als in der Baseline-Bedingung ($M = 1.61$, $SD = 0.47$). Die Haupteffekte von Valenz ($F(1, 41) = 0.51$, $p = .480$, $\eta^2 = .01$) und Urteilsdimension ($F(1, 41) = 3.06$, $p = .088$, $\eta^2 = .07$) wurden nicht signifikant: Ratings auf der Grundlage peripherer positiver ($M = 1.29$, $SD = 0.54$) und negativer Verhaltensweisen ($M = 1.38$, $SD = 0.61$) waren durchschnittlich gleich hoch. Auch unterschieden sich die Urteile auf der Basis peripherer Agency- ($M = 1.18$, $SD = 0.76$) und Communion-Verhaltensweisen ($M = 1.48$, $SD = 0.58$) nicht voneinander.

Bezüglich der Interaktionen gab es einen signifikanten Effekt von Valenz x Urteilsdimension ($F(1, 41) = 4.28$, $p = .045$, $\eta^2 = .09$): Urteile auf der Basis peripherer positiver Agency- ($M = 1.22$, $SD = 0.91$) und Communion-Verhaltensweisen ($M = 1.35$, $SD = 0.74$) waren ähnlicher zueinander als Urteile auf der Basis negativer Agency- ($M = 1.15$, $SD = 0.85$) und Communion-Verhaltensweisen ($M = 1.61$, $SD = 0.85$). Die Interaktion von Darbietungsart und Valenz erreichte kein signifikantes Maß ($F(1, 41) = 2.05$, $p = .159$, $\eta^2 = .05$): Positive ($M = 1.63$, $SD = 0.60$) und negative Ratings ($M = 1.58$, $SD = 0.62$) in der Baseline-Bedingung hatten die gleiche Differenz zueinander wie positive ($M = 0.94$, $SD = 0.74$) und negative Ratings ($M = 1.17$, $SD = 0.81$) in der Verzerrungs-Bedingung. Ebenfalls nicht signifikant wurde die Interaktion von Darbietungsart und Urteilsdimension ($F(1, 41) = 1.93$, $p = .172$, $\eta^2 = .05$): Urteile auf der Basis peripherer Agency- ($M = 1.41$, $SD = 0.94$) und Communion-Verhaltensweisen ($M = 1.80$, $SD = 0.67$) in der Bedingung ohne Gegensatz

waren nicht unterschiedlicher zueinander als Ratings auf der Grundlage von Agency- ($M = 0.96$, $SD = 0.67$) und Communion-Verhaltensweisen ($M = 1.16$, $SD = 0.66$) in der Bedingung mit einer Verhaltensweise von gegensätzlicher Valenz und gleicher Dimension. Es wurde keine signifikante Dreifachinteraktion von Darbietungsart, Informationsvalenz und Urteilsdimension sichtbar ($F(1, 41) = 3.21$, $p = .081$, $\eta^2 = .07$).

Einfluss von Verhaltensweisen gegensätzlicher Valenz und Dimension

Um den Einfluss von zentralen Verhaltensweisen mit gegensätzlicher Valenz und Dimension untersuchen zu können, wurde eine $2 \times 2 \times 2$ ANOVA (Darbietungsart \times Informationsvalenz \times Urteilsdimension) mit Messwiederholung vorgenommen, bei der nur die zwei Bedingungen *ohne Gegensatz* und *andere Valenz / andere Dimension* betrachtet wurden. Diese Manipulation war identisch zu der in Experiment 5. Der Haupteffekt der Darbietungsart wurde signifikant ($F(1, 41) = 81.56$, $p < .001$, $\eta^2 = .67$): In der Bedingung mit zentraler gegensätzlicher Verhaltensweise ($M = 1.05$, $SD = 0.39$) ergaben sich neutralere Ratings als in der Baseline-Bedingung ($M = 1.61$, $SD = 0.47$). Des Weiteren zeigte sich ein signifikanter Haupteffekt der Urteilsdimension ($F(1, 41) = 7.98$, $p = .007$, $\eta^2 = .16$): Auf der Basis peripherer Communion-Verhaltensweisen ($M = 1.65$, $SD = 0.68$) wurden extremere Urteile abgegeben als auf der Grundlage von Agency-Verhaltensweisen ($M = 1.01$, $SD = 0.97$). Der Haupteffekt für Valenz erreichte kein signifikantes Niveau ($F(1, 41) = 0.24$, $p = .624$, $\eta^2 = .01$): Bewertungen auf der Basis peripherer positiver ($M = 1.30$, $SD = 0.51$) und negativer Verhaltensweisen ($M = 1.36$, $SD = 0.56$) waren im Mittel gleich hoch.

Es gab keine signifikante Interaktion von Valenz und Urteilsdimension ($F(1, 41) = 0.73$, $p = .398$, $\eta^2 = .02$): Bewertungen auf der Grundlage peripherer positiver Agency- ($M = 1.01$, $SD = 1.07$) und Communion-Verhaltensweisen ($M = 1.59$, $SD = 0.79$) waren nicht unterschiedlicher zueinander als Ratings auf der Basis negativer Agency- ($M = 1.00$, $SD = 1.06$) und Communion-Verhaltensweisen ($M = 1.72$, $SD = 0.87$). Auch die Interaktion von Darbietungsart und Valenz erreichte kein signifikantes Maß ($F(1, 41) = 2.75$, $p = .105$, $\eta^2 = .06$): Positive ($M = 1.63$, $SD = 0.60$) und negative Ratings ($M = 1.58$, $SD = 0.62$) in der Baseline-Bedingung hatten keine andere Differenz zueinander als positive ($M = 0.97$, $SD = 0.58$) und negative Ratings ($M = 1.13$, $SD = 0.64$) in der Bedingung mit Gegensatz. Es konnte allerdings eine signifikante Interaktion der Faktoren *Darbietungsart* und *Urteilsdimension* gezeigt werden ($F(1, 41) = 10.60$, $p = .002$, $\eta^2 = .21$): Urteile, die auf peripheren Agency-

Verhaltensweisen basierten ($M = 1.41, SD = 0.94$), wurden durch die zusätzliche Darbietung einer gegensätzlich valenten Communion-Verhaltensweise ($M = 0.60, SD = 1.12$) stark beeinflusst. Auf der anderen Seite wurden Urteile auf der Basis peripherer Communion-Verhaltensweisen ($M = 1.80, SD = 0.67$) durch die zusätzliche Präsentation einer gegensätzlichen Agency-Verhaltensweise ($M = 1.51, SD = 0.81$) nur schwach verzerrt. Communion-Halo-Effekte waren folglich stärker als Agency-Halo-Effekte. Die Interaktion aller drei Faktoren wurde nicht signifikant ($F(1, 41) = 0.01, p = .917, \eta^2 < .01$).

Einfluss der Position der zentralen Verhaltensweise

Nach den paarweisen Vergleichen der vier Darbietungsbedingungen wurde untersucht, welchen Einfluss die Position der zentralen Verhaltensweise auf die Stärke der resultierenden Halo-Effekte haben kann. Da sich nur in der Bedingung *andere Valenz / andere Dimension* signifikant unterschiedliche Halo-Effekte gezeigt hatten (Interaktion von Darbietungsart und Urteilsdimension), wurde auch nur diese Bedingung fokussiert. Zur Analyse der Daten wurde eine $2 \times 2 \times 2$ ANOVA (Position des zentralen Attributs \times Informationsvalenz \times Urteilsdimension) mit Messwiederholung verwendet. Es wurde ein signifikanter Haupteffekt der Position deutlich ($F(2, 40) = 8.77, p = .001, \eta^2 = .31$): Befand sich die zentrale Verhaltensweise am Anfang ($M = 1.25, SD = 0.57$), wurden extremere Ratings abgegeben als bei einer Position in der Mitte ($M = 1.04, SD = 0.48$). Am neutralsten waren die Urteile, wenn die zentrale Verhaltensweise am Ende auftauchte ($M = 0.87, SD = 0.51$). Der Haupteffekt des Faktors *Urteilsdimension* wurde ebenfalls signifikant ($F(1, 41) = 10.81, p = .002, \eta^2 = .21$): Ratings auf der Basis von vier Communion- und einer zentralen Agency-Verhaltensweise ($M = 1.51, SD = 0.81$) waren extremer als Urteile auf der Basis von vier Agency- und einer Communion-Verhaltensweise ($M = 0.60, SD = 1.12$). Es konnte kein signifikanter Haupteffekt für Valenz gezeigt werden ($F(1, 41) = 1.27, p = .266, \eta^2 = .03$): Ratings auf der Grundlage peripherer positiver ($M = 0.97, SD = 0.58$) und negativer Verhaltensweisen ($M = 1.13, SD = 0.64$) unterschieden sich nicht voneinander.

Bei den Interaktionen gab es keinen signifikanten Effekt von Valenz \times Urteilsdimension ($F(1, 41) = 0.55, p = .461, \eta^2 = .01$): Urteile auf der Basis peripherer positiver Agency- ($M = 0.56, SD = 1.24$) und Communion-Verhaltensweisen ($M = 1.39, SD = 0.98$) hatten die gleiche Differenz zueinander wie Urteile auf der Grundlage negativer Agency- ($M = 0.64, SD = 1.22$) und Communion-Verhaltensweisen ($M = 1.63, SD = 1.07$). Die

Interaktion von Position und Urteilsdimension wurde ebenfalls nicht signifikant ($F(2, 40) = 0.16, p = .854, \eta^2 = .01$): Ratings auf der Basis peripherer Agency- ($M = 0.79, SD = 1.32$) und Communion-Verhaltensweisen ($M = 1.71, SD = 0.93$) mit zentralem Attribut am Anfang, Ratings auf der Grundlage peripherer Agency- ($M = 0.62, SD = 1.14$) und Communion-Verhaltensweisen ($M = 1.46, SD = 0.90$) mit zentralem Attribut in der Mitte und Ratings auf der Basis peripherer Agency- ($M = 0.39, SD = 1.30$) und Communion-Verhaltensweisen ($M = 1.35, SD = 1.00$) mit zentralem Attribut am Ende hatten jeweils die gleiche Differenz zueinander. Demgegenüber wurde eine signifikante Interaktion von Position und Valenz deutlich ($F(2, 40) = 3.87, p = .029, \eta^2 = .16$): Bei der Kombination von vier negativen mit einer positiven Verhaltensweise, führte die Position der positiven Verhaltensweise am Anfang ($M = 1.21, SD = 0.82$), in der Mitte ($M = 1.25, SD = 0.82$) und am Ende ($M = 0.94, SD = 0.77$) nur zu geringfügig unterschiedlichen Urteilen. Dagegen konnten Ratings auf der Basis peripherer positiver Verhaltensweisen stärker neutralisiert werden, wenn die zentrale negative Verhaltensweise am Ende ($M = 0.80, SD = 0.84$) oder in der Mitte ($M = 0.83, SD = 0.70$) statt am Anfang ($M = 1.29, SD = 0.82$) der Darbietung stand. Mit anderen Worten gesagt, waren negative Halo-Effekte stärker von der Position der entsprechenden Verhaltensweise abhängig als positive Halo-Effekte. Die Interaktion aller drei Faktoren erreichte kein signifikantes Maß ($F(2, 40) = 0.50, p = .611, \eta^2 = .02$).

6.3.3 Diskussion

Gegenstand von Experiment 6 war die Untersuchung qualitativer und quantitativer Unterschiede von Halo-Effekten in Abhängigkeit von Informationsvalenz und Urteilsdimension. Wieder wurden Halo-Effekte als Verzerrungen eindeutiger Urteile betrachtet. Zwar waren in der Baseline-Bedingung nicht alle Urteile objektiv eindeutig, jedoch war dies nicht problematisch für die Ergebnisse. Im Vergleich zum fünften Experiment wurden einige Aspekte des Untersuchungsdesigns geändert: Die Verhaltensweisen wurden nicht mehr gleichzeitig, sondern sequentiell präsentiert. Dabei variierte die Position der zentralen Verhaltensweise. Des Weiteren wurden Verhaltensweisen beider Valenzen und Dimensionen jeweils miteinander verbunden. Schließlich erfolgte die Auswahl der dargebotenen und abgefragten Attribute nicht mehr nach einem festen Schema, sondern komplett randomisiert.

Die Ergebnisse bestätigen die Dimensionshypothese: Communion-Halo-Effekte sind stärker als Agency-Halo-Effekte. Auch werden Halo-Effekte von Agency- und Communion-Attributen gleichermaßen von deren Position beeinflusst. Das spricht dafür, dass keinerlei qualitative Unterschiede zwischen diesen zwei Halo-Effekten liegen. Die Valenzhypothese kann auf qualitativer Ebene belegt werden: Die Position positiver Verhaltensweisen ist irrelevant für die Stärke der resultierenden Halo-Effekte. Demgegenüber zeigen sich bei negativen Halo-Effekten klare Recency-Effekte. Negative Attribute können Urteile umso stärker verzerren, je später sie erscheinen. Diese Befunde sprechen deutlich für die Globalität positiver Halo-Effekte im Kontrast zu den rein spezifischen negativen Halo-Effekten. Dass sich für negative Halo-Effekte gerade Recency- und nicht etwa Primacy-Effekte ergeben, ist durch den Zeitpunkt der Urteilsabfrage zu erklären: Recency-Effekte treten besonders dann auf, wenn die einzelnen Items unmittelbar nach der Präsentation erinnert werden müssen (Howard & Kahana, 2002). Das war in diesem Experiment der Fall. Die Ergebnisse werden zudem dadurch gestärkt, dass Recency-Effekte bei dieser Art der Abfrage unabhängig von der Anzahl der Items (Murdock, 1962) und der Präsentationsrate (Murdock & Metcalfe, 1978) sind.

Bezüglich der quantitativen Valenzunterschiede können die Analysen keine signifikanten Befunde aufzeigen: Positive Halo-Effekte waren in dem Experiment nicht stärker als negativen Halo-Effekte. Das ist höchstwahrscheinlich auf die Instruktion für die Probanden zurückzuführen. Da die Dimension des beurteilten Attributs zur Vermeidung zu vieler Durchgänge zwischen den Versuchsteilnehmern variiert wurde, wurde auch die Instruktion in den beiden Bedingungen angepasst: Es wurde mitgeteilt, dass sich alle abgefragten Skalen auf die Kompetenz beziehungsweise Wärme von Personen beziehen. Dies hatte möglicherweise zur Folge, dass die Probanden in den einzelnen Durchgängen nicht nur die jeweils präsentierten Skalen, sondern auch die übergeordnete Dimension im Gesamten bewerteten. Statt spezifischer Urteile fällten die Teilnehmer also in gewissem Maße große Gesamturteile. Hierfür gelten andere Unterschiede als bei Halo-Effekten. Bereits in Experiment 4 wurde gezeigt, dass positive und negative Gesamteindrücke durchschnittlich gleich stark sind. Um diese Theorie zu bestätigen, muss die gleiche Manipulation mit einer veränderten Instruktion durchgeführt werden.

In Experiment 5 wurde die Frage aufgeworfen, inwiefern Halo-Effekte von der Verzerrungsanfälligkeit der Ratings abzugrenzen sind. Hierfür können die Befunde keine

eindeutige Antwort liefern, da sich signifikante Unterschiede von verschiedenen Halo-Effekten nur bei der zusätzlichen Präsentation von Verhaltensweisen mit gegensätzlicher Valenz und Dimension ergaben. Diese Manipulation war identisch zu der des fünften Experiments. Dass sich beim Vergleich der anderen Darbietungsbedingungen keine Unterschiede zeigten, kann zum Teil auf das allgemeine Fehlen des Valenzeffekts zurückgeführt werden. Auch ist es möglich, dass Attribute, die zu starken Halo-Effekten führen, gleichzeitig selbst sehr robust gegenüber Verzerrungen sind. Setzt man einen solchen Zusammenhang voraus, erscheint es logisch, dass sich die größten Verzerrungen dann ergeben, wenn eine Verhaltensweise mit einem starken Halo-Effekt die Beurteilung einer wenig robusten Eigenschaft beeinflussen soll. Die geringste Verzerrung würde dementsprechend im umgekehrten Fall auftreten. Genau diese Manipulationen sind gegeben, wenn mehrere periphere Verhaltensweisen mit einer zentralen Verhaltensweise von gegensätzlicher Valenz und Dimension kombiniert werden. Somit kann erklärt werden, dass sich gerade in einer solchen Bedingung signifikante Effekte zeigen.

6.4 Zusammenfassung der Experimente 4 bis 6

Im zweiten Experimentalblock wurde der Einfluss von Informationsvalenz und Urteilsdimension auf die Qualität und Quantität von Halo-Effekten fokussiert. Ebenfalls wurde die Robustheit von Halo-Effekten analysiert. Experiment 4 betrachtete Halo-Effekte noch einmal als Vorannahmen, während in den Experimenten 5 und 6 Halo-Effekte als Verzerrungen eindeutiger Urteile untersucht wurden. Die Studien präsentierten fiktive Personen mit verschiedenen Verhaltensweisen, die von den Versuchsteilnehmern hinsichtlich einer Reihe bipolarer Persönlichkeitsskalen bewertet wurden. Die Darbietung der Verhaltensweisen erfolgte parallel oder sequentiell, wobei in letzterem Fall die Position der Verhaltensweisen variiert werden konnte. Des Weiteren wurde die Präsentation von gegensätzlich valenten Bildern oder unterschiedlichen peripheren Verhaltensweisen manipuliert. In Experiment 4 wurden die Halo-Effekte schlicht aus den Mittelwerten von Einzelratings berechnet. In den Experimenten 5 und 6 wurden die Urteile in einer nicht verzerrten Baseline-Bedingung sowie einer Verzerrungs-Bedingung erfasst und die Halo-Effekte über die Differenz der beiden Ratings ermittelt.

Die Valenzhypothese kann anhand von zweien der drei Experimente bestätigt werden: Positive Halo-Effekte sind von größerem Ausmaß und wirken globaler als negative Halo-Effekte. Der qualitative Unterschied ist dadurch erkennbar, dass die Darbietung gegensätzlicher Bilder und die Position der Verhaltensweise einen großen Einfluss auf die Stärke negativer Halo-Effekte haben, jedoch nur zu geringfügigen Unterschieden bei positiven Halo-Effekten führen. Aus den Daten ist auch ersichtlich, dass positive Halo-Effekte, so wie positive Urteile im Allgemeinen, robuster gegenüber Verzerrungen sind als negative Halo-Effekte. Die fehlende Bestätigung der Valenzhypothese in einem der Experimente ist mit einem ungewollten Effekt der Instruktion erklärbar.

Zwei der drei Experimente belegen die Dimensionshypothese: Communion-Halo-Effekte sind im Mittel stärker als Agency-Halo-Effekte. Bei positiver Valenz ist dieser Unterschied, aufgrund der Globalität positiver Informationen, jedoch sehr schwach. Halo-Effekte von Agency- und Communion-Verhaltensweisen weisen aber keine qualitativen Unterschiede zueinander auf, da sie gleichermaßen durch die Darbietung gegensätzlicher Bilder und die Position der Verhaltensweise beeinflusst werden. Jedoch gibt es Anzeichen dafür, dass Communion-Attribute zu robusteren Urteilen beziehungsweise Halo-Effekten führen als Agency-Attribute. Der fehlende Effekt in einem der drei Experimente ist vermutlich auf eine zu geringe Stichprobengröße zurückzuführen.

Anhand eines weiteren abhängigen Maßes ist erkennbar, dass sich in Bezug auf Valenz und Urteilsdimension ein anderes Muster zeigt, wenn statt spezifischer Skalen der Gesamteindruck auf Basis verschiedener Verhaltensweisen gebildet werden muss: Wie bei Halo-Effekten sind Gesamteindrücke auf der Grundlage von Communion-Verhaltensweisen extremer als Gesamteindrücke, die auf der Basis von Agency-Verhaltensweisen gebildet werden. Jedoch sind positive und negative Gesamteindrücke gleichermaßen stark und robust. Das bedeutet, dass zwar auch bei negativer Valenz ein starker Gesamteindruck entstehen kann, dieser jedoch nicht bei Halo-Effekten wirksam wird.

7 Einflüsse von Kontextfaktoren auf Vorannahmen und Verzerrungen

Der Fokus des dritten Experimentalblocks lag auf dem Einfluss von Kontextfaktoren auf die Effekte der Urteilsdimension. Abhängig vom Experiment wurden Halo-Effekte als Vorannahmen oder Verzerrungen betrachtet. Gemäß der dritten zentralen Hypothese sollten Agency-Halo-Effekte im beruflichen Bereich größer sein als in privaten Situationen. Auch wurde postuliert, dass Agency-Halo-Effekte bei naturwissenschaftlich-technischen Berufsgruppen höher sind als bei sozialen und Pflegeberufen. Für Communion-Halo-Effekte sollten umgekehrte Zusammenhänge gelten. Zur Überprüfung der Hypothesen wurden drei Experimente durchgeführt.

7.1 Bedeutung des Berufs der beurteilten Person: Experiment 7

Im siebten Experiment wurde zunächst nur der Beruf der beurteilten Personen als Kontextvariable betrachtet. Dabei wurden naturwissenschaftliche und technische von sozialen und Pflegeberufen unterschieden. Die theoretische Grundlage der Untersuchung war die Tatsache, dass Agency- und Communion-Attribute für Sympathie und Respekt unterschiedlich bedeutsam sind (Wojciszke et al., 2009). Zudem ist die Bedeutsamkeit der beiden Dimensionen von der Beziehung zwischen der eigenen und der beurteilten Person abhängig (Wojciszke & Abele, 2008).

Es wurde vermutet, dass Respekt in naturwissenschaftlichen und technischen Berufen einen wichtigeren Faktor darstellt als in sozialen und Pflegeberufen und für Sympathie der umgekehrte Fall gilt. Zudem sollten gemeinsame Ziele und Erfolg in naturwissenschaftlich-technischen Berufen einen höheren Stellenwert einnehmen und folglich eine stärkere Interdependenz zwischen einzelnen Personen herrschen. Ausgehend von diesen Überlegungen wurde postuliert, dass Agency-Halo-Effekte bei der Beurteilung von Naturwissenschaftlern größer sind als bei der Bewertung von Personen mit sozialen oder Pflegeberufen. Communion-Halo-Effekte sollten dagegen im Kontext von sozialen und Pflegeberufen größer sein als im naturwissenschaftlich-technischen Bereich.

7.1.1 Methode

Teilnehmer

An dem Experiment nahmen 46 (18 weiblich, 28 männlich; 23 in jeder Zwischensubjekt-Bedingung) Mitglieder des Fitnessclubs *Elan* in Eisenberg (Thüringen) teil. Das Alter der Probanden betrug im Mittel 30.59 Jahre ($SD = 11.53$, $Min = 18$, $Max = 63$). Deutsch war für alle 46 Teilnehmer die Muttersprache. Die Stichprobe teilte sich auf in 23 (eine weiblich, 22 männlich; elf in der einen und zwölf in der anderen Zwischensubjekt-Bedingung) Personen mit naturwissenschaftlichen oder technischen Berufen und 23 (17 weiblich, sechs männlich; zwölf in der einen und elf in der anderen Zwischensubjekt-Bedingung) Personen mit sozialen oder Pflegeberufen. Für ihr Mitwirken an der Untersuchung wurden die Probanden mit drei Euro vergütet.

Materialien

Das Experiment basierte auf einem 2 (Berufsgruppe der beurteilten Person) x 2 (Informationsvalenz) x 2 (Urteilsdimension) Design. In diesem Zusammenhang wurde die Berufsgruppe der beurteilten Person als Zwischensubjekt-Variable mit den Ausprägungen *naturwissenschaftliche und technische Berufe* und *soziale und Pflegeberufe* festgelegt. Informationsvalenz (positiv versus negativ) und Urteilsdimension (Agency versus Communion) wurden innerhalb der Versuchsteilnehmer manipuliert.

Als Items wurden die gleichen 24 Verhaltenspaare wie in den Experimenten 3 bis 5 verwendet (siehe Anhang F). Jeweils drei Verhaltensweisen konnten einer Eigenschaft zugeordnet werden, so dass insgesamt acht positive und acht negative Eigenschaften (jeweils vier für Agency und Communion) repräsentiert waren.

Zur Zusammenstellung von geeigneten Berufsbezeichnungen, wurden vorab jeweils 20 Personen gebeten, typische naturwissenschaftliche und technische beziehungsweise soziale und Pflegeberufe zu generieren. Die einzelnen Berufe wurden nach der Häufigkeit ihrer Nennung ausgewertet. Insgesamt wurden jeweils 16 Berufe beider Kategorien für die Untersuchung ausgewählt. Die generierten Berufsbezeichnungen mit den dazugehörigen Häufigkeiten finden sich in Anhang I.

Die Halo-Effekte, in dieser Untersuchung als Vorannahmen betrachtet, wurden aus den Mittelwerten der verschiedenen Einzelratings berechnet. Die Einzelratings selbst

erfolgten auf acht unterschiedlichen bipolaren Skalen, die von -3 bis 3 (zum Beispiel „egoistisch ... hilfsbereit“) reichten. Pro Durchgang wurde folgende Frage gestellt: „Wie würden Sie diese Person auf folgender Dimension einschätzen?“. Jedes der 16 Attribute (jeweils durch eine zufällige Verhaltensweise repräsentiert) wurde auf allen acht Skalen beurteilt. Damit hatte jede Zwischensubjekt-Bedingung 128 (16 x 8) Durchgänge. Wie auch schon in einigen Experimenten zuvor wurde jede Verhaltensweise auch auf der Skala bewertet, auf deren Basis sie ursprünglich generiert worden war. Diese Urteile dienten der Kontrolle der Items und wurden nicht in weitere Berechnungen einbezogen.

Prozedur

Die Probanden wurden in den Räumlichkeiten des Fitnessclubs *Elan* persönlich für das Experiment geworben. Die Durchführung der Untersuchung erfolgte in Form von Einzelsitzungen. Zu diesem Zweck wurde ein gesonderter Raum als Labor eingerichtet. Am Anfang jeder Sitzung unterschrieben die Teilnehmer eine Einverständniserklärung und begannen danach das Experiment. Dabei wurden sie zufällig einer der beiden Zwischensubjekt-Bedingungen zugeteilt.

In der Instruktion wurde den Probanden mitgeteilt, dass sie im Folgenden an einer Bewertungsaufgabe teilnehmen würden, in der ihnen verschiedene Personen dargeboten werden, die jeweils eine bestimmte Verhaltensweise zeigen. Diese Personen würden auf acht Skalen diverser Persönlichkeitsdimensionen bewertet werden müssen. Je nach Zwischensubjekt-Bedingung wurde zudem beschrieben, dass alle Personen verschiedene naturwissenschaftliche oder technische beziehungsweise soziale oder Pflegeberufe haben. Die Probanden wurden darauf hingewiesen, sich bei noch offenen Fragen an die Versuchsleitung zu wenden und anderenfalls das Experiment zu starten.

Der Experimententeil umfasste 128 Durchgänge, die in zufälliger Reihenfolge und ohne Zeitbeschränkung präsentiert wurden. Die Teilnehmer benötigten im Mittel 20 Minuten für die Beantwortung aller Fragen. Jedes der 16 Attribute war zufällig durch eine von drei Verhaltensweisen aus der Datenbasis repräsentiert. Zudem wurde jedes Attribut, entsprechend der Zwischensubjekt-Bedingung, mit einem der 16 naturwissenschaftlich-technischen oder sozialen und Pflegeberufe kombiniert.

Am Ende der Untersuchung gaben die Probanden ihr Geschlecht, ihr Alter und ihren Beruf an. Zudem wurde erfragt, ob Deutsch die Muttersprache, die erste oder die zweite

Fremdsprache sei. Nach Beendigung des Experiments wurden die Teilnehmer über Inhalt und Ziel der Untersuchung aufgeklärt und erhielten ihre Vergütung.

7.1.2 Ergebnisse

Vorab-Analysen

Vor der detaillierten Analyse der Daten wurden die verwendeten Verhaltensweisen einer Kontrolle unterzogen. Hierzu wurde für jede Verhaltensweise das Urteil auf der Dimension betrachtet, aus der die Verhaltensweise ursprünglich generiert worden war. Die Bewertungen sollten bei positiven Verhaltensweisen im positiven Bereich (Skalenwert 1 bis 3) und bei negativen Verhaltensweisen im negativen Bereich (Skalenwert -3 bis -1) liegen. Die Bedingung war bei allen 46 Probanden für alle Verhaltensweisen erfüllt. Somit konnte die komplette Stichprobe in die weiteren Analysen einbezogen werden.

Anschließend wurde getestet, ob die abhängigen Variablen mit Halo-Effekten gleichgesetzt werden können. Ein Urteil wurde dann als Halo-Effekt betrachtet, wenn es signifikant vom Mittelpunkt der Skala abwich. Zu diesem Zweck wurden alle abhängigen Variablen durch t-Tests gegen den Wert 0 getestet. Die Variable *negativer Halo-Effekt bei ungleicher Dimension* unterschied sich nicht signifikant vom Wert 0 ($p > .050$). Alle anderen abhängigen Variablen wiesen signifikante Unterschiede zum Skalenmittelpunkt auf. Aus diesem Grund wurde der Faktor *Gleichheit von dargebotener und beurteilter Dimension* (gleiche versus ungleiche Dimension) in diesem Experiment nicht betrachtet.

Da positive und negative Halo-Effekte auf derselben Skala erfasst wurden, wurden vor der Überprüfung der Hypothesen alle Ratings, die auf negativen Verhaltensweisen basierten, durch eine Multiplikation mit -1 umkodiert.

Generalisierbarkeit der bisherigen Ergebnisse

Die erste Analyse betraf noch einmal die Effekte von Informationsvalenz und Urteilsdimension auf die Stärke von Halo-Effekten. Da in diesem Experiment erstmals keine Studenten, sondern im Berufsleben stehende Probanden untersucht wurden, sollten die bisher gezeigten Ergebnisse auf diese Weise über verschiedene Personengruppen hinweg generalisiert werden. Zu diesem Zweck wurde eine 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung durchgeführt. Der Haupteffekt für Valenz konnte

signifikant bestätigt werden ($F(1, 45) = 121.21, p < .001, \eta^2 = .73$): Positive Halo-Effekte ($M = 1.11, SD = 0.42$) waren über die Zwischensubjekt-Bedingungen hinweg stärker als negative Halo-Effekte ($M = 0.31, SD = 0.39$). Es war dagegen kein signifikanter Haupteffekt der Urteilsdimension ersichtlich ($F(1, 45) = 3.22, p = .079, \eta^2 = .07$): Agency- ($M = 0.67, SD = 0.36$) und Communion-Halo-Effekte ($M = 0.75, SD = 0.35$) unterschieden sich nicht voneinander. Die Interaktion von Informationsvalenz und Urteilsdimension erreichte ein signifikantes Niveau ($F(1, 45) = 14.40, p < .001, \eta^2 = .24$). Zur genaueren Analyse der Daten wurden Agency- und Communion-Halo-Effekte anschließend, getrennt für beide Valenzbedingungen, mit Hilfe von t-Tests miteinander verglichen: Halo-Effekte positiver Agency- ($M = 1.15, SD = 0.45$) und Communion-Verhaltensweisen ($M = 1.07, SD = 0.48$) waren im Mittel gleich stark ($t(45) = 1.21, p = .232$). Im Gegensatz dazu waren negative Communion-Halo-Effekte ($M = 0.43, SD = 0.41$) gegenüber negativen Agency-Halo-Effekten ($M = 0.19, SD = 0.47$) von signifikant höherem Ausmaß ($t(45) = 3.96, p < .001$).

Auch die Beeinflussbarkeit der Vorannahmen über Agency- und Communion-Eigenschaften wurde an dieser Stichprobe untersucht. Dies geschah unter der Verwendung einer 2×2 ANOVA (Informationsvalenz \times Dimension der beurteilten Eigenschaft) mit Messwiederholung. Neben dem schon geschilderten Valenzeffekt ergab sich auch ein signifikanter Haupteffekt für die Dimension der beurteilten Eigenschaft ($F(1, 45) = 9.87, p = .003, \eta^2 = .18$): Vorannahmen über Agency-Eigenschaften ($M = 0.77, SD = 0.35$) wurden durchschnittlich stärker durch Halo-Effekte geprägt als Vorannahmen über Communion-Eigenschaften ($M = 0.65, SD = 0.34$). Ebenfalls signifikant wurde die Interaktion der beiden Faktoren ($F(1, 45) = 101.36, p < .001, \eta^2 = .69$): Positive Halo-Effekte prägten Agency-Attribute ($M = 1.35, SD = 0.46$) stärker als Communion-Attribute ($M = 0.87, SD = 0.44$). Negative Halo-Effekte hatten dagegen auf Agency-Attribute ($M = 0.19, SD = 0.45$) geringere Einflüsse als auf Communion-Attribute ($M = 0.43, SD = 0.42$).

Einfluss der Berufsgruppe der beurteilten Person

Um den Einfluss der Berufsgruppe der beurteilten Person auf Halo-Effekte zu untersuchen, wurde eine $2 \times 2 \times 2$ ANOVA (Berufsgruppe der beurteilten Person \times Informationsvalenz \times Urteilsdimension) mit Messwiederholung durchgeführt. Die Haupteffekte von Informationsvalenz und Urteilsdimension und die Interaktion dieser Faktoren waren aufgrund des neu hinzugekommenen Zwischensubjekt-Faktors nicht

identisch mit den bereits geschilderten Befunden. Da die Unterschiede jedoch nur geringfügiger Natur waren, wurden die drei Effekte nicht erneut betrachtet.

Es zeigte sich kein signifikanter Haupteffekt der Berufsgruppe ($F(1, 44) = 0.12, p = .733, \eta^2 < .01$): Halo-Effekte bei der Beurteilung von Naturwissenschaftlern und Technikern ($M = 0.72, SD = 0.39$) waren genauso stark wie die Halo-Effekte bei der Bewertung von Personen mit sozialen und Pflegeberufen ($M = 0.69, SD = 0.24$). Die Interaktion von Berufsgruppe und Valenz wurde auch nicht signifikant ($F(1, 44) = 0.86, p = .358, \eta^2 = .02$): Positive ($M = 1.09, SD = 0.45$) und negative Halo-Effekte ($M = 0.36, SD = 0.45$) in der naturwissenschaftlichen Berufsbedingung waren nicht unterschiedlicher zueinander als positive ($M = 1.13, SD = 0.39$) und negative Halo-Effekte ($M = 0.26, SD = 0.33$) in der Bedingung mit sozialen und Pflegeberufen. Die Interaktion von Berufsgruppe und Urteilsdimension erreichte kein signifikantes Maß ($F(1, 44) = 0.77, p = .385, \eta^2 = .02$): Agency-Halo-Effekte bei der Beurteilung naturwissenschaftlicher ($M = 0.66, SD = 0.45$) und sozialer Berufe ($M = 0.67, SD = 0.25$) wiesen keine andere Differenz zueinander auf als Communion-Halo-Effekte bei der Bewertung naturwissenschaftlicher ($M = 0.78, SD = 0.37$) und sozialer Berufe ($M = 0.71, SD = 0.34$). Schließlich wurde auch die Interaktion aller drei Faktoren nicht signifikant ($F(1, 44) = 0.14, p = .706, \eta^2 < .01$).

Einfluss der Berufsgruppe des Beurteilenden

Da sich keine signifikanten Ergebnisse bezüglich des Berufs der beurteilten Person gezeigt hatten, wurde in einer zusätzlichen Analyse die Berufsgruppe des Beurteilenden als unabhängige Variable betrachtet. Dazu wurde eine $2 \times 2 \times 2$ ANOVA (Berufsgruppe des Beurteilenden \times Informationsvalenz \times Urteilsdimension) mit Messwiederholung durchgeführt. Erneut unterschieden sich die Haupteffekte von Informationsvalenz und Urteilsdimension und die Interaktion der beiden Faktoren nur geringfügig von den bereits geschilderten Ergebnissen und wurden daher nicht noch einmal betrachtet.

Es war kein signifikanter Haupteffekt der Berufsgruppe des Beurteilenden zu erkennen ($F(1, 44) = 0.16, p = .691, \eta^2 < .01$): Teilnehmer mit naturwissenschaftlichen und technischen Berufen ($M = 0.69, SD = 0.25$) waren im selben Maße für Halo-Effekte anfällig wie Probanden mit sozialen und Pflegeberufen ($M = 0.73, SD = 0.39$). Die Interaktion von Berufsgruppe und Valenz wurde ebenfalls nicht signifikant ($F(1, 44) = 1.89, p = .176, \eta^2 = .04$): Positive ($M = 1.04, SD = 0.44$) und negative Halo-Effekte ($M = 0.34, SD = 0.21$) bei

Probanden mit naturwissenschaftlich-technischen Berufen wiesen die gleiche Differenz zueinander auf wie positive ($M = 1.18, SD = 0.39$) und negative Halo-Effekte ($M = 0.28, SD = 0.52$) bei Teilnehmern mit sozialen und Pflegeberufen. Auch konnte keine signifikante Interaktion von Berufsgruppe und Urteilsdimension gezeigt werden ($F(1, 44) = 0.48, p = .490, \eta^2 = .01$): Agency-Halo-Effekte bei Probanden mit naturwissenschaftlichen ($M = 0.63, SD = 0.27$) und sozialen Berufen ($M = 0.70, SD = 0.44$) waren genauso unterschiedlich zueinander wie Communion-Halo-Effekte bei Probanden mit naturwissenschaftlichen ($M = 0.74, SD = 0.32$) und sozialen Berufen ($M = 0.75, SD = 0.39$). Die Interaktion aller drei Faktoren wurde nicht signifikant ($F(1, 44) = 2.69, p = .108, \eta^2 = .06$).

7.1.3 Diskussion

Experiment 7 untersuchte den Einfluss der Berufsgruppe der beurteilten Person auf die Stärke von Agency- und Communion-Halo-Effekten. Zudem sollten die Ergebnisse der bisherigen Experimente mit der neuen Stichprobe repliziert werden. Die Daten bestätigen erneut, dass positive Halo-Effekte stärker sind als negative Halo-Effekte. Auch sind negative Communion-Halo-Effekte von größerem Ausmaß als negative Agency-Halo-Effekte. Bei positiver Valenz existiert dagegen kein quantitativer Unterschied in Abhängigkeit von der Urteilsdimension. In Bezug auf die Beeinflussbarkeit können Agency-Urteile leichter positiv und Communion-Urteile leichter negativ geprägt werden. Im Vergleich zu den bisherigen Ergebnissen waren Agency-Urteile in dieser Studie besonders extrem, so dass sie im Mittel sogar signifikant extremer waren als Communion-Urteile. Dies ist vermutlich auf die Darbietung der Berufsbezeichnungen zurückzuführen, die mit einer allgemein höheren Bedeutsamkeit von Agency-Attributen einherging. Dass alle Befunde der ersten beiden Experimentalblöcke repliziert werden konnten, spricht für die Generalisierbarkeit der bisherigen Ergebnisse über verschiedenartige Stichproben hinweg.

Für den Einfluss der Berufsgruppe konnte die Kontexthypothese nicht bestätigt werden: In den Daten waren Halo-Effekte von Agency- und Communion-Verhaltensweisen bei der Bewertung von Personen mit naturwissenschaftlichen und technischen Berufen genauso groß wie bei der Beurteilung von Menschen in sozialen und Pflegeberufen. Es existierte folglich keine signifikante Interaktion der beiden Faktoren. Das Ergebnis kann möglicherweise darauf zurückgeführt werden, dass sich unterschiedliche Halo-Effekte für

Agency- und Communion-Verhaltensweisen nur im unmittelbaren Vergleich verschiedener Berufsgruppen zeigen. Zudem könnte die Teststärke eines Zwischensubjekt-Designs schlichtweg zu gering sein, um eventuelle Unterschiede aufzuzeigen. Um diese Frage zu klären, ist es notwendig, die Berufsbezeichnungen der fiktiven Personen als Innersubjekt-Faktor zu konzipieren. Eine weitere Erklärung für den fehlenden Effekt wäre die Tatsache, dass Berufsbezeichnungen allein keinen beruflichen Kontext induzieren. Die Probanden könnten die Berufe der fiktiven Personen zwar zur Kenntnis genommen, aber als irrelevant für die Interaktion betrachtet haben. Daher sollte die Situation, in der die verschiedenen Urteile erfolgen sollen, künftig explizit manipuliert werden. Auch ist es notwendig, die Berufsgruppe der Versuchsteilnehmer noch einmal näher zu betrachten. Zwar wurde der Beruf auch in dieser Untersuchung schon erfasst und als Faktor in die Datenanalyse integriert, jedoch wurde er nicht gezielt manipuliert. So wurden beispielsweise auch kaufmännische und Dienstleistungsberufe, aufgrund des häufigen Umgangs mit Menschen, den sozialen Berufen zugeordnet. Um diesen Faktor einschließen zu können, sollte dementsprechend eine sehr spezifische Auswahl der Probanden erfolgen.

7.2 Bedeutung der Situation und des Beurteilenden: Experiment 8

Auch im achten Experimente wurde der Einfluss unterschiedlicher Kontextfaktoren auf die Stärke von Agency- und Communion-Halo-Effekten untersucht. Neben der Berufsgruppe der beurteilten Person wurde dieses Mal auch die Situation, in der sich die beschriebenen Verhaltensweisen abspielen, variiert. Zusätzlich erfolgte eine gezielte Auswahl der Stichprobe: Probanden mit naturwissenschaftlichen und technischen Berufen wurden von Versuchsteilnehmer mit sozialen und Pflegeberufen unterschieden.

Wie bereits in Experiment 7 wurde vermutet, dass in naturwissenschaftlichen und technischen Berufen eine größere Interdependenz zwischen den Personen herrscht und Respekt eine größere Rolle spielt als in sozialen und Pflegeberufen. Umgekehrt sollte in sozialen und Pflegeberufen zwischenmenschliche Sympathie von höherer Bedeutung sein. Auch zwischen verschiedenen Situationen sollte ein derartiger Unterschied existieren: In einer beruflichen Situation sollte eine stärkere Interdependenz herrschen und Respekt wichtiger sein als in einer rein privaten Situation. In dieser Situation sollte wiederum

Sympathie von größerer Bedeutsamkeit sein. Aufgrund dieser Unterschiede wurde angenommen, dass Halo-Effekte von Agency-Verhaltensweisen in beruflichen Situationen und bei der Bewertung von Menschen mit naturwissenschaftlichen oder technischen Berufen größer sind als in privaten Situationen beziehungsweise bei der Beurteilung von Personen mit einem sozialen oder Pflegeberuf. Für Halo-Effekte von Communion-Verhaltensweisen sollte der umgekehrte Fall gelten. Eine weitere Hypothese war, dass Probanden, die selbst einen naturwissenschaftlichen oder technischen Beruf ausüben, für Agency-Halo-Effekte anfälliger und für Communion-Halo-Effekte weniger anfällig sind als Probanden mit sozialen oder Pflegeberufen.

7.2.1 Methode

Teilnehmer

Insgesamt 84 (40 weiblich, 44 männlich; 21 in jeder Zwischensubjekt-Bedingung) Mitglieder des Fitnessclubs *Elan* in Eisenberg (Thüringen) nahmen an der Studie teil. Das Alter der Probanden betrug im Durchschnitt 39.50 Jahre ($SD = 13.40$, $Min = 18$, $Max = 64$). Deutsch war für alle 84 Teilnehmer die Muttersprache. Keiner der Probanden hatte an Experiment 7 teilgenommen. Die Stichprobe unterteilte sich in 42 (fünf weiblich, 37 männlich; 21 in jeder Situations-Bedingung) Personen mit naturwissenschaftlichen und technischen sowie 42 (35 weiblich, sieben männlich; 21 in jeder Situations-Bedingung) Personen mit sozialen und Pflegeberufen. Die Probanden mit naturwissenschaftlichen und technischen Berufen waren durchschnittlich 35.55 Jahre ($SD = 12.93$, $Min = 18$, $Max = 62$), die Probanden mit sozialen und Pflegeberufen im Mittel 43.45 Jahre ($SD = 12.81$, $Min = 20$, $Max = 64$) alt. Für ihre Teilnahme erhielten die Probanden drei Euro.

Materialien

Die Grundlage der Untersuchung war ein 2 (Situation) x 2 (Berufsgruppe des Beurteilenden) x 2 (Berufsgruppe der beurteilten Person) x 2 (Informationsvalenz) x 2 (Urteilsdimension) Design. Die Situation wurde zwischen den Versuchsteilnehmern manipuliert und hatte die Ausprägungen *beruflich* und *privat*. Die Berufsgruppe des Beurteilenden war ein Zwischensubjekt-Faktor mit den Stufen *naturwissenschaftliche und technische Berufe* und *soziale und Pflegeberufe*. Die Berufsgruppe der beurteilten Person

(naturwissenschaftliche und technische Berufe versus soziale und Pflegeberufe) sowie Informationsvalenz (positiv versus negativ) und Urteilsdimension (Agency versus Communion) wurden innerhalb der Teilnehmer manipuliert.

Als Items wurden 18 Verhaltenspaare aus der Datenbasis verwendet, die insgesamt sechs positive und sechs negative Eigenschaften (jeweils drei für Agency und Communion) repräsentierten. Gegenüber dem vorherigen Experiment wurden zwei Eigenschaften (und damit sechs Verhaltenspaare) aus dem Design ausgeschlossen, um eine zu hohe Anzahl an Untersuchungsdurchgängen (da die Berufsgruppe der beurteilten Person nun ein Innersubjekt-Faktor war) zu vermeiden. Die Wahl erfolgte zufällig durch das Ziehen zweier Zahlen und fiel auf die Paare *selbständig-abhängig* und *gesellig-eigenbrötlerisch*. Die Berufsbezeichnungen waren identisch zu denen in Experiment 7 (siehe Anhang I).

Die einzelnen Ratings, die zur Berechnung der Halo-Effekte dienten, wurden auf sechs bipolaren Skalen von -3 bis 3 (zum Beispiel „*chaotisch ... ordentlich*“) erfasst. In allen Durchgängen lautete die Frage „*Wie würden Sie diese Person auf folgender Dimension einschätzen?*“. Jedes der zwölf Attribute (jeweils in Form einer zufälligen Verhaltensweise repräsentiert) wurde je einmal mit einem naturwissenschaftlichen oder technischen Beruf und einem sozialen oder Pflegeberuf gepaart. Die resultierenden 24 Verhaltensweisen wurden auf allen sechs Skalen beurteilt. Somit hatte jede Zwischensubjekt-Bedingung 144 (12 x 6 x 2) Durchgänge. Wie schon zuvor wurde jede Verhaltensweise auch auf der Skala bewertet, auf deren Basis sie ursprünglich generiert worden war. Diese Urteile dienten der Kontrolle der Items und wurden nicht in weitere Berechnungen einbezogen.

Prozedur

Die Teilnehmer wurden in den Räumlichkeiten des Fitnessclubs *Elan* persönlich für die Untersuchung geworben. Dabei wurde der Beruf des Probanden vorab erfragt, um sie der richtigen Gruppe zuordnen zu können. Die Durchführung des Experiments erfolgte in Form von Einzelsitzungen. Zu diesem Zweck wurde ein gesonderter Raum als Labor eingerichtet. Zu Beginn jeder Experimentalsitzung unterschrieben die Teilnehmer eine Einverständniserklärung und starteten danach das Experiment. Die Zuteilung zu einer der beiden Bedingungen des Zwischensubjekt-Faktors *Situation* erfolgte zufällig.

In der Instruktion wurde den Probanden mitgeteilt, dass sie im Folgenden an einer Bewertungsaufgabe teilnehmen würden, in der ihnen unterschiedliche Personen mit

verschiedenen Berufen präsentiert werden, die jeweils eine bestimmte Verhaltensweise zeigen. Diese Personen würden auf sechs Persönlichkeitsskalen bewertet werden müssen. Abhängig von der Zwischensubjekt-Bedingung wurden die Probanden gebeten, sich vorzustellen, dass sie die Personen beruflich (als Kollegen) beziehungsweise privat (als Bekannte) kennenlernen und die jeweils dargebotenen Verhaltensweisen an ihnen beobachten. Die Teilnehmer wurden darauf hingewiesen, sich bei noch offenen Fragen an die Versuchsleitung zu wenden und anderenfalls das Experiment zu starten.

Der Experimentaltel bestand aus 144 Durchgängen, die in einer zufälligen Reihenfolge und ohne Zeitlimit präsentiert wurden. Die Teilnehmer benötigten im Durchschnitt 25 Minuten für die Bearbeitung des Experimentaltels. Alle zwölf Attribute wurden jeweils zufällig mit einem der 16 naturwissenschaftlichen und technischen sowie einem der 16 sozialen und Pflegeberufe kombiniert. Bei beiden Kombinationen war das Attribut durch eine zufällige Verhaltensweise aus der Datenbasis repräsentiert. Die Programmierung verhinderte, dass die gleiche Verhaltensweise eines Attributs gleichzeitig mit beiden Berufsgruppen in Verbindung stand.

Am Ende der Untersuchung gaben die Teilnehmer ihr Geschlecht und ihr Alter an. Zudem wurde erfragt, ob Deutsch die Muttersprache, die erste oder die zweite Fremdsprache sei. Nach Beendigung des Experiments wurden die Teilnehmer über Inhalt und Ziel der Untersuchung aufgeklärt und erhielten ihre Vergütung.

7.2.2 Ergebnisse

Vorab-Analysen

Bevor die Daten genauen Analysen unterzogen wurden, erfolgte zunächst eine Kontrolle der verwendeten Verhaltensweisen. Hierzu wurde für jede Verhaltensweise das Urteil auf der Dimension betrachtet, aus der die Verhaltensweise ursprünglich generiert worden war. Die Bewertungen sollten bei positiven Verhaltensweisen im positiven Bereich (Skalenwert 1 bis 3) und bei negativen Verhaltensweisen im Negativbereich (Skalenwert -3 bis -1) liegen. Die Bedingung war bei allen Probanden für alle Verhaltensweisen erfüllt. Damit mussten keine Teilnehmer von der Datenanalyse ausgeschlossen werden.

Anschließend wurde getestet, ob die abhängigen Variablen mit Halo-Effekten gleichgesetzt werden können. Ein Urteil wurde als Halo-Effekt betrachtet, wenn es sich

signifikant vom Mittelpunkt der Skala unterschied. Dafür wurden alle abhängigen Variablen mit Hilfe von t-Tests gegen den Wert 0 getestet. Die Variable *negativer Halo-Effekt bei ungleicher Dimension* unterschied sich in allen Zwischensubjekt-Kombinationen nicht signifikant vom Wert 0 ($p > .050$). Alle anderen abhängigen Variablen zeigten signifikante Unterschiede zum Skalenmittelpunkt. Aus diesem Grund wurde der Faktor *Gleichheit von dargebotener und beurteilter Dimension* (gleiche versus ungleiche Dimension) in dieser Untersuchung nicht betrachtet.

Da positive und negative Halo-Effekte auf der gleichen Skala erhoben wurden, wurden vor der Überprüfung der Hypothesen alle Ratings, die auf negativen Verhaltensweisen basierten, durch eine Multiplikation mit -1 umkodiert.

Einflüsse von Informationsvalenz und Urteilsdimension

Die Analyse der Effekte von Informationsvalenz und Urteilsdimension geschah mit Hilfe einer 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung. Es ergab sich ein signifikanter Haupteffekt für Valenz ($F(1, 83) = 88.90, p < .001, \eta^2 = .52$): Positive Halo-Effekte ($M = 1.10, SD = 0.42$) waren über alle Zwischensubjekt-Bedingungen hinweg stärker als negative Halo-Effekte ($M = 0.32, SD = 0.46$). Der Haupteffekt der Urteilsdimension erreichte kein signifikantes Niveau ($F(1, 83) = 2.93, p = .091, \eta^2 = .03$): Agency- ($M = 0.68, SD = 0.25$) und Communion-Halo-Effekte ($M = 0.74, SD = 0.29$) unterschieden sich in ihrer Stärke nicht voneinander. Auch die Interaktion von Valenz und Urteilsdimension wurde nicht signifikant ($F(1, 83) = 1.66, p = .202, \eta^2 = .02$). Aus diesem Grund wurden Agency- und Communion-Halo-Effekte anschließend, getrennt für beide Valenzbedingungen, auf Basis von t-Tests miteinander verglichen: Positive Agency- ($M = 1.09, SD = 0.54$) und Communion-Halo-Effekte ($M = 1.10, SD = 0.39$) waren im Durchschnitt gleich stark ($t(83) = 0.39, p = .697$). Demgegenüber waren negative Communion-Halo-Effekte ($M = 0.37, SD = 0.54$) gegenüber negativen Agency-Halo-Effekten ($M = 0.27, SD = 0.46$) von signifikant höherem Ausmaß ($t(83) = 2.11, p = .038$).

Einfluss der Berufsgruppe der beurteilten Person

Zur Untersuchung des Einflusses der Berufsgruppe der zu beurteilenden Person auf die unterschiedlichen Halo-Effekte wurde eine 2 x 2 x 2 ANOVA (Berufsgruppe der beurteilten Person x Informationsvalenz x Urteilsdimension) mit Messwiederholung

durchgeführt. Die Haupteffekte von Informationsvalenz und Urteilsdimension und deren Interaktion wurden bereits im letzten Abschnitt betrachtet. Der Haupteffekt der Berufsgruppe wurde nicht signifikant bestätigt ($F(1, 83) = 0.21, p = .647, \eta^2 < .01$): Halo-Effekte bei der Beurteilung von Menschen mit naturwissenschaftlichen und technischen Berufen ($M = 0.72, SD = 0.26$) und bei der Bewertung von Personen in sozialen und Pflegeberufen ($M = 0.70, SD = 0.27$) unterschieden sich nicht voneinander.

Es konnte keine signifikante Interaktion von Berufsgruppe und Valenz deutlich gemacht werden ($F(1, 83) = 0.86, p = .357, \eta^2 = .01$): Positive ($M = 1.09, SD = 0.38$) und negative Halo-Effekte ($M = 0.34, SD = 0.54$) bei der Bewertung von Naturwissenschaftlern und Technikern hatten die gleiche Differenz zueinander wie positive ($M = 1.10, SD = 0.53$) und negative Halo-Effekte ($M = 0.30, SD = 0.45$) bei der Beurteilung von Personen mit sozialen Berufen. Jedoch zeigte sich eine signifikante Interaktion von Berufsgruppe und Urteilsdimension ($F(1, 83) = 14.20, p < .001, \eta^2 = .15$): Agency-Halo-Effekte waren bei der Beurteilung naturwissenschaftlich-technischer Berufe ($M = 0.75, SD = 0.30$) stärker als bei der Bewertung von sozialen und Pflegeberufen ($M = 0.61, SD = 0.37$). Im Gegensatz dazu waren Communion-Halo-Effekte bei der Beurteilung naturwissenschaftlicher Berufe ($M = 0.68, SD = 0.33$) schwächer als bei der Bewertung sozialer Berufe ($M = 0.79, SD = 0.37$). Die Dreifachinteraktion wurde nicht signifikant ($F(1, 83) = 1.41, p = .239, \eta^2 = .02$).

Einfluss der Situation

Die Untersuchung potentieller Einflüsse der Situation, in der die verschiedenen Einschätzungen vorgenommen werden sollen, wurde anhand einer $2 \times 2 \times 2$ ANOVA (Situation x Informationsvalenz x Urteilsdimension) durchgeführt. Die Haupteffekte von Informationsvalenz und Urteilsdimension sowie die Interaktion der beiden Faktoren waren aufgrund des neu hinzugekommenen Zwischensubjekt-Faktors nicht identisch mit den bereits geschilderten Befunden. Da die Unterschiede jedoch nur gering waren, wurden die drei Effekte nicht erneut betrachtet.

Es war kein signifikanter Haupteffekt der Situation erkennbar ($F(1, 82) = 0.11, p = .744, \eta^2 < .01$): Die durchschnittlichen Halo-Effekte in beruflichen ($M = 0.70, SD = 0.21$) und privaten Situationen ($M = 0.72, SD = 0.23$) hatten das gleiche Ausmaß. Die Interaktion von Berufsgruppe und Valenz wurde ebenfalls nicht signifikant ($F(1, 82) < 0.01, p = .970, \eta^2 < .01$): Positive ($M = 1.09, SD = 0.39$) und negative Halo-Effekte ($M = 0.31, SD = 0.42$) im

Kontext beruflicher Situationen hatten die gleiche Differenz zueinander wie positive ($M = 1.10$, $SD = 0.45$) und negative Halo-Effekte ($M = 0.33$, $SD = 0.49$) in privaten Situationen. Es gab auch keinen signifikanten Effekt der Interaktion *Situation x Urteilsdimension* ($F(1, 82) = 0.10$, $p = .756$, $\eta^2 < .01$): Agency-Halo-Effekte in beruflichen ($M = 0.68$, $SD = 0.24$) und privaten Situationen ($M = 0.68$, $SD = 0.26$) waren nicht unterschiedlicher zueinander als Communion-Halo-Effekte in beruflichen ($M = 0.72$, $SD = 0.31$) und privaten Situationen ($M = 0.75$, $SD = 0.27$). Die Interaktion der drei Faktoren *Situation, Informationsvalenz und Urteilsdimension* wurde nicht signifikant ($F(1, 82) = 0.61$, $p = .436$, $\eta^2 = .01$).

Einfluss der Berufsgruppe des Beurteilenden

Um den Einfluss des Berufs des Probanden zu untersuchen, wurde eine $2 \times 2 \times 2$ ANOVA (Berufsgruppe des Beurteilenden x Informationsvalenz x Urteilsdimension) mit Messwiederholung berechnet. Erneut unterschieden sich die Haupteffekte und die Interaktion der Faktoren *Informationsvalenz und Urteilsdimension* nur geringfügig von den bereits geschilderten Ergebnissen und wurden daher nicht noch einmal betrachtet.

Der Haupteffekt der Berufsgruppe des Beurteilenden erreichte kein signifikantes Niveau ($F(1, 82) = 0.01$, $p = .919$, $\eta^2 < .01$): Probanden mit naturwissenschaftlichen und technischen Berufen ($M = 0.71$, $SD = 0.25$) zeigten dasselbe Ausmaß an Halo-Effekten wie Teilnehmer mit sozialen und Pflegeberufen ($M = 0.71$, $SD = 0.19$). Es war keine signifikante Interaktion von Berufsgruppe und Valenz erkennbar ($F(1, 82) = 0.01$, $p = .935$, $\eta^2 < .01$): Positive ($M = 1.09$, $SD = 0.44$) und negative Halo-Effekte ($M = 0.33$, $SD = 0.48$) bei Teilnehmern mit naturwissenschaftlichen und technischen Berufen wiesen keine andere Differenz zueinander auf als positive ($M = 1.10$, $SD = 0.39$) und negative Halo-Effekte ($M = 0.32$, $SD = 0.44$) bei Probanden mit sozialen und Pflegeberufen. Die Interaktion von Berufsgruppe und Urteilsdimension wurde ebenfalls nicht signifikant ($F(1, 82) = 0.20$, $p = .657$, $\eta^2 < .01$): Agency-Halo-Effekte bei Probanden mit naturwissenschaftlich-technischen ($M = 0.67$, $SD = 0.26$) und sozialen beziehungsweise Pflegeberufen ($M = 0.68$, $SD = 0.24$) waren im Mittel nicht unterschiedlicher zueinander als Communion-Halo-Effekte bei Teilnehmern mit naturwissenschaftlichen ($M = 0.75$, $SD = 0.32$) und sozialen Berufen ($M = 0.73$, $SD = 0.26$). Es ergab sich keine signifikante Dreifachinteraktion von Berufsgruppe, Informationsvalenz und Urteilsdimension ($F(1, 82) < 0.01$, $p = .975$, $\eta^2 < .01$).

Interaktionen aller Faktoren

Da sich in den Daten keine signifikanten Einflüsse der beiden Zwischensubjekt-Faktoren gezeigt hatten, wurden alle unabhängigen Variablen in einer einzigen Analyse zusammengefasst. Das Ziel dieses Vorgehens war es, mögliche Einflüsse der zwei Faktoren aufzuzeigen, die in den vorherigen Analysen noch nicht berücksichtigt worden waren. Zu diesem Zweck wurde eine 2 x 2 x 2 x 2 x 2 ANOVA (Situation x Berufsgruppe des Beurteilenden x Berufsgruppe der beurteilten Person x Informationsvalenz x Urteilsdimension) mit Messwiederholung durchgeführt. Da sich nun zwei Zwischensubjekt-Faktoren in dem Design befanden, waren die einzelnen Ergebnisse nicht identisch zu den bereits geschilderten Befunden. Jedoch waren die Abweichungen sehr gering, so dass auf eine erneute Betrachtung dieser Effekte verzichtet wurde.

Neben den schon aufgeführten Befunden gab es keine signifikante Interaktion von Situation und Berufsgruppe des Beurteilenden ($F(1, 80) = 0.01, p = .941, \eta^2 < .01$): Halo-Effekte in beruflichen ($M = 0.70, SD = 0.26$) und privaten Situationen ($M = 0.72, SD = 0.26$) bei Teilnehmern mit naturwissenschaftlichen und technischen Berufen waren nicht unterschiedlicher zueinander als die Halo-Effekte in beruflichen ($M = 0.70, SD = 0.16$) und privaten Situationen ($M = 0.71, SD = 0.21$) bei Probanden mit sozialen Berufen. Die Interaktion von Situation und Berufsgruppe der beurteilten Person erreichte auch kein signifikantes Niveau ($F(1, 80) = 0.02, p = .897, \eta^2 < .01$): Halo-Effekte in beruflichen ($M = 0.71, SD = 0.21$) und privaten Situationen ($M = 0.73, SD = 0.31$) bei der Bewertung von Naturwissenschaftlern wiesen die gleiche Differenz zueinander auf wie die Halo-Effekte in beruflichen ($M = 0.70, SD = 0.28$) und privaten Situationen ($M = 0.71, SD = 0.27$) bei der Bewertung von Personen mit sozialen Berufen. Schließlich wurde ebenfalls die Interaktion von Berufsgruppe des Beurteilenden und Berufsgruppe der beurteilten Person nicht signifikant ($F(1, 80) = 0.03, p = .873, \eta^2 < .01$): Halo-Effekte bei der Beurteilung naturwissenschaftlicher ($M = 0.72, SD = 0.29$) und sozialer Berufe ($M = 0.71, SD = 0.30$) bei Naturwissenschaftlern und Technikern waren nicht unterschiedlicher zueinander als die Halo-Effekte bei der Bewertung naturwissenschaftlicher ($M = 0.72, SD = 0.23$) und sozialer Berufe ($M = 0.70, SD = 0.25$) bei Teilnehmern mit sozialen und Pflegeberufen. Alle Vierfachinteraktionen sowie auch die Fünffachinteraktion erreichten kein signifikantes Niveau ($p > .050$). Daher wurde auf eine weitere Analyse der Daten verzichtet.

7.2.3 Diskussion

Gegenstand von Experiment 8 war die Analyse von Agency- und Communion-Halo-Effekten in Abhängigkeit von verschiedenen Kontexteffekten. Dabei wurden sowohl die Situation, in der sich die Verhaltensweisen und Beurteilungen abspielen, als auch die Berufsgruppe der beurteilten Person sowie des Probanden betrachtet. Auch die Einflüsse von Informationsvalenz und Urteilsdimension wurden noch einmal untersucht. Die Daten machen deutlich, dass positive Halo-Effekte über alle Bedingungen hinweg von größerem Ausmaß sind als negative Halo-Effekte. Zudem sind negative Communion-Halo-Effekte im Mittel stärker als negative Agency-Halo-Effekte, wogegen es bei positiver Valenz keinen Unterschied gibt. Alle bisherigen Ergebnisse wurden damit erneut repliziert.

Die Analysen zeigen, dass die Berufsgruppe des Beurteilten signifikant mit der Urteilsdimension interagiert: Agency-Halo-Effekte sind bei der Einschätzung von Personen mit naturwissenschaftlichen und technischen Berufen stärker als bei der Beurteilung von Menschen mit sozialen und Pflegeberufen. Im Gegensatz dazu gilt für Communion-Halo-Effekte das umgekehrte Muster. Die Kontexthypothese kann somit für diese Variable bestätigt werden. In dem Zusammenhang stellt sich die Frage, warum der Effekt zwar in diesem Experiment, nicht aber in Experiment 7 zu erkennen war. Zum einen kann diese Diskrepanz auf die höhere Teststärke eines Innersubjekt-Designs gegenüber einem Zwischensubjekt-Design zurückzuführen sein. Zum anderen ist es möglich, dass die verschiedenen Berufskategorien nur im direkten Vergleich salient werden und zu unterschiedlichen Effekten führen. Diese Erklärung wird durch die Tatsache gestützt, dass Urteile von Natur aus Vergleiche beinhalten (Festinger, 1954).

In den Daten wurde kein signifikanter Einfluss der Situation gefunden: Halo-Effekte von Agency- und Communion-Verhaltensweisen waren in beruflichen Situationen genauso groß wie in privaten Situationen. Angesichts der Ergebnisse für die Berufsgruppe der beurteilten Person, kann auch dieses Ergebnis potentiell damit erklärt werden, dass die jeweiligen Kontexte nur im Kontrast zueinander salient werden.

Letztendlich zeigte sich in der Analyse auch kein Effekt für die Berufsgruppe des Probanden: Teilnehmer mit naturwissenschaftlichen und technischen Berufen waren im gleichen Maße für die unterschiedlichen Halo-Effekte anfällig wie Teilnehmer mit sozialen und Pflegeberufen. Sofern interindividuelle Unterschiede hinsichtlich der Anfälligkeit für

Halo-Effekte existieren, scheinen diese zumindest nicht durch die Berufsgruppe des Beurteilenden bedingt zu sein. Das Ergebnis ist dahingehend von Vorteil, dass es die Generalisierbarkeit der Valenz- und Dimensionseffekte nicht in Frage stellt. Wären Effekte des Berufs gefunden worden, müssten in den anderen Experimenten konsequenterweise auch die Studiengänge der Probanden berücksichtigt werden. Da es sich bei einem Großteil aller Teilnehmer um Psychologie-Studenten handelte, wäre die Repräsentativität der Befunde in diesem Fall eingeschränkt. Der fehlende Effekt schließt jedoch nicht aus, dass eventuell andere Charakteristika der Beurteilenden, wie etwa der Status in der beruflichen Hierarchie, zu signifikant unterschiedlichen Halo-Effekten führen.

7.3 Generalisierbarkeit der Ergebnisse: Experiment 9

Nachdem der Einfluss von Kontextfaktoren auf Vorannahmen untersucht wurde, sollten Halo-Effekte im neunten Experiment noch einmal als Verzerrungen eindeutiger Urteile betrachtet werden. Wie bei den Experimenten 5 und 6 wurden hierzu zwei experimentelle Bedingungen, eine Baseline- und eine Verzerrungs-Bedingung, verwendet. Bezüglich der Kontextfaktoren wurde versucht, Berufsgruppe und Situation gleichzeitig zu manipulieren. Zu diesem Zweck wurde, neben naturwissenschaftlichen und technischen Berufen und sozialen und Pflegeberufen, eine dritte Innersubjekt-Bedingung in das Untersuchungsdesign integriert. In dieser Bedingung war gar keine Berufsbezeichnung präsent, so dass ein eher privater Kontext induziert wurde.

Ausgehend von der Kontexthypothese wurde angenommen, dass Agency-Halo-Effekte bei der Beurteilung von Naturwissenschaftlern und Technikern stärker sind als bei der Bewertung von Personen mit sozialen oder Pflegeberufen. Am schwächsten sollten Agency-Halo-Effekte sein, wenn überhaupt keine Berufsbezeichnung vorgegeben wird. Für Halo-Effekte von Communion-Verhaltensweisen sollte sich ein umgekehrtes Muster ergeben. Das übergeordnete Ziel dieser Studie war es, alle bisherigen Ergebnisse noch einmal zu generalisieren. Aus diesem Grund lag der Untersuchung, analog zu Experiment 6, die komplette Datenbasis zugrunde. Zusätzlich wurde die Auswahl der präsentierten Verhaltensweisen sowie der abgefragten Eigenschaften randomisiert.

7.3.1 Methode

Teilnehmer

Teilnehmer des Experiments waren insgesamt 40 (24 weiblich, 16 männlich; 20 in jeder Zwischensubjekt-Bedingung) Studenten der Universität Heidelberg. Das Alter der Versuchsteilnehmer betrug im Durchschnitt 22.70 Jahre ($SD = 2.36$, $Min = 19$, $Max = 29$). Deutsch war für 36 Teilnehmer die Muttersprache, für drei Teilnehmer die erste und für einen Teilnehmer die zweite Fremdsprache. Für ihr Mitwirken an der Studie wurden die Probanden mit drei Euro oder einer halben Versuchspersonenstunde vergütet.

Materialien

Der Studie lag ein 2 (Darbietungsart) x 3 (Berufsgruppe der beurteilten Person) x 2 (Informationsvalenz) x 2 (Urteilsdimension) Design zugrunde. Die Darbietungsart war ein Zwischensubjekt-Faktor mit den Stufen *ohne Gegensatz* und *mit Gegensatz*. Die Berufsgruppe der beurteilten Person wurde innerhalb der Versuchsteilnehmer manipuliert und hatte die Ausprägungen *naturwissenschaftliche und technische Berufe, soziale und Pflegeberufe* und *keine Berufsbezeichnung*. Informationsvalenz (positiv versus negativ) und Urteilsdimension (Agency versus Communion) wurden ebenfalls als Innersubjekt-Faktoren konzipiert. Die Variable *Gleichheit von dargebotener und beurteilter Dimension* war aus Gründen der Übersichtlichkeit nicht Teil des Untersuchungsdesigns.

Als Items dienten alle 28 Verhaltenspaare aus der Datenbasis (siehe Anhang F), die zusammen 24 Attribute repräsentierten. In der Bedingung *ohne Gegensatz* wurden jeweils drei zufällig ausgewählte Attribute gleicher Valenz und Dimension (stets in Form einer zufällig ausgewählten Verhaltensweise) miteinander kombiniert. In der Bedingung *mit Gegensatz* kam zu jeder dieser Kombinationen ein zufälliges zentrales Attribut der gegensätzlichen Valenz und Dimension hinzu (zum Beispiel ein positives Agency- zusätzlich zu drei negativen Communion-Attributen). Das grundlegende Schema der beiden Bedingungen war somit analog zu dem von Experiment 5. Die Berufsbezeichnungen waren identisch zu denen in den Experimenten 7 und 8 (siehe Anhang I).

Die Ratings wurden auf zufälligen bipolaren Skalen von -3 bis 3 erfasst. Die Frage lautet hierbei stets wie folgt: „*Wie würden Sie diese Person auf folgender Dimension einschätzen?*“. Jeder der vier Valenz-Dimensions-Kombinationen wurde zufällig mit acht

abhängigen Skalen (vier für Agency- und vier für Communion-Eigenschaften) gepaart. Jede der daraus resultierenden 32 Kombinationen wurde wiederum mit den drei Berufsgruppen verbunden. Damit gab es in jeder Zwischensubjekt-Bedingung insgesamt 96 ($4 \times 8 \times 3$) Durchgänge. Die Verhaltensweisen, welche die Attribute repräsentierten, waren in jedem Durchgang randomisiert. Wie bereits zuvor wurde es nicht als problematisch betrachtet, dass nicht alle Urteile in der Baseline-Bedingung objektiv eindeutig waren.

Prozedur

Alle Probanden wurden in der Mensa der Universität Heidelberg durch studentische Hilfskräfte für die Untersuchung geworben. Das Labor wurde in einem gesonderten Raum über der Mensa eingerichtet. Nachdem die Teilnehmer eine Einverständniserklärung unterschrieben hatten, begannen sie mit dem Experiment. Dabei wurden sie zufällig einer der beiden Zwischensubjekt-Bedingungen zugeteilt. An einem Termin konnten maximal sechs Personen gleichzeitig teilnehmen.

In der Instruktion wurde den Probanden, abhängig von der Zwischensubjekt-Bedingung, mitgeteilt, dass sie im Folgenden an einer Bewertungsaufgabe teilnehmen würden, in der ihnen verschiedene Personen präsentiert werden, die jeweils drei beziehungsweise vier bestimmte Verhaltensweisen zeigen. Diese Personen würden auf unterschiedlichen Skalen bewertet werden müssen. Die Teilnehmer wurden darauf hingewiesen, sich bei noch offenen Fragen an die Versuchsleitung zu wenden und anderenfalls das Experiment zu starten.

Der Experimententeil umfasste 96 Durchgänge, die randomisiert und ohne Zeitlimit präsentiert wurden. Die Probanden benötigten im Mittel 15 Minuten zur Bearbeitung aller Fragen. In jedem Durchgang wurden alle drei oder vier Verhaltensweisen gleichzeitig dargeboten, wobei die Positionen stets randomisiert waren. Sofern ein Beruf präsentiert wurde, war auch dieser randomisiert. Stets sollte auf Basis der Verhaltensweisen eine zufällig aus der Datenbasis ausgewählte Agency- oder Communion-Eigenschaft beurteilt werden. Durch die Programmierung wurde verhindert, dass in der Verzerrungs-Bedingung die zentrale Verhaltensweise und die abgefragte Skala dem gleichen Attribut (beziehungsweise Attributpaar) entsprachen und das Urteil somit eindeutig war.

Am Ende der Untersuchung gaben die Probanden ihr Geschlecht und ihr Alter an. Ebenfalls wurde erfragt, ob Deutsch die Muttersprache, die erste oder die zweite

Fremdsprache sei. Nach Beendigung des Experiments wurden die Teilnehmer über Inhalt und Ziel der Untersuchung aufgeklärt und erhielten ihre Vergütung.

7.3.2 Ergebnisse

Vorab-Analysen

In den Vorab-Analysen wurde überprüft, ob die Urteile in der Baseline-Bedingung nicht neutral sind. Nur dann wären Verzerrungen durch gegensätzliche Verhaltensweisen auch als Halo-Effekte interpretierbar. Ein Rating wurde dann als nicht neutral angesehen, wenn es sich signifikant vom Mittelpunkt der Skala unterschied. Zu diesem Zweck wurden alle abhängigen Maße in der Baseline-Bedingung mit Hilfe von t-Tests gegen den Wert 0 getestet. Alle abhängigen Variablen wiesen signifikante Unterschiede zum Mittelpunkt der Skala auf ($p < .050$) und wurden daher in die weiteren Analysen integriert.

Da positive und negative Halo-Effekte auf derselben Skala erfasst wurden, wurden vor der Überprüfung der Hypothesen alle Ratings, die auf peripheren negativen Verhaltensweisen basierten, durch eine Multiplikation mit -1 umkodiert.

Einflüsse von Informationsvalenz und Urteilsdimension ohne Gegensatz

Zunächst wurden die Effekte von Informationsvalenz und Urteilsdimension in der Zwischensubjekt-Bedingung *ohne Gegensatz* analysiert. Dazu wurde eine 2 x 2 ANOVA (Informationsvalenz x Urteilsdimension) mit Messwiederholung durchgeführt. Es war ein signifikanter Haupteffekt für Valenz erkennbar ($F(1, 19) = 4.48, p = .048, \eta^2 = .19$): Urteile auf der Grundlage positiver Verhaltensweisen ($M = 1.37, SD = 0.47$) waren extremer als Urteile auf der Basis negativer Verhaltensweisen ($M = 1.17, SD = 0.30$). Der Haupteffekt der Urteilsdimension wurde nicht signifikant ($F(1, 19) = 2.24, p = .151, \eta^2 = .11$): Urteile auf der Basis von Agency- ($M = 1.23, SD = 0.35$) und Communion-Verhaltensweisen ($M = 1.31, SD = 0.36$) hatten das gleiche Ausmaß. Auch die Interaktion der beiden Faktoren erreichte kein signifikantes Maß ($F(1, 19) = 1.21, p = .285, \eta^2 = .06$): Urteile auf der Grundlage positiver Agency- ($M = 1.36, SD = 0.53$) und Communion-Verhaltensweisen ($M = 1.38, SD = 0.47$) waren nicht unterschiedlicher zueinander als Urteile auf der Basis negativer Agency- ($M = 1.10, SD = 0.26$) und Communion-Verhaltensweisen ($M = 1.23, SD = 0.39$).

Einfluss der Berufsgruppe der beurteilten Person ohne Gegensatz

Um den Einfluss der Berufsgruppe der beurteilten Person auf die einfachen Ratings zu untersuchen, wurde in der Zwischensubjekt-Bedingung *ohne Gegensatz* eine 2 x 2 x 2 ANOVA (Berufsgruppe der beurteilten Person x Informationsvalenz x Urteilsdimension) mit Messwiederholung berechnet. Die Haupteffekte der Faktoren *Informationsvalenz* und *Urteilsdimension* sowie deren Interaktion wurden schon im letzten Abschnitt betrachtet. Es zeigte sich kein signifikanter Haupteffekt der Berufsgruppe ($F(2, 18) = 0.13, p = .878, \eta^2 = .01$): Ratings bei der Beurteilung von Menschen mit naturwissenschaftlichen und technischen Berufen ($M = 1.27, SD = 0.39$), bei der Beurteilung von Menschen mit sozialen und Pflegeberufen ($M = 1.28, SD = 0.44$) und bei der Bewertung von Personen ohne Berufsbezeichnung ($M = 1.26, SD = 0.34$) unterschieden sich nicht voneinander.

Die Interaktion von Berufsgruppe und Valenz erreichte kein signifikantes Maß ($F(2, 18) = 0.07, p = .936, \eta^2 = .01$): Ratings auf der Grundlage von positiven ($M = 1.36, SD = 0.51$) und negativen Verhaltensweisen ($M = 1.18, SD = 0.36$) bei der Einschätzung von Naturwissenschaftlern, Ratings auf der Basis von positiven ($M = 1.38, SD = 0.56$) und negativen Verhaltensweisen ($M = 1.18, SD = 0.41$) bei der Bewertung von Menschen in sozialen Berufen und Ratings auf der Grundlage von positiven ($M = 1.36, SD = 0.52$) und negativen Verhaltensweisen ($M = 1.15, SD = 0.35$) bei der Beurteilung von Personen ohne Berufsbezeichnung hatten jeweils die gleiche Differenz zueinander. Es war allerdings eine signifikante Interaktion von Berufsgruppe und Urteilsdimension zu erkennen ($F(2, 18) = 4.83, p = .021, \eta^2 = .35$): Ratings auf der Basis von Agency-Verhaltensweisen waren bei der Einschätzung von Menschen mit naturwissenschaftlichen und technischen Berufen ($M = 1.31, SD = 0.48$) extremer als bei der Beurteilung von Menschen mit sozialen und Pflegeberufen ($M = 1.24, SD = 0.46$) und der Bewertung von Personen ohne jegliche Berufsbezeichnung ($M = 1.14, SD = 0.38$). Im Gegensatz dazu waren Ratings auf der Grundlage von Communion-Verhaltensweisen bei der Beurteilung von Menschen mit naturwissenschaftlich-technischen Berufen ($M = 1.23, SD = 0.35$) neutraler als bei der Einschätzung von Personen mit sozialen und Pflegeberufen ($M = 1.33, SD = 0.49$) und der Bewertung von Menschen ohne jegliche Berufsbezeichnung ($M = 1.37, SD = 0.43$). Eine signifikante Dreifachinteraktion von Berufsgruppe, Valenz und Urteilsdimension war nicht ersichtlich ($F(2, 18) = 0.78, p = .475, \eta^2 = .08$).

Einfluss gegensätzlicher Verhaltensweisen in Bezug auf Valenz und Dimension

Der Einfluss der zentralen gegensätzlichen Verhaltensweisen auf die einzelnen Ratings wurde anhand einer 2 x 2 x 2 ANOVA (Darbietungsart x Informationsvalenz x Urteilsdimension) mit Messwiederholung ermittelt. Der Haupteffekt des Faktors *Darbietungsart* erreichte ein signifikantes Maß ($F(1, 38) = 24.97, p < .001, \eta^2 = .40$): In der Bedingung mit einer zentralen gegensätzlichen Verhaltensweise ($M = 0.86, SD = 0.14$) ergaben sich neutralere Ratings als ohne Gegensatz ($M = 1.27, SD = 0.33$). Ebenso wurden die Haupteffekte für Valenz ($F(1, 38) = 24.44, p < .001, \eta^2 = .39$) und Urteilsdimension ($F(1, 38) = 5.87, p = .020, \eta^2 = .13$) signifikant: Hatten die peripheren Verhaltensweisen eine positive Valenz ($M = 1.24, SD = 0.40$), wurden extremere Ratings abgegeben als bei einer negativen Valenz ($M = 0.89, SD = 0.40$). Des Weiteren waren Ratings, die auf peripheren Communion-Verhaltensweisen basierten ($M = 1.13, SD = 0.36$), höher als Ratings auf der Grundlage von Agency-Verhaltensweisen ($M = 1.01, SD = 0.36$).

Bei den Interaktionen wurde kein signifikanter Effekt von Valenz x Urteilsdimension gefunden ($F(1, 38) = 0.90, p = .349, \eta^2 = .02$): Urteile auf der Basis peripherer positiver Agency- ($M = 1.20, SD = 0.44$) und Communion-Verhaltensweisen ($M = 1.28, SD = 0.44$) zeigten keine andere Differenz zueinander als Urteile auf der Basis negativer Agency- ($M = 0.81, SD = 0.41$) und Communion-Verhaltensweisen ($M = 0.97, SD = 0.50$). Auch die Interaktion von Darbietungsart und Urteilsdimension wurde nicht signifikant ($F(1, 38) = 0.71, p = .404, \eta^2 = .02$): Ratings auf der Basis peripherer Agency- ($M = 1.23, SD = 0.35$) und Communion-Verhaltensweisen ($M = 1.31, SD = 0.36$) in der Baseline-Bedingung wiesen die gleiche Differenz zueinander auf wie Ratings auf der Grundlage peripherer Agency- ($M = 0.78, SD = 0.20$) und Communion-Verhaltensweisen ($M = 0.94, SD = 0.26$) in der Bedingung mit Gegensatz. Ein signifikantes Ergebnis ergab sich dagegen für die Interaktion von Darbietungsart und Valenz ($F(1, 38) = 4.29, p = .045, \eta^2 = .10$): Ratings auf der Grundlage dreier negativer Verhaltensweisen ($M = 1.17, SD = 0.30$) wurden durch die zusätzliche Präsentation einer positiven Verhaltensweise ($M = 0.62, SD = 0.29$) stark geschwächt. Demgegenüber wurden positive Urteile ($M = 1.37, SD = 0.47$) durch die zusätzliche Darbietung einer negativen Verhaltensweise ($M = 1.11, SD = 0.25$) nur wenig beeinflusst. Positive Halo-Effekte waren folglich stärker als negative Halo-Effekte. Es gab keine signifikante Dreifachinteraktion ($F(1, 38) = 0.05, p = .825, \eta^2 < .01$).

Einfluss gegensätzlicher Verhaltensweisen in Bezug auf die Berufsgruppe

Abschließend wurde der Einfluss der zentralen gegensätzlichen Verhaltensweisen in Bezug auf die verschiedenen Berufsgruppen untersucht. Dies geschah mit Hilfe einer 2 x 2 x 2 x 2 ANOVA (Darbietungsart x Berufsgruppe der beurteilten Person x Informationsvalenz x Urteilsdimension) mit Messwiederholung. Die Haupteffekte und Interaktionen von Darbietungsart, Informationsvalenz und Urteilsdimension wurden bereits im letzten Abschnitt ausführlich geschildert. Es war kein signifikanter Haupteffekt der Berufsgruppe erkennbar ($F(2, 37) = 1.03, p = .368, \eta^2 = .05$): Ratings bei der Bewertung von Personen mit naturwissenschaftlich-technischen Berufen ($M = 1.07, SD = 0.37$), bei der Beurteilung von Menschen mit sozialen und Pflegeberufen ($M = 1.09, SD = 0.40$) und bei der Einschätzung von Personen ohne jegliche Berufsbezeichnung ($M = 1.04, SD = 0.34$) waren über beide Zwischensubjekt-Bedingungen hinweg gleich hoch.

Die Interaktion von Darbietungsart und Berufsgruppe wurde nicht signifikant ($F(2, 37) = 0.35, p = .705, \eta^2 = .02$): Beurteilungen von Menschen mit naturwissenschaftlichen Berufen ohne ($M = 1.27, SD = 0.39$) und mit Gegensatz ($M = 0.88, SD = 0.22$), Bewertungen von Personen mit einem sozialen Beruf ohne ($M = 1.28, SD = 0.44$) und mit Gegensatz ($M = 0.90, SD = 0.24$) und Bewertungen von Personen ohne Berufsbezeichnung ohne ($M = 1.26, SD = 0.34$) und mit Gegensatz ($M = 0.81, SD = 0.14$) hatten jeweils die gleiche Differenz zueinander. Auch die Interaktion von Berufsgruppe und Valenz erreichte kein signifikantes Niveau ($F(2, 37) = 0.27, p = .765, \eta^2 = .01$): Ratings auf der Basis peripherer positiver ($M = 1.23, SD = 0.45$) und negativer Verhaltensweisen ($M = 0.92, SD = 0.41$) bei der Beurteilung von Naturwissenschaftlern, Ratings auf der Basis peripherer positiver ($M = 1.27, SD = 0.51$) und negativer Verhaltensweisen ($M = 0.91, SD = 0.49$) bei der Bewertung von sozialen Berufen und Ratings auf der Grundlage peripherer positiver ($M = 1.22, SD = 0.48$) und negativer Verhaltensweisen ($M = 0.85, SD = 0.45$) bei der Beurteilung von Personen ohne Berufsbezeichnung wiesen jeweils die gleiche Differenz zueinander auf.

Ein signifikantes Ergebnis zeigte sich für die Interaktion von Berufsgruppe und Urteilsdimension ($F(2, 37) = 31.23, p < .001, \eta^2 = .63$): Ratings auf der Grundlage von peripheren Agency-Verhaltensweisen waren bei der Einschätzung von Menschen mit naturwissenschaftlich-technischen Berufen ($M = 1.15, SD = 0.42$) extremer als bei der Bewertung von Personen mit sozialen und Pflegeberufen ($M = 1.00, SD = 0.46$) und der Beurteilung von Menschen ohne jegliche Berufsbezeichnung ($M = 0.87, SD = 0.42$). Im

Gegensatz dazu waren Ratings auf der Basis peripherer Communion-Verhaltensweisen bei der Bewertung von Menschen mit naturwissenschaftlichen und technischen Berufen ($M = 1.00$, $SD = 0.40$) neutraler als bei der Einschätzung von Menschen mit sozialen und Pflegeberufen ($M = 1.17$, $SD = 0.45$) und der Beurteilung von Menschen ohne jegliche Berufsbezeichnung ($M = 1.20$, $SD = 0.40$). Die Dreifachinteraktion von Darbietungsart, Berufsgruppe und Urteilsdimension wurde auch signifikant ($F(2, 37) = 4.48$, $p = .018$, $\eta^2 = .20$): Zentrale Agency-Verhaltensweisen führten bei der Einschätzung von Personen mit naturwissenschaftlichen Berufen zu größeren Verzerrungen als bei der Bewertung von Menschen ohne Berufsbezeichnung und der Beurteilung von Menschen mit sozialen und Pflegeberufen. Demgegenüber führten zentrale Communion-Verhaltensweisen bei der Beurteilung von Naturwissenschaftlern zu geringeren Verzerrungen als bei der Bewertung von Personen mit sozialen und Pflegeberufen und der Einschätzung von Menschen ohne Berufsbezeichnung. Mit anderen Worten ausgedrückt, waren Agency-Halo-Effekte im Kontext von Naturwissenschaftlern stärker als im Kontext von sozialen Berufen und im privaten Kontext. Für Communion-Halo-Effekte ergab sich das umgekehrte Muster. Die Dreifachinteraktionen von Darbietungsart, Berufsgruppe und Valenz ($F(2, 37) = 0.07$, $p = .931$, $\eta^2 < .01$) sowie von Berufsgruppe, Valenz und Urteilsdimension ($F(2, 37) = 0.98$, $p = .385$, $\eta^2 = .05$) wurden nicht signifikant. Das Gleiche galt für die Interaktion aller vier Faktoren ($F(2, 37) = 0.01$, $p = .988$, $\eta^2 < .01$). Die Mittelwerte und Standardabweichungen der signifikanten Dreifachinteraktion sind in Tabelle 9 dargestellt.

Tabelle 9. Einfluss von Berufsgruppe und Urteilsdimension auf Halo-Effekte

	Darbietungsart		Differenz
	ohne Gegensatz	mit Gegensatz	
3 x Agency (+ 1 x Communion)			
naturwiss. und technische Berufe	1.31 (0.48)	0.99 (0.29)	0.32
soziale und Pflegeberufe	1.24 (0.46)	0.77 (0.33)	0.47
keine Berufsbezeichnung	1.14 (0.38)	0.59 (0.26)	0.55
3 x Communion (+ 1 x Agency)			
naturwiss. und technische Berufe	1.23 (0.35)	0.77 (0.32)	0.46
soziale und Pflegeberufe	1.33 (0.49)	1.02 (0.35)	0.31
keine Berufsbezeichnung	1.37 (0.43)	1.04 (0.29)	0.33

7.3.3 Diskussion

In Experiment 9 wurde der Einfluss von Kontexteffekten auf die Halo-Effekte von Agency- und Communion-Verhaltensweisen untersucht. Anders als bei den vorherigen beiden Studien wurden Halo-Effekte dieses Mal als Verzerrungen eindeutiger Urteile betrachtet. Die Manipulation erfolgte durch die Präsentation von Verhaltensweisen mit gegensätzlicher Valenz und Dimension zusätzlich zu einer Gruppe von peripheren Verhaltensweisen. Diese Methode war bereits in den Experimenten 5 und 6 erfolgreich angewandt worden. Als Beispiel für einen typischen Kontextfaktor wurden erneut zwei verschiedene Berufsgruppen verwendet. Zusätzlich gab es eine Bedingung, in der gar keine Auskunft über den Beruf einer Person gegeben wurde. Dadurch sollte die Situation, in der Verhaltensweisen beobachtet und die Urteile abgegeben werden, indirekt variiert werden. Um die Befunde der vorherigen Experimente zu generalisieren, erfolgte die Auswahl der dargebotenen und abgefragten Attribute nicht nach einem vorher festgelegten Schema, sondern durchgehend randomisiert.

Die Daten der Erhebung bestätigen die Kontexthypothese: Werden Personen mit naturwissenschaftlichen und technischen Berufen bewertet, sind Agency-Halo-Effekte von größerem Ausmaß als bei der Beurteilung von Menschen mit sozialen und Pflegeberufen. Am schwächsten sind Halo-Effekte von Agency-Verhaltensweisen, wenn es um die Einschätzung von Personen geht, deren Beruf nicht bekannt ist. Für Communion-Halo-Effekte findet sich das umgekehrte Muster: Es zeigen sich die stärksten Effekte bei der Beurteilung von Personen ohne Berufsbezeichnung und die schwächsten Effekte, wenn Menschen mit naturwissenschaftlichen und technischen Berufen bewertet werden sollen. Die gleichen Ergebnisse treten auch dann auf, wenn statt Halo-Effekten nur die Ratings in der Baseline-Bedingung betrachtet werden. Unter der Voraussetzung, dass durch das Weglassen von Berufsbezeichnungen ein privater Kontext vermittelt wird, kann die Kontexthypothese somit auch für unterschiedliche Situationen belegt werden: In privaten Situationen sind Halo-Effekte von Communion-Verhaltensweisen größer als in rein beruflichen Situationen. Für Halo-Effekte von Agency-Verhaltensweisen gilt dagegen das umgekehrte Prinzip. Die Befunde sind konform mit der Tatsache, dass die Bedeutsamkeit von Agency- und Communion-Attributen von der Interdependenz zu anderen Personen abhängt (Wojciszke & Abele, 2008) und sich für Sympathie und Respekt unterscheidet

(Wojciszke et al., 2009). Da die Berufsgruppe des Beurteilten innerhalb der Teilnehmer manipuliert wurde, kann erneut vermutet werden, dass verschiedene Berufsgruppen und Situationen nur im direkten Vergleich salient werden und zu unterschiedlichen Mustern bezüglich Agency- und Communion-Halo-Effekten führen.

Zusätzlich zu der Kontexthypothese können auch die Valenzhypothese und, zumindest indirekt, auch die Dimensionshypothese noch einmal belegt werden: Positive Halo-Effekte sind stärker als negative Halo-Effekte. Des Weiteren sind Communion-Halo-Effekte, sofern keine Personen mit naturwissenschaftlichen oder technischen Berufen beurteilt werden, von größerem Ausmaß als Agency-Halo-Effekte. Werden statt Halo-Effekten nur die Ratings in der Baseline-Bedingung analysiert, werden die gleichen Ergebnisse deutlich. Aufgrund der randomisierten Auswahl der dargebotenen Verhaltensweisen und abgefragten Eigenschaften, kann das Zutreffen der Hypothesen folglich für eine große Anzahl an Attributen generalisiert werden.

7.4 Zusammenfassung der Experimente 7 bis 9

Der Schwerpunkt des dritten Experimentalblocks lag in der Untersuchung des Einflusses diverser Kontextfaktoren auf die Halo-Effekte von Agency- und Communion-Verhaltensweisen. In diesem Zusammenhang wurden sowohl verschiedene Berufsgruppen des Beurteilenden und der beurteilten Person sowie die Situation, in der die einzelnen Urteile gefällt werden, betrachtet. In den Experimenten 7 und 8 wurden Halo-Effekte als Vorannahmen, in Experiment 9 dagegen als Verzerrungen eindeutiger Urteile untersucht. Die Experimentaldurchgänge boten fiktive Personen mit verschiedenen Berufen dar, die unterschiedliche Verhaltensweisen zeigten. Die fiktiven Personen wurden von den Probanden hinsichtlich einer Reihe bipolarer Persönlichkeitsskalen bewertet. Es wurden naturwissenschaftliche und technische Berufe von sozialen und Pflegeberufen abgegrenzt. Zusätzlich waren die Situationen entweder beruflicher oder privater Natur.

Zwei der drei Experimente belegen die Kontexthypothese für die Berufsgruppe der beurteilten Person: Halo-Effekte von Agency-Verhaltensweisen sind bei der Bewertung von Menschen mit naturwissenschaftlichen und technischen Berufen stärker als bei der Beurteilung von Menschen mit sozialen und Pflegeberufen. Für Halo-Effekte von

Communion-Verhaltensweisen ergibt sich das entsprechend umgekehrte Muster. Die fehlende Bestätigung der Hypothese in einem der Experimente ist wahrscheinlich auf die Verwendung einer Zwischensubjekt-Manipulation zurückzuführen. Neben der höheren Teststärke von Innersubjekt-Designs ist es denkbar, dass verschiedene berufliche Gruppen nur im unmittelbaren Kontrast zueinander salient werden und zu unterschiedlich starken Agency- und Communion-Halo-Effekten führen.

Der Einfluss der Situation kann durch eines der Experimente bestätigt werden: Agency-Halo-Effekte sind in beruflichen Situationen stärker als in privaten Situationen, Communion-Halo-Effekte sind dagegen in privaten Situationen von größerem Ausmaß. Jedoch scheint sich auch diese Interaktion nur dann zu zeigen, wenn unterschiedliche Situationen in einem Innersubjekt-Design im direkten Vergleich miteinander stehen. Ein signifikanter Einfluss der Berufsgruppe des Beurteilenden kann durch keine der Untersuchungen belegt werden: Probanden mit naturwissenschaftlichen und technischen Berufen sind im gleichen Maße für Agency- und Communion-Halo-Effekte anfällig wie Teilnehmer mit sozialen und Pflegeberufen. Möglicherweise können jedoch andere Eigenschaften der Beurteilenden signifikante Effekte mit sich bringen.

Neben verschiedenen Kontextfaktoren wurden in den Untersuchungen auch die Einflüsse von Informationsvalenz und Urteilsdimension noch einmal betrachtet. Die Valenzhypothese kann erneut bestätigt werden: Positive Halo-Effekte sind stärker als negative Halo-Effekte. Sofern keine naturwissenschaftlichen und technischen Berufe analysiert werden, wird auch die Dimensionshypothese belegt: Halo-Effekte von Communion-Verhaltensweisen sind von stärkerem Ausmaß als Halo-Effekte von Agency-Verhaltensweisen. Da in zweien der drei Studien keine Studenten, sondern Menschen im Berufsleben untersucht wurden, sprechen diese Befunde für die Generalisierbarkeit der Ergebnisse über verschiedenartige Stichproben hinweg.

8 Allgemeine Diskussion

Die vorliegende Arbeit untersuchte den Einfluss von Informationsvalenz, Urteilsdimension und verschiedener Kontextfaktoren auf die Qualität und Quantität von Halo-Effekten. Aufgrund der großen Anzahl an durchgeführten Experimenten erfolgt zunächst eine Zusammenfassung der wesentlichen Ergebnisse. Danach werden die Ergebnisse unter Einbettung in die aktuelle Befundlage diskutiert. Abschließend wird ein Ausblick auf zukünftige Forschungsansätze gegeben.

8.1 Zusammenfassung der Ergebnisse

Zur Untersuchung der zentralen Hypothesen wurden insgesamt neun Experimente, gegliedert in drei Experimentalblöcke, durchgeführt. Vorher wurde in einer allgemeinen Voruntersuchung überprüft, ob Informationsvalenz, Urteilsdimension und Kontextfaktoren auch relevant für Halo-Effekte sind. Die Ergebnisse machten deutlich, dass positive Persönlichkeitsattribute im Vergleich zu negativen Attributen extremer valent sind, stärker auf den Gesamteindruck über eine Person generalisiert werden und als bedeutsamer für zwischenmenschliche Interaktion empfunden werden. Ähnliche Vorteile wurden für Communion-Eigenschaften im Vergleich zu Agency-Eigenschaften gefunden. Die Unterschiede zwischen Agency- und Communion-Eigenschaften waren besonders bei der Beurteilung von Menschen im Allgemeinen und Personen in sozialen Berufen zu finden. Wurden stattdessen Personen in naturwissenschaftlichen Berufen betrachtet, relativierten sich die Vorteile für Communion-Eigenschaften und kehrten sich zum Teil sogar um. Die Befunde wurden als Belege für die zentralen Hypothesen betrachtet.

Die Valenzhypothese wurde auf quantitativer Ebene in allen neun Experimenten betrachtet. Bis auf eine Untersuchung (Experiment 6) konnten alle Experimente die postulierte Asymmetrie bestätigen: Positive Halo-Effekte sind stärker als negative Halo-Effekte. Darüber hinaus wurde die Valenzhypothese auch auf qualitativer Ebene belegt: Positive Halo-Effekte sind globaler Natur, wogegen negative Halo-Effekte spezifisch sind. Dieser Unterschied wurde bereits auf quantitativer Ebene dadurch deutlich, dass positive Halo-Effekte einander ähnlicher sind als negative Halo-Effekte (Experiment 1 und 2). Ein

weiterer Beleg liegt in dem Befund, dass die zusätzliche Präsentation von Bildern der gleichen (Experiment 2 und 3) oder gegensätzlichen Valenz (Experiment 4) zwar negative Halo-Effekte, nicht aber positive Halo-Effekte verstärken beziehungsweise abschwächen kann. Die gleiche Asymmetrie wurde gefunden, wenn die Position positiver und negativer Verhaltensweisen (Experiment 6) bei einer sequentiellen Präsentation der einzelnen Informationen variiert wurde: Während die Position positiver Verhaltensweisen irrelevant für das Ausmaß der daraus resultierenden Halo-Effekte war, ergaben sich für negative Verhaltensweisen starke Recency-Effekte. Zusätzlich war der qualitative Unterschied zwischen positiven und negativen Halo-Effekten dadurch erkennbar, dass Inter-Rating-Korrelationen auf der Basis positiver Informationen höher waren als Inter-Rating-Korrelationen auf der Grundlage negativer Informationen (Experiment 3).

Die Dimensionshypothese wurde in allen neun Experimenten überprüft. In den Untersuchungen, in denen Halo-Effekte als Vorannahmen betrachtet wurden, erfolgte die Datenanalyse getrennt nach Valenz. Bei negativer Informationsvalenz konnte eine Untersuchung (Experiment 4) die Vermutung nur tendenziell bestätigen, in sämtlichen anderen Experimenten wurden signifikante Unterschiede gefunden: Negative Communion-Halo-Effekte sind stärker als negative Agency-Halo-Effekte. In den positiven Bedingungen war dagegen in keiner der Untersuchungen ein signifikanter Unterschied zu erkennen: Positive Agency- und Communion-Halo-Effekte sind im Durchschnitt gleich stark. In den Experimenten, in denen Halo-Effekte als Verzerrungen untersucht wurden, wurde die Dimensionshypothese über beide Valenzbedingungen hinweg betrachtet. Auch hier wurde die Hypothese signifikant bestätigt. In den Experimenten wurde zudem belegt, dass sich Agency- und Communion-Halo-Effekte nur in ihrer Stärke und nicht qualitativ voneinander unterscheiden: Agency- und Communion-Halo-Effekte wurden im gleichen Maße durch die Präsentation von Bildern der gleichen (Experiment 2 und 3) oder gegensätzlichen Valenz (Experiment 4) sowie von der Position einer Verhaltensweise (Experiment 6) beeinflusst. Ebenso ergaben sich ähnlich hohe Inter-Rating-Korrelationen (Experiment 3).

Die Kontexthypothese war in drei Studien der Fokus der Untersuchungen. Als Kontextvariablen wurden der Beruf der beurteilten Person (Experiment 7, 8 und 9), der Beruf des Beurteilenden (Experiment 8) und die Situation (Experiment 8 und 9) betrachtet. Bis auf eine Untersuchung (Experiment 7) bestätigten alle Experimente den postulierten Einfluss der Berufsgruppe der beurteilten Person: Agency-Halo-Effekte sind bei der

Bewertung von Personen mit naturwissenschaftlichen und technischen Berufen größer als bei der Beurteilung von Menschen mit sozialen und Pflegeberufen. Für Communion-Halo-Effekte gilt das umgekehrte Muster. Der Einfluss der Situation wurde in einer der Untersuchungen (Experiment 9) nachgewiesen: Agency-Halo-Effekte sind in beruflichen Situationen stärker als in privaten Situationen, Communion-Halo-Effekte sind in privaten Situationen von größerem Ausmaß. Ein signifikanter Einfluss der Berufsgruppe des Beurteilenden konnte nicht gezeigt werden: Menschen mit naturwissenschaftlichen und technischen Berufen sind im gleichen Maße für Agency- und Communion-Halo-Effekte anfällig wie Personen mit sozialen und Pflegeberufen.

Abgesehen von den Kontextvariablen waren alle Befunde über verschiedene Manipulationen und Stichproben hinweg stabil und zeigten sich für Vorannahmen und Verzerrungen gleichermaßen. Die Effekte von Informationsvalenz und Urteilsdimension waren auch dann präsent, wenn Ratings unabhängig von Halo-Effekten betrachtet wurden (Experiment 5 und 9). In einigen Untersuchungen wurden weitere Erkenntnisse gewonnen, die nicht Bestandteil der Hypothesen waren. So wurde beispielsweise deutlich, dass Vorannahmen über Communion-Eigenschaften stärker durch negative und Vorannahmen über Agency-Eigenschaften stärker durch positive Halo-Effekte geprägt werden können, wenn die Ratings auf bipolaren Skalen erfolgen (Experiment 3, 4 und 7). Auch gab es Anzeichen dafür, dass Communion-Attribute im direkten Vergleich zu robusteren Urteilen und Halo-Effekten führen als Agency-Attribute (Experiment 5). Ein dritter bedeutsamer Befund war zu erkennen, wenn statt spezifischer Eigenschaften der Gesamteindruck bewertet werden sollte (Experiment 4). Im Gegensatz zu Halo-Effekten ergaben sich für Gesamteindrücke keine Valenzunterschiede: Gesamteindrücke auf der Basis positiver und negativer Verhaltensweisen sind gleichermaßen stark.

8.2 Einordnung und Implikationen der Befunde

Die in der vorliegenden Arbeit gezeigten Ergebnisse sind von hoher Bedeutung für das bessere Verständnis von Halo-Effekten. Zwar wurden Halo-Effekte bereits in vielen Bereichen umfassend erforscht, jedoch existieren nach wie vor unklare und zum Teil widersprüchliche Befunde. Dies betrifft vor allem die grundlegende Frage, ob es sich bei

Halo-Effekten um ein einheitliches Phänomen handelt, oder aber qualitativ verschiedene Formen von Halo-Effekten existieren. Auch bezüglich der Ursachen, Einflussfaktoren und Vermeidung von Halo-Effekten gibt es viele offene Fragen. Die Befunde der vorliegenden Arbeit können einige dieser Fragen klären und haben somit wichtige Implikationen für den aktuellen Stand der Forschung und die weitere Untersuchung von Halo-Effekten.

Einheitlicher Halo-Effekt versus vielfältige Halo-Effekte

In der bisherigen Literatur existieren unterschiedliche Standpunkte bezüglich der Einheitlichkeit von Halo-Effekten, so dass die Problematik noch nicht eindeutig geklärt werden konnte. Die Befunde der vorliegenden Arbeit sprechen eindeutig dafür, dass Halo-Effekte sehr vielfältig sind und hinsichtlich ihres Ausmaßes und ihrer zugrunde liegenden Prozesse variieren können. Verschiedene Annahmen, nach denen Halo-Effekte universell, einheitlich und frei von Widersprüchen seien, konnten somit widerlegt werden (Cooper, 1981b; Feldman, 1986; Kozlowski et al., 1986; Nathan, 1986; Newcomb, 1931; Pike, 1999; Pulakos et al., 1986; Saal et al., 1980). Gleichzeitig reiht sich die vorliegende Arbeit in die wenige bisherige Literatur ein, die Belege für die Vielfältigkeit von Halo-Effekten lieferte (Balzer & Sulsky, 1992; Fox et al., 1983; Fox et al., 1989; Murphy et al., 1993).

Die gezeigten Ergebnisse haben bedeutende Implikationen für einige existierende Untersuchungen. In diesem Zusammenhang sind vor allem die drei Prozessmodelle von Halo-Effekten nach Fiscaro und Lance (1990) zu nennen, die bisher nur in wenigen Studien betrachtet und miteinander verglichen wurden (Dennis, 2007; Lance, LaPointe, & Fiscaro, 1994; Lance, LaPointe, & Stewart, 1994; Oh & Ramaprasad, 2003). Dabei stand meist die Frage im Vordergrund, welches der Modelle am besten in der Lage ist, Halo-Effekte zu beschreiben. Die vorliegenden Befunde machen deutlich, dass nicht etwa nur ein Modell, sondern mindestens zwei der Modelle in einem gewissen Maße gültig sind und, abhängig von verschiedenen Faktoren, mehr oder weniger zutreffen können. Hierbei handelt es sich zum einen um das General Impression Model, welches Halo-Effekte auf der Basis globaler Eindrücke beschreibt. Als zweites ist das Salient Dimension Model zu nennen, das Halo-Effekte auf den Einfluss spezifischer salienter Eigenschaften zurückführt. Ein gutes Modell zur Beschreibung von Halo-Effekten sollte beide dieser Prozesse berücksichtigen. Auch das dritte Modell, Inadequate Discrimination Model genannt, sollte nicht gänzlich unbeachtet bleiben. Gemäß diesem Modell werden Halo-Effekte dadurch definiert, dass bestimmte

Merkmale oder Verhaltensweisen zwar korrekt wahrgenommen, aber auf eine falsche Eigenschaft zurückgeführt werden. In der vorliegenden Arbeit wurde versucht, diese Verzerrung durch eine gründliche Vorauswahl und eine klare Definition der dargebotenen Informationen weitgehend auszuschließen. Außerhalb eines experimentellen Designs kann diese Bedingung jedoch nicht mit Sicherheit erfüllt werden. Damit ist anzunehmen, dass in der Praxis auch dieses Modell unter bestimmten Umständen gilt.

Auf Basis der Befundlage dieser Arbeit kann geschlussfolgert werden, dass Halo-Effekte auf mindestens zwei verschiedene Arten geprägt werden können: Zum einen beeinflusst die Struktur der zugrunde liegenden Informationen, ob globale oder aber spezifische Halo-Effekte wirken. Während bei einer hohen Density ein starker globaler Eindruck entsteht, können bei einer geringen Density nur Einflüsse einzelner spezifischer Stimuli wirken. Da globale Halo-Effekte unmittelbar durch das parallele Auftreten vieler spezifischer Halo-Effekte hervorgerufen werden, sind sie auch die stärkere der beiden Formen. Die zweite Art der Beeinflussung von Halo-Effekten wird über die Bedeutsamkeit der zugrunde liegenden Information im aktuellen Urteilskontext vermittelt. Attribute, die zum Zeitpunkt der Beurteilung besonders wichtig sind beziehungsweise als sehr wichtig erachtet werden, führen zu stärkeren Halo-Effekten als wenig bedeutsame Attribute. Im Gegensatz zum Einfluss der Density ist dieser Effekt rein quantitativer Art, da er nicht auf strukturellen Eigenschaften der Informationen beruht. In Anbetracht dieser Vielfältigkeit lassen sich auch die verschiedenen Probleme bei der Erfassung von Halo-Effekten in das Gesamtbild einordnen. In der Vergangenheit zeigten mehrere Autoren auf, dass die verschiedenen Methoden zur Erfassung von Halo-Effekten häufig zu unterschiedlichen Resultaten führen (Alliger & Williams, 1989; Balzer & Sulsky, 1992; Fiscaro & Vance, 1994; Fox et al., 1983; Fox et al., 1989; Pulakos et al., 1986; Saal et al., 1980). Die vorliegende Arbeit kann nachweisen, dass nicht alle Maße auch die gleiche Art von Halo-Effekten messen. Beispielsweise erfassen Inter-Rating-Korrelationen zwar qualitative, nicht jedoch rein quantitative Unterschiede von Halo-Effekten. Das heißt, dass globale und spezifische Halo-Effekte aufgrund unterschiedlich hoher Inter-Rating-Korrelationen unterschieden werden können. Zwei globale oder zwei spezifische Halo-Effekte, die sich lediglich in ihrer Größe unterscheiden, haben dagegen keine unterschiedlichen Inter-Rating-Korrelationen und können somit nicht voneinander abgegrenzt werden. Legt man diese Erkenntnis zugrunde, lassen sich bei einer erneuten Betrachtung möglicherweise einige der bisher

aufgetretenen Widersprüche bei der Erfassung von Halo-Effekten auflösen. Für zukünftige Erhebungen sollten stets mehrere Maße verwendet werden.

Einfluss der Informationsvalenz

Der Einfluss von Valenz auf verschiedene Aspekte von Urteilen und Entscheidungen war in der Vergangenheit der Gegenstand vieler Untersuchungen. Durch die vorliegende Arbeit wird dieses Spektrum um Urteilsverzerrungen durch Halo-Effekte erweitert. Die Befunde lassen darauf schließen, dass positive Halo-Effekte globaler Natur und negative Halo-Effekte rein spezifisch sind. Aus diesem Grund sind positive Halo-Effekte auch größer als negative Halo-Effekte. Dieser Unterschied lässt sich anhand der Density Hypothesis erklären, der zufolge positive Informationen einander sehr ähnlich, negative Informationen hingegen viel distinkter voneinander sind (Unkelbach et al., 2008). Die Ergebnisse dieser Arbeit sind konform mit einer Reihe anderer Befunde, die ebenfalls durch die Density Hypothesis erklärt werden können (Boucher & Osgood, 1969; H. H. Clark & Clark, 1977; Ekman & Friesen, 1971; Potter et al., 2007; Semin & Fiedler, 1992).

Eine bedeutende Konsequenz der Globalität von positiven Halo-Effekten besteht in der starken Robustheit gegenüber Veränderungen. So lassen sich positive Halo-Effekte nur geringfügig durch die Präsentation von Bildern, zusätzlichen Informationen oder durch die Variierung der Darbietungsreihenfolge verstärken oder abschwächen. Demgegenüber sind negative Halo-Effekte in relativ starkem Maße beeinflussbar. Das bedeutet, dass bei einem gleichzeitigen Vorhandensein von positiven und negativen Stimuli, was in der Realität normalerweise der Fall ist, die positiven Halo-Effekte überwiegen. Dies trifft sogar dann zu, wenn die Anzahl der negativen Stimuli die Anzahl der positiven Stimuli übersteigt. In der Praxis ist es folglich wahrscheinlicher, dass Menschen zu positiv statt zu negativ beurteilt werden. Von besonderem Interesse ist der Einfluss der Präsentationsreihenfolge im Falle einer sequentiellen Informationsdarbietung, da auch im realen Leben unterschiedliche Verhaltensweisen ausschließlich nacheinander und nicht gleichzeitig beobachtet werden können. Steht eine einzige positive Verhaltensweise mehreren negativen Handlungen gegenüber, führt sie stets zu einem relativ starken positiven Halo-Effekt. Dabei ist es nicht von Bedeutung, ob das positive Verhalten am Anfang, in der Mitte, oder am Ende der Verhaltenskette auftritt. Wird jedoch eine negative Verhaltensweise inmitten mehrerer positiver Handlungen beobachtet, gibt es die stärksten negativen Halo-Effekte dann, wenn

die negative Verhaltensweise zum Schluss erfolgt. In den durchgeführten Erhebungen ist dieser Recency-Effekt damit erklärbar, dass die Urteilsabfrage direkt im Anschluss an die Präsentation aller Verhaltensweisen geschah (Howard & Kahana, 2002; Murdock, 1962; Murdock & Metcalfe, 1978). Ein möglicher Kritikpunkt wäre die Tatsache, dass in dem Experiment lediglich Gruppen von fünf Verhaltensweisen präsentiert wurden, die unter Umständen alle gleichzeitig im Kurzzeitgedächtnis gespeichert werden können (George A. Miller, 1956). Sollte dies der Fall sein, dürften gar keine Primacy- oder Recency-Effekte auftreten, wodurch es sich bei den Befunden um statistische Artefakte handeln würde. Da jedoch alle Verhaltensbeschreibungen sehr ausführlich und mit mehreren Beispielen formuliert waren, handelte es sich faktisch um mehr als nur fünf Items. Aus diesem Grund kann der Kritikpunkt zum Teil entkräftet werden.

Es ist anzumerken, dass sich die beschriebene Valenzasymmetrie im Wesentlichen auf die Darbietung von Einzelinformationen bezieht. Eine einzige positive Eigenschaft kann bereits einen positiven globalen Eindruck hervorrufen, während ein einzelnes negatives Attribut nicht zu einem negativen globalen Eindruck führt. Je mehr negative Eigenschaften einer Person dargeboten werden, desto eher kann auch ein negativer Gesamteindruck entstehen. Die Asymmetrie von positiven und negativen Halo-Effekten sinkt daher mit zunehmender Anzahl an Informationen. Anhand der Befunde ist ebenfalls erkennbar, dass auch dann kein Valenzunterschied auftritt, wenn statt der Beurteilung einzelner Attribute lediglich ein Gesamteindruck gebildet wird. Auf rein deskriptiver Ebene sind negative Gesamteindrücke sogar tendenziell stärker als positive Gesamteindrücke. Dieses Ergebnis kann durch die Tatsache erklärt werden, dass negative Merkmale oder Verhaltensweisen diagnostischer für die dahinterstehenden Attribute sind als positive Merkmale (Reeder & Brewer, 1979; Skowronski & Carlston, 1989). In Bezug auf Halo-Effekte lässt sich damit schlussfolgern, dass einzelne negative Attribute nur dann zu globalen Eindrücken führen, wenn diese direkt erfragt werden. Da automatisch keine negativen Gesamteindrücke gebildet werden, können auch keine negativen globalen Halo-Effekte entstehen.

Einfluss von Urteilsdimension und Kontextfaktoren

Agency und Communion wurden in der Vergangenheit bereits in einer Vielzahl von Studien untersucht, jedoch noch nicht gezielt mit Halo-Effekten in Verbindung gebracht. Diese Forschungslücke wird in der vorliegenden Arbeit aufgegriffen. Die Ergebnisse der

Experimente lassen darauf schließen, dass Communion-Halo-Effekte im Durchschnitt ein stärkeres Ausmaß haben als Agency-Halo-Effekte. Dieser Befund ist dadurch zu erklären, dass Communion-Attribute durchschnittlich bevorzugt verarbeitet werden, bei der Beurteilung anderer Personen wichtiger sind und stärker mit Sympathie in Verbindung stehen als Agency-Attribute (Abele & Bruckmüller, 2011; Asch, 1946; Scholer & Higgins, 2008; Suitner & Maass, 2008; Wojciszke & Abele, 2008; Wojciszke et al., 2009). Jedoch ist erkennbar, dass die Stärke von Agency- und Communion-Halo-Effekten auch mit einigen Kontextfaktoren interagiert. So sind etwa Communion-Halo-Effekte in privaten Situationen stärker als in beruflichen Situationen. Auch treten bei der Beurteilung von Personen mit sozialen und Pflegeberufen größere Communion-Halo-Effekte auf als bei der Einschätzung von Menschen mit naturwissenschaftlichen und technischen Berufen. Für Agency-Halo-Effekte gilt jeweils das umgekehrte Muster. Begründungen für diese Unterschiede lassen sich in den Befunden finden, dass die Bedeutung von Agency- und Communion-Attributen durch die Nähe und Interdependenz zwischen dem Beurteiler und der beurteilten Person beeinflusst wird und sich für Sympathie und Respekt unterscheidet (Abele & Wojciszke, 2007; Wojciszke & Abele, 2008; Wojciszke et al., 2009).

Die Ergebnisse der vorliegenden Arbeit lassen darauf schließen, dass sich Agency- und Communion-Halo-Effekte lediglich in ihrem Ausmaß und nicht ihren grundlegenden Prozessen voneinander unterscheiden. Das bedeutet, dass Agency- und Communion-Halo-Effekte bei konstanter Valenz gleichermaßen global oder spezifisch sind. Diese Befunde implizieren auch, dass sich Agency und Communion nicht in ihrer grundlegenden Struktur voneinander unterscheiden, also dieselbe Density aufweisen. In der Vergangenheit wurde eine solche Fragestellung noch nicht formuliert. Jedoch besagen die Daten einer in Druck befindlichen Untersuchung, dass sich Agency und Communion sehr wohl in ihrer Density unterscheiden (Bruckmüller & Abele, in press). Die beiden Autoren belegen, dass positive Agency-Attribute eine höhere Density aufweisen als positive Communion-Attribute. Bei negativer Valenz haben dagegen Communion-Attribute eine höhere Density als Agency-Attribute. Diese Asymmetrie widerspricht den Ergebnissen der vorliegenden Arbeit. Es ist allerdings anzumerken, dass die Interaktion von Dimension und Valenz im Vergleich zum Haupteffekt für Valenz einen relativ geringen Effekt hatte. Aufgrund der Gesamtheit aller Befunde kann vermutet werden, dass Agency und Communion zwar eine unterschiedliche Density aufweisen, diese Asymmetrie jedoch so geringfügig ist, dass die Qualität der Halo-

Effekte davon nicht erkennbar beeinflusst wird. Möglicherweise gibt es jedoch Methoden, die auch solche kleinen qualitativen Unterschiede sichtbar machen können.

Alle aufgeführten Effekte der Urteilsdimension sowie der Kontextfaktoren beruhen auf der Annahme, dass die Bedeutsamkeit eines Attributs entscheidend für die Stärke der daraus resultierenden Halo-Effekte ist. Zwar sprechen viele Befunde für diese Tatsache, jedoch wurde die Bedeutsamkeit der Attribute nur in einer allgemeinen Voruntersuchung, nicht aber in den einzelnen Experimenten als Mediator untersucht. Das Gleiche gilt für die Grundlage der Kontexthypothese, der zufolge Agency- und Communion-Eigenschaften in verschiedenen Situationen und Berufen von unterschiedlicher Bedeutung sind. Es ist daher nicht vollständig auszuschließen, dass neben oder anstelle der Bedeutsamkeit andere Mediatorvariablen wirken. Ein weiterer Ansatzpunkt angesichts der vorliegenden Befunde liegt in der Erforschung der Baseline von Agency- und Communion-Attributen. Zwar wurde bereits gezeigt, dass die durchschnittliche Sicht auf die Welt eher positiver Natur ist (H. H. Clark & Clark, 1977), jedoch gibt es diesbezüglich noch keine Vergleiche von Agency und Communion. Die vorliegende Arbeit lässt darauf schließen, dass Agency-Eigenschaften tendenziell positiver eingeschätzt werden als Communion-Eigenschaften. Damit stellt sie einen wichtigen Schritt in der Forschung dar. Zukünftige Experimente können an diesem Punkt ansetzen und die genannte Asymmetrie genauer untersuchen.

Ursachen und Vermeidung von Halo-Effekten

Nach dem aktuellen Stand der Forschung existieren bereits viele verschiedene Vermutungen und Theorien über die Ursachen von Halo-Effekten und eine Reihe von Strategien, um derartige Urteilsverzerrungen zu vermeiden. Legt man die Annahme eines einheitlichen Halo-Effekts zugrunde, sind diese Ansätze häufig mit Widersprüchen und Unklarheiten verbunden. Auf Basis der vorliegenden Befunde, denen zufolge Halo-Effekte verschiedene Formen annehmen können, ist es möglich, diese Widersprüche teilweise zu entkräften. Es muss nicht mehr angenommen werden, dass Halo-Effekte ausschließlich auf Aspekte der Person oder der Situation zurückzuführen sind. Beispielsweise können Halo-Effekte einerseits auf impliziten Persönlichkeitstheorien basieren (Schneider, 2004; Wolfradt, 2008), gleichzeitig jedoch auch durch Merkmale der Urteilsituation verursacht werden (Murphy & Anhalt, 1992). Welche der Ursachen gerade vorherrschend ist, sollte beeinflussen, welche Formen und Mechanismen von Halo-Effekten auftreten. Sind zum

Beispiel die zu beurteilenden Eigenschaften nicht klar definiert (Cooper, 1981b; Guilford, 1954; Symonds, 1925), wirken Halo-Effekte so, wie sie durch das Inadequate Discrimination Model beschrieben werden (Fiscaro & Lance, 1990).

Ein interessanter und bisher ungeklärter Punkt ist die Frage, wie Halo-Effekte von Diskriminierung abgegrenzt werden können. Zwar wurde der Zusammenhang zwischen beiden Phänomenen schon in wenigen frühen Studien thematisiert, (Caldwell & Mowry, 1934; Zebrowitz et al., 1993), später jedoch nicht mehr aufgegriffen. Allport (1954) beschreibt soziale Diskriminierung als Verweigerung, Individuen oder Gruppen die gleiche Behandlung zukommen zu lassen, die sie sich auch wünschen würden. Mummendey und Wenzel (1999) definieren soziale Diskriminierung dagegen als ein negatives, abwertendes Verhalten gegenüber einer Fremdgruppe, welches von dieser als ungerecht und illegitim empfunden, von der Eigengruppe jedoch als legitim und gerechtfertigt wahrgenommen wird. Beide Definitionen schließen nicht aus, dass Diskriminierung eine unmittelbare Folge negativer Halo-Effekte sein kann. So ist es etwa denkbar, dass ein ausländischer Name oder das Tragen eines Kopftuchs für bestimmte Personen ein negatives salientes Merkmal darstellt, welches unbewusst die Einschätzung anderer Eigenschaften negativ verzerrt und damit auch zu abwertenden Verhaltensweisen führt. Sollte dies zum Teil der Fall sein, wäre es notwendig, im Rahmen von Antidiskriminierungskampagnen auch über negative Halo-Effekte aufzuklären. Da Diskriminierung, wie zum Beispiel anhand der aktuellen Thematik der anonymisierten Bewerbungen deutlich wird, ein großes und allgegenwärtiges Problem darstellt, ist es wichtig, auf diesem Gebiet weitere Forschung zu leisten.

Die vorliegende Arbeit hat auch bedeutende Implikationen für die Vermeidung unerwünschter Halo-Effekte. In der bisherigen Forschung wiesen mehrere Autoren nach, dass Trainings und insbesondere das ausführliche Erklären von Halo-Effekten derartige Verzerrungen reduzieren oder sogar gänzlich verhindern können (Baltes & Parker, 2000; Bernardin & Walter, 1977; Brown, 1968; McIntyre et al., 1984; Pulakos, 1984). Andere Studien konnten diese Effekte dagegen nicht zeigen (Bernardin, 1978; Edwards & Waters, 1984; Franke et al., 1974; Murphy & Anhalt, 1992; Wetzels et al., 1981). Wie im Falle der Ursachen von Halo-Effekten können auch diese Widersprüche auf die Vielfältigkeit des Phänomens zurückgeführt werden. Möglicherweise wirken Trainings am besten, wenn die aktuell vorherrschende Form von Halo-Effekten diejenige ist, die auch zuvor im Training thematisiert wurde. Ein Verbesserungsvorschlag kann folglich dahingehend formuliert

werden, dass zukünftige Trainings über die Vielfältigkeit, die verschiedenen Formen und eventuelle Einflussfaktoren von Halo-Effekten aufklären sollten.

8.3 Ausblick auf zukünftige Forschungsmöglichkeiten

Auch in Zukunft wird es von hoher Relevanz sein, die Ursachen und Prozesse von Halo-Effekten in verschiedenen Kontexten zu erforschen. Die vorliegende Arbeit stellt in diesem Zusammenhang einen wichtigen Schritt dar, da sie die Vielfältigkeit von Halo-Effekten hervorhebt und den Einfluss verschiedener Faktoren bestätigt. Für zukünftige Untersuchungen wird es von großer Bedeutung sein, an dieser Stelle anzusetzen und zu versuchen, die gefundenen Ergebnisse zu replizieren und zu erweitern.

Um die Befunde der vorliegenden Arbeit generalisieren zu können, ist es vor allem notwendig, eine große Basis an verschiedenen Eigenschaften und Verhaltensweisen zu haben. Zwar wurden in der vorliegenden Arbeit umfangreiche Voruntersuchungen zur Generierung geeigneter Daten durchgeführt, jedoch ergaben sich daraus vergleichsweise wenig Materialien. Für künftige Studien ist es daher von Relevanz, die hier verwendete Datenbasis um zusätzliche Attribute zu erweitern. Ein weiterer bedeutender Punkt zur Generalisierung der Ergebnisse ist die Anwendung des Untersuchungsdesigns auf eine Vielzahl von Kontexten. Die durchgeführten Experimente zu dieser Thematik fokussierten ausschließlich die Beurteilung von Personen mit naturwissenschaftlichen und technischen beziehungsweise sozialen und Pflegeberufen. In zukünftigen Erhebungen sollten weitere Kontextfaktoren, wie beispielsweise das Alter, das Geschlecht oder der soziale Status der beurteilten Personen, betrachtet werden. Des Weiteren ist es von Interesse, die Rolle der Urteilssituation und auch die Merkmale der Beurteilenden, die in der vorliegenden Arbeit keine signifikanten Ergebnisse lieferten, noch einmal genauer zu untersuchen. Statt der Unterscheidung beruflicher und privater Situationen könnte es etwa angebrachter sein, formelle und informelle Interaktionssituationen voneinander abzugrenzen. Auch ist es möglich, dass Halo-Effekte zwar nicht von dem Beruf, jedoch von den grundlegenden Einstellungen, Persönlichkeitseigenschaften oder anderen Merkmalen der beurteilenden Personen beeinflusst werden, die es zu berücksichtigen gilt. Neben der Erweiterung des Materials und Designs ist es für die Aussagekraft der vorliegenden Arbeit von großer

Bedeutung, die gezeigten Ergebnisse mit anderen Maßen von Halo-Effekten zu replizieren. Dies betrifft insbesondere die Verwendung von Strukturgleichungsmodellen, welche es ebenso wie die hier vorgezogene Methode ermöglichen, globale und spezifische Halo-Effekte gleichzeitig zu erfassen. Auch können die anderen Maße dazu verwendet werden, eventuelle signifikante Befunde zusätzlich zu stützen.

Zusätzlich zu den genannten Ansätzen zur Generalisierung der Ergebnisse sollten zukünftige Untersuchungen dazu beitragen, die gezeigten Befunde auf andere Bereiche anzuwenden und zu vertiefen. Dazu gehört zum Beispiel die Betrachtung potentieller Moderator- und Mediatorvariablen. Die Hypothesen der vorliegenden Arbeit stützen sich unter anderem darauf, dass die Bedeutsamkeit eines Persönlichkeitsattributs das Ausmaß der daraus resultierenden Halo-Effekte mitbestimmt. Dies wurde hier jedoch nur indirekt anhand einer allgemeinen Voruntersuchung kontrolliert. Zukünftige Studien sollten diese und weitere Variablen in ihr Experimentaldesign aufnehmen. Des Weiteren kann es von Interesse sein, die Hypothesen dieser Arbeit nicht nur auf die Beurteilung von Personen, sondern auch auf die Bewertung von Unternehmen, Produkten und Dienstleistungen anzuwenden. Daraus folgt auch, dass nicht nur Agency und Communion, sondern auch andere Urteilsdimensionen in den Fokus der Aufmerksamkeit gelangen. Eine weitere Forschungsmöglichkeit besteht darin, nicht nur grundlegende Eigenschaftsdimensionen, sondern die einzelnen Eigenschaften selbst zu fokussieren. Dies ist insbesondere in der Praxis von hoher Relevanz. Wird beispielsweise ein Bewerber als viel kompetenter eingeschätzt als er in Wirklichkeit ist, ist es von großer Wichtigkeit zu erkennen, welche Eigenschaften oder Verhaltensweisen diesen Halo-Effekt im Detail verursacht haben. Ein abschließender Ansatzpunkt der vorliegenden Arbeit besteht darin, Halo-Effekte nicht nur anhand fiktiver Szenarien zu untersuchen, sondern Interaktionssituationen zu schaffen, in denen reale Verhaltensweisen beobachtet und daraufhin Urteile und Entscheidungen getroffen werden. Werden all die genannten Punkte berücksichtigt, wird es zukünftig möglich sein, Halo-Effekte in ihrer Vielfältigkeit noch besser zu verstehen.

9 Fazit

Die vorliegende Arbeit hatte das Ziel, den Einfluss verschiedener Faktoren auf die Prozesse und das Ausmaß von Halo-Effekten nachzuweisen. Auf diese Weise sollte gezeigt werden, dass kein einheitlicher Halo-Effekt existiert, sondern qualitativ und quantitativ unterschiedliche Halo-Effekte auftreten können. In diesem Zusammenhang wurde ein Modell entwickelt, das verschiedene Arten von Halo-Effekten gleichzeitig berücksichtigt. Basierend auf diesem Modell wurde postuliert, dass die Art und das Ausmaß von Halo-Effekten unabhängig voneinander durch die Valenz und Dimension der vorhandenen Informationen sowie verschiedene Kontextfaktoren bestimmt werden. Auf Grundlage der durchgeführten Experimente konnten diese Vermutungen bestätigt werden.

Wie die Literatur zeigt, sind Halo-Effekte in der Vergangenheit zwar gut erforscht worden, jedoch nicht frei von Unklarheiten und Widersprüchen. Insbesondere die Frage nach der Eindämmung und Vermeidung unerwünschter Halo-Effekte konnte aus diesem Grund bisher nicht zufriedenstellend beantwortet werden. Dies ist ein großes Problem, da Halo-Effekte allgegenwärtige Phänomene darstellen, welche zu schlechten Urteilen und Entscheidungen und im schlimmsten Fall zu weiteren negativen Konsequenzen führen können. Die vorliegende Arbeit stellt diesbezüglich einen wichtigen Schritt dar, da sie zu einem besseren Verständnis der vielfältigen Einflussfaktoren, Prozesse und Auswirkungen von Halo-Effekten beiträgt. Aufgrund der Berücksichtigung verschiedener Situationen und Personengruppen in den einzelnen Experimenten können diese Erkenntnisse zudem auf viele Kontexte angewandt werden. Damit ist eine wichtige Voraussetzung geschaffen, um erfolgreiche Maßnahmen gegen unerwünschte Halo-Effekte entwickeln zu können. Durch die weitere Erforschung dieser Thematik kann es ermöglicht werden, optimale Urteile und Entscheidungen in einer Vielzahl praktischer Bereiche weiter zu fördern.

Literaturverzeichnis

- Abele, A. E., & Bruckmüller, S. (2011). The bigger one of the "Big Two"? Preferential processing of communal information. *Journal of Experimental Social Psychology, 47*(5), 935-948.
- Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology, 38*(7), 1202-1217.
- Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*(5), 751-763.
- Abikoff, H., Courtney, M., Pelham, W. E., & Koplewicz, H. S. (1993). Teachers' ratings of disruptive behaviors: The influence of halo effects. *Journal of Abnormal Child Psychology, 21*(5), 519-533.
- Ailawadi, K. L., Harlam, B. A., César, J., & Trounce, D. (2006). Promotion profitability for a retailer: The role of promotion, brand, category, and store characteristics. *Journal of Marketing Research, 43*(4), 518-535.
- Alliger, G. M., & Williams, K. J. (1989). Confounding among measures of leniency and halo. *Educational and Psychological Measurement, 49*(1), 1-10.
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge: Addison-Wesley.
- Amishav, R., & Kimchi, R. (2010). Perceptual integrality of componential and configural information in faces. *Psychonomic Bulletin & Review, 17*(5), 743-748.
- Anastasiow, N. J. (1963). "Success" in first grade as seen by teachers: Gough's Adjective Check List and teachers' ratings. *Psychological Reports, 13*(2), 403-407.
- Anderson, N. H. (1971). Two more tests against change of meaning in adjective combinations. *Journal of Verbal Learning and Verbal Behavior, 10*(1), 75-85.

- Andrews, J. C., Netemeyer, R. G., & Burton, S. (1998). Consumer generalization of nutrient content claims in advertising. *Journal of Marketing*, *62*(4), 62-75.
- Anusic, I., Schimmack, U., Pinkus, R. T., & Lockwood, P. (2009). The nature and structure of correlations among Big Five ratings: The halo-alpha-beta model. *Journal of Personality and Social Psychology*, *97*(6), 1142-1156.
- Archer, J., & McCarthy, B. (1988). Personal biases in student assessment. *Educational Research*, *30*(2), 142-145.
- Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*(3), 258-290.
- Athey, T. R., & McIntyre, R. M. (1987). Effect of rater training on rater accuracy: Levels-of-processing theory and social facilitation theory perspectives. *Journal of Applied Psychology*, *72*(4), 567-572.
- Babad, E. (2005). The psychological price of media bias. *Journal of Experimental Psychology: Applied*, *11*(4), 245-255.
- Bagozzi, R. P. (1996). The role of arousal in the creation and control of the halo effect in attitude models. *Psychology and Marketing*, *13*(3), 235-264.
- Bakan, D. (1966). *The duality of human existence: Isolation and communion in Western man*. Boston: Beacon Press.
- Baltes, B. B., & Parker, C. P. (2000). Reducing the effects of performance expectations on behavioral ratings. *Organizational Behavior and Human Decision Processes*, *82*(2), 237-267.
- Balzer, W. K., & Sulsky, L. M. (1992). Halo and performance appraisal research: A critical examination. *Journal of Applied Psychology*, *77*(6), 975-985.
- Banks, W. P., & White, H. (1982). Single ordering as a processing limitation. *Journal of Verbal Learning and Verbal Behavior*, *21*(1), 39-54.

- Bartlett, C. J. (1983). What's the difference between valid and invalid halo? Forced-choice measurement without forcing a choice. *Journal of Applied Psychology, 68*(2), 218-226.
- Becker, B. E., & Cardy, R. L. (1986). Influence of halo error on appraisal effectiveness: A conceptual and empirical reconsideration. *Journal of Applied Psychology, 71*(4), 662-671.
- Beckwith, N. E., & Lehmann, D. R. (1975). The importance of halo effects in multi-attribute attitude models. *Journal of Marketing Research, 12*(3), 265-275.
- Bernardin, H. J. (1977). Behavioral expectation scales versus summated scales: A fairer comparison. *Journal of Applied Psychology, 62*(4), 422-427.
- Bernardin, H. J. (1978). Effects of rater training on leniency and halo errors in student ratings of instructors. *Journal of Applied Psychology, 63*(3), 301-308.
- Bernardin, H. J., Cardy, R. L., & Carlyle, J. J. (1982). Cognitive complexity and appraisal effectiveness: Back to the drawing board? *Journal of Applied Psychology, 67*(2), 151-160.
- Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology, 65*(1), 60-66.
- Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. *Journal of Applied Psychology, 62*(1), 64-69.
- Bingham, W. V. (1939). Halo, invalid and valid. *Journal of Applied Psychology, 23*(2), 221-228.
- Blum, M. L., & Naylor, J. C. (1968). *Industrial psychology: Its theoretical and social foundations*. New York: Harper & Row.
- Boatwright, P., Kalra, A., & Zhang, W. (2008). Should consumers use the halo to form product evaluations? *Management Science, 54*(1), 217-223.

- Borkenau, P., & Ostendorf, F. (1987). Ein Versuch der Trennung deskriptiver und evaluativer Bedeutungskomponenten in Eigenschafts-beschreibenden Begriffen. *Archiv für Psychologie*, 139(3), 189-207.
- Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology*, 60(5), 556-560.
- Bortz, J. (2005). Überprüfung von Korrelationshypothesen. In J. Bortz (Ed.), *Statistik für Human- und Sozialwissenschaftler* (pp. 213-224). Heidelberg: Springer Medizin Verlag.
- Boucher, J., & Osgood, C. E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 8(1), 1-8.
- Bowden, A. O., Caldwell, F. F., & West, G. A. (1934). Halo prestige. *Journal of Abnormal and Social Psychology*, 28(4), 400-406.
- Bownas, D. A., & Bernardin, H. J. (1991). Suppressing illusory halo with forced-choice items. *Journal of Applied Psychology*, 76(4), 592-594.
- Bradshaw, J. L. (1969). The information conveyed by varying the dimensions of features in human outline faces. *Perception & Psychophysics*, 6(1), 5-9.
- Brown, E. M. (1968). Influence of training, method, and relationship on the halo effect. *Journal of Applied Psychology*, 52(3), 195-199.
- Bruckmüller, S., & Abele, A. E. (in press). The density of the "Big Two": How are agency and communion structurally represented? *Social Psychology, Special Issue on the BIG Two*.
- Bruner, J. S., & Taguiri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 634-654). Cambridge: Addison-Wesley.
- Buffardi, L., & Gibson, J. F. (1980). Relation between raters' characteristics and halo error. *Perceptual and Motor Skills*, 51(3), 1003-1011.

- Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science, 11*(5), 429-433.
- Caldwell, F. F., & Mowry, M. D. (1934). Teachers' grades as criteria of achievement of bilingual children. *Journal of Applied Psychology, 18*(2), 288-292.
- Chandon, P., & Wansink, B. (2007). The biasing health halos of fast-food restaurant health claims: Lower calorie estimates and higher side-dish consumption intentions. *Journal of Consumer Research, 34*(3), 301-314.
- Clark, C. C., & Lawless, H. T. (1994). Limiting response alternatives in time-intensity scaling: An examination of the halo-dumping effect. *Chemical Senses, 19*(6), 583-594.
- Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- Cook, G. I., Marsh, R. L., & Hicks, J. L. (2003). Halo and devil effects demonstrate valenced-based influences on source-monitoring decisions. *Consciousness and Cognition, 12*(2), 257-278.
- Coombs, W. T., & Holladay, S. J. (2006). Unpacking the halo effect: Reputation and crisis management. *Journal of Communication Management, 10*(2), 123-137.
- Cooper, W. H. (1981a). Conceptual similarity as a source of illusory halo in job performance ratings. *Journal of Applied Psychology, 66*(3), 302-307.
- Cooper, W. H. (1981b). Ubiquitous halo. *Psychological Bulletin, 90*(2), 218-244.
- Cooper, W. H. (1983). Internal homogeneity, descriptiveness, and halo: Resurrecting some answers and questions about the structure of job performance rating categories. *Personnel Psychology, 36*(3), 489-502.
- Coren, S. (1998). Student evaluations of an instructor's racism and sexism: Truth or expedience? *Ethics & Behavior, 8*(3), 201-213.
- Crissy, W. J. E., & Regan, J. J. (1951). Halo in the employment interview. *Journal of Applied Psychology, 35*(5), 338-341.

- Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44*(1), 20-33.
- Davenport, W. G., Brooker, G., & Munro, N. (1971). Factors in social perception: Seating position. *Perceptual and Motor Skills, 33*(3), 747-752.
- Davison, D. C. (1973). Perception of instructor in relation to self and evaluation of instructor's performance. *Perceptual and Motor Skills, 36*(2), 533-534.
- De Soto, C. B. (1961). The predilection for single orderings. *Journal of Abnormal and Social Psychology, 62*(1), 16-23.
- Dennis, I. (2007). Halo effects in grading student projects. *Journal of Applied Psychology, 92*(4), 1169-1176.
- DeRath, G., & Carp, F. M. (1959). The picture-choice test as an indirect measure of attitudes. *Journal of Applied Psychology, 43*(1), 12-15.
- DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology, 91*(6), 1138-1151.
- Dickinson, T. L., & Glebocki, G. G. (1990). Modifications in the format of the mixed standard scale. *Organizational Behavior and Human Decision Processes, 47*(1), 124-137.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*(1), 417-440.
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*(6), 1246-1256.
- Dillon, W. R., Mulani, N., & Frederick, D. G. (1984). Removing perceptual distortions in product space analysis. *Journal of Marketing Research, 21*(2), 184-193.
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285-290.

- Dompnier, B., Pansu, P., & Bressoux, P. (2006). An integrative model of scholastic judgments: Pupils' characteristics, class context, halo effect and internal attributions. *European Journal of Psychology of Education, 21*(2), 119-133.
- Edwards, J. E., & Waters, L. K. (1984). Halo and leniency control in ratings as influenced by format, training, and rater characteristic differences. *Managerial Psychology, 5*(1), 1-16.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124-129.
- Engelhard, G. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement, 31*(2), 93-112.
- Epkins, C. C. (1994). Peer ratings of depression, anxiety, and aggression in inpatient and elementary school children: Rating biases and influence of rater's self-reported depression, anxiety, and aggression. *Journal of Abnormal Child Psychology, 22*(5), 611-628.
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482-498.
- Feeley, T. H. (2002a). Comment on halo effects in rating and evaluation research. *Human Communication Research, 28*(4), 578-586.
- Feeley, T. H. (2002b). Evidence of halo effects in student evaluations of communication instruction. *Communication Education, 51*(3), 225-236.
- Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111*(2), 304-341.
- Feldman, J. M. (1986). A note on the statistical correction of halo error. *Journal of Applied Psychology, 71*(1), 173-176.
- Fensterheim, H., & Tresselt, M. E. (1953). The influence of value systems on the perception of people. *Journal of Abnormal and Social Psychology, 48*(1), 93-98.

- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117-140.
- Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, 107(4), 659-676.
- Fiedler, K., & Armbruster, T. (1994). Two halves may be more than one whole: Category-split effects on frequency illusions. *Journal of Personality and Social Psychology*, 66(4), 633-645.
- Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition and Emotion*, 25(4), 639-656.
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52(3), 399-433.
- Fisicaro, S. A. (1988). A reexamination of the relation between halo error and accuracy. *Journal of Applied Psychology*, 73(2), 239-244.
- Fisicaro, S. A., & Lance, C. E. (1990). Implications of three causal models for the measurement of halo error. *Applied Psychological Measurement*, 14(4), 419-429.
- Fisicaro, S. A., & Vance, R. J. (1994). Comments on the measurement of halo. *Educational and Psychological Measurement*, 54(2), 366-371.
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77-83.
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878-902.
- Follman, J., Wiley, R., Geiger, G., & Lavelly, C. (1974). Effects of halo-effect instructions on level of ratings and intercorrelations. *Psychological Reports*, 35(1), 392.
- Foon, A. E. (1986). Social psychology as science or history: An experimental approach. *The Journal of Social Psychology*, 126(4), 431-435.

- Foster, G., & Ysseldyke, J. (1976). Expectancy and halo effects as a result of artificially induced teacher bias. *Contemporary Educational Psychology, 1*(1), 37-45.
- Fox, S., Bizman, A., & Herrmann, E. (1983). The halo effect: Is it a unitary concept? *Journal of Occupational Psychology, 56*(4), 289-296.
- Fox, S., Bizman, A., & Hoffman, M. (1989). The halo effect: It really isn't unitary: A rejoinder to Nathan (1986). *Journal of Occupational Psychology, 62*(2), 183-188.
- Fox, S., Caspy, T., & Reisler, A. (1994). Variables affecting leniency, halo and validity of self-appraisal. *Journal of Occupational and Organizational Psychology, 67*(1), 45-56.
- Franke, J., Hoffmann, K., & Kamphaus, G. (1974). On possible training against "halo effects" in judging people. *Psychologische Beiträge, 16*(2), 188-202.
- Frone, M. R., Adams, J., Rice, R. W., & Instone-Noonan, D. (1986). Halo error: A field study comparison of self- and subordinate evaluations of leadership process and leader effectiveness. *Personality and Social Psychology Bulletin, 12*(4), 454-461.
- Gardner, D. C. (1974). Halo effects in employer ratings of career education students and coordinators. *Psychology in the Schools, 11*(3), 363-366.
- Geiselman, R. E., Haight, N. A., & Kimata, L. G. (1984). Context effects on the perceived physical attractiveness of faces. *Journal of Experimental Social Psychology, 20*(5), 409-424.
- Gibb, G. D. (1983). Influence of "halo" and "demon" effects in subjective grading. *Perceptual and Motor Skills, 56*(1), 67-70.
- Godoy, R., Reyes-García, V., Huanca, T., Tanner, S., Leonard, W. R., McDade, T., & Vadez, V. (2005). Do smiles have a face value? Panel evidence from Amazonian Indians. *Journal of Economic Psychology, 26*(4), 469-490.
- Goffin, R. D., Jolley, R. B., & Wagner, S. H. (2003). Is halo helpful? Effects of inducing halo on performance rating accuracy. *Social Behavior and Personality, 31*(6), 625-636.

- Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229.
- Goldman, M., Cowles, M. D., & Florez, C. A. (1983). The halo effect of an initial impression upon speaker and audience. *The Journal of Social Psychology, 120*(2), 197-201.
- Grove, B. A., & Kerr, W. A. (1951). Specific evidence on origin of halo effect in measurement of employee morale. *Journal of Social Psychology, 34*(2), 165-170.
- Gruenfeld, L., & Arbuthnot, J. (1969). Field independence as a conceptual framework for prediction of variability in ratings of others. *Perceptual and Motor Skills, 28*(1), 31-44.
- Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.
- Hallworth, H. J. (1961). Teachers' personality ratings of high school pupils. *Journal of Educational Psychology, 52*(6), 297-302.
- Harvey, R. J. (1982). The future of partial correlation as a means to reduce halo in performance ratings. *Journal of Applied Psychology, 67*(2), 171-176.
- Henik, A., & Tzelgov, J. (1985). Control of halo error: A multiple regression approach. *Journal of Applied Psychology, 70*(3), 577-580.
- Henrikson, E. H. (1943). Comparison of ratings of voice and teaching ability. *Journal of Educational Psychology, 34*(2), 121-123.
- Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology, 88*(3), 297-306.
- Holbrook, M. B. (1983). Using a structural model of halo effect to assess perceptual distortion due to affective overtones. *Journal of Consumer Research, 10*(2), 247-252.
- Holzbach, R. L. (1978). Rater bias in performance ratings: Superior, self-, and peer ratings. *Journal of Applied Psychology, 63*(5), 579-588.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*(3), 269-299.

- Hulin, C. L. (1982). Some reflections on general performance dimensions and halo rating error. *Journal of Applied Psychology, 67*(2), 165-170.
- Intons-Peterson, M. J., & Samuels, A. K. (1978). The cultural halo effect: Black and White women rate Black and White men. *Bulletin of the Psychonomic Society, 11*(5), 309-312.
- Iramaneerat, C., & Yudkowsky, R. (2007). Rater errors in a clinical skills assessment of medical students. *Evaluation & the Health Professions, 30*(3), 266-283.
- Jackson, C. J. (1996). An individual differences approach to the halo-accuracy paradox. *Personality and Individual Differences, 21*(6), 947-957.
- Jackson, C. J., & Furnham, A. (2001). Appraisal ratings, halo, and selection: A study using sales staff. *European Journal of Psychological Assessment, 17*(1), 17-24.
- Jacobs, R., & Kozlowski, S. W. J. (1985). A closer look at halo error in performance ratings. *Academy of Management Journal, 28*(1), 201-212.
- Jaeger, R. M., & Freijo, T. D. (1975). Race and sex as concomitants of composite halo in teachers' evaluative rating of pupils. *Journal of Educational Psychology, 67*(2), 226-237.
- Jennings, T., Palmer, J. K., & Thomas, A. (2004). Effects of performance context on processing speed and performance ratings. *Journal of Business and Psychology, 18*(4), 453-463.
- Johnson, D. M. (1963). Reanalysis of experimental halo effects. *Journal of Applied Psychology, 47*(1), 46-47.
- Johnson, D. M., & Vidulich, R. N. (1956). Experimental manipulation of the halo effect. *Journal of Applied Psychology, 40*(2), 130-134.
- Jones, L., & Fletcher, C. (2002). Self-assessment in a selection situation: An evaluation of different measurement approaches. *Journal of Occupational and Organizational Psychology, 75*(2), 145-161.

- Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology, 89*(6), 899-913.
- Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review, 104*(2), 344-366.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341-350.
- Kammann, R., Smith, R., Martin, C., & McQueen, M. (1984). Low accuracy in judgments of others' psychological well-being as seen from a phenomenological perspective. *Journal of Personality, 52*(2), 107-123.
- Kaplan, M. F. (1974). Context-induced shifts in personality trait evaluation: A comment on the evaluative halo effect and the meaning change interpretations. *Psychological Bulletin, 81*(11), 891-895.
- Kaplan, R. M. (1978). Is beauty talent? Sex interaction in the attractiveness halo effect. *Sex Roles, 4*(2), 195-204.
- Kasten, R., & Weintraub, Z. (1999). Rating errors and rating accuracy: A field experiment. *Human Performance, 12*(2), 137-153.
- Katz, S. (2003). Physical appearance: The importance of being beautiful. In J. M. Henslin (Ed.), *Down to earth sociology: Introductory readings* (pp. 313-320). New York: Free Press.
- Kiker, D. S., & Motowidlo, S. J. (1998). Effects of rating strategy on interdimensional variance, reliability, and validity of interview ratings. *Journal of Applied Psychology, 83*(5), 763-768.
- King, L. M., Hunter, J. E., & Schmidt, F. L. (1980). Halo in a multidimensional forced-choice performance evaluation scale. *Journal of Applied Psychology, 65*(5), 507-516.

- Klauer, K. C. (1988). Die Consider-the-opposite-Strategie für Korrektur und Analyse von Halo-Fehlern. *Zeitschrift für experimentelle und angewandte Psychologie*, 35(4), 597-627.
- Klauer, K. C., & Schmeling, A. (1990). Sind Halo-Fehler Flüchtigkeitsfehler? *Zeitschrift für experimentelle und angewandte Psychologie*, 37(4), 594-607.
- Kleiter, G. D., Doherty, M. E., & Brake, G. L. (2002). The psychophysics metaphor in calibration research. In P. Sedlmeier & T. Betsch (Eds.), *Frequency processing and cognition* (pp. 239-255). Oxford: Oxford University Press.
- Kozlowski, S. W. J., & Kirsch, M. P. (1987). The systematic distortion hypothesis, halo, and accuracy: An individual-level analysis. *Journal of Applied Psychology*, 72(2), 252-261.
- Kozlowski, S. W. J., Kirsch, M. P., & Chao, G. T. (1986). Job knowledge, rater familiarity, conceptual similarity and halo error: An exploration. *Journal of Applied Psychology*, 71(1), 45-49.
- Kozup, J. C., Creyer, E. H., & Burton, S. (2003). Making healthful food choices: The influence of health claims and nutrition information on consumers' evaluations of packaged food products and restaurant menu items. *Journal of Marketing*, 67(2), 19-34.
- Lachman, S. J., & Bass, A. R. (1985). A direct study of halo effect. *The Journal of Psychology: Interdisciplinary and Applied*, 119(6), 535-540.
- Lance, C. E., Fiscaro, S. A., & LaPointe, J. A. (1990). An examination of negative halo error in ratings. *Educational and Psychological Measurement*, 50(3), 545-554.
- Lance, C. E., LaPointe, J. A., & Fiscaro, S. A. (1994). Tests of three causal models of halo rater error. *Organizational Behavior and Human Decision Processes*, 57(1), 83-96.
- Lance, C. E., LaPointe, J. A., & Stewart, A. M. (1994). A test of the context dependency of three causal models of halo rating error. *Journal of Applied Psychology*, 79(3), 332-340.
- Lance, C. E., & Woehr, D. J. (1986). Statistical control of halo: Clarification from two cognitive models of the performance appraisal process. *Journal of Applied Psychology*, 71(4), 679-685.

- Landy, D., & Aronson, E. (1969). The influence of the character of the criminal and his victim on the decisions of simulated jurors. *Journal of Experimental Social Psychology, 5*(2), 141-152.
- Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology, 29*(3), 299-304.
- Landy, F. J., Vance, R. J., & Barnes-Farrell, J. L. (1982). Statistical control of halo: A response. *Journal of Applied Psychology, 67*(2), 177-180.
- Landy, F. J., Vance, R. J., Barnes-Farrell, J. L., & Steele, J. W. (1980). Statistical control of halo error in performance ratings. *Journal of Applied Psychology, 65*(5), 501-506.
- Larose, H., & Standing, L. (1998). Does the halo effect occur in the elderly? *Social Behavior and Personality, 26*(2), 147-150.
- Latham, G. P., Wexley, K. N., & Pursell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology, 60*(5), 550-555.
- Lau, S. (1982). The effect of smiling on person perception. *Journal of Social Psychology, 117*(1), 63-67.
- Lee, R., Malone, M., & Greco, S. (1981). Multitrait-multimethod-multirater analysis of performance ratings for law enforcement personnel. *Journal of Applied Psychology, 66*(5), 625-632.
- Leitner, L. A. (1972). Discrimination of counselor interpersonal skills in self and others. *Journal of Counseling Psychology, 19*(6), 509-511.
- Lucker, G. W., Beane, W. E., & Helmreich, R. L. (1981). The strength of the halo effect in physical attractiveness research. *The Journal of Psychology: Interdisciplinary and Applied, 107*(1), 69-75.
- McIntyre, R. M., Smith, D. E., & Hasset, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology, 69*(1), 147-156.

- McNeil, D. W., & Janzen, W. B. (1987). Attitudes toward problem drinkers: Cognitive, affective, and behavioral intention dimensions. *International Journal of the Addictions, 22*(10), 993-1017.
- Meindl, J. R., & Ehrlich, S. B. (1987). The romance of leadership and the evaluation of organizational performance. *Academy of Management Journal, 30*(1), 91-109.
- Michelson, L., Mannarino, A., Marchione, K., Kazdin, A. E., & Costello, A. (1985). Expectancy bias in behavioral observations of therapeutic outcome: An experimental analysis of treatment and halo effects. *Behaviour Research and Therapy, 23*(4), 407-414.
- Miller, A. G. (1970). Role of physical attractiveness in impression formation. *Psychonomic Science, 19*(4), 241-243.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.
- Miller, G. A., & Chapman, L. J. (1968). Response bias and schizophrenic beliefs. *Journal of Abnormal Psychology, 73*(3), 252-255.
- Mills, J., & O'Neal, E. (1971). Anticipated choice, attention, and halo effect. *Psychonomic Science, 22*(4), 231-233.
- Min Han, C. (1989). Country image: Halo or summary construct? *Journal of Marketing Research, 26*(2), 222-229.
- Mossholder, K. W., & Giles, W. F. (1983). The use of partial correlation to control halo in performance ratings. *Educational and Psychological Measurement, 43*(4), 977-984.
- Mount, M. K. (1984). Psychometric properties of subordinate ratings of managerial performance. *Personnel Psychology, 37*(4), 687-702.
- Mount, M. K., & Thompson, D. E. (1987). Cognitive categorization and quality of performance ratings. *Journal of Applied Psychology, 72*(2), 240-246.
- Mulaik, S. A., & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling, 7*(1), 36-73.

- Mumma, G. H. (2002). Effects of three types of potentially biasing information on symptom severity judgments for major depressive episode. *Journal of Clinical Psychology, 58*(10), 1327-1345.
- Mummendey, A., & Wenzel, M. (1999). Social discrimination and tolerance in intergroup relations: Reactions to intergroup difference. *Personality and Social Psychology Review, 3*(2), 158-174.
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*(5), 482-488.
- Murdock, B. B., & Metcalfe, J. (1978). Controlled rehearsal in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior, 17*(3), 309-324.
- Murphy, K. R. (1982). Difficulties in the statistical control of halo. *Journal of Applied Psychology, 67*(2), 161-164.
- Murphy, K. R., & Anhalt, R. L. (1992). Is halo error a property of the rater, ratees, or the specific behaviors observed? *Journal of Applied Psychology, 77*(4), 494-500.
- Murphy, K. R., & Balzer, W. K. (1986). Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy. *Journal of Applied Psychology, 71*(1), 39-44.
- Murphy, K. R., & Jako, R. A. (1989). Under what conditions are observed intercorrelations greater or smaller than true intercorrelations? *Journal of Applied Psychology, 74*(5), 827-830.
- Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and consequences of halo error: A critical analysis. *Journal of Applied Psychology, 78*(2), 218-225.
- Murphy, K. R., & Reynolds, D. H. (1988). Does true halo affect observed halo? *Journal of Applied Psychology, 73*(2), 235-238.
- Murray, S. L., & Holmes, J. G. (1997). A leap of faith? Positive illusions in romantic relationships. *Personality and Social Psychology Bulletin, 23*(6), 586-604.

- Myers, J. H. (1965). Removing halo from job evaluation factor structure. *Journal of Applied Psychology, 49*(3), 217-221.
- Nathan, B. R. (1986). The halo effect: It is a unitary concept! A reply to Fox et al. (1983). *Journal of Occupational Psychology, 59*(1), 41-44.
- Nathan, B. R., & Lord, R. G. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. *Journal of Applied Psychology, 68*(1), 102-114.
- Nathan, B. R., & Tippins, N. (1990). The consequences of halo "error" in performance ratings: A field study of the moderating effect of halo on test validation results. *Journal of Applied Psychology, 75*(3), 290-296.
- Newcomb, T. (1931). An experiment designed to test the validity of a rating technique. *Journal of Educational Psychology, 22*(4), 279-289.
- Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*(4), 250-256.
- O'Neal, E. (1971). Influence of future choice importance and arousal upon the halo effect. *Journal of Personality and Social Psychology, 19*(3), 334-340.
- O'Neal, E., & Mills, J. (1969). The influence of anticipated choice on the halo effect. *Journal of Experimental Social Psychology, 5*(3), 347-351.
- O'Sullivan, M. (2003). The fundamental attribution error in detecting deception: The boy-who-cried-wolf effect. *Personality and Social Psychology Bulletin, 29*(10), 1316-1327.
- Oh, M.-Y., & Ramaprasad, J. (2003). Halo effect: Conceptual definition and empirical exploration with regard to South Korean subsidiaries of US and Japanese multinational corporations. *Journal of Communication Management, 7*(4), 317-330.
- Ostroff, C. (1993). Rater perceptions, satisfaction and performance ratings. *Journal of Occupational and Organizational Psychology, 66*(4), 345-356.

- Ostroff, C., & Ilgen, D. R. (1992). Cognitive categories of raters and rating accuracy. *Journal of Business and Psychology, 7*(1), 3-26.
- Otta, E., Barcellos Pereira Lira, B., Delevati, N. M., Cesar, O. P., & Salati Guirello Pires, C. (1993). The effects of smiling and of head tilting on person perception. *The Journal of Psychology: Interdisciplinary and Applied, 128*(3), 323-331.
- Paine, F. T., Carroll, S. J., & Leete, B. A. (1966). Need satisfactions of managerial level personnel in a government agency. *Journal of Applied Psychology, 50*(3), 247-249.
- Palmer, J. K., & Feldman, J. M. (2005). Accountability and need for cognition effects on contrast, halo, and accuracy in performance ratings. *The Journal of Psychology: Interdisciplinary and Applied, 139*(2), 119-137.
- Palmer, J. K., & Loveland, J. M. (2008). The influence of group discussion on performance judgments: Rating accuracy, contrast effects, and halo. *The Journal of Psychology: Interdisciplinary and Applied, 142*(2), 117-130.
- Palmer, J. K., Thomas, A., & Maurer, T. (2003). Moderating effects of context on the relationship between behavioral diaries and performance rating halo and accuracy. *North American Journal of Psychology, 5*(1), 81-90.
- Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin, 92*(2), 382-402.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534-552.
- Payne, J. W., Bettman, J. R., & Luce, M. F. (1996). When time is money: Decision behavior under opportunity-cost time pressure. *Organizational Behavior and Human Decision Processes, 66*(2), 131-152.
- Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology, 7*(4), 1-18.

- Peeters, G. (1991). Evaluative inference in social cognition: The roles of direct versus indirect evaluation and positive-negative asymmetry. *European Journal of Social Psychology, 21*(2), 131-146.
- Pellegrini, R. J. (1971). Some effects of seating position on social perception. *Psychological Reports, 28*(3), 887-893.
- Petzold, P. (1992). Context effects in judgments of attributes: An information-integration approach. In H.-G. Geissler, S. W. Link & J. T. Townsend (Eds.), *Cognition, information processing, and psychophysics: Basic issues* (pp. 175-205). Hillsdale: Lawrence Erlbaum Associates, Inc.
- Phelps, L., Schmitz, C. D., & Boatright, B. (1986). The effects of halo and leniency on cooperating teacher reports using Likert-type rating scales. *Journal of Educational Research, 79*(3), 151-154.
- Pike, G. R. (1999). The constant error of the halo in educational outcomes research. *Research in Higher Education, 40*(1), 61-86.
- Potter, T., Corneille, O., Ruys, K. I., & Rhodes, G. (2007). "Just another pretty face": A multidimensional scaling approach to face attractiveness and variability. *Psychonomic Bulletin & Review, 14*(2), 368-372.
- Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology, 69*(4), 581-588.
- Pulakos, E. D., Schmitt, N., & Ostroff, C. (1986). A warning about the use of a standard deviation across dimensions within rates to measure halo. *Journal of Applied Psychology, 71*(1), 29-32.
- Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86*(1), 61-79.
- Remmers, H. H. (1934). Reliability and halo effect of high school and college students' judgments of their teachers. *Journal of Applied Psychology, 18*(5), 619-630.

- Remmers, H. H., & Martin, R. D. (1944). Halo effect in reverse - are teachers' ratings of high-school pupils valid? *Journal of Educational Psychology, 35*(4), 193-200.
- Rizzo, W. A., & Frank, F. D. (1977). Influence of irrelevant cues and alternate forms of graphic rating scales on the halo effect. *Personnel Psychology, 30*(3), 405-417.
- Rotshtein, P., Geng, J. J., Driver, J., & Dolan, R. J. (2007). Role of features and second-order spatial relations in face discrimination, face recognition, and individual face skills: Behavioral and magnetic resonance imaging data. *Journal of Cognitive Neuroscience, 19*(9), 1435-1452.
- Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin, 88*(2), 413-428.
- Sappenfield, B. R. (1971). Social desirability, the halo effect, and stereotypical perception in person perception and self-perception. *Perceptual and Motor Skills, 33*(3), 683-689.
- Sattler, J. M., Hillix, W. A., & Neher, L. A. (1970). Halo effect in examiner scoring of intelligence test responses. *Journal of Consulting and Clinical Psychology, 34*(2), 172-176.
- Sattler, J. M., & Winget, B. M. (1970). Intelligence testing procedures as affected by expectancy and IQ. *Journal of Clinical Psychology, 26*(4), 446-448.
- Sawyer, C. E., & Raybould, E. C. (1981). The visual halo effect in rating scales for infants. *Educational Studies, 7*(1), 47-53.
- Schachar, R., Sandberg, S., & Rutter, M. (1986). Agreement between teachers' ratings and observations of hyperactivity, inattentiveness, and defiance. *Journal of Abnormal Child Psychology, 14*(2), 331-345.
- Schmitt, M. (1992). Schönheit und Talent: Untersuchungen zum Verschwinden des Halo-Effekts. *Zeitschrift für experimentelle und angewandte Psychologie, 39*(3), 475-492.
- Schneider, D. J. (2004). *The psychology of stereotyping*. New York: Guilford Press.

- Scholer, A. A., & Higgins, E. T. (2008). People as resources: Exploring the functionality of warm and cold. *European Journal of Social Psychology, 38*(7), 1111-1120.
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*(2), 195-202.
- Schwarzer, G., & Massaro, D. W. (2001). Modeling face identification processing in children and adults. *Journal of Experimental Child Psychology, 79*(2), 139-161.
- Semin, G. R., & Fiedler, K. (1992). *Language, interaction and social cognition*. London: Sage Publications.
- Sinclair, R. C. (1988). Mood, categorization breadth, and performance appraisal: The effects of order of information acquisition and affective state on halo, accuracy, information retrieval, and evaluations. *Organizational Behavior and Human Decision Processes, 42*(1), 22-46.
- Sine, W. D., Shane, S., & Di Gregorio, D. (2003). The halo effect and technology licensing: The influence of institutional prestige on the licensing of university inventions. *Management Science, 49*(4), 478-496.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*(1), 131-142.
- Solomonson, A. L., & Lance, C. E. (1997). Examination of the relationship between true halo and halo error in performance ratings. *Journal of Applied Psychology, 82*(5), 665-674.
- Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology, 98*(3), 520-534.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*(2), 245-251.

- Stevens, J., Quittner, A. L., & Abikoff, H. (1998). Factors influencing elementary school teachers' ratings of ADHD and ODD behaviors. *Journal of Clinical Child Psychology, 27*(4), 406-414.
- Suitner, C., & Maass, A. (2008). The role of valence in the perception of agency and communion. *European Journal of Social Psychology, 38*(7), 1073-1082.
- Surawski, M. K., & Ossoff, E. P. (2006). The effects of physical and vocal attractiveness on impression formation of politicians. *Current Psychology: Developmental, Learning, Personality, Social, 25*(1), 15-27.
- Suter, W. N., & Roberts, W. L. (1987). An experimental investigation of the beliefs-of-relatedness source of halo. *Contemporary Educational Psychology, 12*(1), 77-85.
- Symonds, P. M. (1925). Notes on rating. *Journal of Applied Psychology, 9*(2), 188-195.
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46*(2), 225-245.
- Taylor, K., Bernardin, H. J., & Riegelhaupt, B. J. (1984). Halo error: An assessment of priming as a reduction technique. *Perceptual and Motor Skills, 59*(2), 447-457.
- Terman, L. M. (1939). The effect of happiness or unhappiness on self-report regarding attitudes, reaction patterns, and facts of personal history. *Psychological Bulletin, 36*(3), 197-202.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology, 4*(1), 25-29.
- Toops, H. A. (1923). The measurement of teachers. *Journal of Educational Psychology, 14*(5), 310-311.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*(3), 299-314.
- Tsui, A. S., & Barry, B. (1986). Interpersonal affect and rating errors. *Academy of Management Journal, 29*(3), 586-599.

- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.
- Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101*(4), 547-567.
- Tziner, A. (1984). A fairer examination of rating scales when used for performance appraisal in a real organizational setting. *Journal of Occupational Behaviour, 5*(2), 103-112.
- Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology, 95*(1), 36-49.
- Van Scotter, J. R., Moustafa, K., Burnett, J. R., & Michael, P. G. (2007). Influence of prior acquaintance with the ratee on rater accuracy and halo. *Journal of Management Development, 26*(8), 790-803.
- Vance, R. J., Winne, P. S., & Wright, E. S. (1983). A longitudinal examination of rater and ratee effects in performance ratings. *Personnel Psychology, 36*(3), 609-620.
- Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*(1), 108-131.
- Vlaev, I., & Chater, N. (2007). Context effects in games: Local versus global sequential effects on choice in the prisoner's dilemma game. *Judgment and Decision Making, 2*(5), 380-389.
- Wade, T. J., Fuller, L., Bresnan, J., Schaefer, S., & Mlynarski, L. (2007). Weight halo effects: Individual differences in personality evaluations and perceived life success of men as a function of weight? *Personality and Individual Differences, 42*(2), 317-324.
- Wade, T. J., Loyden, J., Renninger, L., & Tobey, L. (2003). Weight halo effects: Individual differences in personality evaluations as a function of weight? *Personality and Individual Differences, 34*(2), 263-268.

- Wansink, B. (2003). Overcoming the taste stigma of soy. *Journal of Food Science*, 68(8), 2604-2606.
- Wells, F. L. (1907). A statistical study of literary merit. *Archives of Psychology*, 1(7), 1-30.
- Wetzel, C. G., Wilson, T. D., & Kort, J. (1981). The halo effect revisited: Forewarned is not forearmed. *Journal of Experimental Social Psychology*, 17(4), 427-439.
- Wicklund, R. A., & Eckert-Nowack, M. (1989). The ascription of self-knowledge as a halo effect. *Basic and Applied Social Psychology*, 10(4), 355-370.
- Wojciszke, B., & Abele, A. E. (2008). The primacy of communion over agency and its reversals in evaluations. *European Journal of Social Psychology*, 38(7), 1139-1147.
- Wojciszke, B., Abele, A. E., & Baryla, W. (2009). Two dimensions of interpersonal attitudes: Liking depends on communion, respect depends on agency. *European Journal of Social Psychology*, 39(6), 973-990.
- Wolfradt, U. (2008). Implizite Persönlichkeitstheorien. In L.-E. Petersen & B. Six (Eds.), *Stereotype, Vorurteile und soziale Diskriminierung* (pp. 71-79). Weinheim: Beltz Verlag.
- Wyer, R. S. (1974). Changes in meaning and halo effects in personality impression formation. *Journal of Personality and Social Psychology*, 29(6), 829-835.
- Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don't all look alike: Individual impressions of other racial groups. *Journal of Personality and Social Psychology*, 65(1), 85-101.
- Zysberg, L., & Nevo, B. (2004). "The smarts that counts?": Psychologists' decision-making in personnel selection. *Journal of Business and Psychology*, 19(1), 117-124.

Abbildungsverzeichnis

Abbildung 1.	Kausale Modelle von Halo-Effekten aus der Literatur	32
Abbildung 2.	Postulierter Einfluss der Valenz auf die Qualität von Halo-Effekten.....	38
Abbildung 3.	Veranschaulichung verschiedener Experimentaldurchgänge	45
Abbildung 4.	Vorgehensweise bei der Überprüfung des grundlegenden Modells	47
Abbildung 5.	Schrittweise Berechnung der positiven abhängigen Variablen	50
Abbildung 6.	Beispiel für die Berechnung von Halo-Effekten als Verzerrung	52
Abbildung 7.	Veranschaulichung der quantitativen Valenzunterschiede	63
Abbildung 8.	Postulierter Einfluss von Bilddarbietungen auf Halo-Effekte.....	70
Abbildung 9.	Veranschaulichung der Beeinflussbarkeit verschiedener Attribute	92

Tabellenverzeichnis

Tabelle 1.	Beispiel für ein in den Experimenten verwendetes Itemset.....	44
Tabelle 2.	Übersicht über die durchgeführten Experimente	53
Tabelle 3.	Für die Voruntersuchung verwendete Persönlichkeitseigenschaften	55
Tabelle 4.	Einfluss der zusätzlichen Darbietung von Bildern auf Halo-Effekte.....	75
Tabelle 5.	Inter-Rating-Korrelationen in Abhängigkeit von Valenz und Dimension.....	87
Tabelle 6.	Beispielhafte Übersicht über die berechneten Einzelkorrelationen.....	89
Tabelle 7.	Beispiel für einen Durchgang in den beiden Experimentalbedingungen	109
Tabelle 8.	Übersicht über die verschiedenen Darbietungsbedingungen	116
Tabelle 9.	Einfluss von Berufsgruppe und Urteilsdimension auf Halo-Effekte.....	150

Anhang

Anhang A: Fragebogen zur Generierung gegensätzlicher Eigenschaften

Liebe Teilnehmer,

im Rahmen einer Voruntersuchung für meine Dissertation an der Universität Heidelberg führe ich eine kurze Befragung durch.



Dazu möchte ich Sie bitten, zu jedem der unten genannten Adjektive das Ihrer Meinung nach **passende Gegenstück** einzutragen. Versuchen Sie dabei möglichst, kein Wort zu verwenden, welches mit der Vorsilbe „un“ oder „in“ beginnt.

<u>Beispiel:</u>		
FALSCH:	höflich → <u>un</u> höflich	aktiv → <u>in</u> aktiv
RICHTIG:	höflich → abweisend	aktiv → faul

Vielen Dank für Ihre Unterstützung!

Michael Gräf

klug	_____
aktiv	_____
durchsetzungsfähig	_____
kreativ	_____
selbständig	_____
intelligent	_____
zuverlässig	_____
kompetent	_____
hartherzig	_____
rechthaberisch	_____
schüchtern	_____
fürsorglich	_____

chaotisch	_____
tatkräftig	_____
dominant	_____
leichtgläubig	_____
strebsam	_____
selbstbewusst	_____
herzlich	_____
rational	_____
fleißig	_____
sicher	_____
ehrgeizig	_____
sensibel	_____
konkurrenzbetont	_____
zielstrebig	_____
höflich	_____
gefühlbetont	_____
verletzlich	_____
gerecht	_____
freundlich	_____
hilfsbereit	_____
ehrlich	_____
verständnisvoll	_____
liebepoll	_____
treu	_____
einfühlsam	_____
selbstsicher	_____
harmoniebedürftig	_____
überheblich	_____
unterstützend	_____
eigennützig	_____
offen	_____

tolerant

gesellig

moralisch

eingebildet

faul

leistungsfähig

egoistisch

ausdauernd

stur

Anhang B: Fragebogen zur Gegenüberprüfung gegensätzlicher Eigenschaften

Liebe Teilnehmer,

im Rahmen einer Voruntersuchung für meine Dissertation an der Universität Heidelberg führe ich eine kurze Befragung durch.



Dazu möchte ich Sie bitten, zu jedem der unten genannten Adjektive das Ihrer Meinung nach **passende Gegenstück** einzutragen.

Vielen Dank für Ihre Unterstützung!

Michael Gräf

verschlossen	_____
weichherzig	_____
eigenbrötlerisch	_____
abhängig	_____
gleichgültig	_____
herzlos	_____
unterwürfig	_____
stark	_____
verständnislos	_____
dumm	_____
barmherzig	_____
ziellos	_____
einfalllos	_____
frech	_____
bescheiden	_____
passiv	_____

hart	_____
misstrauisch	_____
verlogen	_____
ideenlos	_____
parteiisch	_____
engstirnig	_____
träge	_____
selbstlos	_____
schwach	_____
zweifelnd	_____
nachgiebig	_____
phantasielos	_____
warmherzig	_____
kalt	_____
streitsüchtig	_____
treulos	_____
verträumt	_____
ordentlich	_____

Anhang C: Ergebnisse der Vorerhebung zur Eigenschaftsauswahl

Dimension	am häufigsten genannte Eigenschaft		nicht distinkte negative Eigenschaft***	Anteil der am häufigsten genannten Eigenschaft (in %)	
	positiv ^a	distinkt negativ ^b		positiv	distinkt negativ
Agency	klug**	dumm**	unklug	50.00	87.50
	aktiv	passiv	nicht aktiv	90.00	37.50
	kreativ	einfalllos	unkreativ	50.00	66.67
	sicher	zweifelnd	unsicher	25.00	20.83
	zielstrebig	ziellos	nicht zielstrebig	45.00	25.00
	fleißig*	faul*	nicht fleißig	79.17 ^c	95.83
	selbständig*	abhängig*	unselbständig	65.00	70.83
	selbstbewusst*	schüchtern*	nicht selbstbewusst	62.50 ^c	66.67
	ordentlich*	chaotisch*	unordentlich	54.17 ^c	95.00 ^d

Communion	herzlich	herzlos	nicht herzlich	30.00	25.00
	ehrllich*	verlogen*	unehrllich	85.00	75.00
	verständnisvoll	verständnislos	nicht verständnisvoll	85.00	33.33
	treu	treulos	untreu	80.00	29.17
	harmonie- bedürftig*	streitsüchtig*	nicht harmonie- bedürftig	60.00	70.83
	offen**	verschlossen**	nicht offen	75.00	62.50
	tolerant	engstirnig	intolerant	55.00	50.00
	gesellig*	eigen- brötlerisch*	ungesellig	70.00	54.17
	weichherzig	hartherzig	nicht weichherzig	29.17 ^c	70.00 ^d
	hilfsbereit*	egoistisch*	nicht hilfsbereit	66.67 ^c	79.17

Anmerkung. Diese Tabelle enthält nur die Eigenschaftspaare, bei denen die Eigenschaften beider Valenzen in mindestens 20 Prozent der Fälle aus ihrem jeweiligen Gegensatz generiert wurden. Nur die Eigenschaften mit den höchsten prozentualen Anteilen wurden für die Experimente 1 und 2 ausgewählt.

^a N = 20, sofern nicht anders verzeichnet

^b N = 24, sofern nicht anders verzeichnet

^c Datengrundlage ist N = 24 (Fragebogen 1)

^d Datengrundlage ist N = 20 (Fragebogen 2)

* Eigenschaft in Experiment 1 und 2 verwendet

** Eigenschaft nur in Experiment 1 verwendet

*** Valenzbedingung nur in Experiment 1 verwendet

Anhang D: Verwendete Bilder und deren Attraktivitätsratings**Bilder attraktiver Personen**

pos_männl_1



pos_männl_2



pos_männl_3



pos_weibl_1



pos_weibl_2



pos_weibl_3



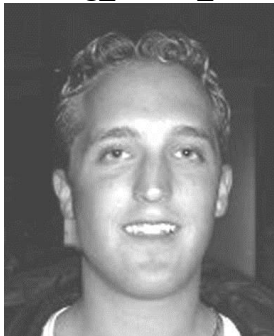
pos_weibl_4*



pos_weibl_5*

**Bilder unattraktiver Personen**

neg_männl_1



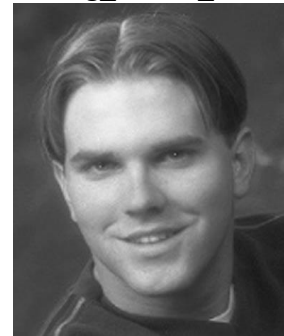
neg_männl_2



neg_männl_3



neg_männl_4*



neg_männl_5*



neg_weibl_1



neg_weibl_2



neg_weibl_3



	Attraktivitätsratings				
	Min	Max	M	SD	zM
Einzelbilder					
pos_männl_1	-93	100	49.81	45.75	1.75
pos_männl_2	-100	100	38.36	50.59	1.35
pos_männl_3	-100	100	28.87	45.19	1.01
pos_weibl_1	-100	100	35.43	45.76	1.24
pos_weibl_2	-100	100	33.86	45.81	1.19
pos_weibl_3	-100	100	32.94	56.79	1.16
pos_weibl_4*	-100	100	42.13	36.33	1.48
pos_weibl_5*	-82	100	28.81	42.31	1.01
neg_männl_1	-100	100	-38.97	53.45	-1.39
neg_männl_2	-100	80	-38.43	41.54	-1.37
neg_männl_3	-100	100	-37.58	48.28	-1.34
neg_männl_4*	-100	100	-34.70	56.77	-1.24
neg_männl_5*	-100	84	-32.12	47.41	-1.14
neg_weibl_1	-100	84	-39.82	47.66	-1.42
neg_weibl_2	-100	58	-36.93	36.99	-1.31
neg_weibl_3	-100	85	-28.25	43.67	-1.01
				M	zM
Zusammenfassung der Einzelbilder					
attraktive Männer (pos_männl_1 bis pos_männl_3)				39.01	1.37
attraktive Frauen (pos_weibl_1 bis pos_weibl_3)				34.08	1.20
attraktive Frauen* (pos_weibl_1 bis pos_weibl_5)				34.63	1.22
unattraktive Männer (neg_männl_1 bis neg_männl_3)				-38.33	-1.36
unattraktive Männer* (neg_männl_1 bis neg_männl_5)				-36.36	-1.29
unattraktive Frauen (neg_weibl_1 bis neg_weibl_3)				-35.00	-1.25

Anmerkung. Min = Minimum; Max = Maximum; M = Mittelwert; SD = Standardabweichung; zM = z-standardisierter Mittelwert; N = 171

* Bild nur in Experiment 3 und 4 verwendet

Anhang E: Ergebnisse der Erhebung zur Generierung von Verhaltensweisen

Eigenschaft		am häufigsten genannte Verhaltensweisen	
positiv	negativ	positiv*	negativ**
klug	un- klug / dumm	<p>... hat eine gute Wortwahl und drückt sich gewählt aus</p> <p>... argumentiert in Diskussionen sachlich und mit Hintergrundwissen</p> <p>... findet gute Lösungsansätze zu allen Problemen</p> <p>... kann anderen auch komplexe Sachverhalte gut erklären</p> <p>... weiß auf Fragen meistens eine gute Antwort</p> <p>... erzielt gute Ergebnisse im Studium und auf Arbeit</p>	<p>... hat eine schlechte Wortwahl und benutzt viel Umgangssprache</p> <p>... argumentiert in Diskussionen unsachlich und ohne Hintergrundwissen</p> <p>... geht Probleme trotz offensichtlicher Lösungen falsch an</p> <p>... kann anderen selbst einfache Sachverhalte nur schlecht erklären</p> <p>... weiß auf Fragen meistens keine Antwort</p> <p>... erzielt schlechte Ergebnisse im Studium und auf Arbeit</p>
aktiv	nicht aktiv / passiv	<p>... ist häufig unterwegs und treibt viel Sport</p> <p>... greift in das Geschehen um sich herum ein und beobachtet ungern nur</p> <p>... plant gern selbst Unternehmungen und lässt nicht alles auf sich zukommen</p>	<p>... ist selten unterwegs und treibt kaum Sport</p> <p>... greift nicht in das Geschehen um sich herum ein und beobachtet lieber nur</p> <p>... lässt lieber andere Unternehmungen planen und alles auf sich zukommen</p>
kreativ	unkrea- tiv / ein- fallslos	<p>... weiß sich immer gut selbst zu beschäftigen</p> <p>... probiert gern Neues aus und hält nicht an Altbewährtem fest</p> <p>... hat viele Ideen und Einfälle zu neuen Fragestellungen und Problematiken</p>	<p>... weiß oft nichts mit sich anzufangen</p> <p>... hält an Altbewährtem fest und probiert nichts Neues aus</p> <p>... hat keine Ideen und Einfälle zu neuen Fragestellungen und Problematiken</p>

sicher	un- sicher / zwei- felnd	... lässt sich nicht aus der Ruhe bringen, wenn etwas nicht so läuft wie geplant ... spricht flüssig und artikuliert ... fragt andere selten um Rat und orientiert sich nicht an deren Verhalten ... hat keine Probleme, Entscheidungen zu treffen und grübelt nicht lang	... lässt sich schnell aus der Ruhe bringen, wenn etwas nicht so läuft wie geplant ... spricht stockend und leise ... fragt andere oft um Rat und orientiert sich an deren Verhalten ... hat Probleme, Entscheidungen zu treffen und grübelt sehr lang
ziel- strebig	nicht ziel- stre- big / ziellos	... hat klare Pläne für die Zukunft und arbeitet konsequent dafür ... gibt trotz Problemen nicht auf und lässt sich nicht vom Weg abbringen	... hat keine Pläne für die Zukunft und bringt nichts Angefangenes zu Ende ... gibt bei Problemen schnell auf und lässt sich vom Weg abbringen
fleißig	nicht fleißig / faul	... wälzt unangenehme Pflichten nicht auf andere ab ... lernt und arbeitet viel und kontinuierlich ... erledigt anstehende Arbeit sofort und vor dem Vergnügen	... wälzt unangenehme Pflichten auf andere ab ... lernt und arbeitet wenig und nicht mehr als notwendig ... zögert anstehende Arbeit hinaus und tut stattdessen nichts
selb- ständig	unselb- stän- dig / ab- hängig	... führt den eigenen Haushalt sehr gut ohne Unterstützung ... setzt sich ohne Hilfe eigene Ziele und verfolgt diese ... bewältigt Hindernisse und wichtige Entscheidungen ohne fremde Hilfe	... führt den eigenen Haushalt ohne Unterstützung nur sehr schlecht ... holt sich immer Hilfe beim Setzen und Verfolgen von Zielen ... bewältigt Hindernisse und wichtige Entscheidungen nur mit fremder Hilfe
selbst- be- wusst	nicht selbst- be- wusst / schüch- tern	... hat keine Probleme, vor anderen aufzutreten und sich zu präsentieren ... hat eine aufrechte Körperhaltung ... beteiligt sich viel an Gesprächen und vertritt die eigene Meinung ... hält stets Blickkontakt zum Gegenüber und schaut nicht auf den Boden	... hat große Probleme, vor anderen aufzutreten und sich zu präsentieren ... hat eine gebeugte Körperhaltung ... beteiligt sich wenig an Gesprächen und hält die eigene Meinung zurück ... hält keinen Blickkontakt zum Gegenüber und schaut oft auf den Boden

ordentlich	unordentlich / chaotisch	<p>... putzt und räumt die eigene Wohnung regelmäßig auf</p> <p>... hält die eigenen Dinge gut sortiert und lässt nichts herumliegen</p> <p>... hat einen guten Überblick über die eigenen Termine</p> <p>... hat ein gepflegtes Äußeres</p>	<p>... putzt und räumt die eigene Wohnung fast nie auf</p> <p>... lässt alles herumliegen und hält die eigenen Dinge nicht sortiert</p> <p>... hat einen schlechten Überblick über die eigenen Termine</p> <p>... hat ein ungepflegtes Äußeres</p>
herzlich	nicht herzlich / herzlos	<p>... zeigt sich interessiert an dem, was andere zu sagen haben</p> <p>... zeigt viel Mitgefühl, wenn es anderen schlecht geht</p>	<p>... zeigt sich nicht interessiert an dem, was andere zu sagen haben</p> <p>... zeigt kein Mitgefühl, wenn es anderen schlecht geht</p>
ehrlich	unehrlich / verlogen	<p>... redet hinter dem Rücken nicht schlecht über andere</p> <p>... sagt anderen immer die Wahrheit</p> <p>... verstellt sich und täuscht andere nicht, um sich Vorteile zu verschaffen</p>	<p>... redet hinter dem Rücken schlecht über andere</p> <p>... sagt anderen selten die Wahrheit</p> <p>... verstellt sich und täuscht andere, um sich Vorteile zu verschaffen</p>
verständnisvoll	nicht verständnisvoll / verständnislos	<p>... verurteilt andere nicht für ihr Handeln und hört einfach nur zu</p> <p>... versucht sich in andere hineinzuversetzen und ihren Standpunkt nachzuvollziehen</p> <p>... geht auf die Gefühle und Bedürfnisse von anderen ein</p>	<p>... verurteilt andere schnell für ihr Handeln und hört nicht einfach nur zu</p> <p>... will sich nicht in andere hineinversetzen und ihren Standpunkt nachvollziehen</p> <p>... geht nicht auf die Gefühle und Bedürfnisse von anderen ein</p>
treu	untreu / treulos	<p>... geht dem Partner nicht fremd</p> <p>... steht anderen gegenüber immer zu Partner, Familie und Freunden</p> <p>... lässt andere nicht hängen, wenn sie einen brauchen</p>	<p>... geht dem Partner fremd</p> <p>... steht anderen gegenüber nicht zu Partner, Familie und Freunden</p> <p>... lässt andere hängen, wenn sie einen brauchen</p>

harmonebedürftig	nicht harmonebedürftig	... versucht Konflikte zu vermeiden	... fängt oft und auch wegen kleiner Dinge Konflikte an
	harmonebedürftig / streitsüchtig	... versucht Auseinandersetzungen schnell zu beenden und nimmt sich bewusst zurück	... fängt Auseinandersetzungen immer wieder neu an und will das letzte Wort haben
		... vermeidet es, andere zu kritisieren und zu provozieren	... provoziert und kritisiert andere oft bewusst
offen	nicht offen / verschlossen	... erzählt anderen auch Privates und gibt viel von sich preis	... erzählt anderen nichts Privates und gibt wenig von sich preis
		... geht von sich aus auf andere Menschen zu	... geht von sich aus nicht auf andere Menschen zu
tolerant	intolerant / engstirnig	... gibt auch zu, im Unrecht gewesen zu sein und beharrt nicht auf der eigenen Meinung	... gibt nicht zu, im Unrecht zu sein und beharrt auf der eigenen Meinung
		... akzeptiert andere Sichtweisen und Weltanschauungen	... akzeptiert keine anderen Sichtweisen und Weltanschauungen
		... äußert keine Vorurteile gegenüber Minderheiten	... äußert Vorurteile gegenüber Minderheiten
gesellig	ungesellig / eigenbrötlerisch	... umgibt sich gern mit anderen Menschen und ist ungern allein	... umgibt sich ungern mit anderen Menschen und ist am liebsten für sich allein
		... hält sich in Gesellschaft mitten im Getümmel auf	... hält sich in Gesellschaft abseits von allen anderen
		... nimmt Einladungen gern an und besucht häufig Gruppenveranstaltungen	... nimmt Einladungen nicht an und besucht keine Gruppenveranstaltungen
weichherzig	nicht weichherzig / hartherzig	... zeigt viel Gefühle und Emotionen	... zeigt keine Gefühle und Emotionen
		... gibt den Bitten anderer nach	... gibt den Bitten anderen nicht nach

hilfs- bereit	nicht hilfs- bereit / ego- istisch	... bietet Unterstützung, wenn sich andere in einer akuten Notlage befinden	... sieht weg, wenn sich andere in einer akuten Notlage befinden
		... setzt sich für andere Menschen ein, wenn diese schlecht behandelt werden	... setzt sich nicht für andere Menschen ein, wenn diese schlecht behandelt werden
		... bringt auch Opfer für andere und kümmert sich nicht nur um sich selbst	... kümmert sich nur um sich selbst und bringt keine Opfer für andere
		... stellt die Bedürfnisse anderer vor die eigenen Interessen	... stellt die eigenen Interessen vor die Bedürfnisse anderer

* N = 20

** Zusammenfassung der aus negativen (N = 20) und distinkt negativen (N = 20) Eigenschaften generierten Verhaltensweisen

Anhang F: Ergebnisse der Gegenüberprüfung der Verhaltensweisen

Verhaltensweise		am häufigsten genannte Eigenschaft		Anteil der am häufigsten genannten Eigenschaft (in %)	
		positiv	negativ	positiv ^a	negativ ^b
Agency					
... hält die eigenen Dinge gut sortiert und lässt nichts herumliegen	... lässt alles herumliegen und hält die eigenen Dinge nicht sortiert	ordentlich	chaotisch (unordentlich)	94.12	93.75
... putzt und räumt die eigene Wohnung regelmäßig auf	... putzt und räumt die eigene Wohnung fast nie auf			88.24	59.38
... hat einen guten Überblick über die eigenen Termine	... hat einen schlechten Überblick über die eigenen Termine			52.94	62.50
... bewältigt Hindernisse und wichtige Entscheidungen ohne fremde Hilfe*	... bewältigt Hindernisse und wichtige Entscheidungen nur mit fremder Hilfe*	selbständig (unabhängig)	abhängig (unselbständig)	91.18	75.00
... setzt sich ohne Hilfe eigene Ziele und verfolgt diese*	... holt sich immer Hilfe beim Setzen und Verfolgen von Zielen*			50.00	75.00
... fragt andere selten um Rat und orientiert sich nicht an deren Verhalten*	... fragt andere oft um Rat und orientiert sich an deren Verhalten*			58.82	56.25
... lernt und arbeitet viel und kontinuierlich	... lernt und arbeitet wenig und nicht mehr als notwendig	fleißig	faul	76.47	68.75
... erledigt anstehende Arbeit sofort und vor dem Vergnügen	... zögert anstehende Arbeit hinaus und tut stattdessen nichts			50.00	68.75
... erzielt gute Ergebnisse im Studium und auf Arbeit	... erzielt schlechte Ergebnisse im Studium und auf Arbeit			52.94	53.13

... hat keine Probleme, vor anderen aufzutreten und sich zu präsentieren	... hat große Probleme, vor anderen aufzutreten und sich zu präsentieren	selbst- be- wusst (sicher)	schüch- tern (un- sicher)	85.29	84.38
... hält stets Blickkontakt zum Gegenüber und schaut nicht auf den Boden	... hält keinen Blickkontakt zum Gegenüber und schaut oft auf den Boden			67.65	81.25
... hat eine aufrechte Körperhaltung	... hat eine gebeugte Körperhaltung			82.35	53.13
... weiß auf Fragen meistens eine gute Antwort**	... weiß auf Fragen meistens keine Antwort**	klug	dumm	64.71	50.00
... hat viele Ideen und Einfälle zu neuen Fragestellungen und Problematiken**	... hat keine Ideen und Einfälle zu neuen Fragestellungen und Problematiken**	kreativ	un- kreativ	73.53	65.63
Communion					
... sagt anderen immer die Wahrheit	... sagt anderen selten die Wahrheit	ehrlich	ver- logen (unehr- lich)	91.18	75.00
... verstellt sich und täuscht andere nicht, um sich Vorteile zu verschaffen	... verstellt sich und täuscht andere, um sich Vorteile zu verschaffen			76.47	53.13
... redet hinter dem Rücken nicht schlecht über andere	... redet hinter dem Rücken schlecht über andere			64.71	50.00
... umgibt sich gern mit anderen Menschen und ist ungern allein*	... umgibt sich ungern mit anderen Menschen und ist am liebsten für sich allein*	gesellig	eigen- bröt- lerisch (unge- sellig)	94.12	71.88
... nimmt Einladungen gern an und besucht häufig Gruppenveranstaltungen*	... nimmt Einladungen nicht an und besucht keine Gruppenveranstaltungen*			91.18	59.38
... hält sich in Gesellschaft mitten im Getümmel auf*	... hält sich in Gesellschaft abseits von allen anderen*			79.41	50.00

... bringt auch Opfer für andere und kümmert sich nicht nur um sich selbst	... kümmert sich nur um sich selbst und bringt keine Opfer für andere	hilfs- bereit	ego- istisch	79.41	81.25
... bietet Unterstützung, wenn sich andere in einer akuten Notlage befinden	... sieht weg, wenn sich andere in einer akuten Notlage befinden			97.06	53.13
... stellt die Bedürfnisse anderer vor die eigenen Interessen	... stellt die eigenen Interessen vor die Bedürfnisse anderer			52.94	68.75
... versucht Konflikte zu vermeiden	... fängt oft und auch wegen kleiner Dinge Konflikte an	harmo- niebe- dürftig	streit- süchtig	67.65	90.63
... vermeidet es andere zu kritisieren und zu provozieren	... provoziert und kritisiert andere oft bewusst			58.82	62.50
... versucht Auseinandersetzungen schnell zu beenden und nimmt sich bewusst zurück	... fängt Auseinandersetzungen immer wieder neu an und will das letzte Wort haben			55.88	59.38
... erzählt anderen auch Privates und gibt viel von sich preis**	... erzählt anderen nichts Privates und gibt wenig von sich preis**	offen	ver- schlos- sen	61.76	50.00
... akzeptiert andere Sichtweisen und Weltanschauungen**	... akzeptiert keine anderen Sichtweisen und Weltanschauungen**	to- lerant	eng- stirnig (into- lerant)	50.00	53.13

Anmerkung. Diese Tabelle enthält nur die Verhaltensweisen, bei denen in beiden Valenzbedingungen in mindestens 50 Prozent der Fälle die ursprüngliche Eigenschaft generiert wurde. Nur diese Verhaltensweisen wurden in den Experimenten verwendet. Für die abhängigen Skalen wurden die negativen Eigenschaften in Klammern außer Acht gelassen.

^a N = 34

^b N = 32

* Item nicht in Experiment 8 verwendet

** Item nur in Experiment 6 und 9 verwendet

Anhang I: Verwendete Berufsbezeichnungen und deren Häufigkeiten

naturwissenschaftliche und technische Berufe		soziale und Pflegeberufe	
in der Vorerhebung genannter Beruf	Anteil (in %)	in der Vorerhebung genannter Beruf	Anteil (in %)
Physiker/in	75.00	Sozialarbeiter/in	80.00
Chemiker/in	65.00	Pfleger/in*	60.00
Biologe/in	60.00	Altenpfleger/in	
Techniker/in*	60.00	Krankenpfleger/in	
Elektrotechniker/in		Kinderpfleger/in	
Maschinenbautechniker/in		Berater/in*	55.00
Mathematiker/in	60.00	Gesundheitsberater/in	
Laborant/in*	55.00	Drogenberater/in	
Physiklaborant/in		Familienberater/in	
Chemielaborant/in		Lehrer/in	40.00
Biologielaborant/in		Kindergärtner/in	35.00
Umweltwissenschaftler/in	45.00	Erzieher/in*	35.00
Ingenieur/in*	35.00	Heimerzieher/in	
Bauingenieur/in		Jugenderzieher/in	
Vermessungsingenieur/in		Psychologe/in	30.00
Architekt/in*	30.00	Bewährungshelfer/in	25.00
Landschaftsarchitekt/in		Pädagoge/in*	25.00
Innenarchitekt/in		Sozialpädagoge/in	
Informatiker/in	30.00	Heilpädagoge/in	
Geologe/in	25.00	Streetworker/in	25.00
Mechaniker/in**	20.00		
Arzt/Ärztin**	10.00		

Anmerkung. Die Häufigkeiten der spezifischen Berufsbezeichnungen sind nicht separat aufgeführt, sondern in die Häufigkeiten des jeweiligen Oberbegriffs integriert.

Mehrfachnennungen möglich; N = 20 pro Bedingung

* Oberbegriff nicht in den Experimenten verwendet

** Beruf nicht in den Experimenten verwendet

Selbständigkeitserklärung

Entsprechend § 8 Abs. 1 Buchst. b) der Promotionsordnung der Fakultät für Verhaltens- und Empirische Kulturwissenschaften an der Ruprecht-Karls-Universität Heidelberg erkläre ich hiermit, dass ich die vorgelegte Dissertation selbständig angefertigt, nur die angegebenen Hilfsmittel benutzt und alle wörtlich oder inhaltlich übernommenen Zitate als solche gekennzeichnet habe.

Des Weiteren erkläre ich gemäß § 8 Abs. 1 Buchst. c) der Promotionsordnung der Fakultät der Verhaltens- und Empirische Kulturwissenschaften an der Ruprecht-Karls-Universität Heidelberg ehrenwörtlich, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.

Heidelberg, 31. Oktober 2012

Michael Gräf