

**Gedächtnisgestützte Urteilsbildung –
der Einfluss der Stimulusinformation und der
Beurteilererfahrung**

**Dissertation zur Erlangung des akademischen Grades
doctor philosophiae (Dr. phil.)**

vorgelegt dem Rat der Fakultät für Sozial- und Verhaltenswissenschaften
der Friedrich-Schiller-Universität Jena
von Dipl.-Psych. Thomas M. Schönburg
geboren am 11.01.1975 in Halle/Saale

Gutachter

1. Frau Prof. Dr. B. Edeler

2. Herr Prof. Dr. W. Frindte

Tag des Kolloquiums: 04.10.2006

Meiner lieben Frau

Danksagung

An dieser Stelle möchte ich allen, die mir bei der Durchführung der wissenschaftlichen Studien und der Fertigstellung dieser Arbeit geholfen haben, danken.

Mein ganz besonderer Dank gilt Frau Prof. Dr. Brigitte Edeler, die es mir ermöglichte, diese Dissertation zu schreiben und die mir stets mit wissenschaftlichem Rat zur Seite stand. Prof. Dr. Frindte danke ich herzlich für seine Bereitschaft, sich als Zweitgutachter zur Verfügung zu stellen.

Weiterhin möchte ich den Versuchspersonen, insbesondere den Experten dafür danken, dass sie einen Teil ihrer kostbaren Freizeit für die Teilnahme an einer wissenschaftlichen Studie opferten. Ohne ihre Bereitschaft, an den Experimenten teilzunehmen, wäre die Durchführung dieses Projektes nicht möglich gewesen.

Inhaltsverzeichnis

	Seite
Theoretischer Teil	
1. Einleitung – Allgemeine Zielstellung, Abgrenzung	1
2. Der Begriff der gedächtnisgestützten Urteilsbildung	4
2.1. Das Gedächtnis	6
2.2. Das Mehrspeichermodell von Pashler & Carrier (1996)	6
2.2.1. Sensorisches Gedächtnis	7
2.2.2. Das Kurzzeitgedächtnis	8
2.2.3. Episodisches versus semantisches Gedächtnis	9
2.3. Die multimodale Gedächtnistheorie von Engelkamp (1991)	11
2.4. Der Zeitpunkt der Urteilsbildung im Gedächtnis	14
3. Modelle und Mechanismen der Urteilsbildung	16
3.1. Das Zweiprozessmodell der Eindrucksbildung von Brewer (1988)	16
3.2. Der Prozess der Eindrucksbildung nach Brewer (1988)	18
3.3. Das Kontinuummodell der Eindrucksbildung von Fiske und Neberg (1990)	20
3.4. Der Prozess der Eindrucksbildung nach Fiske und Neberg (1990)	22
3.5. Reduktion der Komplexität bei der Urteilsbildung	24
3.6. Vergleichsprozesse in der Urteilsbildung	26
4. Einfluss der Stimulusinformation auf die Urteilsbildung	27
4.1. Stimulusinformation: Text	28
4.1.1. Personenbeschreibende Information als Textverstehen	28
4.1.2. Situationsmodelle beim textverstehenden Lesen	30
4.1.3. Urteilsbildung über Protagonisten aus schriftlichen Informationen	34
4.2. Stimulusinformation: Text mit Bild	35
4.2.1. Gesicht und Persönlichkeitseindruck – Erste Forschungsansätze	35
4.2.2. Die Rolle des Gesichtes bei bildhaften Darstellungen	37
4.2.3. Das Fünf-Faktoren-Modell der Persönlichkeit – die Big Five	40
4.2.4. Das Fünf-Faktoren-Modell bei der Beurteilung von Gesichtern	42

	Seite
4.2.5. Physiognomie und Persönlichkeitseindruck	43
4.2.6. Geschwindigkeit der visuellen Eindrucksbildung	44
4.2.7. Folgerungen für die Beurteilung von Protagonisten eines Textes bei zusätzlicher Vorlage eines Bildes	45
4.2.8. Die Beurteilung von Personen aufgrund verschiedener Stimulusmodi	46
4.3. Stimulusinformation: Film	48
4.3.1. Bewegte Bilder	48
4.3.2. Annahmen für die Beurteilung von Personen anhand eines Filmausschnittes	52
5. Einfluss der Beurteilererfahrung	54
5.1. Zur Abgrenzung von Experten und Novizen	54
5.2. Die Messung des Expertisegrades	57
5.3. Anfänger und Experten im Vergleich	58
5.3.1. Die Rolle bereichsspezifischen Wissens	58
5.3.2. Organisation des Wissens	61
5.3.3. Kognitive Flexibilität	62
5.4. Ausgewählte empirische Befunde in der Urteilsbildung bei Probanden mit unterschiedlicher Erfahrung	65
5.4.1. Klinische Urteilsbildung	65
5.4.2. Urteilsbildung von Pädagogen	70
5.4.3. Urteilsbildung in anderen Domänen	72
5.5. Zusammenfassung und Schlussfolgerungen für die eigene Arbeit	74
5.6. Training als Möglichkeit der Veränderung der Personenbeurteilung	77
Experimenteller Teil	80
6. Überblick	80
6.1. Anliegen der Untersuchung	80
6.2. Herleitung der Hypothesen hinsichtlich der Variation der Stimulusinformation	81
6.3. Herleitung der Hypothesen hinsichtlich der Variation der Beurteilererfahrung	83
6.4. Operationalisierung	85

	Seite
7. Hypothesen	86
7.1. Hypothesen hinsichtlich der Variation der Stimulusinformation	86
7.2. Hypothesen hinsichtlich der Variation der Beurteilererfahrung	87
8. Methode	88
8.1. Versuchspersonen	88
8.2. Training der Probanden in der Personenbeurteilung	90
8.3. Untersuchungsdesign	92
8.4. Stimulusmaterial	94
8.5. Begründung der Auswahl des Stimulusmaterials	95
9. Durchführung	99
9.1. Voruntersuchungen	99
9.2. Stimulusinformation: Text	101
9.3. Stimulusinformation: Text mit Bild	105
9.4. Stimulusinformation: Film	105
9.5. Aufbereitung der Protokolle der Gedächtnisabfrage und zur freien Beurteilung der Geschworenen	106
10. Ergebnisse	108
10.1. Ergebnisse hinsichtlich der Variation des Stimulusmaterials	108
10.1.1. Ergebnisse zur Reproduktion von Gedächtnisinhalten	108
10.1.2. Ergebnisse zur Beurteilung der Geschworenen	110
10.1.3. Ergebnisse zur subjektiven Urteilssicherheit	113
10.1.4. Hypothesenbezug und Zusammenfassung der Ergebnisse hinsichtlich der Variation der Stimulusinformation	113
10.2. Ergebnisse hinsichtlich der Variation der Beurteilererfahrung	116
10.2.1. Ergebnisse zur Reproduktion von Gedächtnisinhalten	116
10.2.2. Ergebnisse zur Beurteilung der Geschworenen	117
10.2.3. Ergebnisse hinsichtlich der Urteilssicherheit	119
10.2.4. Hypothesenbezug und Zusammenfassung der Ergebnisse hinsichtlich der Variation der Beurteilererfahrung	120

	Seite
11. Diskussion der Ergebnisse	122
11.1. Diskussion hinsichtlich der Variation der Stimulusinformation	122
11.2. Diskussion hinsichtlich der Variation der Beurteilererfahrung	129
12. Zusammenfassung und Ausblick	134
Literaturangaben	137
Anhang	157
Ergebnisse der Voruntersuchungen	158
Vorinformation	160
Instruktionen	161
Ergebnisse Reproduktion der Gedächtnisinhalte	167
Ergebnisse Beurteilung der Geschworenen	188

Theoretischer Teil

1. Einleitung

“on ne voit bien qu'avec le coeur. L'essentiel est invisible pour les yeux.”

Antoine de Saint-Exupéry (1943)

“Man sieht nur mit dem Herzen gut. Das Wesentliche ist für die Augen unsichtbar.“

Antoine de Saint-Exupéry (1943),
ins Deutsche übertragen von Grete und Josef
Leitgeb, S. 54

In dem Buch „Der kleine Prinz“ von de Saint-Exupéry (1943), dem das o.g. Zitat entstammt, macht der Erzähler die Bekanntschaft des kleinen Prinzen. Der kleine Prinz berichtet ihm von seiner Reise, bei der er auf der Erde unter anderem einen Fuchs trifft. Der Fuchs wird der Freund des kleinen Prinzen und zum Abschied schenkt ihm der Fuchs ein wichtiges Geheimnis: „Man sieht nur mit dem Herzen gut. Das Wesentliche ist für die Augen unsichtbar.“ (de Saint-Exupéry, 1943, S. 54).

Das wunderbare Geheimnis des Fuchses, welches er dem kleinen Prinzen schenkt, ist eine Erkenntnis, die uns anrührt. Gerade wenn wir an einen Menschen denken, den wir lieben, scheint das oben genannte Zitat von de Saint-Exupéry in besonderem Maße zuzutreffen. Nicht nur das Äußere ist wichtig, nein, das Gefühl, welches wir für diesen Menschen empfinden, bestimmt unsere Einschätzung von ihm und die Einschätzung der Beziehung zu diesem Menschen.

Begegnen wir einem Menschen, den wir noch nicht kennen, in einem anderen Kontext, beispielsweise, wenn ein potentieller Kunde ein Geschäft betritt, ist es weit schwieriger, „mit dem Herzen“ zu sehen. Möglicherweise werden wir, obwohl wir einen Kauf abschließen und mit ihm kommunizieren, diesen Menschen nie näher kennen lernen und nie erfahren, welche Bücher er gern liest, ob er gern spazieren geht, ob er tierlieb ist oder Vanilleeiskrem mag. Gleichwohl werden wir ihn, wenn das Geschäft betritt, einschätzen und überlegen, wie wir uns ihm gegenüber in dieser Situation am besten verhalten, damit er etwas kauft. Wir nehmen

also eine Beurteilung einer Person vor, ohne viele Informationen über diese Person zu haben. Anhand welcher Informationen nehmen wir eine solche Einschätzung vor? Wie beeinflussen verschiedene Informationen unsere Beurteilung von Personen? Und wie beeinflusst eine professionelle Ausbildung von Beurteilern deren Meinungsbildung?

Diese Fragen werden in der vorliegenden Arbeit aufgegriffen. Die Beurteilung von Personen ist im Alltag von zentraler Bedeutung. Zu Hause, im beruflichen Alltag, beim Einkauf, in der Freizeit – überall begegnen wir Menschen, mit denen wir in Kontakt treten, die uns beurteilen und über die wir eine Einschätzung treffen. Oftmals wird uns noch nicht einmal bewusst, dass wir Bewertungen von anderen Personen vorgenommen haben. Oberflächlich betrachtet, erscheinen diese Situationen von geringer Wichtigkeit, aber gerade im beruflichen Kontext kommt Beurteilungen zuweilen eine große Bedeutung zu. Beispielsweise ist es wichtig, für eine neu zu besetzende Stelle einen geeigneten Bewerber zu finden, weil personelle Fehlentscheidungen den Unternehmenserfolg empfindlich schmälern können (Obermann, 2002; Püschel, 2005). Polizisten müssen einschätzen, ob ein an ihnen vorbei rennender Passant möglicherweise ein Straftäter sein könnte, ein Richter muss darüber urteilen, ob eine Person eine Straftat begangen hat und welches Strafmaß angemessen ist. Es gibt zahlreiche weitere Beispiele, die belegen, dass die Urteilsbildung und die Erforschung dieses Prozesses von großer Relevanz sind. Auf verschiedenen Forschungsfeldern werden große Anstrengungen unternommen, einzelne Prozesse bei der Urteilsbildung und zugrunde liegende Mechanismen zu beschreiben und zu erklären. Zahlreiche Untersuchungen zum Thema Personenbeurteilung finden außerhalb der Wissenschaft kaum Beachtung und interessante Ergebnisse werden in der Praxis oft nicht umgesetzt, weil der Bezug zwischen Wissenschaft und Praxis oft nur sehr schwer hergestellt werden kann (Kanning, 1999), weil wissenschaftliche Fragestellungen und Praxisinteressen auseinander gehen und unterschiedliche Schwerpunktsetzungen des Praxis- und des Wissenschaftszuganges einander näher gebracht werden müssen (Höft, 2005) oder weil Ergebnisse aufgrund des gewählten Versuchsmaterials nur eine mangelnde ökologische Validität haben (z.B. Spiro, 1980, Krolak-Schwerdt, 2000).

Zentrales Anliegen der vorliegenden Arbeit ist es, Ergebnisse mit einer möglichst hohen ökologischen Validität zu erhalten und in der Diskussion mögliche Ansatzpunkte für die praktische Nutzung dieser Ergebnisse zu geben. Thematisch geht es zum einen darum, wie unterschiedliche Stimulusinformationen die Beurteilung von Personen beeinflussen und zum

anderen darum, wie sich eine unterschiedliche Beurteilererfahrung auf die Beurteilung von Personen auswirkt.

Im theoretischen Teil der Arbeit wird zuerst der Begriff der gedächtnisgestützten Urteilsbildung in sozialen Situationen abgegrenzt und es werden Gedächtnisprozesse bei der Urteilsbildung beschrieben. Im Abschnitt Modelle und Mechanismen der Urteilsbildung werden anerkannte Urteilsbildungsmodelle vorgestellt und danach die Reduktion der Komplexität von Informationen und ablaufende Vergleichsprozesse bei der Beurteilung von Personen als zwei Kernmechanismen erläutert.

Für die Variation des Stimulusmaterials wurde Material ausgewählt, an dem der Einfluss der unterschiedlichen Stimuli gut deutlich werden sollte. Es wurde zum einen ein Text, der das Gespräch von mehreren Personen wiedergibt ausgewählt, zum Zweiten wurde diesem Text ein Foto der am Gespräch beteiligten Personen beigelegt und zum Dritten wurde diese Gesprächsszene auf Video als Film dargestellt. Die Wirkung der einzelnen Stimuli und der derzeitige Erkenntnisstand auf diesen Gebieten wird im Abschnitt Einfluss der Stimulusinformation auf die Urteilsbildung ausführlich beschrieben. So werden die Prozesse des Aufbaus eines mentalen Modells nach Darbietung eines Textes, der Personinformationen enthält, die Eindruckswirkung von Gesichtern bei bildhaften Darstellungen und die Erkenntnisse zur Verarbeitung und Bewertung bewegter Bilder geschildert.

Zur Untersuchung der Frage, wie sich die unterschiedliche Erfahrung von Beurteilern auf die Beurteilung von Personen unterscheidet, werden im Theorieteil der vorliegenden Arbeit im Abschnitt Einfluss der Beurteilererfahrung zunächst Abgrenzungskriterien zwischen Experten und Novizen erläutert und empirische Befunde in der Urteilsbildung bei Probanden mit unterschiedlicher Erfahrung vorgestellt. Ein weiterer Teil dieses Abschnittes befasst sich mit dem Training in der Personenbeurteilung als Möglichkeit, die Urteilsbildung von Personen zu verändern und damit zu objektiveren Ergebnissen in der Beurteilung anderer Menschen zu gelangen.

Im experimentellen Teil wird untersucht, ob und gegebenenfalls wie sich der Eindruck einer Person unterscheidet, wenn der Beurteiler zusätzlich zu einem Text ein Foto erhält und ob sich der Persönlichkeitseindruck nochmals unterscheidet, wenn die Darstellung des Textes als Filmszene erfolgt.

Die empirische Umsetzung zur Beantwortung der Frage nach dem Einfluss der Beurteilererfahrung erfolgte zum einen durch die Auswahl von Kontrastgruppen, zum anderen durch ein aufwändiges, umfassendes Training von Probanden in der Personenbeurteilung vor der Durchführung des Experimentes. Die Kontrastgruppen bestanden zum einen aus Probanden, die keine Ausbildung in der Personenbeurteilung haben und die sich im beruflichen Kontext nicht mit Personenbeurteilung beschäftigen und zum anderen aus Probanden, die eine Ausbildung mit Inhalten zur Personenbeurteilung absolviert haben und die im Beruf täglich eine Vielzahl von professionellen Personenbeurteilungen vornehmen müssen. Es wird untersucht, wie sich ein Training und eine professionelle Ausbildung und langjährige Berufserfahrung im Vergleich zu Novizen auf die Beurteilung von Personen auswirken.

In der abschließenden Diskussion der Ergebnisse geht es vor allem um die Einordnung der Ergebnisse in die bisherige Forschungslandschaft und um die ökologische Validität und damit um die praktische Relevanz der Ergebnisse im Alltag der Menschen, die sich professionell mit Personenbeurteilung beschäftigen.

2. Der Begriff der gedächtnisgestützten Urteilsbildung

Die Beurteilung von Personen erfolgt zumeist in komplexen sozialen Situationen, an denen Personen direkt oder indirekt beteiligt sind. Der Begriff „sozial“ kennzeichnet in diesem Zusammenhang, dass es sich um von Menschen geschaffene Situationen handelt (z.B. Börse) oder zumindest um Situationen, die durch Verhaltensregeln, Werte und Konventionen gestaltet wurden (z.B. Straßenverkehr) (Kanning, 1999). Jeder Mensch ist, wenn er sich in einer solchen Situation befindet, darauf angewiesen, die Komplexität der Situation zu reduzieren und die Situation und die darin handelnden Personen fortwährend zu beurteilen. Der Prozess der sozialen Wahrnehmung – das Einschätzen anderer Personen – (Greenberg & Baron, 2000, Kenny, 1994) verlangt vom Wahrnehmer, dass er Informationen über die Zielperson kombiniert, integriert und interpretiert (Werth, 2004). Der erste Eindruck basiert vor allem auf Informationen, die leicht nach außen hin sichtbar sind, wie z.B. die Kleidung, Körpergröße, Haarfarbe, Brille etc. (z.B. Alley, 1988, Young et al., 1993, Zebrowitz, 1997). Gerade die Bildung des ersten Eindrucks von einer Person läuft weitgehend automatisch ab (Smith & Mackie, 2000) und häufig ist dem Urteiler nicht einmal bewusst, welche Informationen und Schlussfolgerungen zu dieser ersten Einschätzung geführt haben. Aronson

et al. (2002) sprechen von „quick attributional snapshots we form when we first meet someone“ (S. 133), die häufig nur in geringem Umfang mit den tatsächlichen Eigenschaften der Zielperson übereinstimmen (Funder & Colvin, 1988). Wie sich der Begriff Urteilsbildung definieren lässt und welche Besonderheiten die gedächtnisgestützte Urteilsbildung aufweist, soll im Folgenden erörtert werden.

Edeler (1992) definiert „soziale Urteilsbildung“ als einen Denkprozess, „der mit einer Stellungnahme über eine Person, meist in Form einer Aussage, abschließt, die die Bewertung dieser Person, d.h. ihre „Gewichtung“ hinsichtlich des Grades ihrer Eignung zur Bewältigung definierter Anforderungen oder einfach ihre Einordnung in eine Klassifikation, ausdrückt und beeinflusst ist von der jeweiligen Aufgabe und ihrem Ziel, von den Handlungsintentionen, Bedürfnissen, Erfahrungen, Motiven, Emotionen, Erwartungen usw. des Urteilers sowie von den Werten und Normen des gesellschaftlichen Kontextes.“ (S. 11f.). Soziale Urteile oder Personenurteile versteht sie als Ergebnisse eines Denkprozesses, der die Beziehung des Urteilers zum Beurteilten darstellt, wobei auch gesellschaftliche Normen mit in diesen Denkprozess einfließen (Edeler, 1992). Im Unterschied zur Beurteilung von diskreten Objekten, geht es also in der sozialen Urteilsbildung beim Beurteilungsobjekt „Person“ um eine Einschätzung des Ausmaßes, in dem eine bestimmte Eigenschaft auf die Zielperson zutrifft (z.B. Snyder, & Swann, 1978; Snyder & Uranowitz, 1978). Walther (1997) spricht im Zusammenhang von Urteilen über Personen von einem dimensionalen „Mehr-oder-Weniger“ (S.9). In Objektbeurteilungen gibt es meist ein experimentell hergestelltes „objektives“ Wahrheitskriterium, während Personenmerkmale nicht objektiv zu beobachten sind. In der Regel sind die Aussagen, Verhaltensweisen und Fragen mehrdeutig (Fiedler et. al., 1996) und müssen erst interpretiert werden. Das heißt, der Beurteiler muss die gegebenen Daten auf ihre Bedeutung untersuchen (Srull & Wyer, 1989) und muss von ihnen unter Unsicherheit auf Persönlichkeitsmerkmale schließen, da eine deterministische, fehlerfreie Regel nicht bekannt ist (Fiedler & Plessner, 2004). Klassifikationen, Bewertungen und Voraussagen werden auch von Hiebsch (1986) als grundlegende Mechanismen der Informationsverarbeitung bezeichnet. Dabei ist die Gedächtnisfunktion die „Grundfunktion“, die in allen diesen Funktionen notwendigerweise enthalten ist (Hiebsch, 1986, Edeler, 1992) und die demzufolge auch bei der Urteilsbildung eine zentrale Rolle spielt.

2.1. Das Gedächtnis

Das Gebiet der Gedächtnisforschung ist sehr komplex, wie die große Anzahl an Theorien und Modellen widerspiegelt. Derzeit gibt es keinen umfassenden theoretischen Rahmen für Gedächtnisprozesse, der möglichst viele der empirischen Daten erklären kann (Perrig, 1988; Krause, 2000). In diesem Zusammenhang spricht Bower (1998) von einer Kapazitäten sprengenden Informationsflut, die eine Ordnung und ein Verständnis der Erkenntnisse erschwert. Aufgrund der enormen Vielzahl an Gedächtnistheorien wird an dieser Stelle ausdrücklich darauf verzichtet, einen weitreichenden Überblick zu geben. Vielmehr sollen aus der Menge der Theorien und Modelle für die vorliegende Arbeit relevante Aspekte der Gedächtnisforschung dargestellt werden.

2.2. Das Mehrspeichermodell von Pashler & Carrier (1996)

Das Modell von Pashler & Carrier (1996) kann als eine Weiterentwicklung des klassischen Mehrspeichermodells von Atkinson & Shiffrin (1968) betrachtet werden. Atkinson & Shiffrin (1968) postulieren in ihrem Modell drei Speichersysteme für das menschliche Gedächtnis – das sensorische Gedächtnis, das Kurzzeitgedächtnis und das Langzeitgedächtnis. In der modernen Konzeption des Mehrspeichermodells von Pashler & Carrier (1996), welches hier kurz erläutert werden soll, lassen sich viele Ergebnisse neuerer Studien integrieren. Nach Pashler & Carrier (1996) fließt nach der Präsentation eines Stimulus die Information nicht automatisch von den sensorischen Gedächtnissystemen zum Kurzzeitgedächtnis. Das sensorische Gedächtnis behält den Input des Stimulus für den Fall, dass der Stimulus verloren geht. Den Einlass der sensorischen Information regelt dabei die selektive Aufmerksamkeit. Bei einer Weiterverarbeitung des Stimulus aktiviert der Stimulus gleichzeitig Repräsentationen im semantischen Gedächtnis – hier erfolgt die Stimulusidentifikation – Repräsentationen in einem Kurzzeitgedächtnis, welches von der Modalität des Inputs abhängt. Es ist aber auch möglich, dass die Information direkt ins episodische Langzeitgedächtnis gelangt. Wenn Spuren im episodischen Langzeitgedächtnis gebildet werden, um Information aus dem episodischen Langzeitgedächtnis abzurufen und ins Kurzzeitgedächtnis zu transferieren, beginnt die Aktivität des zentralen Prozessors aktiv. Weiterhin ist der zentrale Prozessor am Transfer zwischen den verschiedenen Formen des Kurzzeitgedächtnisses beteiligt, wenn Informationen in einer anderen Modalität rekodiert werden sollen. Er ist auch

an anderen mentalen Operationen beteiligt, wie bei der Auswahl von Handlungsalternativen. Im Gegensatz zum Modell von Atkinson & Shiffrin (1968) erfolgt also bei der Informationsverarbeitung nicht zwingend ein serielles Durchlaufen der Gedächtnisspeicher. Ein serieller Informationsfluss ist zwar möglich, aber nicht zwingend. So kann nach dem Modell von Pashler & Carrier (1996) Information ins Kurzzeitgedächtnis gelangen, ohne vorher im sensorischen Gedächtnis zu verweilen, oder Information kann ins Langzeitgedächtnis gelangen, ohne vorher im Kurzzeitgedächtnis gewesen zu sein. Ein weiterer immanenter Unterschied zum Modell von Atkinson und Shiffrin (1968) ist die Annahme eines zentralen Prozessors. Durch diese Annahme wird Ergebnissen zur zentralen Interferenz Rechnung getragen. Nach Pashler und Carrier (1996) entsteht zentrale Interferenz dann, wenn mehr als ein zentraler Prozess zur gleichen Zeit aktiv wird, wie beispielsweise bei Antwortselektion, Abruf aus dem Langzeitgedächtnis oder Speicherung im Langzeitgedächtnis.

2.2.1. Sensorisches Gedächtnis

Nach kurzer Präsentation eines visuellen Stimulus bleibt im visuellen sensorischen Gedächtnis eine sensorische Spur des Stimulus mehrere hundert Millisekunden bestehen. Phänomenologisch stellt diese Spur eine zerfallende Version des gegebenen Stimulus dar. Die exakte Speicherkapazität ist nicht bekannt, jedoch ist sie größer als die Kapazität des Kurzzeitgedächtnisses und beträgt ca. 12 Buchstaben. Das auditive sensorische Gedächtnis kann Informationen länger als das visuell sensorische Gedächtnis speichern (ein bis zwei Sekunden). Die Speicherkapazität ist jedoch begrenzter. Dieses System zu erklären, bereitet bislang Schwierigkeiten. Dies liegt zum einen daran, dass in Laborexperimenten zum Gedächtnis die Darbietung von Stimulusmaterial zeitlich begrenzt wird und dass in der realen Welt Stimuli meist nicht so schnell „verschwinden“. Als eine mögliche Funktion des sensorischen Gedächtnisses sehen Pashler und Carrier (1996), dass es visuelle Eindrücke gerade so lange hält, um die Zeit während des Augenblinzeln zu überbrücken.

2.2.2. Das Kurzzeitgedächtnis

Das Kurzzeitgedächtnis enthält mehrere Subsysteme, um Information für kurze Zeit zu behalten. Das Speichern von Informationen in diesen Systemen schützt diese davor, von neuen sensorischen Informationen gelöscht zu werden. Da die Systeme spezialisierte kognitive Funktionen, wie z.B. Sprache, visuelles Problemlösen oder motorische Kontrolle repräsentieren, wird angeführt, dass das Kurzzeitgedächtnis auch Ausdruck eines Prozesses sein könnte und nicht unbedingt eine Struktur repräsentieren muss (Pashler & Carrier, 1996). Große Aufmerksamkeit im vorliegenden Modell wird der Erklärung des Transfers von einem Speicher in einen anderen gewidmet. Zu diesem Transfer von einem Gedächtnissystem in ein anderes ist Aufmerksamkeit und willentliche Kontrolle notwendig (Pashler & Carrier, 1996). Auf das Thema der Aufmerksamkeit und der willentlichen Kontrolle wird im Abschnitt Modelle der Eindrucksbildung im Zusammenhang mit automatischen und kontrollierten Prozessen bei der Informationsverarbeitung noch näher eingegangen. Im vorliegenden Zusammenhang bezieht sich der Begriff der Aufmerksamkeit auf Prozesse, die an der Stimulusselektion, der Kapazitätsbegrenzung der Speicher und der Variation von kognitiven Ressourcen beteiligt sind. Aufmerksamkeit spielt bereits bei der frühen Verarbeitung von Informationen eine Rolle. Beispielsweise kann aktives Ignorieren einer Information das Ausmaß oder die Qualität der perzeptuellen Verarbeitung beeinträchtigen. Auch paralleles Verarbeiten ist möglich, jedoch nur in begrenztem Ausmaß, da die Kapazität des sensorischen Gedächtnisses, vor allem innerhalb derselben sensorischen Modalität, schnell ausgeschöpft ist (Duncan, 1987).

Im Normalfall besteht willentliche Kontrolle darüber, welche Information gespeichert werden soll. Für die visuelle und auditorische Modalität wurde dies nachgewiesen. Loftus und Ginn (1984) konnten jedoch zeigen, dass auch nicht willentlich selektierte Information ins Kurzzeitgedächtnis transferiert werden kann, und zwar dann, wenn kurze Zeit nach der wahrzunehmenden Information irrelevante Information präsentiert wird.

Auch beim Transfer in das Langzeitgedächtnis spielt die willentliche Kontrolle eine wichtige Rolle. Eine der effektivsten Strategien, um Information in das Langzeitgedächtnis zu überführen, ist Elaboration. Dieser Begriff der Elaboration von Informationen ist eng mit dem Modell der Verarbeitungstiefe von Craik und Lockhardt (1972) verbunden. Craik und Lockhardt (1972) postulieren für präsentierte Information eine Verarbeitung in mehreren Stufen. Auf der ersten Stufe erfolgt eine Analyse der physikalisch-sensorischen Merkmale

eines Wortes, dann findet eine Mustererkennung statt und als letzte Stufe wird das Wort semantisch verarbeitet. Die Gedächtnisleistung ist umso besser, je tiefer die Verarbeitung erfolgt. Elaboration kann den Abruf von assoziierten Informationen enthalten, so dass die neue Information mit bereits bestehenden Wissensstrukturen vernetzt werden kann. Ebenso kann ein Abruf von Gedächtnisinhalten bestehende Gedächtnisspuren stärken.

Zusammenfassend stellen Pashler & Carrier (1996) fest, dass wenn sich Information im Kurzzeitgedächtnis befindet, dies weder eine notwendige noch hinreichende Bedingung dafür darstellt, um ins Langzeitgedächtnis transferiert zu werden. Konkurrierende Aufgaben, die Aufmerksamkeit und damit kognitive Ressourcen erfordern, zeigen große Effekte auf die Gedächtnisleistung. Dies ist selbst dann der Fall, wenn verschiedene sensorische Modalitäten beteiligt sind und keine offene Reaktion gefordert wird.

2.2.3. Episodisches versus semantisches Gedächtnis

Die Kapazität des Langzeitgedächtnisses ist umfassend, eine genaue Quantifizierung ist jedoch kaum möglich. Pashler & Carrier (1996) postulieren für ihr Mehrspeichermodell ein episodisches und ein semantisches Gedächtnis. Die Unterscheidung von episodischem und semantischem Gedächtnis geht auf Tulving (1972) zurück. Das semantische Gedächtnis beantwortet Fragen über allgemeines Wissen, es enthält also Wissen, Regeln, Tatbestände und Bedeutungen, die nicht an einen spezifischen Kontext gebunden sind. Das episodische Gedächtnis erlaubt räumlich-zeitliche Unterscheidungen von Episoden, die eine Person selbst erlebt hat. Die Grenzen zwischen episodischem und semantischem Gedächtnis sind jedoch unscharf (Wessells, 1994). Es gibt Autoren, die die episodische und semantische Information als ein Kontinuum beschreiben, an dessen einen Ende eine auf einen räumlich-zeitlichen Kontext bezogene Information und am anderen Ende eine kontextfreie Information steht (z.B. Kihlstrom, 1980). Wessells (1994) nennt als Grundlage für das Verständnis des semantischen Gedächtnisses die Bestimmung der Repräsentation von Wissen im System der menschlichen Informationsverarbeitung.

Im Zusammenhang mit der Repräsentation von Wissen im semantischen Gedächtnis soll nachfolgend der propositionale Ansatz dargestellt werden, da im empirischen Teil dieser Arbeit die Stimulusinformation systematisch variiert wurde und es sich bei einer Modalität des Stimulusinputs um Text handelt. Um Gedächtnisleistungen zu operationalisieren, werden komplexe Texte in Propositionen unterteilt, so dass man ein für die Operationalisierung

quantitatives Maß erhält. Die Proposition ist ein grundlegender Bestandteil der Wissensstruktur (Kintsch, 1974). Es wird angenommen, dass Konzepte zu Propositionen zusammengefasst werden können (z.B. Anderson & Bower, 1973, Engelkamp, 1973). Propositionen bestehen aus funktional differenzierten Konzepten: aus dem Prädikatkonzept und mindestens einem Argumentkonzept. Das Prädikat stiftet eine Relation zwischen den Argumenten. Es ist möglich, einen ganzen Text in Propositionen aufzugliedern, die dann die semantische Struktur eines Textes wiedergeben (Kintsch, 1974). Ein weiteres mögliches Prinzip zur Strukturbildung ist die Hierarchie. Im Unterschied zu einer zufälligen Anhäufung von Wörtern ist ein Text ein kohärentes Ganzes. Der Text bildet ein Netzwerk semantischer Einheiten, die alle so miteinander verknüpft sind, dass kein Element unverbunden bleibt (Grabowski, 1991). Kohärenz kann nach Kintsch (1974) durch zwei Mittel entstehen: durch Argumentüberlappung, bei der eine Proposition ein Argument mit einer anderen Proposition gemeinsam hat und durch Einbettung, bei der eine ganze Proposition in einer anderen als Argument wiederholt wird. Der Kohärenzgraph eines Textes expliziert die Verbindungen zwischen den am Text beteiligten Propositionen und bringt dessen hierarchische Struktur zum Ausdruck. (Kintsch, 1974). In Studien konnte gezeigt werden, dass Propositionen für die semantisch-kognitive Verarbeitung tatsächlich von Bedeutung sind. So konnte Engelkamp (1973) zeigen, dass Sätze, die sich aus zwei Propositionen zusammensetzen, schlechter behalten werden als Sätze, die nur eine Proposition bilden, obwohl beide über die gleiche Oberflächenstruktur (gleiche Wortanzahl) verfügten. Auch die Lesezeit von Sätzen steigt mit der Anzahl ihrer Propositionen bei wiederum gleicher Wortanzahl (Kintsch & Keenan, 1973). Ratcliff & McKoon (1978) konnten nachweisen, dass es sich bei Propositionen um Einheiten handelt, die assoziativ miteinander verbunden sind. Ihre Versuchspersonen entschieden schneller, dass ein dargebotenes Wort in einem zuvor gelernten Satz vorkam, wenn das zuvor dargebotene Wort zur selben Proposition gehörte wie das Zielwort. Beyer (1987) konnte einen Hierarchieeffekt bei längeren Texten (mehr als 50 Propositionen) aufzeigen. Die Wiedergabe für hierarchisch hoch stehende Propositionen ist dabei besser als für hierarchisch niedriger stehende. Des Weiteren werden hierarchisch hoch stehende Propositionen länger erinnert.

Die Konzeption eines propositional-amodalen Gedächtnisses als Basis für Sprachverständnis und für Denk- und Entscheidungsprozesse wurde jedoch auch kritisiert. Dörner (1997) beispielsweise führt an, dass „die Annahme eines solchen Gedächtnis schlecht begründet ist“ und im Weiteren „die Annahme eines amodalen Gedächtnis, als Gedächtnis für begriffliche Relationen unnötig, da sich solche begrifflichen Relationen auch schon in einem modalen

Gedächtnis auffinden lassen“ (S. 172). Auch Grabowski (1991) übt inhaltliche Kritik. Schwächen sieht er z.B. in der Anwendung bei der Überprüfung von Behaltensleistungen von längeren Texten. Wenn Versuchspersonen gelesene Texte reproduzieren, so haben die Reproduktionen auf der sprachlichen Oberfläche oft wenig mit dem Originaltext gemein. Durch Paraphrasieren, Zusammenfassen von Inhalten und Konfabulationen ergeben sich oft Schwierigkeiten, die Propositionen der Reproduktionen denen des Originaltextes zuzuordnen. Dieses Problem stellt sich jedoch nicht nur im Zusammenhang mit der Theorie des propositionalen Ansatzes, sondern dürfte – unabhängig von einer Gedächtnistheorie – grundsätzlich eines der methodischen Probleme wieder spiegeln, die sich beim Wiedergeben von längeren Texten ergeben. Trotz der Kritik hat sich das Modell in der Prognose von Behaltensleistungen als robust erwiesen.

2.3. Die multimodale Gedächtnistheorie von Engelkamp (1991)

Die multimodale Gedächtnistheorie von Engelkamp (1991) trägt der Annahme Rechnung, dass Gedächtnisleistungen auf Informationen verschiedener Modalitäten beruhen und in Abhängigkeit von diesen variieren können. Engelkamp (1991) unterscheidet einerseits sprachliche versus nichtsprachliche Informationsverarbeitungsprozesse sowie die Gedächtnisprozesse bei der Wahrnehmung von motorischen Handlungen und beim Ausführen motorischer Handlungen andererseits. Der Einfluss des Ausführens von motorischen Handlungen auf Gedächtnisleistungen soll in der vorliegenden Arbeit nicht thematisiert werden, so dass sich die Ausführungen an dieser Stelle auf die Unterschiede zwischen der Verarbeitung von Bildern und Worten und die Gedächtnisleistung beim Wahrnehmen von motorischen Handlungen beschränken.

Ein gesicherter Befund in der Gedächtnisforschung ist, dass Bilder von Objekten besser erinnert und wieder erkannt werden als die verbalen Bezeichnungen (Madigan, 1983). Wenn das Behalten durch Rekognition geprüft wird, findet sich ein klarer Bildüberlegenheitseffekt (Madigan, 1983). Nach Paivio (1971) ist der Bildüberlegenheitseffekt primär eine Folge der dualen Enkodierung von Bildern, das heißt, es wird angenommen, dass bei der Darbietung von Bildern automatisch eine Aktivierung von Wortmarken erfolgt. Im Gegensatz dazu wird für die Darbietung von Wörtern nur eine geringe Wahrscheinlichkeit angenommen, dass sie automatisch zur Aktivierung der Bildmarken führen.

Dagegen ist nach der multimodalen Theorie (Engelkamp, 1991) die entscheidende Grundlage für den Bildüberlegenheitseffekt in den unterschiedlichen Markeninformationen von Wörtern und Bildern zu suchen. Engelkamp (1991) postuliert für die Verarbeitung von Bildern und Wörtern eigenständige Systeme. Bei der Wortenkodierung stehen Wortmarken und Konzepte im Vordergrund. Die Verarbeitung erfolgt primär über das verbale System.

Die Darbietung von Bildern aktiviert automatisch nur ihre Bildmarken und Konzepte, die Darbietung von Wörtern nur deren Wortmarken und Konzepte. Auch experimentelle Befunde stützen die Annahme, dass bei der Wahrnehmung von Bildern weder Wortmarken noch Sprechprogramme aktiviert werden (z.B. Intraub, 1979, Roediger & Weldon, 1987), so dass der Ort des Bildüberlegenheitseffektes in der Aktivierung von Bildmarken zu suchen ist (Engelkamp, 1991, 1998). Dass den Bildmarken eine große Rolle beim Wiedererkennen von Bildern zukommt, ist bereits an der großen Wiedererkennungsleistung für Bilder zu sehen, die in einer Vielzahl von Experimenten beobachtet wurde (z.B. Standing, 1973, Brown & Scott, 1971). Es wird diskutiert, inwiefern die Wirksamkeit der Bildmarkeninformation an die Verfügbarkeit unterschiedlicher Konzepte gebunden ist. In einer Untersuchung von Wiseman & Neisser (1974) beispielsweise wurden den Probanden unvollständige Bildkonturen dargeboten. Als Resultat wurde gefunden, dass nur diejenigen Bilder gut wieder erkannt wurden, deren Bedeutung in der Enkodierphase korrekt erkannt worden war. Die Bedeutung der Verfügbarkeit von semantischen Kategorien wird auch an vielen anderen Untersuchungen – z.B. zur Rekognitionsleistung bei Schachpositionen beim Vergleich zwischen Novizen und Experten (z.B. de Groot, 1965; Chi et. al., 1988) - deutlich. Das Thema bereichsspezifisches Wissen und Verfügbarkeit von semantischen Kategorien für die Gedächtnisleistung wird im Kapitel Einfluss der Beurteilererfahrung noch aufgegriffen.

Für das Behalten von wahrgenommenen Handlungen liegen kaum Untersuchungen vor (Engelkamp, 1991). Es wird ausgeführt, dass auch wahrgenommene Ereignisse Gedächtnisspuren im visuellen und konzeptuellen System hinterlassen. Dies wird jedoch nur sichtbar, wenn entsprechende Prüfmethode angewendet werden, damit das Behalten sensorischer Information auch sichtbar wird. So interpretiert Engelkamp (1991) Befunde von Brewer & Dupree (1983), dass die gezeigten Handlungssequenzen mit episodischem Charakter von den Probanden nach um-zu-Relationen interpretiert werden und dass diese konzeptuelle Information behalten und im Free Recall wiedergegeben wird. Mit dieser Versuchsanordnung ist das Messen der Behaltensleistung von Oberflächeninformationen von Handlungen nicht möglich, das heißt jedoch nicht, dass die Oberflächeninformation von den

Probanden vergessen wird. Leibrich & White (1983) zeigten, dass auch die Oberflächeninformation von Handlungen längerfristig verfügbar ist und dass diese Verfügbarkeit mit der konzeptuellen Enkodierung der Handlungsinformation interagiert. Als weitere Erkenntnis führt Engelkamp (1991) aus, dass auch wahrgenommene Handlungen besser als ihre verbalen Bezeichnungen erinnert werden. Empirische Belege finden sich bei Jung (1989). Er präsentierte einer Gruppe von Versuchspersonen 48 Handlungen über Video, einer weiteren Gruppe von Versuchspersonen wurden die entsprechenden Handlungssphrasen über eine Tonbandaufnahme verbal wiedergegeben. Jung (1989) konnte einen klaren Bildüberlegenheitseffekt beobachten. Die Probanden, die das Video gesehen hatten, erinnerten 42 % der Handlungen, die Probanden, die die Tonbandaufnahme hörten, nur 26 %. Als Schlussfolgerungen aus den wenigen vorliegenden Untersuchungen zu wahrgenommenen Handlungsereignissen zieht Engelkamp (1991) neben den oben genannten Befunden, dass Handlungsereignisse in einer besonderen Art enkodiert werden. Wahrgenommene Handlungen werden unter ein gemeinsames Handlungsziel subsumiert, sofern die Listenstruktur diese Enkodierung vorsieht. Engelkamp (1991) postuliert, dass wahrgenommene Handlungen, immer wo es möglich ist, als Mittel zur Erreichung von bestimmten Zielen interpretiert und auch enkodiert werden.

Zusammenfassend lässt sich festhalten, dass das menschliche Gedächtnis bei allen Wahrnehmungs- und Urteilsprozessen eine grundlegende Rolle spielt (Hiebsch, 1986). Erklärungsversuche für die Ergebnisse, dass Bilder und Worte unterschiedlich gut behalten werden, gibt z.B. Engelkamp (1991, 1998) mit seiner multimodalen Gedächtnistheorie. Auch werden Befunde für das Behalten einfacher wahrgenommener Handlungen diskutiert. Die Wahrnehmung und Erinnerung von hoch komplexen Situationen in der realen Welt zum jetzigen Zeitpunkt erschöpfend zu erklären, ist praktisch unmöglich. Allen Untersuchungen zum Gedächtnis ist gemein, dass die Fragestellung jeweils auf wenige Variablen reduziert wurde, um elementare Konstrukte zu selektieren, mit denen sich Gedächtnisphänomene erklären lassen. Mit diesen „Ausschnitten“ lässt sich die komplexe Arbeitsweise des Gedächtnisses nicht vollständig abbilden. Auch Engelkamp (1991) führt aus, dass „der Stand der Gedächtnistheorie noch nicht sehr ausgereift [ist]“ (S.11). Im Zusammenhang für die vorliegende Arbeit bleibt festzuhalten, dass bei der Darbietung von Objekten in Bildform diese besser behalten werden als deren verbale Bezeichnungen und dass nach Engelkamp (1991, 1998) dies auf die unterschiedliche Verarbeitung von Wort- und Bildmaterial im Gedächtnis zurückzuführen ist. Auch für wahrgenommene Handlungen gilt, dass diese besser

behalten werden als ein verbales Aufzählen von Handlungen. Im Kontext der sozialen Urteilsbildung können solche Gedächtnisprozesse jedoch nicht völlig isoliert betrachtet werden, vielmehr müssen die Phänomene in ihrer eigenen Komplexität erforscht werden (Strack, 2000).

2.4. Der Zeitpunkt der Urteilsbildung im Gedächtnis

Die Annahme eines Prozesses für das Phänomen der Urteilsbildung wurde von Strack (1988) als „Standardsequenz der Informationsverarbeitung“ bezeichnet. Die Urteilsbildung wird im zeitlichen Verlauf betrachtet. Hierbei gelten Enkodierung der Information, ihre interne Repräsentation mit dazugehörigen Transformationen und der Abruf aus dem Gedächtnis als minimal-verbundene Prozessphasen (Strack, 2000).

Im Bereich der sozialen Wahrnehmung unterscheiden Hertel & Bless (2000) zwei verschiedene Urteilsprozesse hinsichtlich des Zeitpunktes der Urteilsbildung. Zum einen können Urteile unmittelbar während der Enkodierung von relevanten Informationen – also „on-line“ gebildet werden. Diese Urteile werden dann als unabhängige Einheiten im Gedächtnis gespeichert. Zum Anderen kann die Urteilsbildung erinnerungsgestützt verlaufen, das heißt, die im Gedächtnis gespeicherte Information dient als Grundlage der Urteilsbildung und muss zum Zeitpunkt der Urteilsbildung erst aus dem Gedächtnis abgerufen, also erinnert werden (z.B. Hastie & Park, 1986). Online-Urteile können zu einem späteren Zeitpunkt unabhängig von Einzelinformationen abgerufen werden und sind später relativ robust gegen nachfolgende Gedächtnisprozesse, wie z.B. Vergessen von einzelnen Informationen, Neustrukturierung des Materials (Mc Conell, Sherman & Hamilton, 1994).

Als Bedingung, welche eine on-line oder eine erinnerungsgestützte Urteilsbildung beeinflusst, wird genannt, ob der Wahrnehmer zum Zeitpunkt der Enkodierung weiß, dass ein Urteil benötigt wird (Hertel & Bless, 2000). Weiß der Wahrnehmer nicht, dass später ein Urteil benötigt wird, kommt es zu diesem späteren Zeitpunkt zur erinnerungsgestützten Urteilsbildung. Wenn der Wahrnehmer jedoch weiß, dass er ein Urteil abgeben soll, z.B. als Beobachter einer Diskussion im Assessment Center, wird er während des Gespräches on-line Urteile hinsichtlich der Fachkompetenz und der Kommunikationsfähigkeit der Bewerber bilden. Hinsichtlich anderer Aspekte muss er jedoch auf erinnerungsgestützte Prozesse

zurückgreifen, z.B. wenn er nach den Gesprächen bemerkt, dass einer der Bewerber sein Handy gestohlen hat.

Als zweite Bedingung für das Auftreten von on-line Urteilen werden Aufmerksamkeitsprozesse thematisiert. So können on-line Urteile spontan auftreten, wenn das Wahrnehmungsobjekt als bedeutsam eingestuft wird und genügend kognitive Ressourcen vorhanden sind, das heißt, dass die Aufmerksamkeit nicht anderweitig gebunden ist (Hertel & Bless, 2000).

Als dritte mögliche relevante Bedingung für die Art der Urteilsbildung wird die Existenz von Urteilen auf ähnlichen Dimensionen benannt (Hastie & Pennington, 1989). Der Wahrnehmer kann, um den kognitiv aufwendigeren Prozess der erinnerungsgestützten Urteilsbildung umgehen, ähnliche Urteile als Urteilsanker (Tversky & Kahnemann, 1974) oder als Orientierung im Sinne von Halo-Effekten nutzen. Der Rückgriff auf vorher gebildete Urteile auf ähnlichen Dimensionen verbraucht weniger kognitive Ressourcen als eine umfangreiche Suche nach relevanten Informationen im Gedächtnis, so dass Urteile unter diesen Bedingungen on-line gebildeten Urteilen hinsichtlich ihrer Eigenschaften ähnlich sind (Hertel & Bless, 2000).

Erinnerungsgestützte Prozesse zur Urteilsbildung sollten also vor allem dann stattfinden, wenn dem Wahrnehmer vor der Informationsaufnahme nicht bekannt ist, dass später ein Urteil verlangt wird, wenn von ihm vorher keine ähnlichen Urteile gebildet wurden, auf die er zurückgreifen könnte und wenn ihm Urteilsdimensionen zum Zeitpunkt der Informationsaufnahme nicht zur Verfügung stehen. Für den Vergleich der Häufigkeit des Auftretens von Online-Urteilen und gedächtnisgestützter Urteile fasst Moser (1992) zusammen, dass „memory-based judgment formation may be more prevalent than the on-line processing literature suggests“ (S. 555), so dass davon auszugehen ist, dass in komplexen sozialen Situationen außerhalb eines experimentellen Settings, aber auch im experimentellen Setting, welches der angestrebten ökologischen Validität der Ergebnisse wegen eine eher extern valide Aufgabenstellung verwendet, häufig erinnerungsgestützte Prozesse auftreten.

Bevor Ausführungen zum Einfluss von unterschiedlichen Stimulusinformationen und der Erfahrung der Beurteiler bei der gedächtnisgestützten Urteilsbildung gemacht werden, werden im nächsten Kapitel führende Modelle der Eindrucksbildung und zwei Kernmechanismen in der Urteilsbildung vorgestellt.

3. Modelle und Mechanismen der Eindrucksbildung

In diesem Abschnitt werden die gegenwärtig führenden Modelle der Eindrucksbildung – das „Zweiprozessmodell der Eindrucksbildung“ von Brewer (1988) und das „Kontinuummodell der Eindrucksbildung von Fiske und Neuberg (1990) - sowie zwei Kernmechanismen in der Eindrucksbildung – Reduktion der Komplexität und Vergleichsprozesse - vorgestellt. Sowohl bei den Modellen als auch bei der Beschreibung der Kernmechanismen wird häufig eine Unterscheidung von kontrollierten und automatischen Informationsverarbeitungsprozessen in der Eindrucksbildung vorgenommen, so dass an dieser Stelle auf diese Unterscheidung kurz eingegangen wird.

Kontrollierte Prozesse sind aufmerksamkeitsabhängig und arbeiten daher sukzessiv oder seriell (Engelkamp, 1991). Auf diesen eher langsamen, auf bestimmten Regeln basierenden Informationsverarbeitungsprozessen beruht die Durchführung wenig geübter oder neuer Aufgaben. Kontrollierte Prozesse beanspruchen Aufmerksamkeit, Bewusstsein und willentliche Kontrolle. Kontrollierte Prozesse haben einen hohen Verbrauch an kognitiven Ressourcen und lassen sich durch viel Übung in automatisieren (Engelkamp, 1991).

Die automatischen Prozesse sind schnelle, assoziative Informationsverarbeitungsmodi, die sich nicht willentlich unterdrücken lassen. Diese Prozesse beruhen auf Heuristiken, laufen unabhängig von der begrenzten Ressource „Aufmerksamkeit“ ab und benötigen daher kaum kognitive Ressourcen. Auf solchen Prozessen beruht die Durchführung von viel geübten und überlernten Aufgaben, wie beispielsweise Gehen oder Fahrrad fahren, hier wird nur keine oder nur wenig kognitive Anstrengung, Bewusstsein oder willentliche Kontrolle gebraucht (Engelkamp, 1991; Chaiken & Trope, 1999).

3.1. Das Zweiprozessmodell der Eindrucksbildung von Brewer (1988)

Das Zweiprozessmodell der Eindrucksbildung geht davon aus, dass die Motive und Ziele einer wahrnehmenden Person die Eindrucksbildung über eine andere Person beeinflussen. In Abhängigkeit von den Zielen und Motiven der wahrnehmenden Person kann dieselbe Personinformation entweder zu einem kategorienbasierten Eindruck oder zu einem personenbasierten Eindruck führen (Brewer, 1988).

Zunächst sollen die kategorienbasierten Informationsverarbeitungsprozesse kurz erläutert werden. Brewer (1988) postuliert für die Verarbeitung von Personeninformationen eine maßgebliche Beeinflussung durch ihre Zugehörigkeit zu bestimmten sozialen Kategorien. Für die Eindrucksbildung werden die Personinformationen auf der Grundlage von vorhandenem bzw. erworbenem Wissen über bestimmte soziale Gruppen interpretiert. Die wahrgenommene Person wird einer bestimmten sozialen Kategorie zugeordnet, der sie offensichtlich (z.B. Geschlecht, Alter) oder wahrscheinlich (z.B. Polizist, Arzt) angehört. Brewer (1988) betont, dass diese kategorienbasierten Informationsverarbeitungsprozesse bei der Eindrucksbildung Vorrang vor personenbasierten Informationsverarbeitungsprozessen haben und untermauert diese Annahmen durch Studien von Pryor und Ostrom (1981) und Ostrom, Pryor und Simpson (1981). Folgt man den Annahmen von Brewer (1988), kommt es bei der Eindrucksbildung von unbekanntem Personen zum größten Teil zur Bildung eines kategorienbasierten Eindrucks anstelle eines personenbasierten Eindrucks von diesen Personen.

Bei den personenbasierten Informationsverarbeitungsprozessen werden die Informationen über eine unbekannte Person zu einem Gesamteindruck verarbeitet, ohne dass bei der Interpretation auf eine soziale Kategorie Bezug genommen wird. In diesem Fall wird von der wahrgenommenen Person ein individueller, personenbasierter Eindruck gebildet (Brewer, 1988, Brewer & Feinstein, 1999).

Das Zweiprozessmodell der Eindrucksbildung von Brewer (1988) postuliert, dass die sozialen Informationen über eine Zielperson in chronologisch aufeinander folgenden Stadien – Identifikation, Kategorisierung, Personalisation und Individuation – verarbeitet werden. Der Eindrucksbildungsprozess läuft als Sequenz dieser Verarbeitungsstadien ab. Spätere Verarbeitungsstufen des Eindrucksbildungsprozesses werden erst durchlaufen, wenn vorhergehende Verarbeitungsstadien abgeschlossen wurden. Ausgehend von der Annahme, dass eine Person bestrebt ist, möglichst schnell und ohne hohen Verbrauch an kognitiven Ressourcen einen Eindruck von der Zielperson zu gewinnen, postuliert Brewer (1988), dass der Wahrnehmer spätere Verarbeitungsstufen zunächst abwehrt, da diese mehr Aufmerksamkeit und kognitive Ressourcen benötigen als vorhergehende Verarbeitungsstufen. Das heißt, der Wahrnehmer bemüht sich, den kategorienbasierten Eindruck der Zielperson möglichst aufrecht zu erhalten. Der Eindrucksbildungsprozess endet auf der

Verarbeitungsstufe, auf der die Personinformation hinsichtlich der Ziele und der Motivation für den Wahrnehmer zufrieden stellend verarbeitet wurde.

3.2. Der Prozess der Eindrucksbildung nach Brewer (1988)

Mit der Wahrnehmung einer unbekannt Person beginnt der Prozess der Eindrucksbildung. Entsprechend der Sequenz der Verarbeitungsstadien erfolgt die Identifikation der Zielperson durch den Wahrnehmer. Die Identifikation erfolgt durch Interpretation und Zuordnung der Zielperson zu physischen Beurteilungskategorien (Geschlecht, Alter und Ethnizität), die häufig als „primitive“ soziale Kategorien (Hamilton & Sherman, 1994) bezeichnet werden. Die Grundlage für die Identifikation einer Zielperson bilden also kategorienbasierte Informationsverarbeitungsprozesse. Eine bewusste, kontrollierte Informationsverarbeitung der Personinformation findet erst in nachfolgenden Verarbeitungsstadien statt. Die Annahmen Brewers (1988) implizieren, dass die Aktivierung einer ersten Kategorie bei der Begegnung mit der Zielperson automatisch geschieht, da die Identifikation einer Zielperson immer der erste Schritt in der Eindrucksbildung ist. Auf dieser Verarbeitungsstufe wird weiterhin zu der Identifikation durch den Wahrnehmer beurteilt, ob die Zielperson für seine derzeitigen Ziele und Bedürfnisse relevant ist oder nicht. Kommt der Wahrnehmer zu der Einschätzung, dass die Zielperson für ihn keine Relevanz besitzt, wird der Eindrucksbildungsprozess in diesem Verarbeitungsstadium beendet. Die aufgrund ihrer Identifikation aktivierte „primitive“ soziale Kategorie wird zur Einschätzung der Zielperson herangezogen und die Person wird entsprechend beurteilt (z.B. Er ist ein alter Mann).

Wird die Zielperson als relevant wahrgenommen, erfolgt eine weitere Verarbeitung sozialer Information über die Zielperson. Der Wahrnehmer stellt nun fest, ob Selbstinvolvierung vorliegt, das heißt, er schätzt ein, ob er sich mit der Zielperson verbunden fühlt (z.B. Empathie beim Anblick eines weinenden Kindes) oder ob er von ihr in irgendeiner Form abhängig ist (z.B. „Ist die Zielperson die für mich zuständige Krankenschwester?“). Kommt der Wahrnehmer zu dem Schluss, dass hier eine Selbstinvolvierung vorliegt, werden die weiteren Personinformationen durch personenbasierte Informationsverarbeitungsprozesse verarbeitet (Brewer, 1988).

Besitzt die Zielperson für den Wahrnehmer Relevanz und kommt er gleichzeitig zu dem Schluss, dass keine Selbstinvolvierung vorliegt, erfolgt die weitere Verarbeitung auf der Stufe

der Kategorisierung. Brewer (1988) postuliert für soziale Kategorien eine Hierarchie, von stark abstrakten bis hin zu konkreten oder spezifischen Kategorien. Es ist zum Beispiel möglich, dass eine Person die recht abstrakte generelle soziale Kategorie „ältere Frauen“ ausgebildet hat und innerhalb dieser eine recht abstrakte Subkategorie „Lehrerin“ und innerhalb dieser eine noch konkretere Subkategorie der „Lehrerin“ wie beispielsweise „freundliche und kreative Lehrerin“. In diesem Verarbeitungsstadium kommt es, so postuliert Brewer (1988), hinsichtlich der genannten Kategorienhierarchie zu kontrollierten kategorienbasierten Informationsverarbeitungsprozessen, die abwärts (top-down) gerichtet sind. Der Wahrnehmer beginnt mit der abstraktesten Kategorienebene (z.B. Geschlecht, Alter) und prüft, ob die gegebenen Personinformationen einer bestimmten sozialen Kategorie zugehörig sind oder nicht. Wird keine Passung zwischen Kategorie und Information gefunden, geht der Wahrnehmer in der Kategorienhierarchie solange abwärts zu den konkreteren Subkategorien, bis eine zufrieden stellende Passung zwischen einer konkreten Kategorie und den Informationen über die Zielperson erreicht wurde. Wenn eine passende Subkategorie für die Zielperson gefunden wurde, werden neue Personinformationen um diese konkrete Kategorie organisiert und weitere Eigenschaften und Persönlichkeitsmerkmale, die erst aufgrund der Kategorisierung aktiviert und der Zielperson zugeschrieben werden, werden aus dem Kategorieninhalt abgeleitet. Die Zielperson wird anhand der aktivierten Kategorie beurteilt (Brewer, 1988).

Wenn aufgrund inkonsistenter oder vielfältiger Informationen über die Zielperson keine als am angemessensten empfundene soziale Kategorie gefunden werden kann, tritt der Prozess der Individuation auf. Der Wahrnehmer interpretiert dann die Zielperson als ein besonderes Beispiel dieser sozialen Kategorie (Billig, 1985).

Trotz der Individuation handelt es sich bei den Informationsprozessen in diesem Verarbeitungsstadium immer noch um kategorienbasierte Informationsverarbeitungsprozesse. Der Wahrnehmer definiert Eigenschaften und Merkmale, die die Zielperson von einer passenden sozialen Kategorie unterscheiden und nutzt die aktivierten Kategorien zur Beurteilung der Zielperson (Brewer, 1988, Brewer & Feinstein, 1999).

Wenn der Wahrnehmer die Zielperson überhaupt keiner bereits aktivierten sozialen Kategorien zuordnen kann oder wenn die Informationen über die Zielperson zu vielfältig sind, um sie kategorienbasiert zu verarbeiten, findet die Bildung eines personenbasierten Eindrucks statt. Personalisation tritt nur dann auf, wenn der Wahrnehmer die Fragen nach der Relevanz

der Zielperson und der Selbstinvolvierung zuvor beide bejaht hat. Grundlage für die Generierung eines Gesamteindrucks über die Zielperson sind in diesem Fall all ihre individuellen Eigenschaften und Merkmale, ohne dass Bezug auf eine soziale Kategorie genommen wird (Brewer, 1988).

Auf dieser Verarbeitungsstufe finden nur personenbasierte Informationsverarbeitungsprozesse statt, die aufwärts (bottom-up) verlaufen. Die personenbasierten Verarbeitungsprozesse beginnen auf der konkretesten Informationsebene (z.B. die Zielperson wirkt unsicher und hektisch) und enden auf der niedrigsten Abstraktionsebene der Kategorienhierarchie, die den Zielen des Wahrnehmers entsprechen (Wyer & Srull, 1986). In diesem Fall werden die Informationen über eine Zielperson sukzessiv verarbeitet, um zu einem Eindruck von ihr zu gelangen. Kategorieninkonsistente Informationen werden hierbei besonders stark gewichtet (Brewer & Feinstein, 1999). Zum Beispiel kann eine Zielperson nach ihrem unsicheren Benehmen, ihren hektischen Bewegungen, ihrer zitternden Finger usw. beurteilt werden und nicht danach, welchen sozialen Kategorien (Mann oder Frau, Europäer oder Asiate etc.) sie angehört. Der nun vom Wahrnehmer gebildete Eindruck von der Zielperson ist personenbasiert und nicht kategorienbasiert (vgl. Brewer, 1988, Wyer & Srull, 1986). Zusammenfassend ist festzuhalten, dass Brewer (1988) eine automatische Aktivierung einer Urteilskategorie, die nicht willentlich kontrollierbar ist, postuliert. Erst die Anwendung dieser Kategorien in späteren Verarbeitungsstadien unterliegt der willentlichen Kontrolle des Wahrnehmers. Dies kann insbesondere durch die Fokussierung und Gewichtung von inkonsistenter Information über die Zielperson erfolgen. Als wichtige Voraussetzung hierfür werden ausreichend kognitive Ressourcen beim Wahrnehmer postuliert (Chaiken & Trope, 1999).

3.3. Das Kontinuummodell der Eindrucksbildung von Fiske und Neuberg (1990)

Das von Fiske und Neuberg postulierte nachfolgend beschriebene Kontinuummodell der Eindrucksbildung (1990) verknüpft kategorienbasierte Informationsverarbeitungsprozesse mit individuierenden Informationsverarbeitungsprozessen (Zebrowitz, 1996; Fiske, Lin & Neuberg, 1999).

Die Informationsverarbeitungsprozesse, die zur Bildung eines Eindrucks von einer Person ablaufen, können nach diesem Modell auf einem Kontinuum dargestellt werden. Auf dem

Kontinuum wird abgebildet, in wie weit eine Person beim Prozess der Eindrucksbildung aufgrund der Zugehörigkeit zu einer Kategorie urteilt bzw. wie sehr individuelle Eigenschaften der Zielperson einbezogen werden. An einem Ende des Kontinuums befinden sich die kategorienbasierten Informationsverarbeitungsprozesse. Diese Informationsverarbeitungsprozesse haben die Beurteilung einer Zielperson anhand sozialer Kategorien, z.B. „Frau“ oder „Afrikaner“ zur Grundlage. Bei der Eindrucksbildung aufgrund der Zuordnung der Zielperson zu bestimmten sozialen Kategorien werden individuelle Eigenschaften wie „ehrlich“ oder „kreativ“ größtenteils außer Acht gelassen, so dass es zu einer kategorienbasierten Beurteilung der Zielperson kommt. Am anderen Ende des Kontinuums befinden sich die individuierenden Informationsverarbeitungsprozesse, die die Zugehörigkeit der Zielperson zu einer sozialen Kategorie vernachlässigen. Als Ergebnis der individuierenden Informationsverarbeitungsprozesse kann dann ein Urteil über die Zielperson entstehen, das die individuellen Persönlichkeitsmerkmale und Eigenschaften berücksichtigt.

Der von Fiske & Neuberg (1990) geprägte Begriff des „Kontinuums“ im Modell unterscheidet die Eindrucksbildungsprozesse (d.h. kategorienbasiert versus individuierend) nicht im Sinne von distinkten Stufen, sondern geht von graduellen Unterschieden aus. Die Eindrucksbildung von einer Zielperson wird also nicht nur durch eine Art von Informationsverarbeitungsprozessen abgebildet, sondern vielmehr sind beide – kategorienbasierte und individuierende Informationsverarbeitungsprozesse – an der Eindrucksbildung beteiligt. Der Einfluss, welcher Informationsverarbeitungsprozess sich bei der Eindrucksbildung stärker auswirkt, wird durch die dargebotene Information über die Zielperson und von der Motivation des Wahrnehmers bestimmt. Hinsichtlich der dargebotenen Information über die Zielperson finden Informationen über ihre Zugehörigkeit zu einer bestimmten sozialen Kategorie eher in kategorienbasierten Informationsverarbeitungsprozessen ihren Niederschlag. Die Darbietung von inkonsistenten Personinformationen führt dagegen eher zu individuierenden Prozessen. Fiske und Neuberg (1990) postulieren, dass die Motivation des Wahrnehmers eine entscheidende Rolle auf die Anwendung von individuierenden Informationsverarbeitungsprozessen bei der Eindrucksbildung spielt. Das Ziel des Wahrnehmers, sich einen möglichst genauen Eindruck von der Zielperson zu bilden, kann dazu führen, dass er eine Einordnung in Kategorien bewusst vermeidet und zu einem individuellen Eindruck von der Zielperson gelangen kann (Fiske und Neuberg, 1990).

Eine Annahme des Modells ist, dass kategorienbasierte Informationsverarbeitungsprozesse Priorität gegenüber individuierenden Informationsverarbeitungsprozessen haben. Um zu einem Eindruck von einer Zielperson zu gelangen, versucht ein Wahrnehmer zuerst die Kategorienzugehörigkeit zu nutzen, bevor er versucht, sich einen genauen Eindruck zu bilden, der auf den individuellen Eigenschaften der Zielperson beruht. Gelingt die Einordnung in eine zugehörige soziale Kategorie und ist der Wahrnehmer nicht weiter motiviert, sich einen genaueren Eindruck von der Zielperson zu verschaffen, wird der Eindrucksbildungsprozess an dieser Stelle gestoppt und es finden keine weiteren Informationsverarbeitungsprozesse mehr statt. Nur, wenn die Zuordnung zu einer bestimmten sozialen Kategorie nicht gelingt, z.B. aufgrund inkonsistenter Informationen, nutzt der Wahrnehmer individuierende Informationsverarbeitungsprozesse, um sich einen Eindruck von der Zielperson zu bilden und zwar auch dann, wenn er nicht die Absicht hatte, sich einen genauen Eindruck über die Zielperson zu verschaffen.

Für die unterschiedlichen Informationsverarbeitungsprozesse postulieren Fiske und Neuberg (1990) Aufmerksamkeit als notwendigen Mediator. Ein Wahrnehmer muss ein erhöhtes Maß an Aufmerksamkeit auf die individuellen Eigenschaften einer Zielperson aufwenden, um einen individuellen Eindruck und einen nicht kategorienbasierten Eindruck von einer Zielperson zu erhalten.

3.4. Der Prozess der Eindrucksbildung nach Fiske und Neuberg (1990)

Fiske und Neuberg (1990) postulieren für das Kontinuummodell der Eindrucksbildung die nachfolgende Sequenz von Eindrucksbildungsprozessen: Anfangskategorisierung (initial categorization), bestätigende Kategorisierung (confirmatory categorization), Rekategorisierung (recategorization) und sukzessive Integration individueller Eigenschaften (piecemeal integration). In jeder dieser Verarbeitungsstadien finden kategorienbasierte und/oder individuierende Informationsverarbeitungsschritte statt, die nachstehend erläutert werden sollen.

Wenn der Wahrnehmer eine (unbekannte) Zielperson wahrgenommen hat, wird sie sofort kategorisiert. Diese erste Kategorisierung (initial categorization) erfordert keine Aufmerksamkeit und benötigt keine kognitiven Ressourcen. Diese Anfangskategorisierung leitet die ersten Reaktionen des Wahrnehmers auf die Zielperson (Fiske und Neuberg, 1990). Der Wahrnehmer prüft dann, ob die Zielperson für ihn in irgendeiner Weise relevant ist. Ist

dies nicht der Fall, wird der Eindrucksbildungsprozess zu diesem Zeitpunkt beendet. Die Zielperson wird beeinflusst durch die aktivierte Anfangskategorie beurteilt. Die Anfangskategorisierung ist ausschließlich durch kategorienbasierte Informationsverarbeitungsprozesse gekennzeichnet.

Schätzt der Wahrnehmer ein, dass die Zielperson für ihn wenigstens minimal relevant ist, wendet er seine Aufmerksamkeit ihren individuellen Eigenschaften zu. Wie bereits geschrieben, wird die Aufmerksamkeit sowohl durch die präsentierte Information über die Zielperson als auch durch die Motivation des Wahrnehmers determiniert (Fiske und Neuberg, 1990). Wenn der Wahrnehmer seine Aufmerksamkeit nun auf die individuellen Eigenschaften der Zielperson richtet, moderiert die Aufmerksamkeit, ob eher kategorienbasierte oder eher individuierende Informationsverarbeitungsprozesse das Urteil über die Zielperson beeinflussen. Der Wahrnehmer hat nun zwei Möglichkeiten, im Eindrucksbildungsprozess weiter zu verfahren. Der Wahrnehmer wird versuchen, wenn es irgendwie möglich ist, seine Anfangskategorie zu bestätigen. Hierzu wird er sich vorzugsweise auf die Informationen über die Zielperson konzentrieren, die kategorienkonform sind und er wird versuchen, kategorieinkonsistente Informationen abzuwerten, indem er sie z.B. als Zufälle interpretiert. Diese Verarbeitungsstufe wird im Modell bestätigende Kategorisierung genannt. Kann die Anfangskategorie für die Zielperson durch den Wahrnehmer bestätigt werden, bildet er sich einen entsprechenden kategorienbasierten Eindruck von ihr.

Die zweite Möglichkeit ist, dass die Anfangskategorie aufgrund der dargebotenen Informationen nicht bestätigt werden kann, weil z.B. die Informationen, die gegen die gebildete Anfangskategorie sprechen, so dominant sind und nicht geleugnet oder übergangen werden können. Dann ist der Wahrnehmer bemüht, die Rekategorisierung der Zielperson vorzunehmen. Die Zielperson wird nun neu einer Kategorie zugeordnet. Der Wahrnehmer sucht für die Einordnung der Zielperson eine neue Subkategorie, eine völlig neue Kategorie, eine Selbstkategorie oder ein Beispiel. Auch bei der Rekategorisierung sind nach Fiske und Neuberg (1990) die kategorienbasierten Informationsverarbeitungsprozesse vorherrschend und der Wahrnehmer bildet sich einen kategorienbasierten Eindruck von der Zielperson. Wenn auch die Rekategorisierung nicht gelingt, weil eine Kategorisierung nicht möglich ist, versucht der Wahrnehmer, die einzelnen Informationen zu einem individuellen Eindruck der Zielperson zusammenzufügen. An dieser Stelle finden nur noch individuierende Informationsverarbeitungsprozesse statt, die Aufmerksamkeit, kognitive Ressourcen und Zeit

erfordern (Fiske und Neuberg, 1990). Nur aus den individuierenden Informationsverarbeitungsprozessen bildet sich der Wahrnehmer einen individuellen, nicht kategorienbasierten Eindruck von der Zielperson. Der Wahrnehmer beurteilt die Zielperson nun als ein Individuum, unabhängig von Kategorienzugehörigkeiten (z.B. „Sie ist ein intelligenter, ehrlicher und geselliger Mensch.“)

Zu diesem Zeitpunkt prüft der Wahrnehmer, ob er eine weitere Beurteilung der Zielperson vornehmen möchte. Falls ja, sucht er nach weiteren Eigenschaften und Persönlichkeitsmerkmalen der Zielperson und der beschriebene Eindrucksbildungsprozess erfolgt erneut.

3.5. Reduktion der Komplexität bei der Urteilsbildung

Den beschriebenen Eindrucksbildungsmodellen ist gemein, dass bei der Gewinnung eines Eindruckes von einer Zielperson der Einordnung in Urteilkategorien und Stereotypisierung eine zentrale Funktion zukommt. Brewer (1988) und Fiske & Neuberg (1990) interpretieren diese als Heuristiken, die die soziale Informationsverarbeitung leiten. Heuristiken sind Urteilstechniken, die zu einer Reduktion der Komplexität der Urteilsaufgabe beitragen (Tversky & Kahnemann, 1974). Als Faustregeln ermöglichen sie eine Verkürzung der Meinungsbildung (Bierhoff, 2000) und machen es dem Wahrnehmer somit möglich, schnell und ohne hohen Verbrauch an kognitiven Ressourcen einen Eindruck zu gewinnen. Gerade bei der Urteilsbildung in komplexen sozialen Situationen, hat die Verwendung von Urteilsstrategien unter Zeitdruck den Vorteil, dass die Urteilsbildung schnell erfolgt und keinen großen Aufwand erfordert und dass der Wahrnehmer somit schnell reagieren kann.

Unter kognitiver Belastung des Wahrnehmers während der Eindrucksbildung wird er die Zielperson stärker kategorienbasiert beurteilen und stereotypisieren, da er für eine genaue individuelle Eindrucksbildung nicht ausreichend kognitive Ressourcen zur Verfügung hat. Vertreter des sozial kognitiven Ansatzes postulieren hier eine positive lineare Beziehung zwischen kognitiver Belastung und kategorienbasierter Beurteilung einer Zielperson (z.B. Macrae et al. (1993), Gordon & Anderson, (1995)). Sie gehen davon aus, dass unter kognitiver Belastung eine willentliche Kontrolle von Urteilkategorien nicht möglich ist (z.B. Brewer, 1988, Devine, 1989, Fiske & Neuberg, 1990).

Die Anwendung von Urteilsstrategien unter Zeitdruck hat neben dem Vorteil einer schnellen Urteilsbildung, die wenig kognitive Ressourcen erfordert, den Nachteil, dass nicht alle Informationen ausgeschöpft und ausreichend beleuchtet (Arkes, 1991, Bierhoff, 2000) oder wahrgenommene Informationen verzerrt werden (Kanning, 1999; Kanning et. al, 2004).

Die Reduktion der Komplexität in sozialen Urteilsituationen beginnt bereits bei der Wahrnehmung. Nach der Hypothesentheorie der sozialen Wahrnehmung (Bruner & Postman, 1951; Bruner, 1951, Postman, 1951) haben Personen aufgrund von Erfahrungen Erwartungen – Hypothesen – gebildet, die den Wahrnehmungsprozess bei einer Beurteilung massiv beeinflussen können. Der Prozess der hypothesengeleiteten Wahrnehmung läuft schnell und für den Urteiler nicht bewusst ab. Die die soziale Situation leitende Hypothese wird ausgewählt und von der wahrnehmenden Person mit den objektiven Informationen verglichen (Postman, 1951). Oswald (1993) konnte zeigen, dass Urteiler vor allem solche Informationen in der objektiven Situation suchen, die hypothesenkonform sind. Bei Übereinstimmung der Hypothese mit der objektiven Information wird die bestehende Hypothese verstärkt und es ergeben sich für den Urteiler keine Schwierigkeiten bei der Integration neuer Informationen. Sind die objektiven Informationen jedoch hypothesenkonträr, versucht der Beurteiler nach der Hypothesentheorie (Bruner & Postman, 1951), diese Diskrepanz dadurch zu überwinden, dass er die objektive Information an die Hypothese anpasst, indem er diese wenn möglich ignoriert, die Ursache für diese Information nicht der zu beurteilenden Person zuschreibt (externale Attribution) oder diese Informationen bei der Urteilsbildung einfach weniger stark gewichtet.

Auch eine Untersuchung von Kanning & Leisten (2004) brachte das Ergebnis, dass unter hoher kognitiver Belastung die Komplexität der vorhandenen Informationen stark reduziert wurde. Den Versuchspersonen wurden Filmszenen aus einer Übung eines Assessment Centers dargeboten. Sie hatten die Aufgabe, die Bewerber auf drei Dimensionen zu beurteilen. Die Stimulusinformation variierte zwischen geringer kognitiver Belastung – hier sollte nacheinander die Leistung von 4 Bewerbern beurteilt werden – und hoher kognitiver Belastung – hier sollten die 4 Bewerber gleichzeitig in einer Gruppendiskussion bewertet werden -. Die Videoaufnahme wurde entweder vollständig (Ton und Bild) dargeboten oder es wurde nur das bewegte Bild dargeboten, wobei der Ton des Films ausgeschaltet war. In der Bedingung der geringen kognitiven Belastung unterscheiden sich die Urteile auf den drei Dimensionen signifikant. In der Bedingung der hohen kognitiven Belastung jedoch kommen

die Urteiler in beiden Stimulusvariationen zu denselben Urteilen. Kanning und Leisten (2004) interpretieren diese Ergebnisse so, dass unter hoher kognitiver Belastung in beiden Stimulusvariationen (Bild und Bild mit Ton) die nonverbalen Informationen im Sinne einer Urteilsheuristik zur Eindrucksbildung herangezogen werden. Die Verarbeitung visueller Informationen scheint weniger kognitive Ressourcen zu binden als die Verarbeitung verbaler Informationen. Das Urteil der Beobachter wird unter hoher kognitiver Belastung vollständig durch die nonverbalen Informationen der Bewerber erklärt. Dies entspricht einer sehr groben Reduzierung der Komplexität der tatsächlich dargebotenen Informationen.

Die Reduzierung der Komplexität der Vielzahl an dargebotenen Informationen und Entscheidungsaufgaben ist auch in der sozialen Urteilsbildung ein zentraler Mechanismus (Strack, 2000). Wenn man bedenkt, wie viele Urteilsbildungen und Bewertungen jeder Mensch im täglichen Leben vornimmt, muss die Urteilsbildung als ein hoch effektiver Prozess angesehen werden, um eine kognitive Überlastung zu vermeiden. Bereits Nisbett und Ross (1980) (zitiert nach Strack, 2000) vertraten die These, dass „Urteilsvereinfachungen unter den Einschränkungen des Alltags eher die Regel als die Ausnahme darstellen“ (S. 230).

3.6. Vergleichsprozesse in der Urteilsbildung

Neben der im vorigen Abschnitt beschriebenen Komplexitätsreduktion bei der Urteilsbildung als einer der zentralen Mechanismen in komplexen Situationen werden in der sozialen Urteilsbildung vergleichende Prozesse als Kernmechanismen betrachtet (Mussweiler, 2003a, Mussweiler et al., 2004b). Jedes Urteil ist relativ und bezieht sich auf einen Vergleich zwischen der Zielperson und einem zugehörigen Standard (Mussweiler, 2003a). Spontane Vergleiche sind in der Urteilsbildung hocheffiziente, allgegenwärtige Prozesse (Mussweiler et al., 2004a). Sie laufen relativ automatisch ab, um den Beurteiler in die Lage zu versetzen, die ständigen Vergleiche vorzunehmen, ohne seine knappen kognitiven Ressourcen aufzubrauchen (Mussweiler et al., 2004a). Jeder Urteilsprozess scheint von den involvierten Vergleichen beeinflusst zu sein, unabhängig davon, ob explizit danach gefragt wurde, oder ob sie spontan vorkommen (Mussweiler, 2003a). Beispielsweise hängt die Einschätzung, wie feindselig eine Zielperson wahrgenommen wird, davon ab, ob man sie mit dem Papst Johannes Paul oder mit Adolf Hitler vergleicht (Herr, 1986).

Mussweiler (2003a, b) postuliert zwei alternative Vergleichsmechanismen: einerseits Ähnlichkeit (similarity) und andererseits Unähnlichkeit (dissimilarity). Der Test auf Ähnlichkeit beinhaltet die selektive Suche nach Hinweisen, die beinhalten, dass das Ziel Ähnlichkeit mit dem Standard hat und führt zur Assimilation. Der Test auf Unähnlichkeit beinhaltet die selektive Suche nach Beweisen, die darauf hinweisen, dass das Ziel unähnlich zum Standard ist und führt zum Kontrast. Dunning & Hayes (1996) fanden Belege dafür, dass Menschen beim Beurteilen von Verhalten von anderen ihr eigenes Verhalten als Vergleichsnorm für das Bewerten heranzogen. 71 % ihrer Probanden berichteten, dass sie ihr eigenes Verhalten während der Beurteilung zum Vergleich nutzten. Als weiteren Hinweis darauf, dass eigenes Verhalten als Vergleichsstandard aktiviert und herangezogen wurde, werteten sie, dass die Probanden eigenes Verhalten nach dem Beurteilen vom Verhalten anderer schneller erinnerten (Dunning & Hayes, 1996).

Auch Mussweiler, Rüter & Epstude (2004b) fanden, dass soziale Vergleiche stark davon beeinflusst werden, wie sich Menschen selbst definieren, sehen und bewerten. Je nachdem, ob sie denken, dass sie selbst arm oder reich, klug oder dumm sind, so scheint diese Bewertung einen primären Einfluss darauf zu haben, wie sie andere einschätzen (Mussweiler et al., 2004b). Welcher Vergleichsstandard also bei der Beurteilung von anderen Personen ausgewählt wird, hängt zum großen Teil davon ab, wie sich Beurteiler selbst wahrnehmen und folglich, was sie denken, wer sie selbst sind (Mussweiler et al., 2004b).

Die dargelegten Befunde liefern somit auch eine mögliche Erklärung dafür, dass verschiedene Urteiler ein und dasselbe Verhalten einer Person unterschiedlich einschätzen und zu einer anderen Beurteilung der Zielperson kommen, da sie bei der Beurteilung möglicherweise andere Vergleichsstandards heranzogen.

4. Einfluss der Stimulusinformation

Im täglichen Leben beurteilen wir ständig Personen, z.B. den Partner, die Kollegin, die Nachbarin von nebenan; selbst von Passanten, denen wir zuvor noch nie begegnet sind, bilden wir zuweilen ein Urteil. Urteile bilden wir auf der Grundlage unterschiedlichster Medien: bei der Begegnung von Personen in sozialen Situationen, aufgrund von schriftlichem Texten, z.B. beim Lesen eines Textes oder einer Bewerbung, aufgrund von Bildern, z.B. beim Betrachten einer Fotoausstellung mit Portraitaufnahmen oder beim Sehen von Filmen.

Wie beeinflussen unterschiedliche Medien die soziale Urteilsbildung? Dieser Frage soll im folgenden Abschnitt nachgegangen werden.

4.1. Stimulusinformation: Text

4.1.1. Personenbeschreibende Information als Textverstehen

Verbale (schriftliche) Personenbeschreibungen haben im beruflichen und privaten Alltag einen zentralen Platz. Die Informationen über eine Person werden in indirekter Form verbal, durch eine Beschreibung als Text dargestellt. Texte werden als satzübergreifende sprachliche Sinnzusammenhänge, deren einzelne Teile thematisch und syntaktisch untereinander verknüpft und aufeinander bezogen sind, bezeichnet. Um einen Text zu repräsentieren, nutzt der Leser phonologische, lexikalische, syntaktische, semantische und pragmatische Informationen (Krolak-Schwerdt & Wintermantel, 2004). Der wohl bedeutsamste Befund in der Textverstehensforschung für die vorliegende Fragestellung ist der, dass die in einem Text verbundenen Aussagen auch in der Gedächtnisrepräsentation miteinander verknüpft sind und dass diese nicht isoliert gespeichert werden. Dies führt zu einer Steigerung der Leistung in der Wiedergabe eines Textes im Vergleich dazu, wenn nur eine Liste von Aussagen dargeboten wird (van den Broek, 1994).

Personenbeschreibungen, als eine spezielle Textart, sind dadurch gekennzeichnet, dass bestimmte Textelemente durch kausale konnektive Präpositionen („weil“) miteinander verbunden sind. Beispielsweise ist dies bei der Angabe von Zielen oder Motiven für Handlungen des Protagonisten der Fall. Aufgrund dieser spezifischen kausalen Verbindungen wird ein besonderer Typ von mentalen Modellen beim Leser aufgebaut, der häufig als „kausales Modell“ bezeichnet wird (Fletcher, 1994). Untersuchungen (Fletcher, 1994) deuten darauf hin, dass kausale Verbindungen in dem mentalen Modell bevorzugt repräsentiert werden. Leser reproduzieren kausal verknüpfte Elemente besser (van den Broek, 1994) und sie werden als wichtiger als andere Aussagen eingeschätzt (Trabasso & van den Broek, 1985).

In Untersuchungen zum Erinnern von Personenbeschreibungen konnten Wintermantel und Christmann (1983) herausfinden, dass beim Wiedergeben einer Personenbeschreibung eine Tendenz zur Zuschreibung von dauerhaften Merkmalen besteht. Probanden formulierten in

der Wiedergabe und auch in der Zusammenfassung Persönlichkeitseigenschaften, die sie aus im Text vorhandenen Verhaltensbeschreibungen generierten. Das Verstehen von Personenbeschreibungen ist unmittelbar mit der Urteilsbildung verbunden. Es ist wohl kaum möglich, eine Personbeschreibung zu lesen und zu verstehen, ohne ein Urteil über die beschriebene Person zu generieren in dem Sinne, dass vom Rezipient des Textes von Verhaltensweisen auf Persönlichkeitseigenschaften geschlossen wird (Krolak-Schwerdt & Wintermantel, 2004).

In einer weiteren Studie konnte Wintermantel (2002) feststellen, dass die Textstruktur für die mentale Repräsentation der beschriebenen Person von Bedeutung ist. Ist der Text der Personenbeschreibung hochgradig thematisch hierarchisch strukturiert, wird er besser behalten, als wenn er thematisch niedrig hierarchisch strukturiert ist. Bevorzugt reproduziert werden thematisch zentrale Propositionen, die auf unterschiedliche Bereiche der Personenbeschreibung hinweisen, z.B. Biographische Daten, Fähigkeiten und Einstellungen (z.B. Livesley & Bromley, 1973).

Den Einfluss einer ausgewählten textuellen Variable, der direkten Kohärenz, auf die soziale Informationsverarbeitung untersuchten Wintermantel und Krolak-Schwerdt (2002). Sie konnten zeigen, dass die direkte Textkohärenz sowohl auf die Behaltensleistung als auch auf die Anzahl und Komplexität von personbezogenen Attributionen einen Einfluss hat. Sie stellten fest, dass kausale und konzessive Verknüpfungen im Text diese Effekte hervorrufen, dass jedoch die einfache Verknüpfung durch die Konjunktion „und“ keinen Einfluss auf die Behaltensleistung oder die Anzahl von personbezogener Attributionen hat. Auch in einer weiteren Experimentalserie konnten Krolak-Schwerdt und Wintermantel (2002) die Bedeutsamkeit kausaler Verknüpfungen in personbeschreibenden Texten herausstellen. Kausal verknüpfte Aussagen werden von Rezipienten schneller gelesen als gleiche Aussagen ohne Verknüpfung oder mit einer Verknüpfung durch die Konjunktion „und“. Sie interpretieren diese Ergebnisse so, dass also kausale Satzstrukturen das Verstehen beim Lesen von Texten erleichtern.

Diese Befunde erlauben den Schluss, dass es für die Verarbeitung einer Personbeschreibung von großer Bedeutung ist, wie die Aussagen über die Person verbal dargeboten werden. In einer Literaturanalyse, in die 433 Arbeiten eingingen, kam Krolak-Schwerdt (2000) zu dem Ergebnis, dass ein Grossteil der Untersuchungen, in denen es um Personbeschreibungen geht,

eine Liste von sprachlich nicht verbundenen personbeschreibenden Attributen verwendet. Weit mehr als die Hälfte der gesichteten Untersuchungen (275) beschränkte sich auf die Vorgabe von Eigenschaftsbegriffen und/oder kurze Verhaltensbeschreibungen in Form einer Liste. Als Begründung für dieses experimentelle Vorgehen werden das hohe Maß an Kontrolle über das Versuchsmaterial, die möglichen vielfältigen semantischen Manipulationen und die gute und ökonomisch einfache Umsetzbarkeit im Labor angeführt. Auch in Anbetracht der oben zitierten Befunde erscheint es jedoch fraglich, inwieweit allein mit dem Einsatz dieses Stimulusmaterials valide Befunde gewonnen werden können. Denn wie bereits oben ausgeführt, sind die kognitiven Repräsentationen aufgrund von Textbeschreibungen von Personen dadurch bestimmt, in welcher semantischen Beziehung die im Text enthaltenen Einzelmerkmale stehen und wie die Aussagen miteinander verknüpft sind (Krolak-Schwerdt & Wintermantel, 2004).

Wie und wann kognitive Repräsentationen beim Lesen von Texten gebildet werden, wird im folgenden Abschnitt dargelegt.

4.1.2. Situationsmodelle beim textverstehenden Lesen

Lesen wir einen interessanten (natürlichen) Text, kann es geschehen, dass aus den wenigen Worten und Sätzen, die der Text enthält, im Kopf des Lesers eine bildhafte und lebendige Vorstellung des Gelesenen entsteht, die viel mehr Details enthält, als im Text zu lesen waren. Wie in einem Film laufen die beschriebenen Ereignisse und Handlungen klar und lebendig ab. Geht es in dem Text um Personen, erwachen die Protagonisten zum Leben, man bildet eine Vorstellung über diese Personen und man kann, wenn es gewünscht wird, ein Urteil über diese Personen abgeben. Dieses Phänomen ist eine Form des „Situationsmodelles“ (van Dijk & Kintsch, 1983), welches die zentrale Annahme enthält, dass das „Verstehen“ eines Textes als das Herstellen einer mentalen Repräsentation angesehen werden kann (u.a. Zwaan, Langston & Graesser, 1995, Zwaan & Radvansky, 1998). Bezüglich dieser Repräsentationen herrscht Einigkeit darüber, dass sie die Repräsentation in drei Ebenen gliedert (Engelkamp & Pechmann, 1993). Auf der ersten, der Oberflächenebene, wird der Wortlaut des Textes repräsentiert. Die zweite Ebene, die so genannte Textbasis, enthält die Bedeutungsstruktur des Textes unabhängig vom Wortlaut. Die dritte Ebene – das Situationsmodell – wird bei einem tieferen Verständnis des Textes gebildet. In einem Situationsmodell wird die vom Text

beschriebene Situation meist viel detailreicher repräsentiert, als diese im Text mit wenigen Worten charakterisiert wurde. Um eine solche reichhaltige Repräsentation zu erzeugen, müssen die Informationen aus dem schriftlichen Material mit Vorwissen aus dem Langzeitgedächtnis verknüpft werden. Der ursprüngliche Text wird also um Informationen, sog. Inferenzen ergänzt, so dass das Situationsmodell Informationen aus sehr unterschiedlichen Quellen enthalten kann und oftmals zumindest in Teilen weit über den gelesenen Text hinausgeht (Rinck, 2000).

Situationsmodelle weisen eine Reihe besonderer Eigenschaften auf: zum einen werden bei der Generierung Textinformationen mit dem Vorwissen durch Inferenzen verknüpft (Singer, 1994). Zum anderen sind sie multidimensional. Johnson-Laird (1983) definiert Situationsmodelle als multidimensionale mentale Repräsentationen so: „A dynamic model of, say, a football game calls for a temporal sequence of events at various locations, for causal relations between the events, and socially, governed by physical laws and constrained by the “laws” of the game and social conventions and motivated by various intentions” (S. 414). Auch Zwaan & van Oostendorp (1993) postulieren Situationsmodelle als mehrdimensionale Repräsentationen, die Vorwissen und Textinformation integrieren und darüber hinaus Interaktionen zwischen den verschiedenen Dimensionen abbilden. Situationsmodelle sind darüber hinaus flexibel und zielabhängig: die Art und der Auflösungsgrad der repräsentierten Info können sehr unterschiedlich sein. Zu den verantwortlichen Faktoren liegen bisher kaum Befunde vor. Im Unterschied zu Schemata und Skripts, die die generellen Eigenschaften von Personen, Objekten, Situationen oder Handlungen beinhalten, sind Situationsmodelle episodisch: Situationsmodelle repräsentieren spezifische Situationen und ihre Veränderungen.

Hinsichtlich der Informationen, die in Situationsmodellen repräsentiert werden können, besteht grundsätzlich keine Einschränkung. Leser ziehen jedoch keineswegs alle Inferenzen, die aufgrund des Textes und des spezifischen Vorwissens gezogen werden könnten, da dies die Kapazität des kognitiven Systems übersteigen würde (de Vega, 1995). Nach Morrow (1994) werden Situationsmodelle nur gebildet, wenn die Bildung den Leserzielen dient. Auch Foertsch und Gernsbacher (1994) fanden, dass Probanden beim Textlesen nur genau diejenigen Inferenzen bildeten, die zur Beantwortung der jeweiligen Testfragen erforderlich waren.

Die Befunde zur Frage, welche Informationen in Situationsmodellen repräsentiert sind, zeigen, dass unterschiedliche Arten von Informationen repräsentiert werden, z.B. räumliche und zeitliche Aspekte der Situation sowie Charaktereigenschaften, Emotionen und Überzeugungen von handelnden Personen. Zwaan et al. (1995) haben hierzu ein Modell vorgeschlagen, in dem die fünf Dimensionen Personen, Ziele, Kausalität, Raum und Zeit spezifiziert werden, während zahlreiche andere Untersuchungen sich darauf beschränken, nur eine Dimension experimentell zu untersuchen. Die Repräsentation räumlicher Relationen ist dabei die am gründlichsten untersuchte (z.B. Glenberg et al., 1987, Rinck & Bower, 1995).

Albrecht und O'Brien (1993) zeigten, dass auch Charaktereigenschaften und Überzeugungen von Protagonisten in Situationsmodellen repräsentiert werden. Gernsbacher, Goldsmith und Robertson (1992) erbrachten Befunde, dass auch die Emotionen von Hauptpersonen eines Textes in Situationsmodellen repräsentiert werden können. Leser von erzählenden Texten können durch das Identifizieren, Inferieren und das weitere Verfolgen von Handlungszielen der im Text auftretenden Personen eine kausale Struktur in Situationsmodellen aufbauen (Dutke, 1998). In diesem Zusammenhang werden Handlungsziele als Ursachen für beschriebene Handlungen betrachtet. Der Leser überprüft die Zielerreichung fortlaufend, so dass sich die kausale Struktur des Situationsmodells verdichtet. Das Bestreben nach Verdichtung soll nach Postulation von Glenberg et al. (1987) dafür verantwortlich sein, dass der Protagonist normalerweise im Vordergrund des Situationsmodells repräsentiert bleibt. Auch andere Untersuchungen belegen, dass der Leser seine Aufmerksamkeit bevorzugt den Lokomotionen der primär handelnden Personen widmet (Morrow et al., 1989, Rinck & Bower, 1995). Nieding & Ohler (1999) konnten zeigen, dass schon 6-jährige Kinder mentale Modelle aufbauen und dass die Kinder den Aufbau mentaler Modelle an globalen Textinferenzen, speziell an Protagonistenzielhierarchien ausrichten. Während es einige Untersuchungen zum Handlungsverlauf von Protagonisten in Situationsmodellen gibt, ist keine Untersuchung bekannt, welche sich auf das updating von Protagonisteneinschätzung bezieht.

Zur Frage, unter welchen Umständen Situationsmodelle gebildet werden, fasst Morrow (1994) einige Voraussetzungen zusammen: Zum einen muss der Text über die Situation ausreichende Informationen bereitstellen. Andererseits müssen die Leser über ein ausreichendes Vorwissen verfügen, bleibt die Repräsentation auf die Textbasis beschränkt. Wenn das Vorwissen, wie bei Experten, sehr umfangreich ist, kann der Leser Texte verstehen,

deren oberflächliche Struktur einem Laien das Verständnis sehr erschwert. Als weitere Voraussetzung zu Situationsmodellbildung wird formuliert, dass der Informationsgehalt des Textes und das Vorwissen des Lesers interagieren. Wenn der Textaufbau nicht dem Wissensgebiet des Lesers entspricht, ist es auch für Experten sehr schwer, ein Situationsmodell zu bilden. Schneider und Körkel (1989) fanden beispielsweise, dass „Fußballexperten“ mit geringen verbalen Fähigkeiten einen Text über ein Fußballspiel besser erinnerten als „Fußballlaien“ mit besseren verbalen Fähigkeiten. Dieser Befund wird so interpretiert, dass die „Experten“ weniger Probleme hatten, ein Situationsmodell zu konstruieren, weil sie auf relevante Wissensstrukturen im Langzeitgedächtnis zurückgreifen konnten (Ericsson & Kintsch, 1995), während die Fußballlaien bei ihrer Modellkonstruktion am Nullpunkt beginnen mussten. Auch andere Untersuchungen haben gezeigt, dass die domänenspezifische Expertise beim Textverstehen einer geringeren verbalen Fähigkeit entgegenwirken kann (z.B. Yekovich, Walker, Ogle & Thompson, 1990). Als vierte Voraussetzung für die Situationsmodellbildung wird angeführt (Morrow, 1994), dass der Leser über eine ausreichende kognitive Leistungsfähigkeit verfügen muss. Hierzu gehören z.B. für die Kapazität des Arbeitsgedächtnisses, das Sprachverständnis und räumliches Vorstellungsvermögen (Dutke, 1998). Als letzte Voraussetzung wird – wie bereits weiter oben berichtet - genannt, dass Situationsmodelle nur gebildet werden, wenn sie den Leserzielen dient.

Rinck (2000) fasst aus den vorgenannten Untersuchungen zusammen, dass „es sich bei der Bildung von Situationsmodellen eher um eine Möglichkeit als um ein gesetzmäßig auftretendes Phänomen handelt“ (S. 118).

Wenn auch bei der Erforschung von Situationsmodellen noch viele Fragen offen sind, so bieten sie doch – gerade wegen der postulierten Multidimensionalität - eine gute Möglichkeit, die beim textverstehenden Lesen ablaufenden Prozesse zu erklären. Dutke (1998) verweist in diesem Zusammenhang auch auf die große praktische Relevanz, die das Verstehen von „geschriebenem oder gesprochenem Text allein oder in unterschiedlichen Medienkombinationen hat“ (S. 56).

4.1.3. Urteilsbildung über Protagonisten aus schriftlichen Informationen

An dieser Stelle setzt auch die vorliegende Arbeit an. Inwieweit sind Personen in der Lage, allein anhand einer verbalen Auseinandersetzung von Protagonisten, die in Textform vorliegt, die Protagonisten zu beurteilen? Wenn der gesamte Text aus einem Gespräch zwischen den Protagonisten besteht, erhält der Leser keine unmittelbaren Informationen über etwaige Verhaltensweisen, aus denen eine Schlussfolgerung auf Persönlichkeitseigenschaften erfolgen könnte (Beispielsweise kann sonst aus dem Verhalten „X hilft einer älteren Frau den Einkauf zu tragen“ auf die Eigenschaft „X ist hilfsbereit“ geschlossen werden).

Wird einer Person ein natürlich vorkommender Text (Zwaan & van Oostendorp, 1993) eines Gespräches zwischen Protagonisten zum Lesen vorgelegt, wird der Leser versuchen, aus den Gesprächsinhalten Aussagen herauszufiltern, die etwas über den Sprechenden selbst aussagen bzw. implizit etwas über ihn mitteilen. Das Ergebnis dieses kognitiv aufwändigen Prozesses könnte ein Situationsmodell sein, welches nicht nur die Bedeutung des Textes selbst, sondern auch ein mentales Modell der auftretenden Personen repräsentiert (Rinck, 2000). Soll aufgrund des gelesenen Textes ein Urteil über die Protagonisten gebildet werden, kann der Leser nicht nur auf die gelesenen Informationen, sondern auch auf die durch die Verknüpfung von Textinformationen mit Vorwissen gebildeten Inferenzen zurückgreifen. Das vom Leser konstruierte mentale Modell mit den Repräsentationen der Protagonisten sollte dann als Grundlage für den Eindrucksbildungsprozess herangezogen werden. Die (unbekannte) Zielperson wird in diesem Fall nicht „plötzlich“ wahrgenommen (z.B. wenn ein neuer Gast ein Restaurant betritt), sondern der Beurteiler stellt sich durch eine tiefe Verarbeitung des zu lesenden Textes die Zielperson bildhaft und lebendig vor, so dass dann für die Zielperson die erste Kategorisierung im Sinne des Kontinuummodells der Eindrucksbildung (Fiske & Neuberg, 1990) und die weiteren Informationsverarbeitungsschritte stattfinden können.

Das Herstellen einer mentalen Repräsentation erfordert einen hohen kognitiven Aufwand, da das Verstehen eines Textes kein im normalen Alltag automatisierter oder hoch geübter Prozess ist (Zwaan & Radvansky, 1998), sondern vom Leser eine bewusste Informationsverarbeitung und ausreichende kognitive Leistungsfähigkeit erfordert (Morrow, 1994), so dass allein durch das Lesen und Verstehen des Textes ein Teil der Aufmerksamkeit und der kognitiven Ressourcen des Lesers gebunden sind. Beim Prozess der Eindrucksbildung über eine Zielperson werden durch die anfänglichen kategorienbasierten

Informationsverarbeitungsprozesse keine Aufmerksamkeit und keine kognitiven Ressourcen benötigt (Chaiken & Trope, 1999; Fiske & Neuberg, 1990), so dass die Aktivierung einer Anfangskategorie für die im mentalen Modell repräsentierte Zielperson trotz des erheblichen kognitiven Aufwandes bei der Herstellung der mentalen Repräsentation schnell und ohne großen Aufwand möglich sein sollte. Die willentliche Kontrolle und Inhibierung des ersten kategorienbasierten Eindruckes der Protagonisten sollte unter der kognitiven Belastung nur schwer möglich sein (Fiske & Neuberg, (1990); Devine, (1989), Gordon & Anderson, (1995)), so dass ein Beurteiler versuchen wird, diesen Eindruck aufrecht zu erhalten (Brewer, 1988) bzw. die anfängliche Kategorisierung zu bestätigen (Fiske & Neuberg, 1990). Im Urteil des Lesers über die Protagonisten sollte sich dies niederschlagen, indem eine recht grobe Kategorie oder Merkmalsdimension gebildet und diese mit bestätigenden Beispielen aus dem Text untermauert wird. Die postulierte tiefe Verarbeitung des Textes sollte sich auch in einer guten Erinnerungsleistung des Lesers beim Wiedergeben von Gedächtnisinhalten zum gelesenen Text niederschlagen (Craik & Lockhardt, 1972).

Welche Veränderungen ergeben sich, wenn zusätzlich zum vorgelegten Text von den Protagonisten Fotos vorgelegt werden?

4.2. Stimulusinformation Text mit Bild

4.2.1. Gesicht und Persönlichkeitseindruck – Erste Forschungsansätze

Der Einsatz von Fotos als zusätzliches Stimulusmaterial hat den Vorteil, dass Personen nicht nur einen Eindruck aufgrund des vorgelegten Textes über die Protagonisten bilden, sondern dass sie die Protagonisten selbst auch sehen können. Die ersten Forschungsansätze über die Eindruckswirkung von visuellen Stimuli reichen bis ins 18. Jahrhundert zurück, wurden aber nicht durchgängig weiter verfolgt.

Bereits in den Jahren 1775-1778 verfassten der Pastor Johann Caspar Lavater in Zusammenarbeit mit Johann Wolfgang von Goethe eine Anleitung zur physiognomischen Charakterdeutung mit dem Titel „Physiognomische Fragmente zur Beförderung der Menschenkenntnis und Menschenliebe“. Dieses Werk wurde in ganz Europa als Literatursensation angesehen (vgl. Frey, 1999a, b).

Der Physiker Georg Christoph Lichtenberg wurde zum bedeutendsten Kritiker dieses Werkes. Er kommentierte die seit dem Erscheinen des Werkes von Lavater und Goethe überall praktizierten physiognomischen Übungen mit dem Worten: „Wenn die Physiognomik das wird, was Lavater von ihr erwartet, so wird man die Kinder aufhängen, ehe sie die Taten getan haben, die den Galgen verdienen“ (Lichtenberg, verlegt 1980, S. 532). Im Jahr 1778 veröffentlichte er in dem Göttinger Taschen=Calender einen Aufsatz mit dem Titel „Über Physiognomik; und am Ende etwas über die Kupferstiche des Almanachs“, welcher eine überwältigende Resonanz fand. In der erweiterten Fassung dieses Aufsatzes betonte er, dass es um die Auseinandersetzung mit erkenntnistheoretischen Fragen gehe, die für das Selbstverständnis des Menschen und für das Bild, welches er sich über seine soziale Umwelt mache, von größter Bedeutung sei (Lichtenberg, verlegt 1972). Lichtenberg erkannte, dass die Merkmale des menschlichen Erscheinungsbildes einen Mechanismus auslösen, der eine unwillkürliche Zuschreibung von Eigenschaften vornimmt und dass diese Eigenschaftszuschreibungen häufig zur Vorurteilsbildung führen. Durch diese Aufsätze von Lichtenberg, die durch ihre gesellschaftliche Schockwirkung eine ungewöhnliche Breitenwirkung erzielten (Frey, 1999b), war der Versuch Lavaters, die physiognomische Charakterdeutung als „Wissenschaft der Wissenschaften“ (Lavater, 1775, S. 55) auszurufen, kläglich gescheitert. Schon kurze Zeit darauf war die „Physiognomik, als Ausspähungskunst des Innern im Menschen vermittelt gewisser äußerer unwillkürlich gegebener Zeichen, ganz aus der Nachfrage gekommen und nichts von ihr übrig geblieben (Kant, verlegt 1980, S. 241).

Rund 100 Jahre später machte Herrmann von Helmholtz auf die große Kraft der visuellen Eindrucksbildung aufmerksam. In äußerst knapper Form legte Helmholtz die Wirkmechanismen von visuellen Stimuli in seinem Werk „Handbuch der Physiologischen Optik“ (1856-67) dar. Helmholtz postulierte für die visuelle Eindrucksbildung einen Mechanismus, der seinen eigenen Gesetzen folge und so eine Macht über den menschlichen Verstand ausüben würde und bezeichnete diesen Mechanismus als „unbewussten Schluß“ (1867, zitiert nach Frey, 1999b). Nach Helmholtz läuft der Prozess der visuellen Eindrucksbildung spontan und automatisch ab. An der Entstehung des visuellen Eindruckes seien die Zentren, die eine bewusste Informationsverarbeitung steuern, nicht beteiligt, so dass der Eindruck das Ergebnis eines Deutungsprozesses ist, der für „die reinigende und prüfende Arbeit des bewussten Denkens“ (Helmholtz, 1867, S. 449, zitiert nach Frey, 1999b) unzugänglich sei. Interessant ist ein Ausschnitt aus diesem Werk, in dem Helmholtz die unterschiedliche Eindruckswirkung von verschiedenen Stimuli – das Erleben einer im Theater

gespielten Geschichte im Vergleich zum Lesen eines geschriebenen Berichtes über dasselbe Ereignis – beschreibt: „Wir halten ihn (den Schauspieler – Anm. des Verfassers) für zornig oder für leidend, je nachdem er uns die eine oder andere Art der Mienen und Geberden zeigt; er erregt Schrecken oder Mitleiden in uns... und die begründete Ueberzeugung, dass dies alles nur Schein und Spiel sei, hilft durchaus nichts gegen unsere Gemüthsbewegungen, so lange der Schauspieler nicht aus seiner Rolle fällt. Im Gegenteil ergreift und foltert uns eine solche lügenhafte Geschichte, der wir scheinbar persönlich beiwohnen, viel mehr, als es eine entsprechende wahre Thun würde, von der wir einen trockenen actenmäßigen Bericht lesen“ (Helmholtz, 1867, S. 450, zitiert nach Frey, 1999b). Auch die Ausführungen zum Wirkmechanismus von visuellen Stimuli von Helmholtz blieben in der psychologischen Forschung weitgehend unbeachtet (Gilbert, 1989).

Erst am Anfang des letzten Jahrhunderts nahm das Interesse an der Erforschung des mimischen und gestischen Verhaltens rapide zu, vor allem weil es durch die technologische Entwicklung möglich war, Filme aufzunehmen und abzuspielen. Die Euphorie für die Untersuchung der Körperbewegung ließ jedoch schnell nach, als die empirischen Befunde in streng kontrollierten experimentellen Untersuchungen die Erwartungen, dass mimisches und gestisches Verhalten psychodiagnostisch analysiert werden kann, nicht unterstützten. In den USA orientierte man sich um und nutzte zur Psychodiagnostik nicht weiter die Kriterien Gestik, Mimik, Körperhaltung, sondern die Sprache – die Entwicklung von Fragebögen und Tests als leistungsfähiges psychodiagnostisches Instrumentarium wurde fokussiert (Frey, 1999b). „Das Studium des Gesichtsausdrucks und seiner Interpretation starb buchstäblich in den vierziger Jahren aus“ (Davis, 1979, S. 53, zitiert nach Frey, 1999b).

Erst in letzter Zeit finden sich wieder Untersuchungen, die sich mit der visuellen Eindrucksbildung und der hochselektiven Verarbeitung nonverbaler Stimuli beim Menschen befassen.

4.2.2. Die Rolle des Gesichtes bei bildhaften Darstellungen

Unser Eindruck über eine Person wird zu einem großen Teil durch deren äußere Erscheinung bestimmt, wobei das Gesicht eine zentrale Rolle spielt (Hess, 1998). Ob die Größe des Gesichtsanteils bei fotografischen Darstellungen einen Einfluss auf die Ergebnisse eines

Eindrucksbildungsprozesses haben kann, wird in einigen Studien deskriptiver und experimenteller Art untersucht (z.B. Hinton, 1995). Hierbei wurde der Frage nachgegangen, ob sich für den Betrachter Unterschiede im vermittelten Eindruck ergeben, wenn bei der fotografischen Darstellung ein und derselben Person entweder das Gesicht oder eher der Körper anteilmäßig dominieren. Für diese Untersuchungen wurde der sog. Face-ism-Index - ein Quotient, der bei Persondarstellungen die Distanz zwischen der obersten Stelle des Kopfes und der untersten Stelle des Kinns ins Verhältnis setzt zu der Distanz zwischen der obersten Stelle des Kopfes und des zu unterst dargestellten Punktes des Körpers – herangezogen (Archer, Iritani, Kimes und Barrios, 1983). Ein Vergleich von Fotos von Frauen und Männern aus amerikanischen und europäischen Zeitungen und Zeitschriften ergab, dass Abbildungen von Männern mit einem signifikant größeren Gesichtsanteil als Abbildungen von Frauen dargestellt werden (Archer et al., 1983).

Nach dieser Analyse werden bei Männerfotos etwa zwei Drittel des Bildes vom Gesicht eingenommen, während bei Fotos von Frauen eher der Körper dominierte und das Gesicht auf dem Bild weniger als die Hälfte einnahm. Beim Vergleich von Gemälden aus sechs Jahrhunderten (15. bis 20. Jahrhundert), auf denen Portraits und Selbstportraits gezeichnet waren, kamen Archer et al. (1983) zu dem Ergebnis, dass Männer tendenziell mit einem größeren Kopfanteil dargestellt werden als Frauen. Dies macht deutlich, dass diese unterschiedliche Darstellungsweise zwischen den Geschlechtern schon seit Jahrhunderten vorhanden ist. Es wird postuliert, dass diese Tendenz sogar zugenommen zu haben scheint (Archer et al., 1983).

In einem weiteren Experiment ergab sich ein vergleichbarer Unterschied, als Archer et al. (1983) Versuchspersonen baten, einen Mann oder eine Frau zu zeichnen. Hinsichtlich der Körperproportionen wurden bewusst keine Vorgaben gemacht. Somit hätte der Gesichtsanteil bei den Zeichnungen jeden beliebigen Wert annehmen können. Es fand sich, dass die Probanden in ihren Zeichnungen bei Männern das Gesicht größer und detailreicher darstellten, während bei den Zeichnungen der Frauen der Bereich des Kopfes eher vage dargestellt wurde. Dieser Unterschied zwischen den Geschlechtern in der bildhaften Darstellung wurde im Sinne eines Ausdruckes und des Aufrechterhaltens stereotyper Geschlechterkonzepte interpretiert (Archer et al., 1983).

Einen ähnlichen Unterschied fanden Zuckermann und Kieffer (1994) bei der bildhaften Darstellung von ethnischen Gruppen. In einer Untersuchung der Face-ism-Indizes von Fotos aus amerikanischen und europäischen Printmedien, gemalten Portraits und

Personendarstellungen auf Briefmarken kamen sie zu dem Ergebnis, dass Menschen mit weißer Hautfarbe mit einem signifikant größeren Gesichtsanteil dargestellt wurden als Menschen mit schwarzer Hautfarbe. Dass die Größe des relativen Gesichtsanteils auf einer bildhaften Darstellung einen Einfluss auf die Eindrucksbildung hat, ist unbestritten, jedoch nicht abschließend erforscht (Dirscherl et al., 2002). Archer et al. (1983) legten Versuchspersonen Fotos von Personen vor, auf denen entweder das Gesicht anteilmäßig dominierte oder auf denen neben dem Kopf auch ein Teil des Oberkörpers zu sehen war, und ließen die abgebildeten Personen hinsichtlich verschiedener Persönlichkeitsmerkmale einschätzen. Für die Dimensionen Intelligenz, Ehrgeiz und körperliche Präsenz wurden die dargestellten Personen signifikant positiver beurteilt, wenn den Probanden Fotos mit einem großen Gesichtsanteil vorgelegt wurden. Der Eindruck von Wärme wurde durch die Darstellungsform des Gesichtes nicht geändert.

Diese Effekte konnten in Untersuchungen von Zuckermann und Kieffer (1994) in diesem Umfang nicht gefunden werden. Hinsichtlich der zu beurteilenden Merkmale Positivität, Intelligenz, Ehrgeiz und Attraktivität wurden keine Unterschiede in der Beurteilung gefunden, lediglich bezüglich der Wahrnehmung von Dominanz zeigte sich eine signifikante Bevorzugung von Fotos mit hohem Gesichtsanteil. Personen, die mit einem hohen Gesichtsanteil dargestellt wurden, wirkten auf die Betrachter dominanter. Zuckermann et al. (1994) diskutiert verschiedene Erklärungsmöglichkeiten. Möglicherweise wird die größere Dominanz bei Fotos mit hohem Gesichtsanteil dadurch empfunden, dass ein großes Gesicht größere Nähe signalisiert und dass daher der Betrachter eher den Eindruck gewinnt, sich im Einflussbereich der dargestellten Person zu befinden. Ein groß dargestelltes Gesicht, welches allein durch die Darstellung sehr detailreich wirkt, könnte implizieren, dass die dargestellte Zielperson Herausforderungen nicht scheut und sich diesen stellt, während ein kleineres, undeutlich dargestelltes Gesicht eher mit Assoziationen verknüpft sein könnte, dass die Person Konfrontationen aus dem Weg geht oder diese nicht bewältigt. Dies sind jedoch nur vage Erklärungsversuche, auch Zuckermann et al. (1994) räumt ein, dass nicht eindeutig geklärt ist, ob das Größenverhältnis von Gesicht und Körper auf einer bildhaften Darstellung tatsächlich für den Eindruck von Dominanz ausschlaggebend ist. Als Schlussfolgerung für die vorliegende Arbeit kann aus diesen Ergebnissen gezogen werden, dass bei der Erstellung der Fotos von den Protagonisten darauf geachtet werden sollte, dass der Anteil des Gesichtes auf den Bildern in etwa gleich gehalten wird, um Urteilsunterschiede, die durch den Face-ism-Index bedingt sein könnten, zu minimieren.

Im Folgenden geht es nun um die Frage, wie Persönlichkeitsmerkmale anhand des Gesichtes beurteilt werden.

4.2.3. Das Fünf-Faktoren-Modell der Persönlichkeit – die Big Five

Bei der Frage nach der Eindruckswirkung von Gesichtern wird vor allem das Fünf-Faktoren-Modell der Persönlichkeit – die so genannten Big Five -, welches die Faktoren Extraversion, Soziale Verträglichkeit, Gewissenhaftigkeit, Emotionale Stabilität und Intellekt/Kultur/Offenheit umfasst, angeführt. Eine wachsende Anzahl von Forschern ist der Meinung, dass mit den Big Five auf einer sehr hohen Abstraktionsebene die zentralen Persönlichkeitsdimensionen identifiziert wurden. Hogan (1991) z.B. postuliert, dass den Big Five kognitive Schemata zugrunde liegen, die deshalb Bestandteil unserer psychischen Ausstattung sind, weil sie sich im Verlauf der Evolution als vorteilhaft erwiesen haben. Es ist ihre Funktion, das Verhalten anderer Personen in sinnvolle Kategorien zu ordnen und das soziale Leben überschaubar und vorhersagbar zu machen. Aufgrund der zunehmenden Bedeutung der Big Five gerade in jüngerer Zeit (z.B. Wiggins & Trapnell, 1997) soll auf diesen Ansatz kurz näher eingegangen werden.

Die Wurzeln des Fünf-Faktoren-Modells führen auf den lexikalischen Ansatz der Persönlichkeitsforschung zurück.¹ Der lexikalische Ansatz fußt auf der Annahme, dass alle wichtigen Aspekte der menschlichen Persönlichkeit Eingang in die Sprache gefunden haben. Goldberg (1981) formuliert dies so: „The more important is an individual difference in human transactions, the more languages will have a term for it“ S. 142). Als Ausgangspunkt des lexikalischen Ansatzes dient eine möglichst vollständige Sammlung aller im Lexikon enthaltenen persönlichkeitsrelevanten Wörter, wofür dann eine Taxonomie persönlichkeitsbeschreibender Begriffe entwickelt wird. Im nächsten Schritt wird der Versuch unternommen, die bedeutsamsten Persönlichkeitsdimensionen zu identifizieren, wobei man sich auf eine recht eng begrenzte taxonomische Klasse, vor allem auf den Temperaments- und Charakterbereich, beschränkt (Allport, 1949, Goldberg, 1981, 1982). Nach Arbeiten von Francis Galton (1884), Klages (1926) und Baumgarten (1933), die sich mit dem Vorschlag

¹ An dieser Stelle soll keineswegs ein vollständiger geschichtlicher Abriss der Entwicklung des Fünf-Faktoren-Modells der Persönlichkeit gegeben werden. Vielmehr sollen nur solche geschichtlichen Stationen kurz skizziert werden, die notwendig erscheinen.

der Systematisierung von Persönlichkeitsunterschieden aufgrund der Sprache und der Klassifikation von Charaktereigenschaften beschäftigt, die jedoch lange ohne Folgen für die Forschung blieben, wurde dieser Ansatz von Allport und Odbert (1936) aufgegriffen, die aus einem Lexikon aus dem Jahre 1925 eine Liste von 17953 persönlichkeitsbeschreibenden Begriffen zusammenstellten. Als „True Fathers“ des Fünf-Faktoren-Modells bezeichnet Goldberg (1993) Tupes und Christal (1961), die bei der Analyse von Datensätzen stets „five relatively strong and recurrent factors and nothing more of any consequence „ (S. 14) fanden. Die Faktoren wurden Surgency, Agreeableness, Dependability, Emotional Stability und Culture genannt und konnten mehrfach repliziert werden (z.B. Borgatta, 1964).

Norman (1967) entwickelte anknüpfend an die Liste von Allport und Odbert (1936) ein differenziertes Kategoriensystem für persönlichkeitsbeschreibende Begriffe. Es bestand aus den vier Hauptkategorien (1) Stable Biophysical Traits, (2) Temporary States and Activities, (3) Social Roles, Relationships and Effects und (4) Excluded Categories, wobei jede der Hauptkategorien drei bis fünf Unterkategorien enthielt.

Dieser Ansatz gewann erst ab der zweiten Hälfte der 80er Jahre wieder an Bedeutung. Im Gegensatz zu den Erwartungen der frühen Forscher des Fünf-Faktoren-Modells (Norman, 1963, Tupes & Christal, 1961), dass bei Nutzung einer breiteren Variablenbasis weitere Persönlichkeitsfaktoren gefunden werden können, konnte in einer großen Anzahl von Untersuchungen gezeigt werden, dass „If a large number of rating scales is used and if the scope of the scales is very broad, the domain of personality descriptors is almost completely accounted for by five robust factors“ (Digman & Inouye, 1986, S. 116).

Die fünf identifizierten robusten Persönlichkeitsfaktoren wurden von Goldberg (1981) als Big Five bezeichnet. Die Robustheit der Big Five erweist sich in vielerlei Hinsicht: sie sind unabhängig von Antwortformat, das heißt, die fünf Faktoren lassen sich sowohl mit Hilfe von Items, die als Aussagesätze formuliert sind, als auch bei Adjektivskalen aufzeigen (z.B. McCrae, 1989). Zweitens sind sie unabhängig von der Urteiler- und Beurteiltenstichprobe. Selbstratings und Einschätzungen durch den Ehepartner oder durch Bekannte oder durch psychologisch geschulte Experten führen immer auf dieselbe Faktorenstruktur zurück (z.B. McCrae, 1989, McCrae & Costa, 1987, Goldberg, 1990, 1992). Die Big Five beziehen sich auf relativ überdauernde Persönlichkeitsmerkmale, wie sich in mehreren Langzeitstudien nachweisen ließ (McCrae & Costa, 1990, Costa & McCrae, 1980, 1994). Des Weiteren zeigt sich, dass die Big Five unabhängig von der zugrunde liegenden faktorenanalytischen Technik sind (Vassend & Skrandal, 1995).

Der Inhalt der Big Five wird häufig durch die Notation von Norman (1963)

I. Extroversion or Surgency, II. Agreeableness, III. Conscientiousness, IV. Emotional Stability und V. Culture oder durch die bei Costa und Mc Crae verwandten Faktoren E: Extraversion, A: Agreeableness, C: Conscientiousness, N: Neuroticism und O: Openness to Experience umrissen. Im Deutschen finden sich – wie bereits oben benannt - für diese Faktoren die Bezeichnungen I. Extraversion, II. Soziale Verträglichkeit, III. Gewissenhaftigkeit, IV. Emotionale Stabilität und V. Offenheit (z.B. Henss, 1995). Um auch feiner differenzierende Kategorien mittlerer Hierarchieebenen zu berücksichtigen, beinhaltet jeder der fünf Faktoren eine Reihe von Facetten, z.B. für den Faktor Extraversion die Facetten Warmherzigkeit, Aktivität, Geltungsbedürfnis, Lebensfreude und Kontaktfreude, die jedoch je nach Forschungstradition zum Teil stark differieren (Backhaus, 2004). Das außerordentlich hohe Abstraktionsniveau des Fünf-Faktoren-Modells bietet für viele spezifische Fragestellungen ein zu grobes Raster „the Big-Five domains are located at the highest level that is still descriptive of behavior, with only general evaluation located at a higher and more abstract level...these broad domains incorporate hundreds, if not thousands, of traits“ (Goldberg, 1993, S. 27). Der Konsens zwischen den Anhängern der Big Five nimmt jedoch ab, wenn man stärker differenziert (Henss, 1998).

4.2.4. Das Fünf-Faktoren-Modell bei der Beurteilung von Gesichtern

In umfangreichen Studien zum Persönlichkeitseindruck beim Betrachten von Gesichtern auf Fotos (Henss, 1998) wurden im Anschluss an Norman (1963) zur Erfassung der Daten die Marker-Items Extraversion, Soziale Verträglichkeit, Gewissenhaftigkeit, Emotionale Stabilität und Kultur/Offenheit/Intellekt ausgewählt. Zusätzlich dazu wurden die physische Attraktivität, das Alter, Gesundheit, Geschlechtstypisches Aussehen, Mode, Soziale Effekte und die aktuelle Stimmungslage durch die Versuchspersonen verbal eingeschätzt.

Die Untersuchungen zeigen, dass die Gesichterbeurteilung fremder Personen aufgrund von Fotos zu einer differenzierten, gut interpretierbaren Persönlichkeitsstruktur führt. Die beiden Persönlichkeitsfaktoren Soziale Verträglichkeit und Extraversion und die Attraktivität bilden zusammen die – „Big Three der Eindrucksbildung“ (Henss, 1998, S.162). Dieses Ergebnis deutet Henss (1998) so, dass die Personenbeurteilung zu einem großen Teil im Dienste der Handlungsvorbereitung auf soziale Interaktionen steht, da es bei der Einschätzung einer

fremden Person sehr wichtig ist, nach Hinweisen zu suchen, wie sich eine persönliche Beziehung zu der betreffenden Person voraussichtlich gestalten könnte.

Bei weiterer Differenzierung bildeten die zwei weiteren Faktoren des Fünf-Faktoren-Modells Gewissenhaftigkeit und Intellekt/Kultur/Offenheit eigenständige Einheiten.

Der Faktor Emotionale Stabilität konnte nicht repliziert werden. Möglicherweise genügen die Informationen, die ein Foto bietet, nicht für die Beurteilung der emotionalen Stabilität, sondern es werden weitere Informationen - wie dynamische Hinweisreize – benötigt.

Ein Faktor, der gefunden wurde, der keine Entsprechung in dem Fünf-Faktoren-Modell hat, der jedoch varianzstark und robust ist, ist der Faktor Selbstsicherheit, der als eine Facette der Emotionalen Stabilität interpretiert wird.

4.2.5. Physiognomie und Persönlichkeitseindruck

Bei der Untersuchung der Frage, auf welche unterschiedlichen Klassen von Hinweisreizen die Persönlichkeitseindrücke beim Anblick von Gesichtsfotos zurückzuführen sind, stellt Henss (1998) die Eindruckswirkung physiognomischer Merkmale in den Vordergrund, obwohl auch andere Stimulusmerkmale wie beispielsweise Mimik, Haltung, Kleidung, Bildhintergrund eine gewisse Rolle spielen könnten. Er untersucht die Zusammenhänge, wie physiognomische Merkmale mit Persönlichkeitseindrücken verknüpft sind, unabhängig davon, ob der gewonnene Persönlichkeitseindruck richtig oder falsch ist. Insbesondere beschäftigt ihn in diesem Kontext die Frage, ob physiognomische Merkmale in systematischer Weise kovariieren und zu einer überschaubaren Anzahl physiognomischer Faktoren zusammengefasst und identifiziert werden können. Die Untersuchungsergebnisse, welche mehrfach repliziert wurden, waren überwiegend negativ. Die Kovariation zwischen den physiognomischen Skalen war so unsystematisch, dass keine gut interpretierbaren und reproduzierbaren globalen physiognomischen Faktoren beschrieben werden konnten, unabhängig davon, ob individuelle Urteile oder Gruppenstandards zugrunde gelegt wurden. Die wenigen signifikanten linearen Zusammenhänge zwischen Gesichtsmerkmalen und Persönlichkeitsmerkmalen, die gefunden wurden, waren meist nicht sehr eng und nicht replizierbar. Die aufgezeigten multiplen Zusammenhänge zwischen physiognomischen Skalen und einem Persönlichkeitsfaktor konnten ebenfalls nur in seltensten Fällen repliziert werden. Die Ursache für diese Ergebnisse sieht Henss darin, „dass die physiognomischen Merkmale nur schwach und in unsystematischer Weise korrelieren“ (Henss, 1998. S. 257). Aufgrund

dieser Ergebnisse sollten Zusammenhänge zwischen physiognomischen Merkmalen und Persönlichkeitseindrücken nicht oder nur mit größter Vorsicht und Zurückhaltung interpretiert werden.

Für die vorliegende Arbeit heißt dies, dass schon allein aufgrund der Vorlage von Fotos ein differenzierter Persönlichkeitseindruck beim Beurteiler entstehen sollte (Henss, 1998), der es ermöglicht, bestimmte Persönlichkeitseinschätzungen vorzunehmen. Eine Vorhersage, welche spezifischen Merkmale eines Gesichtes welchen Eindruck beim Betrachter hinterlassen, ist jedoch in keiner Weise möglich.

4.2.6. Geschwindigkeit der visuellen Eindrucksbildung

Dass das Betrachten von Bildern mit Gesichtern beim Betrachter einen differenzierten Eindruck der Persönlichkeit hervorruft, wurde oben ausführlich beschrieben (Henss, 1998). Innerhalb welcher kurzer Zeit nach der Vorlage eines Fotos ein visueller Eindruck gebildet wird, und den Betrachter in die Lage versetzt, eine Meinung über die Persönlichkeitseigenschaften der Stimulusperson zu bilden, konnte Jendraczyk (1991) zeigen. Sie legte einer Gruppe von Beurteilern 16 Bewerbungsfotos vor und variierte die Darbietungszeit. Bereits eine Darbietungszeit von 250 Millisekunden reichte aus, um ein differenziertes Bild der Stimuluspersonen entstehen zu lassen. Die Versuchspersonen waren z.B. in der Lage, zu entscheiden, ob jemand als „autoritär“, „sympathisch“, „gefühlbetont“, „intelligent“, „langweilig“ eingeschätzt wird. Die Versuchspersonen konnten bei dieser kurzen Darbietungszeit weiterhin angeben, ob sie die Stimulusperson gern als „Kollegen“, „Vorgesetzten“, „Partner“ oder „Bekanntem“ haben möchten. Die hoch selektive Verarbeitung visueller Stimuli wird durch Ergebnisse untermauert, die zeigen, dass sich die in Bruchteilen einer Sekunde entstandenen Eindrücke kaum ändern, wenn man den Urteilern die Möglichkeit gibt, die Bilder beliebig lange anzusehen. Jendraczyk (1991) konnte zeigen, dass die Eindrucksurteile bei einer Darbietungszeit von 250 Millisekunden mit den Urteilen hoch korrelierten ($r=.92$), die von den Probanden nach einer beliebig langen Expositionszeit abgegeben wurden. Diese Ergebnisse konnten von Frey (1999a) repliziert werden, der Probanden 45 Politikerportraits beurteilen ließ. Auch er konnte zeigen, dass die Persönlichkeitsurteile, die nach einer Darbietungszeit von 250 Millisekunden abgegeben wurden, mit den Urteilen, bei welchen die Probanden die Bilder permanent betrachten konnten, korrelierten ($r=.89$).

4.2.7. Folgerungen für die Beurteilung von Protagonisten eines Textes bei zusätzlicher Vorlage eines Bildes

Wie im vorigen Abschnitt beschrieben, entsteht beim Betrachten von Bildern innerhalb kürzester Zeit ein differenzierter Eindruck über die abgebildete Person (Jendraczyk, 1991; Frey, 1999a). Eine Person, die neben einem Text Bilder der Protagonisten vorgelegt bekommt, wird zuerst die Bilder betrachten und sofort – noch vor dem Lesen des Textes, das eine tiefere Verarbeitung erzwingen und damit kognitiven Ressourcen erfordern würde - eine anfängliche Kategorisierung, die keine großen kognitiven Ressourcen und Aufmerksamkeit erfordert (Chaiken & Trope, 1999; Fiske & Neuberg, 1990), vornehmen. Die für die erste Kategorisierung notwendige Identifikation der Zielperson (Brewer, 1988; Fiske & Neuberg, 1990) erfolgt hier durch das Betrachten der Fotos der Protagonisten. Beginnt die Person nun mit dem Lesen des Textes wird sie versuchen, die Anfangskategorie zu bestätigen und diejenigen Informationen über die Zielperson aus den Gesprächsinhalten herauszufiltern, die kategorienkonform sind (Fiske & Neuberg, 1990). Der Beurteiler wird gemäß der Hypothesentheorie der Wahrnehmung Informationen aus dem Text, die nicht zum ersten aus dem Bild gebildeten Eindruck des Protagonisten passen, weniger gewichten und somit die aufgrund der Fotos gebildete Anfangskategorie durch Informationen, welche aus dem Text stammen, bestätigen (Oswald, 1993). Der Eindrucksbildungsprozess ist an dieser Stelle beendet und weitere Informationsverarbeitungsprozesse beeinflussen die Urteilsbildung nicht mehr.

Eine Rekategorisierung, wie im Abschnitt „Der Prozess der Eindrucksbildung nach Fiske und Neuberg, 1990“ beschrieben, sollte nicht stattfinden, da dieses Vorgehen nur dann gewählt wird, wenn der Beurteiler die gewählte Anfangskategorie aufgrund der Textinformationen nicht bestätigen kann, weil diese so stark und objektiv sind, dass sie nicht mehr als Zufall interpretiert und abgewertet werden können (Kanning, 1999).

Bei einer freien Beurteilung der Protagonisten sollten sich im Unterschied zum Stimulus „Text“ bei zusätzlichem Vorliegen eines Bildes des Protagonisten die Eigenschaftszuschreibungen wieder finden, welche den durch die anfängliche Betrachtung des Bildes gebildeten Kategorien entsprechen. Die Textinformationen nutzt der Beurteiler zur Bestätigung der bereits gebildeten Urteilkategorien und zur Integration weiterer kategorienkonformer Informationen (Oswald, 1993). Die Urteilkategorien sollten größtenteils der anfänglichen visuellen Eindrucksbildung bei der Bildbetrachtung und nicht

wie bei alleiniger Vorlage des Textes einem Situationsmodell mit den Repräsentationen der Protagonisten entstammen. Da die Urteilkategorien in beiden Bedingungen aus verschiedenen Quellen generiert werden, kann angenommen werden, dass sich diese voneinander unterscheiden sollten.

Betrachtet man die unter dieser Stimulusdarbietung zu erbringende Gedächtnisleistung des Lesers, sollte sich zum einen der Bildüberlegenheitseffekt (Engelkamp, 1991) insofern auswirken, dass Bildinformationen über die Zielperson gut behalten werden und zum anderen, dass das Bild des Protagonisten als Anker zum besseren Behalten für die mit dem Bildeindruck konformen Informationen des Textes dient (Baltes et al., 2002). Das Leserziel Beurteilung der Protagonisten ist hier durch das Betrachten der Bilder verbunden mit einer ersten automatischen Kategorisierung bereits zum großen Teil erreicht. Der erneute Aufbau eines umfassenden Situationsmodelles durch eine intensive Informationsverarbeitung der Textinhalte wäre für den Beurteiler kognitiv sehr aufwändig und nicht unbedingt notwendig (Morrow, 1994), so dass der Leser kein erneutes vorwiegend auf Textmaterial basierendes mentales Modell aufbauen sollte. Diese weniger tiefe Textverarbeitung könnte sich in einer weniger detailreichen Wiedergabeleistung des Textes in einer Gedächtnisabfrage niederschlagen (Craik & Lockhardt, 1972). Jedoch sollten aufgrund des vorliegenden Bildankers Textinformationen, die die Urteilkategorie bestätigen und untermauern, gut wiedergegeben werden können (Baltes et al., 2002). Für die vorliegende Arbeit wird angenommen, dass sich der in den vorigen Abschnitten beschriebene Eindruck beim Betrachten eines Fotos auf die Gedächtnisleistung dergestalt auswirkt, dass das Bild als Anker genutzt wird und zu einer verbesserten Wiedergabeleistung der dargebotenen Informationen führt.

4.2.8. Die Beurteilung von Personen aufgrund verschiedener Stimulusmodi

Für die vorliegende Arbeit sind die Untersuchungen von Borkenau und Liebler (z.B. 1992, 1993, 1994) zum einen interessant, da in der Untersuchung die Stimulusinformationen systematisch variiert wurden. Zum anderen sind inhaltliche Bezüge zur Thematik der vorliegenden Arbeit relevant, da Borkenau und Liebler die Merkmalsnutzung bei Personenbeurteilungen, die Merkmalsvalidität und die funktionale Validität untersuchten. Es wurden Urteile über persönlich gut bekannte Personen, über völlig fremde Menschen und

Selbsteinschätzungen erhoben. Statische und dynamische sowie visuelle und akustische Hinweisreize wurden einander gegenübergestellt.

In einer der Untersuchungen handelte es sich um Paare, die zusammenlebten und einander sehr gut kannten (Borkenau & Liebler, 1993). Die Untersuchungen waren mit kleineren Variationen nach dem folgenden Muster aufgebaut. 50 Männer und 50 Frauen im Alter zwischen 20 und 50 Jahren, die nicht an der Universität Bielefeld studiert hatten,² und die über Zeitungsannoncen geworben wurden, dienten als Stimuluspersonen. Von jeder der Stimuluspersonen wurde eine Videoaufnahme in einer standardisierten Situation aufgenommen: Die Stimulusperson betritt einen Raum, geht um einen Tisch herum und setzt sich auf den Stuhl hinter dem Tisch. Dann schaut sie in die Kamera und liest eine Wettervorhersage vor. Danach steht die Person auf, geht um den Tisch herum und verlässt den Raum. Die Sequenz dauerte etwa 90 Sekunden.

Nach der Videoaufzeichnung beurteilten sich die Stimuluspersonen selbst und ihren Partner im Hinblick auf 45 physische Merkmale und 20 Personmerkmale. Von den Videoaufzeichnungen wurden 4 verschiedenen Varianten angefertigt, die dann als Stimulusmaterial verwendet wurden. Es handelt sich um

1. die vollständige Videoaufzeichnung mit Bild und Ton
2. die Aufzeichnung ohne Ton
3. ein Standbild, welches die Stimulusperson zeigt, während sie in die Kamera blickt und
4. eine Tonaufzeichnung.

Versuchspersonen (Urteiler) waren Studenten der Universität Bielefeld, die die Stimuluspersonen nicht kannten. Jede Versuchsperson beurteilte nur eine einzige Modalität, dort jedoch alle 100 Stimuluspersonen. Die Versuchspersonen mussten auf Adjektivskalen zur Erfassung der Big Five nach Norman (1963) die Persönlichkeitsmerkmale einschätzen und in den 4 Kategorien 1.) Merkmale, die den allgemeinen Eindruck wiedergeben (z.B. angenommenes Alter, Sympathie, Attraktivität), 2.) Akustische Merkmale (z.B. tiefe/hohe Stimmlage), 3.) Statisch-visuelle Merkmale (z.B. kurze/lange Haare) und 4.) Dynamisch-visuelle Merkmale (z.B. kleine/große Schritte) die Ausprägung des jeweils genannten physischen Merkmals beurteilen.

Borkenau und Liebler (1992) legten im Hinblick auf die Frage nach der Struktur des Persönlichkeitseindrucks das Fünf-Faktoren-Modell nach Norman (1963), welches bereits im

² Das Forschungsprojekt wurde an der Universität Bielefeld durchgeführt. Um zu gewährleisten, dass sich die Urteiler und die zu Beurteilenden möglichst noch nicht begegnet waren, war es für die Stimuluspersonen Bedingung, dass sie nicht an der Universität Bielefeld studiert hatten.

Abschnitt 4.2.4. „Das Fünf-Faktoren-Modell bei der Gesichterbeurteilung“ erläutert wurde, zugrunde. Unter allen Bedingungen waren die Beurteiler in der Lage, zwischen den fünf Persönlichkeitsfaktoren zu differenzieren. Der Persönlichkeitseindruck war umso weniger differenziert, je weniger Informationen über die Stimuluspersonen zur Verfügung standen. Borkenau und Liebler (1992, 1993) interpretieren ihre Befunde so, dass die Differenziertheit des Persönlichkeitseindrucks grundlegend davon abhängt, wie viele und welche Informationen über die Stimuluspersonen zur Verfügung stehen. Bei der Beantwortung der Frage nach der Urteilerübereinstimmung kommen Borkenau und Liebler (1992) zu dem Ergebnis, dass alle Stimulusmodalitäten Informationen liefern, die von den Beurteilern bis zu einem gewissen Grad ähnlich genutzt wurden. Bis auf eine Ausnahme konnten keine bedeutenden Einflüsse der Stimulusbedingung auf das Ausmaß der Urteilerübereinstimmung nachgewiesen werden.

Dieser Befund wird als ein empirischer Beleg für das von Kenny (1991) postulierte „General model of consensus and accuracy in interpersonal perception“ interpretiert. Dieses Modell nimmt an, dass wenn Urteiler identische Informationen erhalten, die Menge der Informationen die Urteilerübereinstimmung nicht beeinflussen. Der Faktor Emotionale Stabilität betrifft die oben erwähnte Ausnahme – für die Stimulusmodalität „Standbild“ war die Urteilerübereinstimmung sehr gering. Es wird postuliert, dass zur Beurteilung der emotionalen Stabilität dynamisch-visuelle und/oder akustische Informationen unbedingt notwendig sind. Für die Vorlage von Fotos mit Gesichtern, welche der Modalität „Standbild“ sehr nahe kommt, wurden bei Henss (1998) – wie bereits im Abschnitt 4.2.4. „Das Fünf-Faktoren-Modell bei der Beurteilung von Gesichtern“ berichtet, sehr ähnliche Befunde erhoben.

4.3. Stimulusinformation: Film

4.3.1. Bewegte Bilder

In den vorangegangenen Abschnitten wurde anhand der Ergebnisse der Forscher Borkenau und Liebler (z.B. 1992, 1993) und Henss (1995, 1998) dargelegt, dass bei der Beurteilung von Personen aufgrund visuell dargebotener Stimuli der Bewegung der Stimuluspersonen eine besondere Bedeutung zukommen sollte. Bei der Beurteilung von Stimuluspersonen hinsichtlich des Faktors Emotionale Stabilität konnte zwischen den Probanden bei Vorlage

von Gesichtsfotos und Betrachtung eines Standbildes weitestgehend keine Übereinstimmung gefunden werden. Es wurde argumentiert, dass ein Foto oder Standbild möglicherweise keine ausreichenden Informationen bietet, und dass wahrscheinlich weitere Informationen, z.B. dynamisch-visuelle notwendig sind, um hinsichtlich dieses Faktors eine Einschätzung treffen zu können. Dass bei Darbietung dynamisch-visueller Informationen eine Einschätzung mit recht hoher Ladung auch für den Faktor Emotionale Stabilität erfolgen kann, konnten bereits Borkenau und Liebler (1992, 1993) zeigen. Die Erforschung der Wirkung bewegter Bilder war – trotz der rasanten Entwicklung von Film und Fernsehen - lange Zeit ein wenig beachtetes Untersuchungsfeld und erst in den achtziger Jahren wurde die Forderung nach neuen Forschungsansätzen, die den Möglichkeiten der Bildkommunikation theoretisch und methodisch Rechnung tragen, lauter (Schulz, 1975, Weiss, 1982, Winterhoff-Spurk, 1986). Es ging darum, sich in der Medienwirkungsforschung „vorrangig den von der angewandten Forschung vernachlässigten Grundlagenproblemen, insbesondere der *theoretischen und methodischen Innovation*“ zu widmen (Schulz, 1982, S. 3, Hervorhebung im Original). Schulz (1982) forderte vor allem, dass ein flexibles Untersuchungsdesign entwickelt werden sollte, womit sich möglichst viele Wirkungen nonverbaler Kommunikation untersuchen lassen. Trotz einiger neuer Ansätze blieb die Erforschung bewegter Bilder in der Psychologie im Großen und Ganzen eine Disziplin, die wenig Beachtung fand.

Im Zusammenhang mit den Forderungen von Schulz (1982) und einem darauf folgenden Forschungsprojekt wurden beispielsweise Untersuchungen durchgeführt, bei denen einer Versuchsperson eine Videosequenz (ohne Ton) eines Politikers vorgeführt wurde und bei denen von der Versuchsperson simultan zur Videodarbietung vegetative, kognitive und mimische Reaktionen aufgezeichnet wurden (Frey und Bente, 1989, Pöhls, 1989). Damit sollten sowohl die Wirkungen des visuellen Inputs die sich im äußeren Verhalten des Probanden niederschlagen als auch die Einflüsse auf die kognitiven Einstellungen und das affektive Erleben des Betrachters erfasst werden. Da jedoch nur ein Videoclip verwendet wurde und es sich um Einzelfallbetrachtungen handelt, kann von den erhaltenen Daten nicht auf systematische Zusammenhänge zwischen Stimulusinput und Reaktion des Probanden geschlossen werden. Gerade die Frage, ob es sich zwischen der Stimulusinformation – bewegte Bilder eines Menschen - und dem während der Rezeption gezeigten Verhalten des Betrachters um systematische Beziehungen handelt oder ob lediglich Zufallsvariationen vorliegen, kann nur beantwortet werden, wenn Stimulusdaten systematisch variiert werden (Frey, 1999b).

Die Ermittlung von spezifischen Stimulusmerkmalen, die eine bestimmte Reaktion des Betrachters auslösen, ist eine Herausforderung, an die sich die Forscher in Anlehnung an die ethologische Verhaltensforschung bei sozialen Tieren, annäherten. Die so genannten Attrappenversuche in der ethologischen Verhaltensforschung eigneten sich als Methodik, um bestimmte „Schlüsselreize“, die bei sozialen Tieren unwillkürliche Verhaltensreaktionen auslösen, zu identifizieren (Lorenz & Leyhausen, 1968). Durch systematische Variation der Körperformen, der Farbgebung und von Bewegungsweisen von Tierattrappen konnte ermittelt werden, welche speziellen Hinweisreize der komplexen Stimulusinformation ein bestimmtes Verhalten auslösen. Auch in diesen Experimenten konnte gezeigt werden, dass innerhalb der Reaktion auslösenden Elemente insbesondere der Bewegung eine bedeutsame Rolle zukommt: „Wenn man die relative Wirkung einzelner Merkmale gegeneinander abwägt, so stellt sich regelmäßig heraus, daß *Bewegungsmerkmale*... quantitativ gewaltig über alle anderen überwiegen, so sehr, daß sie unter Umständen unentbehrlich werden (Lorenz et al., 1968, S. 45, Hervorhebung im Original).

Nach dem Vorbild der ethologischen Attrappenversuche wurde ein Computeranimationsverfahren, das als „Skriptanimation“ bezeichnet wurde, entwickelt (Kempster, 1998), welches mit der Hilfe einer Puppe das komplexe Bewegungsverhalten einer Stimulusperson wiedergeben kann. Als Grundlage für die Skriptanimation wurde das von Frey et al. (1981, 1983) entwickelte Berner System – ein Notationssystem zur Beschreibung natürlicher menschlicher Bewegung -, welches es gestattet, jede noch so komplexe Körperbewegung nonsemantisch zu kodieren, genutzt. Der Bewegungsablauf eines Menschen wird in Form einer Zeitreihe, in der die einzelnen statischen Positionskonfigurationen enthalten sind, in einem Datenprotokoll niedergelegt. Anhand dieses Protokolls kann das natürliche menschliche Bewegungsverhalten mittels der Puppe exakt auf dem Bildschirm nachgebildet werden. Somit können das Verhalten der Stimulusperson als auch das physiognomische Erscheinungsbild experimentell kontrolliert variiert werden, um die Elemente zu ermitteln, die den Eindruck des Betrachters prägen. (Frey, 1999b). Die erstmals durch die Entwicklung der Skriptanimation möglich gewordenen Ergebnisse deuten darauf hin, dass Urteiler den einzelnen Merkmalen des menschlichen Erscheinungsbildes ein sehr unterschiedliches Gewicht beimessen. Frey (1999b) postuliert für den Prozess der spontanen Eigenschaftszuschreibungen, dass dem Bewegungsverhalten der Stimulusperson mehr Bedeutung zukommt als deren statischen, physiognomischen Aussehen.

Frey (1999b) eliminierte in einer Untersuchung den Einfluss der Physiognomik, in dem er das Bewegungsverhalten von verschiedenen Politikern von derselben Computer animierten Puppe nachbilden ließ. Allein aufgrund der Bewegungsinformation – dargestellt durch die Puppe - gelangten die Beurteiler zum selben Urteil, wie wenn sie den Politiker in Person sahen. Beim Vergleich der Persönlichkeitsprofile findet sich das Eigenschaftsprofil der Originalperson beinahe ohne Veränderung im Profil für die Puppe wieder. Die Übereinstimmung ist sogar für die Beurteilung der „Attraktivität“ der Stimulusperson zu finden. Die Puppe, die sich wie eine als attraktiv beurteilte Politikerin bewegte, wurde auch als attraktiv von den Probanden beurteilt, während die Puppe, die sich wie ein Politiker bewegte, der in natura als hässlich eingeschätzt worden war, als genauso hässlich beurteilt wurde. Frey (1999b) führt die „spezifischen Elemente der Gesamtbewegung, die für die Unterschiede in der unbewussten Beschlussfassung verantwortlich zeichnen...“ auf die „... von Lorenz, Tinbergen und Leyhausen immer wieder betonte, charakteristische Merkmalsarmut der Stimuli, auf die die im subkortikalen Bereich lokalisierten affektiven Steuerungszentren reagieren“ (S. 138), zurück.

Der Vergleich der Bewegungsmuster von Politikern, die besonders positiv oder besonders negativ beurteilt wurden, ergab, dass die Beurteilung stark von der Art und Weise, wie die Stimulusperson eine seitliche Kippbewegung des Kopfes vornahm, abhing. Weshalb eine Person bei der visuellen Eindrucksbildung dieser Dimension der Kopfhaltung bei der Personbeurteilung eine so große Bedeutung beimisst, ist kaum geklärt (Frey, 1999b).

Dass die Kopfhaltung – nur eine von vielen Dimensionen beim Betrachten eines Gesichtes – eine große Wirkung hat, konnten Frey et al. bereits 1983 anhand von gemalten Bildern berühmter Künstler (z.B. Sandro Botticelli) zeigen. Ein Bildausschnitt, der jeweils den Kopf und den Schulterbereich einer Person umfasst, wurde lediglich ein wenig geometrisch gedreht. Aufgrund dieser marginalen Veränderung der Kopfhaltung wurden dieselben Stimuluspersonen, die erst als sympathisch, empfindsam, zärtlich beurteilt wurden, plötzlich als unsympathisch, kalt, hinterlistig usw. eingeschätzt. Dieser Effekt war für Beurteiler aus verschiedenen Ländern und Kulturen robust (Frey et. al, 1983).

In einer Untersuchung von Kanning & Leisten (2004) mit variierendem Stimulusmaterial wurde den Versuchspersonen (Beobachter eines Assessment Centers) entweder eine vollständige Videoaufnahme einer AC-Übung oder nur das bewegte Bild, wobei der Ton des Films ausgeschaltet wurde, präsentiert. Kanning und Leisten (2004) fanden heraus, dass in der

Bedingung einer hohen kognitiven Belastung (gleichzeitige Beobachtung und Bewertung von 4 Bewerbern in einer Gruppendiskussion) die Urteiler unabhängig vom dargebotenen Stimulusmaterial zu denselben Urteilen gelangten. Signifikante Unterschiede zeigten sich in der Bedingung der geringen kognitiven Belastung (Beobachtung und Bewertung von 4 Bewerbern, die nacheinander in einem Rollenspiel mitwirkten). Im Fall der geringen kognitiven Belastung fiel die Bewertung der Bewerber, für die nur die nonverbalen Informationen dargeboten wurden, weniger positiv aus als für die Bewerber, in der sowohl verbale als auch nonverbale Informationen zur Verfügung standen. Kanning und Leisten (2004) interpretieren diese Ergebnisse so, dass unter hoher kognitiver Belastung das Urteil der Beobachter vollständig durch die nonverbalen Informationen der Bewerber erklärt werden kann. Diese Ergebnisse entsprechen einer sehr groben Reduzierung der Komplexität der tatsächlich dargebotenen Informationen.

4.3.2. Annahmen für die Beurteilung von Personen anhand eines Filmausschnittes

Die Eindrucksbildung bei Darbietung eines Filmausschnittes kann als ein routinierter Prozess bezeichnet werden, da das Schauen eines in Echtzeit dargebotenen Ausschnittes dem alltäglichen Erleben von sozialen Situationen in der realen Umwelt sehr nahe kommt und in diesen Situationen ständig Urteile über Personen gebildet werden (Kanning, 1999; Kanning et. al, 2004). Der Stimulus Film unterscheidet sich in zwei wesentlichen Merkmalen von den Stimuli Text und Text mit Bild. Zum ersten sind die Komplexität und die Menge der dargebotenen Informationen sehr viel höher als in den ersten beiden Bedingungen. Zweitens kommen zum semantischen Inhalt des Textes und zum statisch optischen Eindruck des Bildes die Modalitäten Bewegungsinformationen und akustische Informationen, die über den semantischen Inhalt des Wortes hinausgehen (z.B. Sprachmelodie, Stimmlage, Lautstärke). Urteiler wenden gewohnheitsmäßig bei hoher Komplexität dargebotener Inhalte Heuristiken und Faustregeln an, um schnell und mit wenig kognitivem Aufwand zu einem Urteil über die Zielpersonen zu gelangen (Bierhoff, 2000; Strack, 2000). Dadurch verkürzt sich der Meinungsbildungsprozess und die Urteiler kommen zu einem schnellen Eindruck über die wahrgenommenen Personen. Diese Prozesse verlaufen meist automatisch und ohne willentliche Kontrolle (Chaiken & Trope, 1999). Es werden dabei nicht alle möglichen Informationen ausgeschöpft und ausreichend beleuchtet (Bierhoff, 2000). Die Komplexität

vorhandener Informationen wird meist stark reduziert und es kann von einer groben Urteilsvereinfachung gesprochen werden (Strack, 2000).

Die Bewegungsinformationen und akustische Merkmale der Sprache (Stimmlage, Lautstärke, Sprechgeschwindigkeit, Artikulation, Modulation, Atmung) haben unabhängig vom Inhalt der Kommunikation und vom optischen Erscheinungsbild der Zielpersonen einen starken Einfluss auf die Urteilsbildung (z.B. Borkenau & Liebler, 1993; Frey, 1999b). Welche konkreten Wirkmechanismen die Urteilsbildung aufgrund dieser Informationen beeinflussen, ist derzeit wenig untersucht und es gibt nach Kenntnis des Autors bislang kein Modell, welches diese Einflüsse beschreibt. Aufgrund dessen, dass es bislang kein Modell gibt, welches die Einflüsse von nicht sprachlichen Informationen bei der sozialen Urteilsbildung ausreichend berücksichtigt, werden die folgenden Annahmen und Schlussfolgerungen mit größter Vorsicht und Zurückhaltung vorgenommen. Trotz der genannten Einschränkungen wird das gegenwärtig führende Kontinuummodell der Eindrucksbildung (Fiske & Neuberg, 1990; Fiske, 1993), welches die Verarbeitung alltagstypischer Personinformation betrifft, zugrunde gelegt.

Danach wird die Zielperson vom Urteiler nach dem Wahrnehmen kategorisiert und anschließend versucht der Urteiler, diese anfängliche Kategorisierung zu bestätigen (Fiske & Neuberg, 1990), in dem er vornehmlich die Informationen aus dem weiteren Situationsablauf selektiert wahrnimmt, die kategorienkonform sind (Oswald, 1993). Da auch die soziale Situation, die durch einen Filmausschnitt wiedergegeben wird, Interpretationsspielräume hinsichtlich der Verwendung verfügbarer Informationen für den Wahrnehmer und Urteiler lässt, wird es häufig möglich sein, die für die Zielperson gewählte Anfangskategorie aus den weiteren Informationen zu bestätigen. Unter diesen Prozessannahmen sollten sich die Urteile bei den Stimulusdarbietungen Text mit Bild und Film nicht unterscheiden.

Die Wiedergabeleistung in einer Gedächtnisabfrage wird bei dieser Darbietungsform im Vergleich zur Darbietung von Text mit Bild im wesentlichen von den zusätzlichen Faktoren starker Anstieg der Komplexität der dargebotenen Informationen und Hinzukommen zweier Informationsmodalitäten (visuell-dynamische und akustische Informationen) beeinflusst. Die physikalischen Oberflächeneigenschaften dargebotenen Informationen sind denen der Alltagsbeobachtungen sehr ähnlich, so dass die Informationsverarbeitung und –speicherung aufgrund der Routine als ein hoch automatisierter Prozess des Betrachters erfolgt. Aufgrund der großen Menge hoch komplexer Informationen und der Begrenztheit menschlicher

Informationsverarbeitungskapazitäten ergibt sich die Notwendigkeit der Komplexitätsreduktion, die sich in einer geringen Behaltensleistung für Detailinformationen niederschlagen sollte. Konträr dazu wird von Engelkamp (1991) nach der multimodalen Gedächtnistheorie eine Verbesserung der Behaltensleistung mit Zunahme der Anzahl der dargebotenen Modalitäten postuliert, so dass demnach bei den hier vorliegenden zusätzlichen zwei Modalitäten bessere Ergebnisse in der Gedächtnisabfrage zu erwarten wären. Jedoch muss hierzu einschränkend gesagt werden, dass die Experimente, die zu diesen Ergebnissen führten, mit konstruiertem Versuchsmaterial durchgeführt wurden, welches die Komplexität eines Filmausschnittes kaum widerspiegelt. Die Speicherung der dargebotenen hoch komplexen Informationen sollte wie oben beschrieben durch eine große Reduktion der Komplexität und damit durch das Vernachlässigen von Detailinformationen gekennzeichnet sein.

5. Einfluss der Beurteilererfahrung

5.1. Zur Abgrenzung von Experten und Novizen

Nachdem im vorigen Abschnitt Ausführungen zu den Auswirkungen verschiedener Stimuli auf Urteilsbildungsprozesse gemacht wurden, soll in diesem Abschnitt beleuchtet werden, wie sich eine unterschiedliche Beurteilererfahrung auf die Urteilsbildung von Personen auswirkt. Wie bereits mehrfach ausgeführt, bewertet jeder Mensch im Alltag unzählige Male (bewusst oder unbewusst) andere Personen und soziale Situationen, nicht wenige Menschen beschäftigen sich beruflich mit der Beurteilung anderer Personen, die sehr oft weitreichende Konsequenzen für die Beurteilten haben. Viele Menschen, die sich professionell mit Personenbeurteilung beschäftigen, wie z. B. Richter, Polizeibeamte, Personalmanager üben diese anspruchsvolle Tätigkeit seit vielen Jahren aus. Sind diese Personen Experten im Beurteilen von Personen? Was unterscheidet Experten von Personen, die sich nicht professionell mit Personenbeurteilung beschäftigen?

In der Umgangssprache werden Personen, die in einem bestimmten Sachgebiet längerfristig mehr Aufgaben auf höherem Niveau mit geringem Aufwand bewältigen als andere, als Experten bezeichnet. Eine allseits anerkannte Definition von Expertise in der

Forschungsliteratur gibt es nicht, wohl jedoch die Annäherung an den Expertenbegriff von mehreren Seiten.

Eine Definition des Begriffes Experte findet sich im Einleitungskapitel „What is to be an expert?“ von Posner (1988) zum Buch „The nature of expertise“ (Chi, Glaser & Farr, 1988): danach ist ein Experte eine Person, die auf einem bestimmten Gebiet dauerhaft, (also nicht zufällig und singulär) herausragende Leistungen erbringt. Auf das in der psychologischen Forschung wichtigste Merkmal zur Operationalisierung von Expertise – Leistungsstärke im relevanten Gegenstandsbereich – nimmt auch Krems (1994) in seiner Definition Bezug. Er bezeichnet als Expertise „die bereichs- und aufgabenspezifische Problemlösefertigkeit einer Person: ihr in einer Aufgabenklasse eines einzelnen Sachgebietes erworbenes Wissen und Können, das sie in die Lage versetzt, dauerhaft Hervorragendes zu leisten,, (S. 53).

Krems (1994) nutzt als Abgrenzungskriterien von Expertisegruppen die Effizienz, bereichsspezifisches Wissen und Können, Erfahrung und Informationsverarbeitung. Effizienz bei Experten drückt sich zum einen in der überdurchschnittlich hohen Qualität der Aufgabenbewältigung, zum anderen im unterdurchschnittlichen Aufwand (Zeitaufwand, Kosten etc.) aus. Darüber hinaus betont Krems (1994) für das Effizienzkriterium die Dauerhaftigkeit der überdurchschnittlichen Leistungshöhe von Experten, beispielsweise wird in der diagnostischen Beurteilung eine Person, die dauerhaft in einer höheren Anzahl von Fällen zu korrekteren Entscheidungen als Personen mit ähnlich intensiver Ausbildung und Erfahrung gelangt, als Experte bezeichnet.

Grundlage einer hohen Expertisestufe sind umfassende und differenzierte Kenntnisse aus einem Fachgebiet. Zu den bereichsspezifischen Kenntnissen gehören zum einen das Wissen über Sachverhalte und Gesetzmäßigkeiten, zum anderen die Fertigkeiten zur Problembewältigung. Mandl, Gruber und Renkl (1993) sprechen in diesem Zusammenhang von einer Kontextualisierung: Expertise ist stets auf eine spezifische Domäne bezogen. In der gleichen Domäne können etwa gleich gute Leistungen von Experten erbracht werden, die ihr Wissen auf unterschiedlichem Wege erworben haben - Hacker (1992) spricht von Praktiker – und Theoretikerwissen - : eine sehr gute klinische Urteilsfähigkeit kann in einer Universitätsausbildung oder durch langjährige praktische Erfahrung oder durch eine Kombination beider erworben werden.

Expertise beruht ganz wesentlich auf einer langjährigen Erfahrung (z.B. Ericsson & Crutcher, 1990). Das Wissen und die Fertigkeiten werden in einer intensiven und langwierigen Auseinandersetzung mit einem Gegenstand gewonnen. Ericsson & Crutcher (1990) beispielsweise schätzen 8 bis 10 Jahre. Gerade in semantisch vielschichtigen Bereichen gilt eine ausgiebige und lange Problemlöseerfahrung als notwendige Voraussetzung von Expertenschaft.

Erfahrung und Übung allein führen nicht zwangsläufig zum Expertenstatus. Zum einen hat jeder Mensch seine eigene Lerngeschichte, zum anderen spielen ganz sicher kovariierende Merkmale, beispielsweise Intelligenz und Motivation eine große Rolle. Auch Gruber und Mandl (1992) betonen die Bedeutung dispositionaler Faktoren, insbesondere im Bereich basaler mentaler Fähigkeiten für die Ausbildung eines überdurchschnittlichen Leistungsvermögens. Des Weiteren sind für den Kompetenzerwerb weitere Einflussfaktoren wie Feedback, Qualität von Informationsquellen und Vorbilder von Bedeutung (Gruber & Mandl, 1992).

Die bislang genannten Abgrenzungskriterien – Effizienz, bereichsspezifisches Wissen und Erfahrung – legen äußere Vergleichsmaßstäbe fest, anhand derer Experten von Nichtexperten eines Sachgebietes normalerweise unterschieden werden. Die kognitiven Grundlagen effizienter Informationsverarbeitung und Merkmale, die eine erfolgreiche Lerngeschichte ermöglichen, bleiben hierbei unberücksichtigt. Das Identifizieren der durch Erfahrungsbildung modifizierbaren Merkmale und Mechanismen, die Grundlage unterschiedlich effizienten Verhaltens in einzelnen Bereichen sind, steht am Anfang. Unterschiedliche empirische Befunde bestätigen diese Aussage. So stellten beispielsweise Simon und Simon (1978) und Larkin et. al. (1980) die Hypothese auf, dass Anfänger beim Lösen eines Problems eher ziel- und rückwärts orientiert vorgehen, wogegen Experten eher daten- und vorwärts orientiert vorgehen. Anfänger sollen zunächst eine abstrakte Zielanalyse betreiben, von der sie dann ausgehen und rückwärts suchend ihr Vorgehen ableiten, während Experten sich unmittelbar der Situation zuwenden und ausgehend von den Fakten zu den Lösungen – vorwärts gerichtet – voranschreiten. „... the novice solved most of the problems by working backward from the unknown problem solution to the given quantities, while the expert usually worked forward from the givens to the desired quantities“ (Larkin et al., 1980, S. 1338). Priest und Lindsay (1992) fanden hingegen Ergebnisse, die ernste Zweifel an der Generalität vom rückwärts zum vorwärts gerichteten Problemlösen bei wachsender

Expertenschaft aufwerfen. Sie fanden, dass beide Gruppen, Experten und Novizen öfter vorwärts gerichtete Schritte nutzen als rückwärts gerichtete.

Auch Zajchowski und Martin (1993) fanden vorwärts gerichtetes Vorgehen bei Novizen und Experten als Hauptstrategie. Clarke und Lamberts (1997) konnten in einer Langzeitstudie für die Anfänger weder die einen noch die anderen Ergebnisse replizieren. Sie konnten zeigen, dass Anfänger überhaupt keine konsistente Strategie nutzen. Allein diese Ergebnisse zeigen, dass „the true differences between novices and experts are much more complicated than suggested by previous analyses“ (Clarke und Lamberts, 1997, S. 287).

5.2. Die Messung des Expertisegrades

Der kontrastive Ansatz (Voss, Fincher-Kiefer, Green & Post, 1986) scheint sich zur Untersuchung individueller Differenzen besonders zu eignen. Mittels des kontrastiven Verfahrens können zum einen Hypothesen bezüglich Ähnlichkeiten von bzw. Differenzen zwischen Experten und Novizen getestet werden. Zum anderen soll ein Experten-Novizen-Vergleich – betrachtet als querschnittliche Approximation von Entwicklung – die Frage beantworten helfen, wie man Experte wird. Drittens können Personen mit einem bestimmten Expertisegrad in einem eingegrenzten Gebiet untersucht werden und es wird versucht, zu klären, was hohe Expertise ausmacht und wodurch sich Leistungen von Experten auszeichnen.

Die Abgrenzung der Expertisestufen kann auf verschiedene Arten erfolgen. Als Einteilungskriterium wird überwiegend die Zugehörigkeit zu qualitativen Klassen innerhalb der Domäne verwendet (z.B. erfahrene Ärzte vs. Medizinstudenten). Die Abgrenzung geschieht hier über externe Kriterien wie Qualität und Intensität der Ausbildung, die Dauer der Berufserfahrung oder beruflicher Erfolg (z.B. Chi et al. 1981). Gruber (1994) spricht in diesem Zusammenhang von einer Konstruktvalidierung der Klassifikation: die Einteilung in Experten und Novizen erfolgt durch Annahme oder Beobachtung, dass sich erbrachte Leistungen von Mitgliedern der verschiedenen Klassen voneinander unterscheiden. In der Regel handelt es sich hierbei um Plausibilitätsannahmen, d.h. man unterstellt Korrelationen dieser Variablen mit der Fähigkeit zu hervorragenden Leistungen.

Nur in seltenen Fällen ist die Abgrenzung der Expertisegruppen durch Anwendung eines direkten Messverfahrens des Leistungsvermögens der Versuchspersonen auf einem einzelnen Gebiet möglich. Als Ausnahme gilt hier die Domäne Schach – nach dem Elo-System (Elo, 1978) kann die Spielstärke unmittelbar festgestellt werden. Wenn – wie in der Domäne Schach - ein valides Messinstrument existiert, können die Probanden, die auf einem Leistungskontinuum angeordnet sind, entsprechend der Leistungsstärke eingeteilt werden. Die Leistungsstärkeren gelten als Experten, die Leistungsschwächeren als Novizen.

Z.B. wandten Goldin (1978) und Gruber und Ziegler (1990) das o.g. Elo-System an und bildeten zwei distinkte Gruppen, deren Mittelwerte maximal auseinander liegen, so dass sich die Verteilungen möglichst wenig oder nicht überschneiden. Diese Operationalisierung sorgt für eine hohe Validität der Expertendefinition innerhalb der Untersuchung.

Eine weitere Möglichkeit der Gruppeneinteilung besteht in der Konstruktion von Leistungs- oder Wissenstests, die als hoch korrelierend mit dem Expertisegrad angesehen werden.

Schneider, Körkel und Weinert (1990) beispielsweise unterscheiden zwischen Experten und Novizen aufgrund der durch eine Fragebogenerhebung objektivierten bereichsspezifischen Vorkenntnisse. Aufwendiger ist die Klassifikation, die Vessey (1985) verwendet hat: sie unterscheidet ihre Probanden anhand von Beurteilungen von Vorgesetzten und anhand von Leistungen, die die Personen in einer experimentellen Aufgabe zeigten.

5.3. Anfänger und Experten im Vergleich

5.3.1. Die Rolle bereichsspezifischen Wissens

Seit Ende der 70er Jahre ist Expertiseforschung im Vergleich zu anderen Kompetenzforschungsansätzen dadurch gekennzeichnet, dass die hohe Leistung, die Experten erbringen, hauptsächlich als erworben angesehen und weniger personeninternen Dispositionen zugeschrieben wird. Einen Schwerpunkt der Expertiseforschung bildet die Untersuchung von Unterschieden in der Wissensbasis von Novizen und Experten, wobei Fragen nach der Organisation, der Repräsentation, dem Abruf oder der Nutzung von Wissen gestellt werden (Mandl, Gruber und Renkl, 1993; Chi, Glaser & Farr, 1988).

De Groot (1965) fand in einer inzwischen wissenschaftshistorischen Untersuchung bei unterschiedlich guten Schachspielern keine Abweichungen in den Suchstrategien oder der Suchtiefe, jedoch in der Qualität der betrachteten Züge: die Experten überlegten und wählten die besseren Züge. In einer Folgeuntersuchung erinnerten Versuchspersonen Schachpositionen, die ihnen zuvor für kurze Zeit gezeigt worden waren. Dabei fand de Groot, dass die Reproduktionsleistung direkt mit der Spielstärke zusammenhing. De Groot begründete den Expertenvorteil durch eine bessere Enkodierfähigkeit aufgrund des Wiedererkennens vertrauter Muster während der kurzzeitigen Situationsanalyse. Bei Darbietung zufälliger Brettkonstellationen sind erfahrungsbedingte Verarbeitungsvorteile in der Gedächtnisleistung nicht mehr zu finden.

Die große Rolle von domänenspezifischem Vorwissen wurde in einer entwicklungspsychologischen Studie von Chi (1978) besonders deutlich. Kinder mit guten Schachkenntnissen und Erwachsene mit wenig Schacherfahrung sollten Schachpositionen rekonstruieren, die ihnen vorher für 10 Sekunden gezeigt wurden. In einer Kontrollaufgabe sollte einfaches Zahlenmaterial (Listen mit 10 Zahlen) reproduziert werden. Die jungen Schachexperten waren den erwachsenen Schachnovizen bei der schachbezogenen Gedächtnisaufgabe überlegen. Bei den nicht domänenspezifischen Gedächtnisaufgaben schnitten die Erwachsenen erwartungsgemäß besser ab. Chi (1978) konnte also zeigen, dass durch domänenspezifisches Vorwissen altersübliche Entwicklungsunterschiede sogar umgekehrt werden können.

Chase und Simon (1973a, 1973 b) konnten zeigen, dass die Ursache des Leistungsvorteils von Schachspielern eine schachspezifische Fähigkeit und kein außergewöhnlich leistungsstarkes visuelles Gedächtnis ist.

Nach der Chunking-Theorie (Miller, 1956) werden Informationen in größere Einheiten zusammengefasst und mit Labels indiziert. Es genügt die Übernahme des Labels in das Kurzzeitgedächtnis anstelle des gesamten Informationsinhaltes. Beim späteren Abruf anhand des Labels werden die Informationen im Langzeitgedächtnis aufgesucht und dekodiert. In Anlehnung an Millers Chunking-Theorie (1956) entwickelten Chase und Simon (1973a, 1973b) die Pattern Recognition-Theorie. Hauptinhalt dieser Theorie war, dass Expertise die Konsequenz der Abspeicherung tausender schachspezifischer Wissensmuster sei und nicht, dass umgekehrt eine hohe Gedächtnisleistung Folgeerscheinung von hoher Expertise sei.

Die Pattern Recognition-Theorie sollte zum einen die Kurzzeitgedächtnisleistung von Schachexperten erklären, zum anderen sollte sie die Entstehung von Expertise erklären. Unter der Voraussetzung des Funktionierens des Gedächtnisses nach dem Chunking-Prinzip wurden die Ergebnisse im freien Erinnern dadurch erklärt, dass Schachmeister viele Chunks in Form von Schachmustern in ihrem Langzeitgedächtnis gespeichert und mit einem Label verknüpft hatten. Über dieses größere Lexikon relevanter Stellungsmuster ist der Experte in der Lage, das jeweils gegebene Muster als Instanz einer Kategorie zu erkennen. Das Label dieses Musters wurde in einen Kurzzeitspeicher überführt, so dass anstelle der Einzelfiguren nur der Name des Labels erinnert werden musste. Der Abruf eines Labels aus dem Kurzzeitgedächtnis führte zu einer Suche im Langzeitgedächtnis. Durch den Abruf aller unter dieser Kategorie verknüpften Schachfiguren war die Leistung der Schachexperten deutlich besser als die der Novizen. Chase und Simon vertraten weiterhin die Auffassung, dass Experten die für eine wiedererkennbare Kategorie im Gedächtnis bewährten Spielzüge beherrschen. Die Pattern Recognition-Theorie postuliert also auch eine prozedurale Komponente, indem sie mit den im Gedächtnis gespeicherten Schachmustern Aktionsvorschläge für auszuführende Spielzüge verknüpft, die automatisch aktiviert werden, wenn ein Muster erkannt wird. Empirische Belege zu den Annahmen einer Existenz von Zugvorschlag-Chunk-Assoziationen fehlen jedoch.

Mit der Skilled Memory-Theorie entwickelten Chase & Ericsson (1981, 1982) den Chunking-Ansatz konsequent fort. Es konnte gezeigt werden, dass eine hohe Gedächtnisleistung (bezogen auf eine spezifische Domäne – hier Erlernen und Erinnern von Ziffernlisten -) durch intensives Training und durch organisierte Strukturierung des Materials trainierbar ist. Ericsson (1985) konnte zeigen, dass Gedächtnisstrukturen, die von der Organisation des normalen Gedächtnisses abweichen, experimentell erzeugt werden können. Während beim Gedächtnis normalerweise von einem Mehrspeichermodell mit getrennter Funktion der einzelnen Speicher und einer Langwierigkeit der Datenübertragung vom Kurzzeit- in das Langzeitgedächtnis ausgegangen wird, verlegt die Skilled Memory-Theorie die entscheidenden Prozesse beim Umgang mit großen Informationsmengen vom Kurzzeit – in das Langzeitgedächtnis. Die Skilled Memory-Theorie nimmt an, dass Experten eine externe Memorisierungsstruktur nutzen, in die die zu speichernde oder abzurufende Information schnell eingefügt wird.

Sowohl die Pattern Recognition-Theorie als auch die Skilled Memory-Theorie beruhen auf der Kumulierung von Wissen. Die Bedeutung bereichsspezifischer Kenntnisse steht außer Zweifel. Als alleinige Erklärung von Expertise ist quantitative Wissenskumulation jedoch nicht ausreichend und wurde mehrfach kritisiert (z.B. Charness, 1976, Mandel, Gruber und Renkl, 1993).

5.3.2. Organisation des Wissens

Zusammenhänge zwischen der Organisation von Wissenseinheiten in komplexeren Strukturen und dem Expertisegrad sind vielfach untersucht worden (z.B. Chase und Simon, 1973a; Murphy & Wright, 1984, Larkin, 1983). Kompetenzbedingte Unterschiede werden im Komplexitätsgrad der Einheiten, das heißt im Umfang und im Abstraktionsgrad von Wissenseinheiten, in der hierarchischen Ordnung und in der Ähnlichkeit und Vernetztheit von Einheiten vermutet (Krems, 1994). Chase und Ericsson (1981) postulieren beispielsweise, dass Unterschiede zwischen Experten und Novizen auf Unterschiede der in der Problempräsentation instantiierten Elemente des Langzeitgedächtnisses zurückzuführen sind. Das mentale Bild der Problemsituation enthält mehr Aspekte und ist inhaltsreicher. Chiesi, Spilich & Voss (1979) fanden auch, dass Experten bei Erinnerungsaufgaben komplexere Einheiten reproduzieren und dass ihnen in der freien Beschreibung von Kategorien mehr Informationen als Novizen zugänglich sind.

Hinsichtlich der Ordnung von Wissenseinheiten fanden z.B. Chi, Feltovich und Glaser (1981) bei der Untersuchung von Physikaufgaben, dass sich mit zunehmender Kompetenz eine mehrstufige Schemahierarchie mit einer Vielzahl unterschiedlicher Ebenen herausbildet, während Anfänger vergleichsweise flache Strukturen besitzen. Anfänger berücksichtigen eher nebensächliche und oberflächliche Merkmale (z.B. gleiche Farbe, gleiche äußere Eigenschaften), während Experten funktionale und strukturbezogene Eigenschaften einer Domäne einbeziehen (z.B. die physikalischen Grundprinzipien, die für die konkrete Aufgabenstellung bedeutsam sind)(Chi et al., 1982).

Mandl, Gruber und Renkl (1993) betonen, dass neben dem domänenabhängigen Kontext die Expertenleistung maßgeblich von der Situation, in der die Kenntnisse erworben und angewendet werden, abhängt. Expertise kann danach nicht nur durch abstrakte Wissensbestände, die ein Experte besitzt, beschrieben werden, vielmehr muss der Experte

eine Vielzahl von konkreten Problemfällen in der Praxis bearbeiten, um eine erfahrungsbedingte Veränderung der Wissensstruktur zu erreichen (Kolodner, 1983). Zusammenfassend beschreibt Krems (1994) das Entstehen von Expertentum wie folgt: „Die Genese von Expertenschaft kann insgesamt wohl eher als eine graduelle Erweiterung – quantitativ und qualitativ – der bestehenden Fähigkeiten beschrieben werden, in der die auf einer niedrigen Kompetenzstufe schon verfügbaren Verarbeitungsprinzipien nicht ersetzt, sondern ergänzt und erweitert werden,, (S. 74)

5.3.3. Kognitive Flexibilität

Bisherige Expertisetheorien wurden insbesondere wegen der hypothetisierten starren Struktur der Wissensrepräsentation kritisiert (z.B. Frey & Adesman, 1976, Holding, 1985). Die ausschließliche Betrachtung der Aneinanderhäufung von statischen Wissens-elementen ist nicht ausreichend, um alle beobachtbaren Expertisephänomene hinreichend zu erklären (z.B. Charness, 1976). Erst durch die Verknüpfung von Wissen mit dynamischen Aspekten wird die flexible Handhabung und Nutzung von Wissen ermöglicht, wie Experten sie zeigen (Gruber, 1994; Krems & Bachmaier, 1991). Holding (1979, 1985, 1992) griff in seinem SEEK-Modell (Search, EvaluatE and Know) Suchprozesse und Evaluationen auf, die in frühen Informationsverarbeitungsansätzen eine bedeutende Rolle für erfolgreiches Problemlösen spielten.

Das Modell nimmt an, dass die Auswahl zwischen verschiedenen Zugmöglichkeiten beim Schachspiel der grundlegende und bedeutsamste Prozess ist. Um die Suche zu leiten und um Schachpositionen korrekt zu beurteilen, nutzt der Experte sein Wissen, um die bestmöglichen Züge herauszufiltern. Demnach – so postuliert Holding (1979) - unterscheiden sich gute Schachspieler von wenig Geübten darin, dass sie zum einen mehr Wissen haben und zum anderen den aktiven Prozess des Ausschauhaltens (Suche) nach den Zügen mit den vorhersehbar positivsten Konsequenzen (Evaluation) besser bewältigen, wofür sich auch empirische Belege finden. Beispielsweise konnte Charness (1981) einen expertisegradabhängigen Unterschied in der Suche finden: durchschnittliche Schachspieler untersuchten etwa 45 % weniger mögliche Schachzüge als Schachmeister. Holding und Mitarbeiter (Holding, 1979; Holding & Pfau, 1985) konnten zeigen, dass die Evaluationen guter Schachspieler 75 % weniger Einschätzungsfehler enthielten als die schwächerer Spieler. Nach Holding (1985) stellen Suche und Evaluation fundamentale Konzepte von

Schachexpertise dar, Wissen spielt in seinem SEEK-Modell zwar eine wesentliche und unterstützende, jedoch nicht eine so fundamentale Rolle wie Suche und Evaluation. Relevantes Wissen kann den Suchprozess abkürzen helfen und/oder Prozeduren zur Transformation von aktuellen Schachpositionen in erstrebenswerte Stellungen liefern. Erst die Verknüpfung von Wissen mit dynamischen Aspekten der Suche im Problemraum und mit evaluativen Komponenten machen eine flexible Handhabung und Nutzung von Wissen möglich (Krems & Bachmaier, 1991, Gruber, 1994).

Der Aspekt der kognitiven Flexibilität wurde auch von Hatano (1982, 1988) aufgegriffen. Er postuliert in diesem Zusammenhang einen Unterschied zwischen „Routineexperten“ und „adaptiven Experten“, der darin besteht, dass nur eine bestimmte Art der Erfahrung tatsächlich Expertise generiert. Er nennt die Expertise, die Wissen/Können auf neue Probleme transferieren kann, adaptive Expertise. Die Konsequenzen dieser Unterscheidung wurden von Holyoak (1991) wie folgt charakterisiert: „Whereas routine experts are able to solve familiar types of problems quickly and accurately, they have only modest capabilities in dealing with novel types of problems. Adaptive experts, on the other hand, may be able to invent new procedures derived from their expert knowledge.“ (S. 310). Holyoak nimmt auch Bezug auf die Unterscheidung zwischen Fähigkeitenerwerb und „echter Expertise“, welche mit sich bringt, dass nur adaptive Experten (und nicht Routineexperten) als echte Experten bezeichnet werden sollten.

Kognitive Flexibilität wird von Krems (1994) als Fähigkeit des kognitiven Systems bezeichnet, welche zu einer besseren Passung der in der Problemlösung aufgrund wechselnder Situationsanforderungen eingesetzten bereichsspezifischen Kenntnisse führen sollte. Flexibilität in der Situationsinterpretation betrachtet er in drei Varianten: 1. Wechsel des Interpretationsschemas, 2. Wechsel des Analysemodus und 3. Wechsel der Strategie. Die Interpretationsvariabilität von Sachverhalten stellt auf die Fähigkeit der Interpretation eines Datensatzes (z.B. Symptome in der klinischen Urteilsbildung, Verhaltenseigenschaften) aus verschiedenen Perspektiven ab. Hierzu gehört z.B. in der klinischen Urteilsbildung die Fähigkeit, verschiedene diagnostische Kategorien und Erklärungen einzubeziehen und einzelne Sachverhaltselemente als Bestandteile unterschiedlicher Schemata betrachten zu können. Bezüglich der multiplen Sachverhaltsrepräsentation geht Krems (1994) davon aus, dass vorliegende Datensätze mit steigender Kompetenz nicht nur inhaltlich unterschiedlich interpretiert werden können, sondern dass eine Interpretation unter wechselnden

Gesichtspunkten leichter möglich ist – Experten sind zu Mehrfachdeutungen oder Mehrfachinterpretationen der Situation in der Lage. Dieser Aspekt wird auch in der Flexibilitätstheorie von Spiro (Spiro et al., 1991) hervorgehoben – danach sind Experten aufgrund ihrer Wissensrepräsentation in der Lage, einen Transfer zwischen unterschiedlichen Kontexten leichter vorzunehmen. Mehrfachinterpretationen, die aus einem Wechsel der Verarbeitungsstrategien erfolgen, sind als dritter Aspekt zu diskutieren. Zum einen differenziert sich mit zunehmender Erfahrung das Strategierepertoire – Anfänger verwenden eher globale, sachgebietsunabhängige Strategien. Mit wachsender Erfahrung werden zunehmend bereichsspezifische Heuristiken angewendet und allgemeine Problemlösestrategien treten in den Hintergrund (Joseph & Patel, 1986, Wontorra & Krems, 1989). Anfänger leiten ihr weiteres Vorgehen aus einer zu Beginn der Problemlösung erfolgten Zielanalyse ab. Experten hingegen versuchen die Situation unmittelbar zu erfassen und das Problem zu kategorisieren, woran sich dann ein vorlageorientiertes Vorgehen - von den Fakten zu den Lösungen hin - anschließt (Chi et al., 1981).

Diese Vorgehensweise ist jedoch nur bei Problemstellungen zu finden, die für Experten relativ einfach zu lösen sind. Sind kognitiv anspruchsvolle und komplexe Aufgaben, die auch für Experten schwierig und ungewohnt sind, zu lösen, kehren auch Experten zu den sachgebietsunabhängigen Strategien der Anfänger zurück (Joseph und Patel, 1986). Dörner et al. (1988) fanden heraus, dass Experten beim Lösen von ungewohnt schwierigen Aufgaben, wenn sie nur ihre allgemeinen Kenntnisse zur Organisation komplexer Systeme zur Verfügung haben, für ihre besseren Entscheidungen mehr Zeit und mehr Informationen als Anfänger benötigten. Auch Putz-Osterloh (1987, 1988) konnte finden, dass erfahrene Problemlöser die gesamte Komplexität der Situation besser berücksichtigen und ihr Vorgehen nicht an einzelnen Situationskriterien ausrichten. Insgesamt betrachtet scheinen erfahrenen Problemlösern quantitativ mehr Strategien zur Verfügung zu stehen, die sie auch flexibel und situationsabhängig einsetzen können. In der Forschungsliteratur sind jedoch auch widersprüchliche Befunde zu finden. Frensch und Sternberg (1989) fanden beispielsweise, dass sich vielschichtiges domänenspezifisches Wissen und eine hohe Automatisierung der angewandten Strategien dergestalt auswirkt, dass Probanden beim Problemlösen in neuen und ungewohnten Situationen wenig flexibel reagieren.

5.4. Ausgewählte empirische Befunde in der Urteilsbildung bei Probanden mit unterschiedlicher Erfahrung

5.4.1. Klinische Urteilsbildung

Es gibt zahlreiche empirische Befunde, die kontrastierende Verfahren in der Expertenforschung in semantisch vielschichtigen Domänen anwenden. Zu den intensiv studierten Domänen gehören vor allem Schach und Physik, daneben aber auch Programmieren, Mathematik und Denkspiele wie Go, Othello und Bridge. Allen diesen Domänen ist gemein, dass es deterministische Regeln und gesetzmäßige Zusammenhänge zur Lösung von Aufgaben in diesen Domänen gibt. Gerade das Phänomen der Schachexperten als ein Grundpfeiler der Forschung über kognitive Fähigkeiten wurde auch in jüngerer Zeit gut untersucht (z.B. Saariluoma & Laine, 2001; Saariluoma, 1995; Laine, Hyötyniemi & Saariluoma, 1998).

Expertenstudien in Domänen, in denen es um Urteils- und Entscheidungsaufgaben geht, in der die Probanden aufgrund von gegebenen Daten (z.B. medizinisches Symptom, Verhalten eines Menschen) auf eine nicht direkt beobachtbare Gegebenheit (Krankheit, Eigenschaft des Menschen) schließen müssen, sind weniger zahlreich. Diese Schlüsse müssen unter Unsicherheit gezogen werden, da eine deterministische, fehlerfreie Regel nicht bekannt ist (Fiedler & Plessner, 2004). Nachfolgend sollen einige empirische Befunde aus Expertenstudien zu solchen Urteils- und Entscheidungsaufgaben dargestellt werden.

Die Verarbeitung von klinisch-diagnostischen Personinformationen wird im generellen Sinne als soziale Informationsverarbeitung betrachtet, so dass für den Prozess der Diagnoseerstellung und der Beurteilung die Wirksamkeit der aus der sozialen Kognition bekannten Prinzipien, von den einige oben erläutert wurden, angenommen wird (Dawson, Zeitz & Wright, 1989, Krolak-Schwerdt & le Coutre, 2002).

In Studien zur klinischen Urteilsbildung mit Medizinerinnen wird üblicherweise ein Ausschnitt aus dem Prozess des Diagnostizierens betrachtet, z.B. werden Röntgenbilder oder Befundberichte vorgelegt, aufgrund dessen die Diagnoseerstellung erfolgt (z.B. Johnson, 1982, Mehle, 1982, Joseph und Patel, 1986).

In der nachfolgend skizzierten Studie von Elstein, Shulman und Sprafka (1978) wird der gesamte diagnostische Prozess – angefangen von der Erhebung anamnestischer Daten zur Vorgeschichte des Patienten über die körperliche Untersuchung durch den Arzt bis hin Entwicklung einer Vorstellung zur Ätiologie des Falles und zur Diagnosestellung – betrachtet. Das Ziel ihrer zwischen 1969 und 1973 in Michigan (USA) mit Klinikärzten und niedergelassenen Ärzten durchgeführten Untersuchungsserie war die Beschreibung von Diagnosen erfahrener Ärzte durch Prozessanalysen. An der Untersuchung nahmen 24 Ärzte mit Berufserfahrung teil, innerhalb dieser Gruppe wurden die besseren Diagnostiker durch Kollegenbefragung identifiziert. Als Patienten traten trainierte Schauspieler mit einer eingeübten Symptomatik auf, denen zusätzlich Antworten auf mögliche Fragen der Ärzte vermittelt wurden. Die medizinische Befunderhebung wurde auf Videoband aufgezeichnet, die Ärzte sollten während der Untersuchung laut denken. Im Anschluss an die Untersuchung wurde der Arzt mit Hilfe des Videobandes zu seinem Gedankengang befragt. In der Auswertung sollten Fragen nach der Menge der erfragten Informationen, nach dem Zeitpunkt der Hypothesenbildung und nach der Anzahl der gebildeten Hypothesen und dem Zusammenhang mit der Häufigkeit von richtigen Diagnosen beantwortet werden.

Die Autoren fanden heraus, dass die Maße der Informationsaufnahme und Hypothesenbildung während des Untersuchens stark zwischen den Personen und je nach medizinischem Fall variierten. Ein Zusammenhang zwischen der Beurteilung der Kollegen als guter Diagnostiker und der Richtigkeit von Diagnosen fand sich nicht. Bei der Untersuchung der Fälle mit korrekter Diagnosestellung unterschied sich jedoch die Bewertung der Informationen durch die untersuchenden Ärzte. Bei korrekter Diagnose wurden im Durchschnitt - und nicht bei allen medizinischen Fällen - die geschilderten Symptome und Befunde genauer interpretiert und es wurden auch mehr Daten einbezogen. Die Erwartung von Elstein et. al (1978), unabhängige Variablen der Informationsverarbeitung zu identifizieren, die eine Unterscheidung von guten und weniger guten Diagnostikern erlaubt, erfüllte sich nicht. Sie fanden, dass die Hypothesenbildung bereits sehr früh, bei der Befunderhebung, einsetzt. Zum Teil wurden als erstes Hypothesen über allgemeine Krankheitskategorien gebildet, die dann mit fortschreitender Untersuchung spezifiziert wurden. In anderen Fällen wurde bereits bei der ersten Hypothese mit einer ganz speziellen Unterform einer Krankheit begonnen. Dies variierte sowohl zwischen den medizinischen Fällen als auch zwischen den diagnostizierenden Ärzten. Die Untersuchung von Elstein et. al (1978) wurde zwar wegen verschiedener erhebungs- und auswertungsmethodischer Schwächen kritisiert (Kuipers &

Kassirer, 1984), die Ergebnisse konnten jedoch mit wenigen Ausnahmen bestätigt werden (Z.B. Barrows et al., 1982; Johnson et al. 1981b).

Johnson et al. (1981b) haben die Entstehung einer Diagnose bei Vorliegen eines schwierigen medizinischen Falles untersucht. Sie konnten zeigen, wie bereits in der Phase der Problemrepräsentation falsche Vermutungen geäußert werden, die dann zu einer falschen Gewichtung von Diagnosedaten führen. Wiederum waren nicht die Anzahl der Hypothesen oder die Globalität der Hypothesen sondern die richtige Gewichtung der Daten und ihre Interpretation sowie die Qualität der Hypothesen das Kriterium, welches erfolgreiche von weniger erfolgreichen Diagnostikern trennte.

In Replikationen der Experimente von Chase und Simon (1973a, b) mit medizinischen Befundberichten, bei denen ärztliches Wissen von Probanden mit verschiedener Erfahrung (Medizinstudenten und praktizierende Internisten) analysiert wurde, erhielten Norman, Jacoby, Feightner & Campell (1979) und Muzzin et al. (1982) interessante Ergebnisse. Den Probanden wurden Befundberichte vorgelegt und sie wurden um eine möglichst genaue Wiedergabe gebeten. Die nach den Ergebnissen von Chase und Simon (1973a, b) zu erwartenden Unterschiede in der Wiedergabemenge der Falldaten traten beim wörtlichen Erinnern nicht auf. Unterschiede zwischen erfahrenen Ärzten und Medizinstudenten zeigten sich erst, nachdem die Schlussfolgerungen, die die Probanden aus den Daten gezogen hatten berücksichtigt wurden und als zwischen diagnostisch wichtigen und wenig bedeutsamen Informationen unterschieden wurde.

Auch Schmidt, Boshuizen & Hobus (1988) und Schmidt & Boshuizen (1990) erhielten ähnliche Befunde wie Muzzin et al. (1982). Anfängern, Medizinstudenten höheren Semesters und erfahrenen Medizinerinnen wurden Befundberichte vorgelegt und sie wurden um eine möglichst genaue Wiedergabe und anschließend um eine Diagnose gebeten. Die Güte der Diagnosen korrelierte mit der Erfahrung. Beim Erinnern an die Fallberichte schnitten jedoch die älteren Medizinstudenten besser ab als die erfahrenen Ärzte. Dieser Effekt wird damit erklärt, dass Medizinstudenten höheren Semesters ihr pathophysiologisches Wissen beim Lesen und Wiedergeben anwenden und deshalb Details besser erinnern, während erfahrene Mediziner den aktuellen Fall als Beispiel eines bestimmten Krankheitsbildes erkennen und nur dieses behalten.

In einem Sammelreferat zur klinischen Urteilsbildung, zum klinischen Training und zur Berufserfahrung betrachtet Garb (1989) Ergebnisse aus 55 klinischen Studien, in denen zwischen erfahrenen Klinikern und weniger erfahrenen Klinikern, Klinikern und Absolventen des Medizinstudiums und Klinikern oder Absolventen des Medizinstudiums und Urteils Laien verglichen wird. Die Ergebnisse für den Zusammenhang von Erfahrung und Validität des Urteils waren enttäuschend. In Studien zur Persönlichkeitseinschätzung waren erfahrene Klinikern nicht präziser als weniger Erfahrene, unabhängig davon, in welcher Form den Probanden die Daten vorgelegt wurden. Zum Beispiel fand Kendell (1973) in einer Studie, in welcher Psychiater Interviews mit psychiatrischen Patienten beobachteten, dass die Fähigkeit, valide Diagnosen zu erstellen, nicht mit der Länge der Berufserfahrung in der Psychiatrie korrelierte. Weitere Studien, in denen die vermutete Expertise nicht mit der Urteilsvalidität korreliert, und in denen erfahrene Klinikern nicht korrekter als weniger erfahrene Klinikern urteilten, werden angeführt (z.B. Werner, Rose & Yesavage, 1983, Levenberg, 1975, Watson, 1967, Wanderer, 1969).

Die Ergebnisse für den Zusammenhang zwischen Training und Urteilsvalidität waren positiver. In drei Studien wurden von fortgeschrittenen Medizinabsolventen signifikant mehr valide Urteile abgegeben als von Anfänger-Medizinabsolventen (Grebstein, 1963, Sines, 1959, Aronson und Akamatsu, 1981). Aronson und Akatmatsu (1981) berichten, dass 12 Medizinstudienabsolventen signifikant mehr valide Zuordnungen von Persönlichkeitsmerkmalen von Patienten machten, nachdem sie einen umfassenden Kurs über den Gebrauch und die Interpretation des MMPI (Minnesota Multiphasic Personality Inventory) absolviert hatten. Garb (1989) kommt zu dem Schluss, dass die Signifikanz dieser Ergebnisse unter bestimmten Bedingungen den Schluss zulässt, dass Training Urteiler in die Lage versetzt, validere Urteile abzugeben. In 10 von 30 Studien fand sich das Ergebnis, dass im Vergleich zwischen Klinikern, Absolventen und Medizinstudenten Urteiler mit Training im medizinischen Feld signifikant validere Urteile, z.B. für das Ausmaß einer vorliegenden Depression bei psychiatrischen Patienten (Waxner, 1976) abgeben können als Laien. Interessant sind jedoch auch die empirischen Befunde von Griswold & Dana (1970) und Oskamp, (1965) die zu diesen Ergebnissen im Widerspruch stehen. Wurden klinischen Psychologen und Studenten biographische Daten von psychisch gesunden Personen zur Einschätzung vorgelegt, fanden sich keine signifikanten Unterschiede in der Korrektheit. Klinikern waren nur dann akkurater als Studenten, wenn sie Urteile für psychiatrische Patienten abgeben sollten oder Einschätzungen des Suizidrisikos vorgenommen werden

sollten. Garb (1989) erklärt diese Befunde damit, dass Kliniker bessere Einschätzungen für psychiatrische Patienten als für psychisch gesunde Menschen vornehmen können, da sie normalerweise im beruflichen Kontext keine Urteile über Personen abgeben müssen, die keine Behandlung benötigen und bei denen kein Risiko einer Selbst- oder Fremdgefährdung infolge Tötungsabsicht besteht. Er selbst fand Belege (1984, 1986), dass Kliniker bessere Konfidenzeinschätzungen abgeben, wenn biographische Daten für psychiatrische Patienten vorgelegt werden, als wenn es sich um Daten von psychisch gesunden Personen handelt.

Unter Bezugnahme auf das Kontinuum-Modell von Fiske & Neuberg (1990) brachte ein Experiment von Krolak-Schwerdt & le Coutre (2002) zum Prozess klinischer Diagnose und Beurteilung Befunde zur unterschiedlichen Verarbeitung von Personeninformationen des Gedächtnisses von Experten (klinische Psychotherapeuten), Novizen (Diplomanden der klinischen Psychologie) und Laien (Studenten der Naturwissenschaften). Danach, so postulieren Krolak-Schwerdt & le Coutre (2002), hat sich bei klinischen Experten infolge der gesammelten Erfahrung eine kognitive Flexibilität herausgebildet, die sie in die Lage versetzt, entsprechend des Verarbeitungszieles (Eindrucksbildung versus Verhaltensvorhersage versus präzise Erinnerung) zwischen kategoriengeleiteten und einer merkmalsbasierten Verarbeitung hin- und herzuwechseln. Unter dem Ziel der Eindrucksbildung reproduzierten Experten erheblich mehr Merkmale, die konsistent zu der mit dem Stimulusmaterial assoziierten psychopathologischen Kategorie sind, als unter den Zielen Verhaltensvorhersage oder präzise Erinnerung. Gleichzeitig trat bei den Experten ein deutlich erhöhter Anteil an Intrusionen, vor allem falsch erinnerte Symptome und nicht dargebotene Diagnosen, auf, die konsistent zur zugrunde liegenden Krankheitskategorie waren. Auch bei den Novizen findet sich eine erhöhte Intrusionsrate für Attribute, die konsistent zur zugrunde liegenden Krankheitskategorie sind gegenüber der Laiengruppe, was auf eine Aktivierung von klinischen Kategorien, die bei den Novizen infolge Wissensaneignung bereits angelegt sind, zurückgeführt wird. Die prognostizierte Flexibilität der Experten zeigt sich an unterschiedlichen Gedächtnisbefunden unter den verschiedenen Instruktionen: unter dem Ziel der Verhaltensvorhersage reproduzierten die Experten vor allem solche Symptome, die inkonsistent zur zugrunde liegenden Krankheitskategorie waren. Diese Befunde untermauern den von Krolak-Schwerdt & le Coutre (2002) postulierten „Schnelligkeits-/Genauigkeits-Abgleich“ (S. 442), da unter anderer Instruktion die Aufmerksamkeit der Experten von kategoriengeleiteter Verarbeitung auf die individuellen Merkmale verlagert wurde und merkmalsorientierte Verarbeitungsprozesse eingeleitet wurden.

5.4.2. Urteilsbildung von Pädagogen

In Expertenstudien zu Lehrern untersuchten Berliner und Mitarbeiter (Berliner & Carter, 1986, Berliner, 1987a, b, Berliner, Stein, Sabers, Clarridge, Cushing & Pinnegar, 1988, Berliner, 1990) die interpretierende Wahrnehmung von Experten und Anfängern. Als Experten wurden 17 von 55 vorgeschlagenen Lehrern mit mindestens 5 Jahren Berufserfahrung, die von ihren Schuldirektoren als herausragend benannt wurden und die zusätzlich von drei Projektmitarbeitern im Unterricht beobachtet und beurteilt wurden, ausgewählt. Als Anfänger wurden Pädagogen im ersten Berufsjahr, die in der Lehrerausbildung gute Beurteilungen erhalten hatten, untersucht. Als Anwärter wurden Naturwissenschaftler und Ingenieure ohne pädagogische Ausbildung, die sich aber für den Beruf des Lehrers interessierten und die aufgrund Lehrermangels tatsächlich in den Schuldienst eingestellt werden sollten, ausgewählt. Berliner und Carter (1986) fertigten Dias von Unterrichtssituationen aus einer Mathematikstunde und einer Chemiestunde an, die beispielsweise Schülergruppen bei der Arbeit zeigten. Die Bilder wurden den Probanden jeweils eine Sekunde dargeboten und die Probanden sollten anschließend das Gesehene beschreiben. Die Beschreibungen der Probanden waren alle recht exakt, jedoch gab es Unterschiede im Allgemeinheitsgrad der Wiedergabe. Anfänger und Anwärter gaben beobachtbare Details der Dias wieder, z.B. die Zahl und das Geschlecht der Schüler, die Haarfarbe, die genaue räumliche Anordnung der Schüler im Klassenzimmer und Unterrichtsmaterialien. Einige der Lehrerexperten antworteten mit Allgemeinbegriffen, die die vermutete Methode der Unterrichtsstunde beschreiben. Die Erinnerungen bezogen sich auf die Schüler und ihre Arbeitsaktivitäten (Gruppenarbeit), während die Anfänger von Schülern, die miteinander reden, sprachen.

In einem weiteren Experiment (Berliner und Carter, 1986) wurden den Probanden die Bilder jeweils drei Sekunden dargeboten, diese Darbietung wurde dreimal wiederholt. Die Probanden sollten nun das Gesehene aufschreiben und bei jedem Durchgang ergänzen. Die Wiedergabe der Experten bezog sich häufig darauf, ob die dargestellten Situationen typisch oder untypisch waren. Sie entwickelten Hypothesen, was in dem dargestellten Unterrichtssituationen gerade passiert sein könnte und bezogen diese Vermutungen auf ihre eigenen Erfahrungen. Diese Befunde fanden sich auch im dritten Untersuchungsschritt. Die Probanden sollten eine Abfolge von 50 Dias, die einen Unterrichtsverlauf abbildete, interpretieren. Die Lehrerexperten sahen und kommentierten typische Ereignisse im

Unterrichtsgeschehen, z.B. kleine Störungen, Gruppenarbeit in großen Gruppen. Auch beim vierten Experiment zeigten sich deutliche Unterschiede zwischen den Probandengruppen hinsichtlich der kognitiven Gliederung des Unterrichtsgeschehens. Bei der Vorführung einer Serie von Bildern bestimmten die Probanden selbst, wann sie die Vorführung stoppen und kommentieren wollen. Die Experten stimmten besser überein, welche Unterrichtsaspekte sie kommentierten (Aufmerksamkeit, Zeit, Abfolge) und wie eine dargestellte Situation zu interpretieren war (z.B. ob die Mitarbeit des Schülers gerade gut oder schlecht war). Neben den gefundenen Unterschieden zwischen den Gruppen, dass Experten aufgrund ihres Erfahrungswissens um typische Ereignisse des Unterrichts Situationen besser erkennen, dass sie verallgemeinern und in ihre Vermutungen, was in der dargestellten Situation gerade passiert sein könnte, eigene Erfahrungen einfließen lassen, fanden sich auch große interindividuelle Unterschiede innerhalb der Gruppen. Bromme (1992) führt dies darauf zurück, dass die dargebotenen Bilder nicht genug Informationen enthalten, um Unterschiede zwischen den Gruppen noch deutlicher zu machen.

In einer weiteren Studie mit Pädagogen (Carter, Sabers, Cushing, Pinnegar & Berliner, 1987), in der die Verarbeitung schriftlicher Informationen über Schüler, die Klasse und den bisherigen Unterrichtsablauf zwischen Experten und Anfängern untersucht wurde, wurde den Pädagogen mitgeteilt, dass sie eine Klasse mitten im Schuljahr übernehmen und fortführen sollten. Die Probanden erhielten Karten mit Schülerdaten (Noten, Anwesenheit, Kommentare, demografische Daten) sowie die Tests, das Lehrbuch der Klasse und Aufgabenblätter und die Aufgabe, innerhalb von 40 Minuten den Unterricht vorzubereiten, danach Auskunft zu den Schülern zu geben, diese zu gruppieren und einen Sitzplan zu entwerfen. Beide Gruppen – Anfänger und Experten – lehnten es ab, Details zu einzelnen Schülern anzusehen, jedoch mit unterschiedlichen Begründungen. Die Experten nannten unterrichtspraktische Gründe für ihre Ablehnung – sie verwiesen darauf, dass in der ersten Stunde die Aufgabe darin bestünde, sich selbst vorzustellen und ihre Erwartungen über den üblichen Unterrichtsverlauf zu vermitteln und nicht darin, sich mit einzelnen Schülern zu befassen. Anfänger nannten als Grund, dass sie den Schülern ohne Vorurteile gegenüberzutreten wollten. Die Kommentare aller Probanden zu den Informationen der Schülern enthielten Kategorisierungen, z.B. „der gute Schüler“ oder „schüchterne Schüler“. Mit diesen Begriffen verbanden Experten eher Vorstellungen über mögliche Verhaltensweisen gegenüber den Schülern, während die Anwarter mit der Kategorisierung keine Prozess- oder Handlungsvorstellungen verknüpften. Zum Abschluss der Untersuchung wurde die Erinnerung der Probanden an das dargebotene Datenmaterial

überprüft, bei dem sich die Anfänger und Anwarter den Experten als klar überlegen zeigten. Die Experten erinnerten sich nur schlecht an Details, jedoch konnten sie Daten, die für die Unterrichtsgestaltung unmittelbar bedeutsam sind, gut erinnern (z.B. dass ein Schüler eine Lernbehinderung hat). Die Anwarter erinnerten sich an viele Details und waren bemüht, keine Information zu vernachlässigen. Daraus folgern die Autoren, dass die Anfänger und Anwarter weniger zwischen den für ihre Aufgabenbewältigung wichtigen und unwichtigen Informationen unterscheiden können. Ähnliche Befunde wurden auch in Studien zur klinischen Urteilsbildung gefunden, über die oben bereits berichtet wurde.

Einen weiteren empirischen Beweis, dass erfahrene Pädagogen besser zwischen wichtigen und unwichtigen Aspekten beim Beurteilen einer Situation trennen können, liefern die Ergebnisse von Morine & Vallance (1975). Sie befragten Lehrer, die als unterschiedlich erfolgreich eingestuft worden waren, anhand einer Videoaufnahme ihres Unterrichtes im Anschluss an den Unterricht zu den Überlegungen, die ihren Entscheidungen während des Unterrichtes zugrunde gelegen haben. Interessanterweise nannten die weniger erfolgreichen Lehrer mehr Aspekte als die erfolgreichen Lehrer.

Calderhead (1981, 1983) fand ebenfalls Unterschiede von erfahrenen und unerfahrenen Lehrern beim Beurteilen einer Unterrichtssituation. Er gab den Probanden Beschreibungen von Situationen im Klassenzimmer (z.B. eine Störung durch einen Schüler während einer Stillarbeitsphase) und fragte dann, welche weiteren Daten sie benötigen würden, um die Lage beurteilen zu können und eine Entscheidung über das weitere Vorgehen zu treffen. Die erfahrenen Pädagogen erfragten Informationen zur Differenzierung des Schülertyps, während die unerfahrenen Lehrer z.B. anführten, dass sie die zu ergreifende Maßnahme vom Lautstärkegrad abhängig machen würden.

5.4.3. Urteilsbildung in anderen Domänen

Abgesehen von gut untersuchten Domänen wie z.B. Schach scheint die Expertiseforschung in jüngerer Zeit etwas in den Hintergrund getreten zu sein. Für spezielle Gebiete finden sich einzelne Untersuchungen, die sich in weite Untersuchungsfelder psychologischer Grundlagenforschung einfügen (z.B. Harvey, 1992; Clarke & Lamberts, 1997).

Zur Entwicklung von adaptiver Expertise (Holyoak, 1991) beispielsweise vergleichen Barnett und Koslowski (2002) zwei Gruppen mit Erfahrung in unterschiedlichen Domäne (Business Consultants und Restaurant-Manager) und nehmen damit auch Bezug auf die oben beschriebene kognitive Flexibilität, die sich bei Experten ausbildet. Bei der Bewältigung von Aufgaben aus dem Bereich eines Restaurants ergab sich, dass die Business Consultants besser abschnitten als die Restaurant-Manager, obwohl sie nicht die einschlägige Erfahrung im Führen eines Restaurants hatten. Barnett und Koslowski (2002) führen dies u.a. darauf zurück, dass die Business Consultants im täglichen Arbeitskontext mit stets wechselnden Problemen und Anforderungen konfrontiert sind, die sie zu meistern haben und zu deren Lösung sie eine analytischere Annäherung an die Aufgaben benötigen, was dann zu einem tieferen Verständnis von wirtschaftlichen Zusammenhängen usw. führt. Im Gegensatz dazu besteht die Führung eines Restaurants aus einer Fülle praktischer, täglich wiederkehrender Aufgaben, die zu erledigen sind, wodurch Routine – aber nicht dieses tiefe Verständnis für eine analytische Annäherung an Probleme ausgebildet wird.

Eine sehr interessante Studie zur Urteilsbildung findet sich bei English & Mussweiler (2001). Aufbauend auf der Forschung über Urteilsanker (Tversky & Kahnemann, 1974) wurde untersucht, ob eine Urteilempfehlung an den Richter einen direkten Einfluss auf das von ihm abgegebene Urteil hat. Es konnte gezeigt werden, dass dieser Ankereffekt bei Richtern tatsächlich existiert: Urteilsentscheidungen werden durch die geforderte Strafe des Staatsanwaltes beeinflusst. In einer weiteren Studie wurde gezeigt, dass dieser Einfluss auch unabhängig von der durch den Urteiler wahrgenommenen Relevanz des empfohlenen Strafmasses ist. Es konnte nachgewiesen werden, dass selbst erfahrene Richter, die ihr Richteramt 15 Jahre ausüben, dem Ankereffekt in der gleichen Größenordnung wie Berufsanfänger unterliegen. Von den 16 untersuchten Richtern gab lediglich ein Richter an, dass das von außen geforderte Urteil (Anker) sein Urteil beeinflusst hätte. Alle anderen nahmen das Urteil eines Außenstehenden als irrelevant für ihr eigenes Urteil wahr, obwohl sie sich nachweislich davon leiten ließen.

Preckel & Schüpbach (2005) konnten in einer aktuellen Studie, die unter Realbedingungen in einem international tätigen Unternehmen durchgeführt wurde, signifikante Zusammenhänge zwischen der Fähigkeit der Assessment Center-Teilnehmer, die im AC beobachteten und beurteilten Anforderungen korrekt zu erkennen und ihr Verhalten im Sinne einer Selbstdarstellung darauf auszurichten und der von den Beobachtern eingeschätzten Leistung

finden. Es konnte nachgewiesen werden, dass die Erkennensfähigkeit der Teilnehmer für die zu zeigenden Leistungen intraindividuell relativ stabil und unabhängig von kognitiven Fähigkeiten und Persönlichkeitsfaktoren ist. Eine gute Erkennensfähigkeit der AC-Teilnehmer und die entsprechende Selbstdarstellung wirkten sich positiv auf die Beobachterurteile aus. Wenn man bedenkt, wie wichtig die Entscheidungen der AC-Beobachter für das Unternehmen (Püschel, 2005, Obermann, 2002) und für den einzelnen Bewerber sind, stimmt es schon nachdenklich, wenn professionelle Beurteiler die Leistung eines Bewerbers unabhängig von seinen kognitiven Fähigkeiten und Persönlichkeitsfaktoren signifikant besser einschätzen, wenn es dem Bewerber gelingt, die erwarteten Verhaltensweisen zu erkennen und diese entsprechend umzusetzen.

5.5. Zusammenfassung und Schlussfolgerungen für die eigene Arbeit

Die Befunde aus Untersuchungen zur Spezifik der Unterschiede in der Urteilsbildung zwischen Laien und Experten in semantisch vielschichtigen Domänen konvergieren darin, dass sich aus der langjährigen formalen Ausbildung und im Verlauf einer umfangreichen Erfahrungssammlung eine Expertise herausbildet, die aus einer gut organisierten und integrierten Wissensbasis über domänenspezifische Kategorien besteht (Krolak-Schwerdt & le Coutre, 2002, Dawson, Zeitz & Wright, 1989). Der Begriff der kategorialen Wahrnehmung (Harnad, 1987) kennzeichnet die spezielle Anwendung von Wissen von Experten sehr gut, die häufig von den Experten als „intuitiv“ und vom Wissen selbst nicht zu trennen, erlebt wird (Bromme, 1992).

Im klinischen Bereich z.B. wird mit der Entwicklung von Expertise die Vielzahl der klinischen Kategorien zunehmend in einem hierarchischen Netzwerk integriert. Nach Genero & Cantor, (1987) ist eine klinische Kategorie wie z.B. „Schizophrenie“ wie ein Schema oder Prototyp in sich strukturiert und vereint ein Bündel typischer Symptome einer psychischen Störung. Auf der Basis dieses Spezialwissens entwickeln sich bei den Experten spezifische Verarbeitungsprozesse, die dadurch gekennzeichnet sind, dass Experten klinische Merkmalsinformationen abstrakter als Novizen enkodieren, weil sie auf übergeordnete Kategorien zurückgreifen können (Dawson, Zeitz & Wright, 1989). Werden klinische Symptome als Information dargeboten, werden beim klinischen Experten schnell verfügbare Kategorien aktiviert, weil sie bereits früher sehr häufig aktiviert wurden (Krolak-Schwerdt &

le Coutre, 2002). Showers & Cantor (1985) postulieren, dass klinische Experten in hohem Maße schemageleitete Inferenzen und Generalisierungen auf der Grundlage der dargebotenen Merkmalsinformation vornehmen. Diese werden dann als abstrakte Konzepte dem Stimulusmaterial hinzugefügt und mit diesen verknüpft gespeichert.

Bei der Verarbeitung des gegebenen Stimulusmaterial setzt die Hypothesenbildung bereits sehr früh, bei der Befunderhebung ein (Elstein et al., 1978). Bereits in der Phase der Problemrepräsentation kommt es auf die Gewichtung der Daten und ihre Interpretation an, um die „richtigen“ Hypothesen zu generieren (Johnson et al. 1981a).

Die Aktivierung einer bestimmten Kategorie bei der Urteilsbildung führt bei den Experten dazu, dass sie sich beim Replizieren nicht an viele Einzelheiten eines Fallberichtes erinnern, sondern nur an ein Exemplar eines Krankheitstyps (Schmidt & Boshuizen, 1990).

Berufsanfänger hingegen sind bemüht, kein Detail auszulassen.

Diese Befunde fanden sich auch bei den Pädagogen. Anfänger und Anwärter konnten nach Darbietung von Dias mit Unterrichtssituationen beobachtbare Details genau wiedergeben, während erfahrene Pädagogen Allgemeinbegriffe verwendeten und vermutete Unterrichtsmethoden und typische Unterrichtssituationen erkannten und beschrieben (Berliner & Carter, 1986).

Klinischen Experten gelingt ein flexibler Wechsel zwischen kategorien- und merkmalsbasierter Verarbeitung von Informationen in Abhängigkeit vom Verarbeitungsziel. (Krolak-Schwerdt & le Coutre, 2002) Unter der Instruktion der Eindrucksbildung wurden von Experten mehr Merkmale, die konsistent mit der aktivierten Krankheitskategorie sind, reproduziert als von den Novizen. Gleichzeitig traten jedoch auch die höchsten Anteile an Intrusionen auf, die konsistent mit der zugrunde liegenden Krankheitskategorie sind. Ein Großteil der Intrusionen waren falsch erinnerte und nicht dargebotene Symptome und Diagnosen, was für eine Dominanz von psychopathologischen Kategorien bei den Experten unter dem Verarbeitungsziel Eindrucksbildung spricht (Krolak-Schwerdt & le Coutre, 2002). Ein wesentliches Merkmal, welches die gesamte Expertiseforschung durchzieht und was bereits am Anfang des Kapitels diskutiert wurde, ist, dass der Vorsprung der Experten vor den Novizen oder Laien stets domänenspezifisch ist (Gruber, Mandl & Renkl, 1993). Experten nutzen zur Lösung von bekannten Aufgabenstellungen ihr bereichsspezifisches Wissen und bereichsspezifische Heuristiken. Sie können auf ein Strategierepertoire zurückgreifen, welches sie flexibel und situationsabhängig einsetzen. Sollen Experten jedoch ungewohnte Aufgaben

bewältigen, reagieren sie in diesen Situationen wenig flexibel (Frensch & Sternberg, 1989). Die Nutzung von bereichsspezifischen Heuristiken als kognitives Werkzeug (Fiedler, 1997) hat aber auch das Problem, dass unter bestimmten Bedingungen Urteile dadurch systematisch verzerrt werden. Diesen Urteilsverzerrungen können Novizen und Experten gleichermaßen unterliegen, wie z.B. die Studie über Ankereffekte bei Richtern mit langjähriger Berufserfahrung zeigt (Englich & Mussweiler, 2001).

Ausgehend von der Befundlage, dass sich bei Experten aus einer langjährigen Ausbildung und der umfangreichen Erfahrung eine domänenspezifische Expertise herausbildet, kann angenommen werden, dass es sich bei Pädagogen aufgrund ihres mehrjährigen Studiums und nach mehreren Jahren Berufspraxis um eine auf den Kontext Unterricht (Gruber, Mandl & Renkl, 1993) bezogene Expertise handelt. Während des Unterrichtes sind Pädagogen ständig gefordert, soziale Beurteilungen vorzunehmen - sie bewerten die Leistung von einzelnen Schülern, die Entwicklung von Schülern über einen längeren Zeitraum hinweg (z.B. Schuljahr), sie schätzen ein, ob der Lernstoff auch von schwächeren Schülern erfasst wurde und ob darauf aufbauend neuer Stoff vermittelt werden kann, sie bewerten Unterrichtssituationen und entscheiden, wie sie darauf reagieren, um nur einige Beispiele zu nennen. Pädagogen verfügen aufgrund ihres mehrjährigen Studiums bereits über ein umfangreiches Wissen, wie Urteile über Personen gebildet werden, wie man möglichst Urteilsfehler ausschließt und wie man auf bestimmte Situationen reagieren kann (Bromme, 1992). Im Laufe des Berufslebens entwickeln sich bei den Pädagogen kognitive Strukturen, die es ihnen ermöglichen, in Beurteilungssituationen wichtige Informationen von unwichtigen Informationen zu trennen (Carter, Sabers, Cushing, Pinnegar & Berliner, 1987).

Es ergibt sich die Frage, ob Pädagogen als Experten auch unabhängig vom Unterrichtskontext anders soziale Urteile bilden als Menschen, die sich nicht täglich professionell mit der Personenbeurteilung beschäftigen. Ausgehend von den Ergebnissen von Barnett & Koslowski (2002) lässt sich postulieren, dass sich bei den Pädagogen eine domänenspezifische Expertise entwickelt hat, die dazu führt, dass die kategoriale Wahrnehmung von Personen im Unterrichtskontext besonders gut ausgebildet ist, dass jedoch die Übertragung des Wissens über Personenbeurteilung auf andere Situationen nur schwer möglich ist. Die von Holyak (1991) postulierte kognitive Flexibilität, die es ermöglicht, Wissen auf andere Bereiche zu übertragen, dürfte bei den Pädagogen nur schwach ausgeprägt sein, weil „the everyday goal is not to build a specific skill but to get a job done“ (Barnett & Koslowski, 2002, S. 258).

5.6. Training als Möglichkeit der Veränderung der Personenbeurteilung

Dass bereichsspezifisches Wissen und zahlreiche, langjährige Erfahrungen bei Experten dazu führen können, in ihrer Domäne valide Urteile abzugeben, wurde in den vorigen Abschnitten ausführlich dargestellt. Jedoch können auch Experten zahlreichen Urteilsfehlern (z.B. English & Mussweiler, 2001) unterliegen und nicht immer führt eine Expertise auch zu richtigeren Urteilen oder Diagnosen (z.B. Elstein et. al, 1978; Garb, 1989). Als Kernpunkt zur Verringerung von Beurteilungsfehlern und –verzerrungen nennt Kanning (1999), „daß *bestehende Automatismen unterbrochen werden und an ihre Stelle eine ganz bewusst vorgenommene Beurteilung gestellt wird*“ (S. 270, Hervorhebung im Original). Ein gezieltes Training verbunden mit der Bereitschaft der Trainingsteilnehmer, sich kritisch mit dem eigenen Beurteilungsverhalten auseinanderzusetzen und dieses zu verändern, kann zu einer objektiveren Beurteilung anderer Menschen beitragen (Kanning et. al, 2004).

Ein solches praktisches Training muss zum einen Kenntnisse über Phänomene der Urteilsbildung und systematischer Urteilsverzerrungen sowie deren zugrunde liegenden allgemeinen Prozesse vermitteln und auf der anderen Seite mögliche Gegenmaßnahmen aufzeigen. Diese bestehen im Wesentlichen darin, dass routinierte, bei alltäglichen Beurteilungen automatisch ablaufende Prozesse hinterfragt und bewusst gesteuert werden. Kennt man die möglichen Fehler bei der Personenbeurteilung und Regulationsmöglichkeiten, kann das Ausmaß von Beurteilungsfehlern als Folge unzureichender Wahrnehmung und Informationsverarbeitung zumindest begrenzt werden (Werth, 2004). Allein die theoretische Aneignung von Wissen über Urteilsverzerrungen wird für eine Verringerung dieser unerwünschten Effekte nicht ausreichen. Um ein bewusstes Hinterfragen und Steuern der Urteilsbildung anzuwenden, müssen Gegenmaßnahmen praktisch geübt werden, damit sie auch im Alltag selbstverständlich und schnell angewendet werden können (Kanning, 1999).

Dass verschiedene Trainingsformen die Personenbeurteilung verbessern können, konnte z.B. Kamphaus (1980) zeigen. Es wurden drei verschiedene Trainingsgruppen gebildet: eine Gestalttherapie-Gruppe, eine Buch-/Lektüreprogramm-Gruppe und eine Seminargruppe. Nach einem Trainingswochenende (Gestaltgruppe) bzw. einer Trainingswoche (Lektüregruppe) verbesserte sich die Urteilsdifferenziertheit von allen drei Gruppen signifikant gegenüber einer Kontrollgruppe, die an keinem Training teilgenommen hatte. Die Lektüregruppe wies die höchste Differenziertheit auf. Es zeigte sich, dass sich die Beobachtung generell

intensivierte und nicht nur bezogen auf die in der Untersuchung erfragten Merkmale beschränkt blieb. Die Trainingsteilnehmer der Gestaltgruppe gaben an, dass der Zuwachs an Differenzierungsvermögen ihrer aktiven Trainingsteilnahme und einem gewissen Maß an Verunsicherung geschuldet war.

Auch Kanning (1999) postuliert für einen Trainingserfolg die Notwendigkeit, dass das subjektive Sicherheitsgefühl in Bezug auf die Korrektheit der eigenen Urteile kritisch hinterfragt werden muss. Untersuchungen zur subjektiven Urteilssicherheit zeigten, dass Menschen von der Richtigkeit ihrer eigenen Urteile auch dann überzeugt sind, wenn sie sich als objektiv falsch erweisen (Hoffrage, 1993). Unabhängig von der tatsächlichen Güte ihres Urteils empfinden Menschen ihre Entscheidung häufig als wahr oder richtig. Diese subjektive übersteigerte Urteilssicherheit findet sich auch bei Experten (z.B. Garb, 1989; Englich & Mussweiler, 2001), sie ist jedoch kein Kriterium für die Wahrheit eines Urteils. Eine Sensibilisierung für dieses Phänomen kann als Regulationsmöglichkeit herangezogen werden (Werth, 2004).

In der Praxis werden Trainings zur Personenbeurteilung vor allem im Rahmen von Beobachtertrainings für Assessment Center durchgeführt, deren Ziele die Sensibilisierung der Beobachter für mögliche Fehlerquellen bei der Beobachtung und Bewertung, die Anleitung zur Trennung zwischen Verhaltensbeschreibung und – beurteilung, die Auseinandersetzung mit den Beurteilungsdimensionen und möglichen Verhaltensoperationalisierungen und die Sicherstellung eines einheitlichen Vorgehens bei der Datenerhebung (Protokollführung) sind (Niermeyer, 1999 zit. nach Schuler, 2001). Die wenigen vorliegenden Befunde (vgl. Woehr & Huffcutt, 1994) bestätigen die Schlussfolgerung, dass für die Trainingspraxis von Assessment Center-Beobachtern eine Kombination eines Beurteilerfehlertrainings mit einem Bezugsrahmentraining oder einem Verhaltensbeobachtenstrainings am sinnvollsten erscheint (Höft & Funke, 2001). Beim Beurteilerfehlertraining liegt der Trainingsschwerpunkt auf der Sensibilisierung der Beobachter für typische Beurteilungsfehler, während beim Bezugsrahmentraining anhand von konkreten Beispielen und Aufgaben die Verwendung der Beurteilungsdimensionen eingeübt wird. Verhaltensbeobachtungstrainings stellen die Trennung von Beobachtung und Beurteilung in den Mittelpunkt, da diesen Beurteilungsphasen unterschiedliche Informationsverarbeitungsphasen zugrunde liegen (Höft & Funke, 2001). Als Trainingseffekte werden die Reduktion von Halofehlern und/oder Mildefehlern und eine Verbesserung der Beobachtungsakkuratheit oder der

Beurteilungsakkuratheit erwartet. Kanning (2002) empfiehlt für Beobachter vor einem Einsatz im Assessment Center eine eintägige Schulung, die vor allem zwei Anliegen in den Vordergrund stellen sollte: den Beobachtern ihre Rolle als Beurteiler und Diagnostiker zu verdeutlichen und den Einsatz der Beobachtungsinstrumente zu üben.

Die wenigen vorliegenden Befunde zur Wirkung von Trainings zur Personenbeurteilung auf die Qualität der Beobachtung und Einschätzung weisen alle in die Richtung, dass ein Training dazu beitragen kann, die Beurteiler zu befähigen, die Beobachtung strikter von der Bewertung zu trennen, Beurteilungsfehler zu reduzieren und die Akkuratheit der Beobachtung und der Beurteilung zu steigern. Entsprechend sollte sich ein Training zur Personenbeurteilung, welches sowohl auf die Trennung von Wahrnehmung und Bewertung von Verhalten anderer Personen als auch auf mögliche Strategien zur Reduzierung typischer Beurteilungsfehler eingeht, auf eine Beurteilung von Personen positiv auswirken. Für Assessment Center-Beobachter, die sich vorrangig aus Linienvorgesetzten zwei Hierarchieebenen über der Zielposition, unterstützt von Psychologen und externen Beratern rekrutieren, konnte eine positive Wirkung von Trainingsmaßnahmen nachgewiesen werden (Schuler, 2001). Inwieweit ein mehrstündiges Training für Probanden, die sich nicht professionell mit Personenbeurteilung beschäftigen – die also keine formalisierte Ausbildung, in der Personenbeurteilung als Lerninhalt vermittelt wird, absolviert haben und die keinen Beruf ausüben, in dem die Beurteilung von Personen alltägliche Praxis ist – eine ähnlich positive Wirkung von Trainingsmaßnahmen, erreicht, soll im empirischen Teil dieser Arbeit untersucht werden.

Experimenteller Teil

6. Überblick

6.1. Anliegen der Untersuchung

Im experimentellen Teil dieser Arbeit soll nun zum einen untersucht werden, inwieweit unterschiedliche Stimulusinformationen einen Einfluss auf die Urteilsbildung über Personen haben. Zum anderen wird der Frage nach einem Einfluss von unterschiedlichen Erfahrungen der Beurteiler bei der Einschätzung von Personen nachgegangen. Um eine möglichst hohe ökologische Validität der Ergebnisse zu erzielen, wurde Versuchsmaterial ausgewählt, das in seinen gewählten Ausschnitten beim Betrachter Prozesse der Wahrnehmung und Speicherung evoziert, die denen, welche im Alltag und im beruflichen Kontext häufig ablaufen, möglichst nahe kommen.

Zur Frage der Verarbeitung unterschiedlicher Stimuli wird untersucht, inwieweit sich bei der Gedächtnisleistung und bei einer Beurteilung von Personen Unterschiede ergeben.

Hierzu wurden semantisch identische von den Protagonisten als wörtliche Rede dargebotene Inhalte einmal als Text vorgelegt, zum zweiten wurde zusätzlich zu diesem Text ein Foto der handelnden Personen gezeigt und als dritte Bedingung wurde genau dieses Gespräch als Filmausschnitt dargeboten. Die verbalen Informationen, die die zu beurteilenden Personen geben, unterscheiden sich in allen drei Stimulusvariationen nicht. Die Informationsverarbeitungsprozesse, die durch die unterschiedlichen Stimuli evoziert werden, sollten jedoch verschieden sein und somit zu unterschiedlichen Leistungen in Bezug auf die Gedächtnisabfrage und in Bezug auf eine Urteilsbildung führen. Da alle Probanden vor der Durchführung des Experimentes darüber informiert wurden, dass sowohl die Gedächtnisleistung abgefragt wird als auch eine Beurteilung der Protagonisten abgegeben werden soll, sollte das Verarbeitungsziel den Informationsverarbeitungsprozess nicht einseitig beeinflussen (Krolak-Schwerdt & le Coutre, 2002).

Zur Untersuchung der Frage, welchen Einfluss die Beurteilererfahrung auf die Gedächtnisleistung von Personinformation und die Beurteilung von Personen hat, wurde ein Kontrastgruppendesign gewählt, welches sich gut zur Untersuchung individueller Differenzen eignet (Krems, 1994). Die Abgrenzung der einzelnen Gruppen erfolgte über folgende externe

Kriterien (Gruber, 1994): Novizengruppe, die keine formalisierte Ausbildung, in der die Personenbeurteilung als Lerninhalt vermittelt wird, absolviert hat und die sich im Berufsalltag nicht mit Personenbeurteilung im professionellen Sinn befasst; Trainingsgruppe, die keine formalisierte Ausbildung mit Personenbeurteilung als Lerninhalt absolviert hat und die keinen Beruf ausübt, in dem die Beurteilung von Personen alltäglich Praxis ist, die jedoch ein intensives mehrstündiges Training zur Personenbeurteilung erhalten hat; Expertengruppe, die eine langjährige formalisierte Ausbildung mit Inhalten zur professionellen Personenbeurteilung absolviert hat und die im Berufsalltag ständig Beurteilungen von Personen und deren Leistungen vornehmen muss. Im vorliegenden experimentellen Teil der Arbeit wird untersucht, ob sich zwischen den Gruppen Unterschiede hinsichtlich der wiedergegebenen Gedächtnisinhalte, der Beurteilungen der einzelnen Personen und in der subjektiven Urteilssicherheit finden lassen, wenn diesen Probandengruppen Informationen über Personen dargeboten werden.

6.2. Herleitung der Hypothesen zur Variation der Stimulusinformation

Es wird angenommen, dass allein das Lesen und Verstehen des Textes und das Bilden eines mentalen Modells, welches die Protagonisten repräsentiert, ein aufwändiger Prozess ist, der zum einen zu einer tiefen Verarbeitung des Textes und damit zu einer elaborierten Abspeicherung der gelesenen Informationen führt und auf der anderen Seite kognitive Ressourcen bindet. Auf der Grundlage des aufgebauten Situationsmodells sollte es dann für die Probanden möglich sein, sich einen Eindruck über mehrere Personen zu bilden und in einer Gedächtnisabfrage die Originalinformationen umfangreich und detailliert zu reproduzieren. Dies sollte sich in der Benennung mehrerer voneinander unabhängiger Persönlichkeitsmerkmale oder –eigenschaften niederschlagen.

Bei der zusätzlichen Vorlage eines Fotos mit einem Gesicht sollten die Versuchspersonen sich dieses als erstes ansehen und innerhalb kürzester Zeit einen ersten Eindruck von den Protagonisten gebildet haben (Frey 1999a; Jendraczyk, 1991). Dass beim Betrachten eines Fotos mit einem Gesicht von einer unbekannt Person innerhalb von 250 ms ein differenzierter Eindruck entsteht, der sich selbst dann nur marginal ändert, wenn man das Bild beliebig lange betrachten kann, ist ein beeindruckendes Ergebnis, gerade wenn man bedenkt, dass Zusammenhänge zwischen physiognomischen Merkmalen des Gesichtes und

Persönlichkeitseigenschaften nicht gefunden wurden (Henss, 1998). Im Unterschied zum Stimulus Text, bei dem sich der Leser aus dem Gelesenen heraus einen Eindruck von der zu beurteilenden Person bildet, hat der Beurteiler hier bereits vor dem Lesen des Textes einen ersten Eindruck von der zu beurteilenden Person, der ausschließlich auf dem visuellen Eindruck des Fotos beruht. Dieser erste visuelle Eindruck entsteht beim Beurteiler schnell und ohne eine objektive Grundlage (Werth, 2004) und steht nur sehr wenig mit überdauernden Personeigenschaften in Zusammenhang, wie z.B. in einer Studie von Borkenau und Liebler (1992) gezeigt werden konnte, dass die Einschätzungen fremder Beurteiler nach dem Betrachten eines Standbildes mit denen der beurteilten Personen selbst nur für die Faktoren Extraversion und Gewissenhaftigkeit annähernd übereinstimmten. Der Beurteiler sollte aufgrund der Betrachtung des Fotos des jeweiligen Protagonisten einen ersten Eindruck von der zu beurteilenden Person haben, der bereits sehr differenziert ist und sich auf unterschiedlichste Bereiche der Persönlichkeit beziehen kann (Henss, 1998).

Es soll untersucht werden, ob aufgrund zusätzlicher visueller Informationen in einer Gedächtnisabfrage zu den dargebotenen Informationen mehr Originalinformationen als bei Darbietung eines Textes reproduziert werden (Engelkamp, 1991; Baltes et. al., 2002) und ob sich bei einer freien Beurteilung die von den Beurteilern in dieser Stimulusvariation gewählten Beurteilungskriterien und die aus den Informationen generierten Persönlichkeitseigenschaften von denen des Stimulus Text unterscheiden.

Die dritte Stimulusvariation – Darbietung eines Filmausschnittes – entspricht hinsichtlich der Menge und in der Komplexität der Informationen am ehesten Urteilsituationen des täglichen Lebens. Die in diesen Situationen ablaufenden Wahrnehmungs- und Urteilsprozesse sind routinierte, im täglichen Leben sehr häufig angewandte Prozesse (Kanning et. al, 2004), die der Wahrnehmungsselektion und Komplexitätsreduktion dienen. Wenn man davon ausgeht, dass der erste Eindruck aufgrund der visuellen Darbietung der zu beurteilenden Personen im Film ähnlich schnell wie bei der Stimulusvariation Text und Bild gebildet wird, sollten die gewählten Beurteilungskriterien für die Protagonisten auch in etwa denen dieser Stimulusvariation entsprechen, da die Fotos einem Standbild aus dem Filmausschnitt entsprechen. Es wird untersucht, ob weitere Merkmalsdimensionen, die durch die zusätzlich dargebotene Bewegungsinformation oder durch akustische Informationen (Stimmlage, Lautstärke, Sprechgeschwindigkeit, Artikulation, Modulation, Atmung) erschlossen werden könnten, bei der Beurteilung der Personen genutzt werden.

Die dargebotenen Informationen sollten hoch automatisiert verarbeitet werden, so dass weiterhin untersucht wird, ob sich die große Komplexitätsreduktion der Informationen in einer geringeren Behaltensleistung für Originalinformationen als in den anderen zwei Stimulusvariationen niederschlägt (Kanning und Leisten, 2004).

6.3. Herleitung der Hypothesen hinsichtlich der Variation der Beurteilererfahrung

Die Gruppe der Novizen sollte sich bei gegebenen Personinformationen – so wie es formal im Abschnitt „Modelle und Mechanismen der Eindrucksbildung“ beschrieben wurde – einen Eindruck von den zu beurteilenden Personen bilden. Ausgehend von den Befunden, dass Personen – insbesondere dann, wenn sie wissen, dass später ein Urteil über eine Person abgegeben werden soll (Hertel & Bless, 2000) - bereits während der Informationsaufnahme ein Urteil über die Person generieren und von dargebotenen Verhaltensweisen auf Persönlichkeitseigenschaften schließen (Krolak-Schwerdt & Wintermantel, 2004), wird für die Reproduktion von dargebotenen Personinformationen in einer Gedächtnisaufgabe bei den Novizen eine Tendenz zur Zuschreibung von dauerhaften Merkmalen und Persönlichkeitseigenschaften, die aus den dargebotenen Verhaltensweisen generiert werden, erwartet (Wintermantel & Christman, 1983; Krolak-Schwerdt & Wintermantel, 2004). Da die Eindrucksbildung über Personen ein hoch automatisierter, täglich unzählige Male ausgeführter Prozess ist und die Novizengruppe weder durch eine spezielle Ausbildung noch durch professionelle Erfahrungen für die Fehleranfälligkeit von sozialen Urteilen sensibilisiert ist, wird untersucht, ob sich dies hier in einer hohen subjektiven Urteilssicherheit niederschlägt (Hoffrage, 1993).

Es soll untersucht werden, ob sich ein vor der Untersuchung durchgeführtes intensives Training, welches die Probanden für die Unterschiede zwischen Beobachten und Bewerten sensibilisiert und in dem dieses Trennen zwischen den Informationsverarbeitungsphasen auch praktisch geübt wurde, die Trainierten im Vergleich zur Novizengruppe in die Lage versetzt, bei der Wiedergabe von Gedächtnisinhalten besser zwischen dargebotenen Originalinformationen und daraus generierten Eigenschaften und Merkmalen zu trennen, mehr Details zu erinnern und weniger dazu zu neigen, Originalinformationen zu verallgemeinern oder grob zusammenzufassen (Schuler, 2001; Kanning, et. al, 2004). Aufgrund des kurz vor der Untersuchung erfolgten Trainings sollte die Trainingsgruppe

zumindest bemüht sein, bei einer freien Beurteilung von Personen eine objektivere Personbeurteilung vorzunehmen und Fehler möglichst auszuschließen. Da allein das Wissen um bestimmte Beurteilungsfehler und die Vermittlung von Strategien zur Reduktion dieser Fehler zu einer objektivieren Personbeurteilung führen kann (Kanning, 1999; Kanning et. al, 2004), soll untersucht werden, ob sich das intensive Training dergestalt auswirkt, dass die Trainingsgruppe bei der Beurteilung von Personen mehr Originalinformationen zur Generierung von Eigenschaften und Personmerkmalen heranzieht und ob sie weniger globale Verallgemeinerungen, die nicht aus Originalinformationen geschlossen werden können, vornimmt. Da das Training ein gewisses Maß an Unsicherheit bei der Personbeurteilung (Kamphaus, 1980) evoziert, es weiterhin dazu anregen sollte, das subjektive Sicherheitsgefühl in Bezug auf die Korrektheit eines eigenen Urteils kritisch zu hinterfragen und den Trainingsteilnehmern durch praktische Übungen und Erfahrungen aufgezeigt haben sollte, wie häufig ein Urteil Fehler behaftet sein kann, soll weiterhin untersucht werden, wie die so sensibilisierten Probanden der Trainingsgruppe ihre subjektive Urteilssicherheit im Vergleich zu den beiden anderen Kontrastgruppen einschätzen.

Die Expertengruppe sollte auf der einen Seite aufgrund ihrer Ausbildung über ein fundiertes Wissen über Personenbeurteilung und durch die langjährige Berufspraxis über routinierte Informationsverarbeitungsstrategien bei der Personenbeurteilung verfügen. Auf der anderen Seite könnten gerade diese hoch automatisierten Prozesse bei der Verarbeitung von Personinformation und der Beurteilung von Personen bei der Expertengruppe dazu führen, in anderen Beurteilungssituationen als in der gewohnten Domäne wenig flexibel zu reagieren. Es soll untersucht werden, inwieweit sich bei einer Gedächtnisabfrage zu den dargebotenen Personinformationen Unterschiede zu den beiden anderen Kontrastgruppen finden lassen, die darauf hinweisen, dass die Experten auch außerhalb der gewohnten kontextbezogenen Expertise (Gruber, 1994) in Urteilsbildungssituationen eher dazu neigen, Verallgemeinerungen zu verwenden und weniger Details wiederzugeben (Schmidt & Boshuizen, 1990; Krolak-Schwerdt & le Coutre, 2002).

Bei der Beurteilung der Personen ist von Interesse, ob die Expertengruppe, die zumindest in ihrer Domäne wichtige von unwichtigen Informationen beim Beurteilen trennen kann (Bromme, 1992), diese Trennung auch in anderen Beurteilungssituationen vornimmt und wie sich dies auf die Beurteilung auswirkt. Es ist beispielsweise denkbar, dass durch die Konzentration der Experten auf Informationen, die in ihrer Domäne eine wesentliche Rolle bei der Urteilsbildung spielen, andere Informationen, die in der Untersuchungssituation bei

der Personenbeurteilung ebenfalls aussagekräftig sind, vernachlässigt werden, so dass ein Urteil über eine Person beispielsweise nur sehr wenige Merkmalsdimensionen beinhaltet. Es soll untersucht werden, inwieweit sich die Nutzung der dargebotenen Stimulusinformationen bei der Urteilsbildung durch die Expertengruppe von den beiden anderen Kontrastgruppen unterscheidet.

Durch das tägliche selbstverständliche Ausführen von Personbeurteilungen im Berufsalltag, oftmals über viele Jahre hinweg, dürften die Experten ihr Können und ihre Fähigkeiten im Bereich der Personenbeurteilung sehr selbstbewusst einschätzen. Eine wiederholte Schulung, die auch erfahrene Beurteiler erneut für mögliche Beurteilungsfehler sensibilisiert, und wie sie beispielsweise auch für langjährige erfahrene Assessment Center-Beobachter empfohlen wird (Kanning, 2002; Schuler 2001), findet im Berufsalltag in den meisten Fällen nicht statt, so dass das Bewusstsein für die Fehleranfälligkeit von Urteilen gerade mit zunehmender Berufspraxis eher wenig ausgeprägt sein sollte. Es soll untersucht werden, wie sich dies auf das subjektive Sicherheitsgefühl der Expertengruppe bei einer Personbeurteilung auswirkt.

6.4. Operationalisierung

In einer freien Reproduktion von dargebotenen Personinformationen haben die wiedergegebenen Inhalte auf der sprachlichen Oberfläche mit den Originalinformationen oft nur wenig gemein (Grabowski, 1991), so dass für die vorliegende Arbeit folgende Kodierungen für die Wiedergabe von Inhalten vorgenommen werden. Diejenigen Informationen, die wörtlich mit den Originalinformationen übereinstimmen, werden als originale Informationen bezeichnet. Wird die Originalinformation synonym oder lediglich mit einer anderen sprachlichen Oberfläche reproduziert, wird dies als synonyme Information beschrieben. Werden Informationen aus den dargebotenen Personinformationen hinsichtlich der Komplexität stark reduziert (z.B. „Im weiteren diskutieren Nr. 7 und Nr. 8.“), ohne dass z.B. sprachliche Inhalte dieser Diskussion im einzelnen wiedergegeben werden, wird dies als zusammengefasste Information oder als verallgemeinerte Information bzw. als Zusammenfassungen oder Verallgemeinerungen von Informationen bezeichnet. Die Kodierung der Reproduktionsprotokolle aller Probanden zur statistischen Auswertung ist im Abschnitt 9.5. Aufbereitung der Protokolle der Gedächtnisabfrage und zur freien Beurteilung der Geschworenen beschrieben.

7. Hypothesen

7.1. Hypothesen hinsichtlich der Variation der Stimulusinformation

Bei Variation der Stimulusinformationen werden folgende Unterschiede zwischen den einzelnen Stimulusvarianten postuliert:

1.

Bei zusätzlicher Vorlage eines Fotos der zu beurteilenden Personen zu einem Text sollte die Reproduktion der Originalinformationen umfangreicher und detaillierter als in der Text-Bedingung möglich sein.

2.

Bei der Beurteilung der Personen sollten neben Merkmalen und Eigenschaften, die aus dem Text generiert werden können, zusätzlich dazu Beurteilungskriterien bzw. Persönlichkeitseigenschaften genannt werden, die aus dem Bildeindruck resultieren. Diese zusätzlichen Nennungen sollten sich qualitativ und quantitativ nachweisen lassen.

3.

Die subjektive Urteilssicherheit der Beurteiler sollte zum einen höher sein als in der Text-Bedingung, jedoch unter der der Film-Bedingung liegen.

4.

Die Reproduktion von Originalinformationen sollte bei Darbietung eines Filmausschnittes geringer ausfallen als in der Text-mit-Bild-Bedingung. Es sollten hier vermehrt zusammengefasste Inhalte wiedergegeben werden.

5.

Bei der Beurteilung der Personen nach Darbietung eines Filmausschnittes sollten die Beurteiler Eigenschaften und Merkmale generieren, die aus den zusätzlichen Modalitäten Bewegungen und akustische Informationen resultieren. Dies sollte sich qualitativ durch die Unterschiedlichkeit der genannten Beurteilungskategorien und quantitativ durch eine höhere Anzahl an Persönlichkeitseigenschaften nachweisen lassen.

6.

Die subjektive Urteilssicherheit sollte in der Stimulusvariation Film im Vergleich zu den Stimulusvariationen Text und Text mit Bild am höchsten ausgeprägt sein.

7.2. Hypothesen hinsichtlich der Variation der Beurteilererfahrung

Für die Variation der Beurteilererfahrung werden für die Kontrastgruppen folgende Gruppenunterschiede postuliert:

1.

Bei der Wiedergabe von Gedächtnisinhalten nach der Darbietung von Personinformationen sollten die Trainierten im Vergleich zu den Novizen mehr Originalinformationen und mehr Details erinnern können. Sie sollten bei der Gedächtnisaufgabe im Vergleich zu den Novizen weniger Verallgemeinerungen und Zusammenfassungen und aus dem Verhalten generierte globale Eigenschaften nennen.

2.

Bei der Beurteilung der Personen sollte sich nachweisen lassen, dass die Trainierten im Vergleich zu den Novizen weniger Eigenschaften und Persönlichkeitsmerkmale benennen, die sehr globale und weitreichende Einschätzungen der Persönlichkeit darstellen und dass die Trainierten für ihre Einschätzungen mehr Details der Originalinformation berücksichtigen.

3.

Die subjektive Urteilssicherheit der Trainiertengruppe sollte im Vergleich zu den Novizen geringer ausgeprägt sein.

4.

Die Gruppe der Experten sollte in einer Gedächtnisabfrage nach den dargebotenen Personinformationen im Vergleich zu der Novizengruppe weniger Originalinformationen und Details, jedoch mehr zusammengefasste und verallgemeinerte Informationen reproduzieren.

5.

Bei der Beurteilung der Personen wird erwartet, dass die Experten weniger Eigenschaften und Merkmale auf insgesamt weniger Dimensionen zur Einschätzung der Personen als Novizen verwenden. Es sollte sich des Weiteren nachweisen lassen, dass die Experten bei der Personbeurteilung weniger globale, über die Information des Stimulus hinausgehende Einschätzungen verwenden als Novizen.

6.

Für die subjektive Sicherheit bei der Beurteilung wird erwartet, dass die Experten ihre subjektive Urteilssicherheit niedriger als die Novizen, jedoch höher als die Trainierten einschätzen.

8. Methode

8.1. Versuchspersonen

An der Untersuchung nahmen insgesamt 93 Probanden teil. Hiervon waren 31 Probanden Pädagogen im aktiven Schuldienst aus insgesamt 3 nordthüringer Regelschulen, davon 24 Frauen und 7 Männer. Die Probanden hatten ein Alter von 39 bis 63 Jahren, im Durchschnitt waren sie 46,7 Jahre alt. Sie verfügten im Mittel über eine Berufserfahrung von 23,9 Jahren mit einer Spannbreite von 17 bis 38 Berufsjahren im Schuldienst ohne Anrechnung der Studienzeit. Sie werden in der vorliegenden Arbeit als Gruppe der Experten in der Personenbeurteilung bezeichnet, da sie zum einen eine lange formalisierte pädagogische Ausbildung (mehrjähriges Pädagogik-Studium) absolviert haben, in der unter anderem theoretische Grundlagen der sozialen Urteilsbildung und Bewertung von Personenmerkmalen vermittelt werden und zum anderen im Laufe ihrer langjährigen Berufspraxis eine auf den Unterrichtskontext bezogene Expertise erworben haben, die eine schnelle und kategoriengleitete Personenbeurteilung erlaubt (Bromme, 1992).

Als weitere Probanden nahmen 62 Studierende der Friedrich-Schiller-Universität Jena aus den Studienrichtungen Psychologie (20), Betriebswirtschaftslehre (8), Germanistik (3), Lehramt (5), Ernährungswissenschaften (7), Volkswirtschaftslehre (4), Medizin (3), Magisterstudium (12) teil. Hiervon waren 46 Frauen und 16 Männer vertreten. Die Probanden hatten ein Alter von durchschnittlich 22,5 Jahren, die Spannbreite betrug 18 bis 36 Jahre.

Etwa die Hälfte davon, 32 Personen, wurden vor Beginn des Experimentes in der Personenbeurteilung, wie im Abschnitt „8.2. Training der Probanden in der Personenbeurteilung“ beschrieben, intensiv trainiert. Im Weiteren werden die Probanden, die ein Training erhielten, als Trainingsgruppe bezeichnet. Die andere Gruppenhälfte, 30 Personen, nahmen ohne besondere Vorbereitung am Experiment teil. Diese Gruppe wird in der Folge als Novizen bezeichnet.

Die Novizen wurden für ihre Teilnahme am Experiment finanziell entlohnt.

Der Trainingsgruppe wurde eine Verbesserung ihres Urteilsvermögens über andere Personen, eine Mappe mit Trainingsunterlagen, ein Zertifikat über die Teilnahme sowie das Durchführen zahlreicher praxisrelevanter Übungen in Aussicht gestellt. Dadurch konnte eine hohe Motivation der Trainingsteilnehmer vorausgesetzt werden.

Der Anreiz zur Gewinnung der Pädagogen für die Teilnahme am Experiment erfolgte durch die Honorierung ihres Expertenstatus in der Personenbeurteilung. Die Rekrutierung erfolgte durch Kontaktaufnahme mit der jeweiligen Schulleitung. Auf freiwilliger Basis haben sich jeweils einige Pädagogen zur Teilnahme bereit erklärt.

Der große Altersunterschied zwischen der Expertengruppe und den Gruppen der Trainierten und der Novizen ist eine der unkontrollierten Variablen im gewählten Kontrastgruppendesign. In Kontrastgruppendesigns lassen sich diese Einflüsse nur schwer ausschließen (Gruber, 1994). Beim Vergleich von Experten und Novizen ist es nicht leicht, dann, wenn Experten mit mehrjähriger Erfahrung untersucht werden, geeignete Probanden zu finden (Krems, 1994). Gerade die Rekrutierung der Experten als Probanden erwies sich in der praktischen Umsetzung als schwierig, so dass hinsichtlich des Alters keinerlei Einschränkung gemacht wurde. Dass infolge des Kontrastgruppendesigns und des damit einhergehenden Altersunterschiedes weitere Faktoren, wie sozialer Status und langjährige Erfahrungen im persönlichen Bereich beim Umgang mit anderen Personen (z.B. Erziehung von eigenen Kindern), einen unkontrollierten Einfluss auf die Urteilsbildung haben können, ist nicht auszuschließen.

Auch innerhalb der Expertengruppe gibt es relevante Altersunterschiede und eine unterschiedliche Anzahl von absolvierten Dienstjahren im Schuldienst. Beim Erwerb der Expertise hat jede einzelne Versuchsperson ihre eigenen Erfahrungen gemacht. Es ist kaum vorstellbar, dass selbst Pädagogen mit gleicher Anzahl an Berufsjahren im Schuldienst mit

annähernd gleichen Problemen annähernd gleich oft konfrontiert worden wären, so dass die Entwicklung in etwa gleich verlaufen wäre. Jede Person hat also ihre Expertise unkontrolliert, außerhalb des Experimentes, erworben. Trotz der Faktoren, die die Ergebnisse der Untersuchung unkontrolliert beeinflussen können, überwiegen die Vorteile beim Vergleich zwischen Experten und Novizen (Gruber, 1994). Ein Experten-Novizen-Vergleich kann Aufschluss über Verarbeitungsprozesse geben, die z.B. durch Erfahrung entstehen. Des Weiteren wird als Vorteil genannt, dass beim Experten-Novizen-Vergleich Domänen des alltäglichen Lebens untersucht werden, denen Personen mehrere Jahre intensiver Beschäftigung widmeten, so dass die Studien oft über die Experimentalsituation hinaus Validität besitzen (Gruber, 1994). Es war unter anderem Anliegen der vorliegenden Untersuchung, eine möglichst hohe ökologische Validität der Ergebnisse zu erreichen.

8.2. Training der Probanden in der Personenbeurteilung

Ziel des Trainings war, den Probanden grundlegende Erkenntnisse der Psychologie der Personenbeurteilung zu vermitteln, um die Versuchspersonen in die Lage zu versetzen, Fehler und Fallen beim Urteilsprozess zu erkennen und den eigenen Beurteilungsprozess zu optimieren. Dabei war dem Autor bewusst, dass eine vollständige Objektivität in der Personenbeurteilung nicht zu erreichen ist, dass jedoch trotzdem durch ein Training die Annäherung an eine möglichst objektive Beurteilung gelingen sollte.

Das Beurteilungstraining war stark angelehnt an das erste und einzige Trainingsmanual zur professionellen Personenbeurteilung im deutschsprachigen Raum von Kanning et al. (2004). Gestützt auf die Befunde von Woehr & Huffcutt (1994), die für einen optimalen Trainingserfolg in der Praxis eine Kombination eines Beurteilerfehlertrainings mit einem Bezugsrahmentraining oder einem Verhaltensbeobachtenstraining empfehlen, wurden als Trainingsinhalte zum einen Methodische Grundlagen – Beobachtungsmethoden und Grundlagen der Wahrnehmung und Urteilsbildung – und zum anderen Systematische Fehler in der Personenbeurteilung und Gegenmaßnahmen ausgewählt.

Das Training beinhaltete sowohl die theoretische Vermittlung von Wissen über die Personenbeurteilung und typische Beurteilungsfehler als auch die praktische Anwendung

dieses Wissens durch vielfältige Übungen und zahlreiche Fallbeispiele. Das Training bestand aus insgesamt 8 Trainingseinheiten:

Im ersten Teil – *Beobachtungsmethoden* – wurden den Probanden die Vor- und Nachteile von systematischer Beobachtung und Beobachtung im Alltag gegenübergestellt. Deutlich gemacht wurde zudem in diesem Zusammenhang der Unterschied zwischen Beobachten und Bewerten. Anhand der Beobachtung und Analyse einer Videosequenz wurde die Schwierigkeit der Trennung von Dokumentation einer Information und des Schlussfolgerns auf zugrunde liegende Eigenschaften und Fähigkeiten aufgezeigt und den Probanden wurden konkrete Hilfestellungen vermittelt, die es ihnen erleichtern sollten, die Prozesse der Informationsaufnahme und der Bewertung auseinander zuhalten.

In der zweiten Trainingseinheit - *Grundlagen der Wahrnehmung und Urteilsbildung* – wurde ausführlich die Selektivität der Wahrnehmung als wichtige Begrenzung bei der Beobachtung von Personen dargestellt. An praktischen Beispielen wurde demonstriert, dass es kaum möglich ist, unsere soziale Umwelt objektiv wahrzunehmen und trainiert, wie die Qualität der Wahrnehmung verbessert werden kann.

Der dritte Trainingsabschnitt – *Bezugssysteme in der Urteilsbildung* – verdeutlichte die Wirkung verschiedener Bezugssysteme auf die Personenbeurteilung. Darüber hinaus wurden Methoden der Verminderung der dadurch entstehenden Urteilsverzerrungen dargestellt und eingeübt.

In der Trainingseinheit – *Gedächtnis und Vergessen* – wurden zunächst kurz theoretische Grundlagen zur Struktur und Funktion des menschlichen Gedächtnisses sowie Gedächtnismodelle und der Prozess der Informationsverarbeitung von Personinformation im Gedächtnis (z.B. Verfügbarkeitsheuristiken, Repräsentativitätsheuristiken, Simulationsheuristiken) behandelt. Im Weiteren wurden gedächtnisbasierte Beurteilungsfehler z.B. bei Augenzeugen (Erdfelder, 2003), durch den Häufigkeits-Validitäts-Effekt (Hertwig, 1993) oder Rückschaufehler (Pohl & Hell, 1996) sowie Möglichkeiten deren effektiver Begegnung aufgezeigt.

Der nächste Teil – *Ursachenzuschreibung/Attribution* – vermittelte den Trainingsteilnehmern zum einen Inhalte wichtiger Attributionstheorien (Heider, 1958; Kelley, 1973) und die

Auswirkungen von verzerrten Attributionen und zum anderen wurden Möglichkeiten zur Verringerung von unzutreffenden Ursachenzuschreibungen und systematischen Attributionsverzerrungen erarbeitet.

Die Trainingseinheit – *Erwartungsgesteuerte Wahrnehmung* – stellte dar, dass die menschliche Wahrnehmung außer durch die objektiv in der Umwelt vorhandenen Stimuli zusätzlich durch im Gedächtnis gespeicherte Kategorien und Schemata beeinflusst wird (Bruner und Postman, 1951, Lilli, 1984). Diese durch subjektive Erfahrungen und Erwartungen verfälschte Wahrnehmung (z.B. der Pygmalion-Effekt, Labeling, Placeboeffekt) sollte hier deutlich gemacht und mögliche Einflussfaktoren zur Erhöhung der Objektivität im Prozess der frühen Informationsverarbeitung diskutiert werden. Darüber hinaus wurden Möglichkeiten der Verringerung selektiver Wahrnehmung infolge sehr früher Hypothesenbildung beispielsweise durch Erwägung von Alternativhypothesen und deren Veri- oder Falsifizierung aufgezeigt (Krems, 1994).

Die Trainingsteilnehmer wurden für die Einflüsse von positiven und negativen Emotionen auf den Prozess der Personenbeurteilung in Abhängigkeit von deren Intensität im Trainingsteil – *Emotionen und Urteilsbildung* – sensibilisiert und durch Übungsaufgaben der Umgang mit emotionsauslösenden Informationen so trainiert, dass die Teilnehmer in die Lage versetzt wurden, die verzerrende Wirkung von Emotionen im Urteilsprozess zu reduzieren.

Im letzten Trainingsteil – *Vorurteile* – wurden den Trainingsteilnehmern zuerst die Unterschiede zwischen Stereotype und Vorurteilen und deren Bedeutung für die Urteilsbildung nahe gebracht. Die Teilnehmer lernten, widersprechende Informationen zu aktivieren, indem sie sich eigene Vorurteile bewusst machten und sie in Frage stellten, und somit in die Lage versetzt wurden, die negative Wirkung von Vorurteilen auf die Urteilsbildung abzuschwächen.

8.3. Untersuchungsdesign

Das Training der Probanden und die Experimente mit der Trainings- und der Novizengruppe fanden in einem ausreichend großen Seminarraum der Friedrich-Schiller-Universität Jena statt. Die Experimente mit den Experten wurden in der jeweiligen Regelschule, in welcher die

Pädagogen ihren Schuldienst verrichten, durchgeführt. Hierfür wurden von der Schule Aufenthalts- oder Lehrerzimmer zur Verfügung gestellt.

Die Trainingsgruppe wurde im Vorfeld des Experimentes über mehrere Stunden trainiert. Das Experiment fand dann nach einer ausreichend großen Pause am selben Tag statt.

Die Experimente mit den Novizen und Trainierten fanden in kleinen Gruppen zu 4 bis maximal 6 Personen statt.

Das vorhergehende Training der Probanden erfolgte entsprechend mit derselben Gruppengröße. Diese Gruppengröße gestattete ein effektives Training mit vielfältigen Möglichkeiten. Wie bereits im vorigen Abschnitt beschrieben, wurde im Training neben der Vermittlung von umfangreichem theoretischem Wissen viel Wert auf die praktische Übung und Umsetzung gelegt. Bei der theoretischen Wissensvermittlung war aufgrund der maximalen Gruppenstärke von 6 Probanden gewährleistet, dass auch auf individuelle Fragen eingegangen werden konnte. Eine intensivere Auseinandersetzung mit den Trainingsinhalten erfolgte in den praktischen Übungen. Hierzu wurden in Ergänzung zum Theorieteil zahlreiche Einzelbeispiele zusammengetragen, analysiert, bearbeitet und reflektiert, Diskussionen angeregt und Übungen in kleineren Gruppen oder im Plenum durchgeführt.

Der für den Trainingserfolg – maximaler Erkenntniszugewinn und bewusstere Steuerung des sozialen Urteilsprozesses - notwendigen Voraussetzung, am jeweiligen Kenntnisstand der Teilnehmer anzusetzen und deren Vorerfahrung einzubeziehen (Kanning et. al, 2004), konnte durch die gewählte kleine Gruppenstärke gut entsprochen werden.

Die Experimente mit den Experten wurden nach Ende des Unterrichtes in den Nachmittagsstunden durchgeführt. Aus organisatorischen Gründen nahmen jeweils 10 oder 11 Pädagogen am Experiment teil.

Abbildung 1 stellt das Untersuchungsdesign tabellarisch dar.

Anzahl der Probanden	Stimulusinformation			
		Text	Text und Bild	Film
Probandengruppe	Novizen	10	10	10
	Trainierten	11	10	11
	Experten	11	10	10

Abbildung 1.

8.4. Stimulusmaterial

Das Stimulusmaterial wurde dreifach variiert.

Eine Variante war ein Filmausschnitt aus dem Film „Die zwölf Geschworenen“ aus dem Jahre 1957. Reginald Rose schrieb das Drehbuch und produzierte diesen Film. Es handelt sich um eine Schwarz-Weiß-Aufnahme. Inhalt des Filmes ist eine Besprechung von 12 Geschworenen unterschiedlichster sozialer Herkunft, die über das Schicksal eines jungen mutmaßlichen puertoricanischen Vater-Mörders im New York der 50`er Jahre zu entscheiden haben. Der gesamte Film spielt in einem kleinen Hinterzimmer eines Gerichtsgebäudes.

Der gewählte Filmausschnitt hat eine Länge von ca. 3 Minuten, es agieren 5 der 12 Geschworenen.

Als nächste Stimulusvariante wurde ein Textausschnitt aus dem Theaterstück „Die zwölf Geschworenen“ (S.21 bis 26) (Originaltitel: „Twelve angry Men“) verwendet. Das Bühnenstück wurde ebenfalls wie das Drehbuch zum gleichnamigen Film von Reginald Rose verfasst. Der Textausschnitt umfasste nach Bearbeitung 2 DinA4 Seiten, insgesamt 109 Sätze. Dieser Textausschnitt gibt denselben Auszug aus dem Gespräch der 5 Geschworenen wieder, wie der Filmausschnitt. Der Text kann aus folgenden Gründen als Äquivalent zum gewählten Filmausschnitt bezeichnet werden: Die Unterhaltung der Geschworenen wird wörtlich – wie im Film – in schriftlicher Form dargeboten. Des Weiteren werden im Text viele Verhaltensweisen der handelnden Personen inklusive Mimik, Gestik und Hinweise darauf, welche Emotion der jeweilige Protagonist ausstrahlt, in Form von Regieanweisungen beschrieben (z.B. „Blickt ihn eine Sekunde verärgert an.“ oder „Lächelt plötzlich versöhnlich“ (Rose, 1954)).

Die dritte Stimulusvariante bildete der gleiche Textausschnitt kombiniert mit der Vorlage von Fotos der 5 agierenden Geschworenen. Die Fotos entstammen dem Filmausschnitt der ersten Stimulusvariante. Für jeden der 5 Geschworenen wurde ein Foto angefertigt, welches das Gesicht und Teile des Oberkörpers zeigt.

8.5. Begründung der Auswahl des Stimulusmaterials

Die Auswahl gerade dieses Stimulusmaterials erschien aus mehreren Gründen optimal. Zum einen handelt es sich sowohl bei dem Film als auch beim Text um natürlich vorkommende Medien (Rinck, 2000), um eine möglichst hohe ökologische Validität zu erreichen. Gerade dies ist der Schwachpunkt vieler anderer Untersuchungen zur Personenbeurteilung (Krolak-Schwerdt, 2000). Sehr häufig wurden von den Versuchsleitern Eigenschafts- oder Verhaltenslisten von Personen vorgelegt, nach denen sich dann die Versuchsperson ein Urteil über die Person, die mit wenigen Worten dargestellt wurde, bilden sollte. Zu recht wird bezweifelt, ob hierbei valide Ergebnisse erhalten werden können (Krolak-Schwerdt, 2000, Krolak-Schwerdt & Wintermantel, 2004), da in der realen Welt Urteile über Personen in aller Regel nicht aufgrund von Wortlisten gebildet werden, sondern aufgrund der Beobachtung von oder Interaktion mit Personen im sozialen Umfeld oder beim Textlesen aufgrund zusammenhängender Sätze. Auch Spiro (1980) kritisierte, dass die in Laborstudien verwendeten Informationen häufig zu künstlich sind und er versuchte daher, eine eher extern valide Aufgabenstellung zu verwenden, indem er einen recht wirklichkeitsnahen Fall konstruierte und manipulierte. So geht Spiro (1980) in seinem Ansatz davon aus, dass neu zu verarbeitende Information bereits vorhandene Wissensstrukturen systematisch verändert, um zum Beispiel Konsistenz herzustellen. Diese Anpassung wird von ihm als Akkomodation bezeichnet. Er führt die Tatsache, dass Akkomodation in der Literatur selten nachgewiesen werden kann, darauf zurück, dass die verwendete Information in Experimenten zu künstlich ist, um mit bereits bestehenden Wissensstrukturen in Verbindung gebracht zu werden. Er versuchte daher, eine extern validere Aufgabenstellung zu verwenden. Versuchspersonen bekamen zunächst die Vorinformation über ein Ehepaar. In der Beziehung von beiden war es der Fall, dass entweder beide Partner keine Kinder wollten oder einer der Partner Kinder wollte und der andere nicht. Nach dieser Vorinformation erhielten die Versuchspersonen in der einen Bedingung die Information, dass die beiden Partner geheiratet hätten, in der anderen jedoch nicht. Diese nachträglich variierte Information war zum Teil konsistent mit der Vorinformation und zum Teil inkonsistent. Die Ergebnisse im Free-Recall-Test zeigten, dass die Versuchsteilnehmer die Vorgeschichte systematisch an das Schema der nachträglichen Information anglich, um Konsistenz herzustellen. Auch bei der Diskussion der Ergebnisse rückt Spiro (1980) die ökologische Validität der Ergebnisse in den Mittelpunkt. So kritisiert Spiro (1980), dass z.B. Gedächtnisinstruktionen, der sozialen Realität nicht angemessen seien, da es außerhalb des Labors darum ginge, eine Situation zu verstehen oder sich einen Eindruck

zu bilden. Bei allen Versuchen, seine Experimente an der sozialen Realität auszurichten, und damit eine möglichst hohe ökologische Validität zu erhalten, bleibt der Kritikpunkt, dass auch das von Spiro (1980) verwendete Versuchsmaterial extra für die durchgeführten Experimente konstruiert und manipuliert worden und somit künstlich für Laborexperimente hergestellt war.

Bei dem in der vorliegenden Arbeit verwendeten Filmmaterial handelt es sich nicht um extra für dieses Experiment konstruiertes Material, sondern um einen bereits existenten Kinofilm. Dieser Film basiert auf einer authentischen Begebenheit, so dass die Situation für den Betrachter als sehr realitätsnah erlebt wird. Geschlechterstereotype in der Personenbeurteilung werden von vornherein ausgeschlossen, da es sich bei allen 12 Protagonisten um männliche Personen handelt. Der uneingeschränkte Fokus liegt auf den Protagonisten. Dadurch bietet sich die Verwendung dieses Stückes für die Personenbeurteilung an. Weitere relevante Dimensionen in einer Handlung, wie Ort und Zeit (Zwaan et al., 1995), treten in den Hintergrund, da das gesamte Geschehen unter ständiger Anwesenheit aller Protagonisten ortsunverändert im selben Raum stattfand. Zudem wurde die Darstellung in Echtzeit gedreht und enthielt keinerlei Zeitsprünge.

Neben dem Film von 1957, in dem Reginald Rose mit Produzent war, gibt es zu dieser Begebenheit eine weitere Verfilmung aus dem Jahr 1997 vom Produzenten Terence A. Donnelly. Der gewählte Film, aus dem das genutzte Stimulusmaterial entstammt, unterscheidet sich in vielerlei Hinsicht von der neueren Verfilmung. An dieser Stelle soll ausdrücklich vermerkt werden, dass es hier keineswegs darum geht, alle Unterschiede zwischen den beiden Filmen detailliert wiederzugeben. Vielmehr sollen nur diejenigen Aspekte genannt werden, die für die Entscheidung, welcher der beiden Filme als Stimulusmaterial genutzt werden sollte, von Bedeutung waren. Zum einen nutzt der Film von 1997 einen wesentlich salopperen Sprachgebrauch, der nicht mit der Textversion des Bühnenstückes von Reginald Rose identisch ist. Der Film von 1997 wurde mit einer Starbesetzung an bekannten Schauspielern (wie z.B. Armin Mueller-Stahl, Tony Danza, Jack Lemmon) gedreht, was Vorerfahrungen bei den Versuchspersonen bezüglich der Charaktere, welche bestimmte Schauspieler typischerweise in anderen Filmen verkörpern, aktiviert und somit die unvoreingenommene Personwahrnehmung verfälschen kann. Im Film von 1957 spielen ausschließlich weiße Schauspieler, während im Film von 1997 vier schwarze und acht weiße Darsteller vertreten sind, was entsprechende Stereotype aktivieren könnte. Des Weiteren sind im Film von 1997 erhebliche Unterschiede in der Kleidung der Protagonisten

zu finden – die Bandbreite reicht von klassischem Anzug mit Weste über Smoking mit Fliege bis hin zu legerem Jeansoutfit und Poloshirt. Die Geschworenen im Film von 1957 hingegen sind wesentlich einheitlicher gekleidet: alle Darsteller tragen ein helles Hemd und 10 von ihnen eine dunkle Krawatte. Einige Darsteller tragen Jackett, andere nicht. Insgesamt ist bei allen Darstellern somit ein klassischer Kleidungsstil vertreten. Die Wahl der Kleidung lässt somit in dieser Fassung kaum Rückschlüsse auf die Persönlichkeit der Protagonisten zu.

Durch die Schwarz-Weiß-Verfilmung von 1957 werden Störeinflüsse auf die Personenbeurteilung insofern minimiert, dass die Probanden weniger eine Gelegenheit erhalten, sich durch Unterschiede in Äußerlichkeiten (z.B. Farbigkeit der Kleidung) von dem Verhalten und den gesprochenen Inhalten der zu Beurteilenden abzulenken.

Neben den hier besprochenen Unterschieden zwischen den Filmen, die nach außen deutlich werden, war wichtigster Beweggrund für die Verwendung des Filmes von 1957, dass derselbe Autor – Reginald Rose – sowohl das Bühnenstück als auch das Drehbuch für den Film verfasst hat und darüber hinaus Produzent des Filmes war. Auf diese Weise – so kann zumindest augenscheinlich erwartet werden – kann davon ausgegangen werden, dass der Verfasser des literarischen Stückes durch seine Mitwirkung auch der filmischen Auseinandersetzung positiv gegenübersteht.

Trotz der semantischen Gleichheit – weitestgehend übereinstimmende wörtliche Rede - zwischen vorgelegtem Text und gezeigtem Filmausschnitt unterscheiden sich die beiden Darbietungen hinsichtlich der Darbietungsmodalitäten und zusätzlicher Informationen, welche im Film erkennbar jedoch im Text nicht explizit – durch Handlungsanweisungen - benannt worden sind. Während der Text nur einen Informationskanal – die geschriebene Sprache – zur Verfügung hat, aktualisiert der Film fünf Informationskanäle: visuelles Bild, Schrift und andere Grafiken, Dialog, Musik und Geräusche (Metz, 2001).

Durch die Eindrücke von Bild (visuelles Bild und dessen Bewegung, Schrift und andere Grafiken) und Ton (Dialog, Musik und Geräusche) ergeben sich fast unendlich viele zusätzliche Informationen, welche im Film, jedoch nicht im Text enthalten sind. Ein Teil dieser Informationen wird durch das Gesicht der Protagonisten vermittelt. Das Gesicht eines Menschen stellt das zentrale Erkennungsmerkmal und die Basis für eine spontane Eindrucksbildung über Personen dar (Henss, 1998). Durch die Vorlage von Fotos der zu

beurteilenden Protagonisten wurden den Versuchspersonen somit über den Text hinausgehende visuelle Informationen zur Verfügung gestellt, welche ebenfalls im Film enthalten sind. Dass Menschen dazu in der Lage sind, sich ein differenziertes Urteil über eine andere Person zu bilden, wenn sie nur für den Bruchteil einer Sekunde ein Bild von dieser Person angeschaut haben, wurde bereits an anderer Stelle angeführt (Jendraczyk, 1991, Frey, 1999a,b).

Dass alles, was in einem literarischen Text erzählt werden kann, „auch in einem Film annähernd verbildlicht oder erzählt werden“ (S. 45) kann, postuliert Monaco (2001) in seiner Einführung in die Filmwissenschaft „Film verstehen“. Darin verweist er auch auf drei Bereiche, in denen sich die Medien voneinander unterscheiden: die zeitliche Spanne, die Darstellung im Raum und die Rolle des Erzählers. Der üblicherweise, z.B. bei der Verfilmung eines Romans auftretende Unterschied, dass der Film die zeitliche Spanne des Romans nicht vollständig reproduzieren kann und daher „Handlungsdetails...fast regelmäßig bei der Übertragung vom Buch in den Film verloren (gehen)“ (Monaco, 2001, S.46) ist bei der Verwendung des Textes „Die zwölf Geschworenen“ und des gleichnamigen Filmes nicht zu finden. Beide Werke, sowohl der Text als auch der Film finden in Echtzeit statt. Die sonst übliche Medien bedingte Begrenztheit des Mediums Film bei Verfilmung von literarischen Vorlagen hinsichtlich der Zeitspanne trifft hier nicht zu und war einer der Gründe, die zur Wahl des Stimulusmaterials führten.

Der wohl grundlegendste Unterschied zwischen beiden Medien, (Waara, 2003) der die Darstellung des Raumes betrifft, ist darin zu finden, dass ein geschriebener Text nie so konkret und detailliert wie der Film beschreiben kann. „Es wäre eine absurde Aufgabe für einen Schriftsteller zu versuchen, eine Szene genauso detailliert zu beschreiben, wie sie im Kino vermittelt wird“ (Monaco, 2001, S. 47).

Zum dritten von Monaco (2001) genannten Unterschied zwischen Text und Film – die Rolle des Erzählers – sind die Meinungen kontrovers. Auf der einen Seite steht die Behauptung Monacos, dass ein Film alles wiedergeben könne, was ein Text kann. Auf der anderen Seite wird ausgeführt, dass das Medium Film nicht abstrakt erzählen könne (z.B. Seitz, 1979). Es würde die Möglichkeit fehlen, den Ablauf einer Handlung auf einer abstrakten oder philosophischen Ebene kommentieren zu können. Waara (2003) führt dazu jedoch an, dass es auch im Medium Film mittels „voice over-Kommentaren“ (S. 8) die Möglichkeit zum

abstrakten Erzählen gibt, ganz gleich, wie es im Filmkontext umgesetzt wird und ungeachtet der Tatsache, dass Seitz (1979) dies als „unfilmisch“ (S. 396) bezeichnet. Da weder im Text „Die zwölf Geschworenen“ noch im Film ein Erzähler agiert, greift in der vorliegenden Arbeit auch dieser von Monaco bezeichnete Unterschied zwischen Text und Film nicht.

9. Durchführung

9.1. Voruntersuchungen

In Vorbereitung auf das Experiment wurden zwei Voruntersuchungen durchgeführt. Um der Intention des Experimentes, die Wirkung verschiedener Stimuli auf die Beurteilung von Personen zu untersuchen, zu entsprechen, sollte für das Hauptexperiment zunächst sichergestellt sein, dass alle Probanden hinsichtlich der Verfügbarkeit der Informationen über die zu beurteilenden Personen im Gedächtnis eine in etwa gleiche Ausgangslage haben. Um zu eruieren, wie oft der Filmausschnitt geschaut werden muss, um eine etwa gleich große Informationsmenge über die Personen im Gedächtnis zu behalten wie die Personen, denen der Textausschnitt vorgelegt wurde, wurde die erste Voruntersuchung durchgeführt. Die Voruntersuchung beinhaltete nach der Darbietung des Stimulusmaterials ausschließlich eine Gedächtnisabfrage.

Hierfür wurden 72 Studenten aller Fachrichtungen der Fachhochschule Nordhausen geworben. Die Studenten wurden zufällig 6 etwa gleich großen Gruppen zugeordnet. Eine Gruppe erhielt den Stimulustext zum Lesen, mit der Instruktion, diesen Text sorgfältig durchzuarbeiten und sich den Inhalt des Textes einzuprägen. Die Textbearbeitung sollte erst dann beendet werden, wenn die Probanden die Geschworenen-Nummer deren Gesprächsinhalten zuordnen können. Den anderen fünf Gruppen wurde der Filmausschnitt des Stimulusmaterials unterschiedlich oft dargeboten. Die erste Gruppe sah den Filmausschnitt einmal, die zweite zweimal mit einer kurzen Pause, die dritte dreimal, die vierte viermal und die fünfte fünfmal hintereinander mit jeweils einer kurzen Pause. Nach der Stimulusdarbietung sollten alle Probanden ihre Gedächtnisinhalte frei wiedergeben. Diese Ergebnisse wurden hinsichtlich der genannten Originalinformationen und der genannten synonymen Informationen, also hinsichtlich Informationen, die eine andere sprachliche Oberfläche haben als die Originalinformation, ausgewertet. Es ergab sich, dass die Gruppen,

die den Filmausschnitt einmal oder zweimal gesehen hatten, nicht in der Lage waren, die Geschworenen-Nummer dem Gesprächsinhalt zuzuordnen und dass die reproduzierten Gedächtnisinhalte weit hinter den von der Textgruppe wiedergegebenen Inhalten zurückblieben. Die Gruppe, die den Filmausschnitt dreimal gesehen hatte, war in der Lage, die Geschworenen-Nummer dem jeweiligen Gesprächsinhalt und ihrem Handeln zuzuordnen und erinnerte in etwa gleiche Gedächtnisinhalte wie die Textgruppe. Bei den Gruppen, die den Filmausschnitt viermal und fünfmal angesehen hatten, verbesserte sich dieses Ergebnis kaum noch. Die quantitativen Ergebnisse der ersten Voruntersuchung finden sich im Anhang. Um die gleiche Informationsbasis über die später zu beurteilenden Personen im Gedächtnis bei allen Probanden vorliegen zu haben, damit Unterschiede in der Beurteilung auch deutlich werden können, wurde aufgrund der Ergebnisse der Voruntersuchung das Untersuchungsdesign in der Hinsicht bestimmt, dass der Filmausschnitt im Experiment dreimal hintereinander mit einer kurzen Pause dargeboten werden sollte.

Die zweite Voruntersuchung diente dazu, für die Beurteilung der Geschworenen relevanten Eigenschaftsitems, die in der Ratingskala des Experimentes genutzt werden sollten, zu generieren. Für diese Voruntersuchung wurden 108 Studierende der Psychologie im 1. Semester der Friedrich-Schiller-Universität Jena rekrutiert. Die Studierenden wurden in zwei gleich große Gruppen unterteilt. Jede Gruppe sollte aus dem Film „Die zwölf Geschworenen“ sechs Geschworene beobachten und im Anschluss an die Darbietung des gesamten Filmes diese Geschworenen frei beurteilen. Dazu sollten die Studierenden zu jedem der sechs zu beobachtenden Geschworenen die drei ihrer Meinung nach hervorstechendsten Eigenschaften notieren und diese anhand von Verhaltensbeispielen begründen. Diese Eigenschaftszuschreibungen wurden inhaltlich ausgewertet. Für jeden der in diesem Experiment zu beurteilenden Geschworenen wurden die sechs am häufigsten genannten Eigenschaften ausgewählt. Die Ergebnisse zur zweiten Voruntersuchung finden sich im Anhang. Um zu einem ausgewogenen Verhältnis zwischen positiven und negativen Eigenschaftszuschreibungen zu gelangen, wurden einige Eigenschaftszuschreibungen negiert, ohne diese inhaltlich zu verfälschen (z.B. höflich zu unhöflich oder intolerant zu tolerant). Die Anordnung der Eigenschaften für die im Experiment verwendete Ratingskala erfolgte zufällig.

9.2. Stimulusinformation: Text

Die Versuchspersonen wurden in einen Raum geführt und der Versuchsleiter stellte sich vor und legte das Anliegen des Experimentes dar. Dann wurde den Versuchspersonen die Durchführung erläutert. Alle Versuchspersonen erhielten die Information, dass in dem Experiment nach Darbietung der Stimulusinformation eine Gedächtnisabfrage zum dargebotenen Stimulus erfolgen wird und dass anschließend eine Beurteilung der handelnden Personen von ihnen vorgenommen werden soll. Diese Information wurde allen Teilnehmern gegeben, um gleiche Verarbeitungsziele (Fiske & Neuberg, 1990; Krolak-Schwerdt & le Coutre, 2002) bei allen Probanden voraussetzen zu können und um einen einseitigen kategoriengeleiteten oder merkmalsbasierten Informationsverarbeitungsprozess, der größtenteils auf die Instruktion zurückzuführen wäre, zu verhindern.

Zuerst erhielten alle Versuchspersonen einen kurzen Fragebogen. Bei den Experimenten, die mit Studenten durchgeführt wurden (Gruppe der Novizen und Gruppe der Trainierten), wurden das Geschlecht, das Alter, die Studienrichtung und die Anzahl der Semester abgefragt. Des Weiteren sollten die Versuchspersonen angeben, ob sie den Film oder das Buch „Die zwölf Geschworenen“ von Reginald Rose oder den Originaltitel „Twelve angry man“ kennen. Die Gruppe der Experten sollte in dem Fragebogen ebenfalls ihr Geschlecht, ihr Alter sowie die Anzahl der Berufsjahre im Schuldienst angeben. Auch die Experten sollten die Frage beantworten, ob sie den Film oder das Buch „Die zwölf Geschworenen“ von Reginald Rose oder den Originaltitel „Twelve angry man“ kennen. Dann begann das eigentliche Experiment.

In diesem Experiment befinden sich die Versuchspersonen in einer so genannten zero acquaintance situation (Kenny, 1994). Kenny (1994) bezeichnet so eine Situation, in der die Urteiler die Beurteilten nicht kennen und keine Gelegenheit haben, mit ihnen zu interagieren. Diese Situationen eignen sich für psychologische Experimente optimal, da die Versuchspersonen die Protagonisten nicht kennen und nicht mit ihnen in Kontakt treten können. Sie können die Protagonisten jedoch wahrnehmen und – frei von Interaktionseinflüssen – beurteilen. Auch im Alltagsleben finden sich solche Situationen, z.B. wenn eine Person in einem Restaurant eine fremde Person beurteilt, die nicht bemerkt, dass sie beobachtet wird. Selbst in Assessment Centern finden sich diese Situationen wieder: der Beurteiler kennt die zu beurteilenden Bewerber in aller Regel nicht und greift auch in das

Geschehen (z.B. Gruppendiskussion zwischen den Bewerbern) nicht ein. Die zero acquaintance situation (Kenny, 1994) wird hier realisiert, indem die Versuchspersonen entweder einen Text lesen, der ein Gespräch von mehreren Personen, die dann zu beurteilen sind, wiedergibt oder zusätzlich zu dem Text Fotos von den zu beurteilenden Personen erhalten. Die dritte Variation zur Realisierung der zero acquaintance situation im vorliegenden Experiment ist das Ansehen eines Filmausschnittes aus dem Film „Die zwölf Geschworenen“ von 1957.

Alle Versuchspersonen erhielten dieselbe Vorinformation, unabhängig davon, welcher Stimulus in der Folge dargeboten wurde. Als Vorinformation wurde eine Tonbandaufnahme abgespielt, die ca. 3 Minuten dauerte. Diese Aufnahme ist dem Film „Die 12 Geschworenen“ von 1957 entnommen. Mit den Worten des Richters, welche die Versuchspersonen hier nur akustisch wahrnehmen konnten, beginnt auch die Handlung des o.g. Filmes. Dieser Einstieg wurde gewählt, um die Versuchspersonen in die Situation, in welcher sie die Protagonisten in dem folgenden dargebotenen Stimulusmaterial wieder finden werden, einzuführen.

Die Aufnahme gibt die Erläuterungen des Richters am Ende einer Gerichtsverhandlung an die 12 Geschworenen wieder, und verdeutlicht damit den Auftrag der Geschworenen, darüber zu befinden, ob der Angeklagte schuldig gesprochen wird oder nicht. Vor der Darbietung wurde den Versuchspersonen mitgeteilt, dass die Aufnahme lediglich dazu dient, ihnen die Situation, in der sich die 12 Geschworenen befinden, zu verdeutlichen und dass der Richter, der hier spricht, später nicht von ihnen beurteilt ist. Der vollständige Wortlaut der Tonbandaufnahme findet sich im Anhang wieder.

Auf eine optische Wiedergabe wurde bewusst verzichtet, da im Film alle 12 Geschworenen bereits durch verschiedene Kameraeinstellungen gezeigt werden. Die gewünschte Variation der Stimulusinformation (Text, Text mit Foto und Film) über die zu beurteilenden Geschworenen würde dadurch verfälscht, da in diesem Fall die Versuchspersonen bereits vor Darbietung des Materials einen optischen Eindruck der Geschworenen erhalten würden.

Die Versuchspersonen erhielten die Instruktion, den nachfolgend vorgelegten Text aufmerksam zu lesen und durchzuarbeiten, ohne sich Notizen anzufertigen. Sie sollten mit der Textbearbeitung erst dann aufhören, wenn sie der Meinung waren, dass sie die Textinformationen den jeweiligen Geschworenen-Nummern zuordnen konnten (Instruktion siehe Anhang). Die Versuchspersonen benötigten ca. 20 bis 25 Minuten für diese Aufgabe. Dann wurde den Versuchspersonen der Text abgenommen.

Die Versuchspersonen wurden im Anschluss daran aufgefordert, alle Informationen, an welche sie sich noch erinnern, niederzuschreiben. Nach der Erinnerungsaufgabe folgte die Anweisung, von den Geschworenen mit den Nummern 1, 3, 7, 8 und 10 eine freie Beurteilung abzugeben. Diese sollte so präzise und detailliert wie möglich abgehalten werden. Nach der Beurteilungsaufgabe folgte die Instruktion, ein Rating für jeden der fünf ausgewählten Geschworenen auf dreißig vorgegebenen Items vorzunehmen. Für jeden der fünf einzuschätzenden Personen erhielten die Versuchsteilnehmer eine Ein-Item Ratingskala.

Die Ein-Item-Ratingskala dient der Messung der Einschätzungen der Versuchsteilnehmer von Ausprägungen von Persönlichkeitsmerkmalen (Stahlberg & Frey, 1990). Die in der Ein-Item-Ratingskala aufgeführten Eigenschaften und Persönlichkeitsmerkmale wurden in einer Voruntersuchung entwickelt. Für jeden der fünf zuvor frei beurteilten Geschworenen sollten die Versuchspersonen ihre Einschätzung für die Ausprägung der dreißig genannten Persönlichkeitseigenschaften angeben. Für die Einschätzung der Ausprägung wurde den Versuchspersonen eine neunstufige Ratingskala vorgegeben. Die einzelnen Abstufungen innerhalb der Ratingskala von 1 bis 9 wurden durch gleich große Abstände dargestellt. Der Wert „1“ auf der Ratingskala bedeutete hinsichtlich der Ausprägung eines Merkmals „sehr gering“, der Wert „9“ bedeutete „sehr hoch“. Somit erhielt die Einschätzung der Versuchspersonen, dass eine Eigenschaft kaum ausgeprägt ist, den Wert 1, die höchste Ausprägung den Wert 9 und eine mittlere Ausprägung entsprechend den Wert 5.

Die Auswahl einer Ein-Item-Ratingskala mit neun Abstufungen erfolgte, um den Versuchspersonen ausreichend Differenzierungsmöglichkeiten bei der Einschätzung der Ausprägungen der einzelnen Persönlichkeitseigenschaften zu geben. Eine Abstufung in neun Intervallen lässt aufgrund der ungeraden Intervallanzahl eine mittlere Ausprägung – den Wert „5“ - für alle Items zu. Des Weiteren besteht durch die Auswahl der neunstufigen Skala die Möglichkeit für die Versuchspersonen, bei der Einschätzung der Eigenschaften eine „Mitte“ zwischen einer Extremausprägung und der mittleren Ausprägung „5“ zu finden – die Werte „3“ und „7“. Um die Möglichkeit der mittleren Eigenschaftsausprägung einzuräumen, wurde auch eine 5 stufige Skalierung in Betracht gezogen. Diese Idee wurde jedoch verworfen, zum einen da diese Skalierung nicht ausreichend Differenzierungsmöglichkeiten hinsichtlich der Eigenschaftseinschätzung bot. Zum anderen bestand bei der Gruppe der Experten – Pädagogen im Schuldienst mit langjähriger Berufserfahrung – die Möglichkeit, dass eigene Erfahrungen, die sie mit einer fünfstufigen Beurteilungs- und Zensurenkala im Schuldienst

der DDR gemacht hatten, bei der Verwendung einer fünfstufigen Ratingskala einfließen (Berliner & Carter, 1986, Bromme, 1992). Dies ist insofern problematisch, als dass die fünf Intervalle der Zensurenkala eher einer Ordinal- als einer Intervallskala entsprachen. Zum Beispiel wurde bei der Leistungsbeurteilung bei Erfüllung der Hälfte der Anforderungen durch den Schüler nicht etwa die Note mit der mittleren Ausprägung „3“ erteilt, sondern die Note „4“. Die Verwendung einer Skalierung, bei der alle Versuchsteilnehmer zumindest augenscheinlich nur wenig bzw. keine Routine entwickeln konnten, kann somit Verzerrungen beim Eigenschaftsrating, die durch Routinen bei einer der Versuchspersonengruppen belegt wären, verringern.

Zum Abschluss der Beurteilung sollten die Versuchspersonen angeben, wie sicher sie sich bei der Beurteilung der einzelnen Geschworenen fühlten. Die Sicherheitseinschätzung wurde als Skala präsentiert, an deren linken Ende die Zahl „0“ stand für „überhaupt nicht sicher“. Am rechten Ende der Skala war die Zahl „1“ notiert und bedeutete „absolut sicher“. Innerhalb dieser Skala gab es keine fest vorgegebenen Abstufungen, so dass die Versuchsteilnehmer ihre subjektive Sicherheit bei der Beurteilung des jeweiligen Geschworenen auf einem Kontinuum zwischen den beiden Extremen angeben sollten. Sie erhielten entsprechend die Instruktion, diejenige Stelle auf der Linie zwischen beiden Polen der Skala anzukreuzen, die ihrem subjektiven Sicherheitsempfinden bei der Beurteilung entspricht. Die Länge der Linie zwischen den beiden Polen betrug jeweils 13,4 cm. Dadurch ergab sich für die Einschätzung der subjektiven Sicherheit ein Wertebereich von Null bis 13,4. Hierbei brachte ein hoher Wert ein hohes Maß an subjektiver Sicherheit und ein niedriger Wert ein entsprechend niedriges Maß an subjektiver Sicherheit zum Ausdruck. Somit erhielt die Einschätzung, dass sich die Versuchsperson bei der Beurteilung eines Geschworenen überhaupt nicht sicher war, den Wert „0“ und bei der Einschätzung, dass sich die Versuchsperson bei der Beurteilung absolut sicher war, die höchste Ausprägung, den Wert 13,4. Um die subjektive Sicherheit im Ergebnis als vorstellbares Maß handhaben zu können, wurde bei der Auswertung die Länge der Linie bis zur markierten Stelle vermessen und in einen prozentualen Wert umgerechnet. Hierauf wird im Abschnitt Ergebnisse eingegangen.

Die Durchführung des gesamten Experimentes dauerte etwa 2 Stunden.

9.3. Stimulusinformation: Text mit Bild

Die Versuchspersonen erhielten gleichzeitig den identischen Text der ersten Stimulusvariante und zusätzlich von den fünf Geschworenen, die in der Textszene agieren, ein Foto vorgelegt. Das Foto jedes Geschworenen war 9 cm x 13 cm groß und bildete jeweils das Gesicht des Geschworenen sowie Teile des Oberkörpers ab. Es wurde darauf geachtet, dass der Anteil des Gesichtes jedes Geschworenen in etwa gleich groß gehalten wurde, um etwaige Einflüsse durch verschieden große Gesichtsdarstellungen (Zuckermann et. al, 1994) zu minimieren. Die insgesamt fünf Fotos waren auf zwei DIN A 4 Seiten aufgeklebt und unter dem Bild war jeweils die Nummer des entsprechenden Geschworenen notiert. Die Bilder entstammen dem Filmausschnitt der dritten Stimulusvariante.

Die Probanden erhielten die Instruktion, den vorgelegten Text aufmerksam zu lesen und durchzuarbeiten und sich die dazugehörenden Bilder genau anzusehen, ohne sich Notizen anzufertigen. Sie sollten mit der Textbearbeitung erst dann aufhören, wenn sie der Meinung waren, dass sie die Textinformationen den jeweiligen Geschworenen-Nummern zuordnen konnten (Instruktion siehe Anhang). Die Versuchspersonen benötigten ca. 20 bis 25 Minuten für diese Aufgabe. Dann wurde den Versuchspersonen der Text abgenommen. Der weitere Ablauf des Experimentes erfolgte ebenso wie bei der Stimulusvariante Text. Zuerst erfolgte die Gedächtnisabfrage zum Text, anschließend sollten die Probanden die fünf Geschworenen mit eigenen Worten beurteilen. Im Weiteren erfolgte das Rating, bei dem die Probanden für jeden Geschworenen die Ausprägung von 30 vorgegebenen Eigenschaften einschätzen sollten. Als letztes sollten die Geschworenen auch hier angeben, wie sicher sie sich bei der Einschätzung jedes Geschworenen gefühlt haben. Die Durchführung des Experimentes dauerte etwa 2 Stunden.

9.4. Stimulusinformation: Film

Die dritte Variation der Stimulusinformation war die Wiedergabe eines Filmausschnittes aus dem Film „Die zwölf Geschworenen“ aus dem Jahre 1957 von Reginald Rose. Der gewählte Filmausschnitt hat eine Länge von ca. 3 Minuten. Der Filmausschnitt wurde auf einem ausreichend großen Monitor mit guter akustischer Wiedergabe präsentiert. Es wurde darauf geachtet, dass alle Probanden eine gute Sicht hatten. Der Filmausschnitt wurde mit einer

kurzen Pause jeweils dreimal hintereinander dargeboten. Der weitere experimentelle Verlauf wurde wie in den beiden anderen beschriebenen Stimulusvariationen gestaltet. Es folgten freier Gedächtnisabruf, freie Beurteilung von jedem der 5 handelnden Geschworenen, Rating auf einer Ratingskala für jeden der 5 Geschworenen und Angabe der Sicherheit bei der Beurteilung durch die Versuchspersonen aufeinander. Die Durchführung des gesamten Experimentes dauerte ungefähr 2 Stunden.

9.5. Aufbereitung der Protokolle der Gedächtnisabfrage und zur freien Beurteilung der Geschworenen

Um statistisch analysierbare Daten aus den Protokollen zur Gedächtnisabfrage und aus den freien Bewertungen zu erhalten, wurden diese zunächst inhaltsanalytisch ausgewertet. Dabei wurden für die Gedächtnisabfrage folgende Analysekatogorien entwickelt: 1.) Originalinformation und alle Reproduktionen, die wörtlich wiedergegeben wurden; 2.) synonyme Wiedergabe von Originalinformation mit anderer sprachlicher Oberfläche oder die Benennung von konkretem Verhalten eines Geschworenen (z.B. „Nr. 7 stimmt Vorschlag von Nr. 1 zu“); 3.) Benennung von Eigenschaften oder Persönlichkeitsmerkmalen der Geschworenen, die die Probanden aus dem Text/Film generierten; 4.) Benennung von grob zusammengefasstem Verhalten (z.B. „Im weiteren diskutieren Nr. 7 und Nr. 8“). Weiterhin wurde gezählt, wie viele falsche Erinnerungen wiedergegeben wurden. Für die Probanden, die als Stimulusmaterial den Filmausschnitt sahen, wurde zusätzlich eine Kategorie eingeführt, in der die Informationen, die ausschließlich in diesem Medium erinnert werden konnten (visuelle Informationen, z.B. „die Geschworenen sitzen um einen ovalen Tisch“ oder „an der Decke des Zimmers befindet sich ein Ventilator“) dargestellt werden konnten. Wenn keine Nennungen zu einer bzw. mehreren Analysekatogorien gemacht wurden, so wurde dieses Ergebnis mit „0“ angegeben und ging ebenfalls in die statistische Auswertung mit ein.

Für die statistische Auswertung der freien Bewertung der Geschworenen wurden als Analysekatogorien 1.) die Nennung von Eigenschaftszuschreibungen und Persönlichkeitsmerkmalen für jeden Geschworenen; 2.) genanntes Verhalten von Geschworenen, um z.B. generierte Eigenschaften oder Merkmale zu begründen und 3.) die Anzahl der Dimensionen der genannten Persönlichkeitseigenschaften und Merkmale, die von den Probanden verwendet wurden. Hierbei wurde das Fünf-Faktoren-Modell der

Persönlichkeit mit den fünf Dimensionen Extraversion, Soziale Verträglichkeit, Gewissenhaftigkeit, Emotionale Stabilität und Intellekt/Offenheit zugrunde gelegt. Jede genannte Eigenschaft wurde einer Dimension zugeordnet und die Zahl der genutzten Dimensionen wurde ausgezählt. Alle Protokolle wurden von drei Beurteilern nach den genannten Kategorien analysiert und quantifiziert. Bei den drei Beurteilern handelt es sich zum einen um einen Diplom-Psychologen, einen Diplom-Pädagogen und einen Verwaltungswirt.

Die Protokolle zur freien Bewertung der Geschworenen wurden des Weiteren einer qualitativen Inhaltsanalyse unterzogen, um zu untersuchen, ob sich die verwendeten Beurteilungskategorien für die verschiedenen Stimuli und für die verschiedenen Probandengruppen unterscheiden. Die Kategorienentwicklung erfolgte induktiv, also schrittweise aus dem Material heraus (Mayring, 2000) und durch Überarbeitung der Kategorien nach Sichtung von ca. der Hälfte der Protokolle. Für die endgültige Auswertung des Materials wurden nur die Kategorien, in denen die Unterschiede deutlich werden, die für die Beantwortung der Fragestellung der vorliegenden Arbeit relevant sind, herangezogen. Folgende Kategorien wurden verwendet: 1.) Nennungen und Bewertungen zum äußeren Erscheinungsbild der Geschworenen; 2.) Nennungen und Bewertungen zum Sprachgebrauch der Geschworenen; 3.) Einschätzungen, bei denen die Geschworenen einer globalen übergeordneten Kategorie zugeordnet wurden (z.B. Mitläufer, mittelbürgerlicher Familienvater); 4.) Nennungen von Eigenschaften und Bewertungen, zu denen die dargebotenen Stimuli keine Informationen geben und die sehr global sind (z.B. unzufrieden mit eigenem Leben, aus intellektueller Mittel- oder Oberklasse); 5.) Bewertungen zur Sympathie / Antipathie. Des Weiteren wurden die Protokolle hinsichtlich der übrigen genutzten Beurteilungskategorien auf Unterschiede zwischen den Stimulusvariationen untersucht. Die Protokolle wurden wiederum alle von den drei genannten Beurteilern qualitativ inhaltsanalytisch ausgewertet.

10. Ergebnisse

10.1. Ergebnisse hinsichtlich der Variation des Stimulusmaterials

10.1.1. Ergebnisse zur Reproduktion von Gedächtnisinhalten

Zunächst sollen die Ergebnisse zur Reproduktion von Gedächtnisinhalten in Abhängigkeit der dargebotenen Stimuli betrachtet werden. Die statistische Auswertung der Ergebnisse zur Reproduktion von Gedächtnisinhalten zur Untersuchung von Unterschieden zwischen den Gruppen bei Darbietung verschiedener Stimuli erfolgte durch Anwendung der einfaktoriellen Varianzanalyse (ANOVA) für die gewählten Analysekategorien Originalinformation, synonyme Information, Eigenschaftsnennungen, Zusammenfassungen und fälschlicherweise wiedergegebene Intrusionen. Alle Ergebnisse, Grafiken und statistischen Auswertungen finden sich detailliert im Anhang.

Insgesamt wurden die meisten Informationen sinngemäß – also synonym mit einer anderen sprachlichen Oberfläche oder als Beschreibung eines konkreten wahrgenommenen Verhaltens – wiedergegeben (Mittelwert $M = 21,95$), während in den anderen Reproduktionskategorien mit Mittelwerten zwischen $M = 3,63$ für die Reproduktion von Originalinformationen und $M = 4,92$ für die Reproduktion von zusammengefassten Verhaltensbeschreibungen weit weniger Nennungen erfolgten.

Ergebnistabelle 1 im Anhang stellt die Mittelwerte und Standardabweichungen für die reproduzierten Originalinformationen unter allen drei Stimulusvariationen dar. Die beste Reproduktionsrate von originalgetreuen Informationen ergab sich für die Stimulusvariation Text mit Bild mit einem Mittelwert $M = 5,52$, während sich die Mittelwerte der beiden anderen Stimulusvarianten mit einem Mittelwert $M = 2,63$ für die Variation Text und mit einem Mittelwert $M = 2,80$ für die Variation Film kaum unterscheiden. Der hier vorhandene Gruppenunterschied zwischen den Stimulusvariationen Text und Text mit Bild ist mit $F = 4,370$; $p < .05$ auf dem 5 % Niveau signifikant.

Mit einem Mittelwert $M = 25,10$ wurden wiederum in der Stimulusvariante Text mit Bild die meisten sinngemäßen Reproduktionen genannt, die Varianten Text und Film liegen mit

Mittelwerten von $M = 21,22$ (Text) und $M = 18,87$ (Film) unter diesen Nennungen (Ergebnisdarstellung siehe Punkt 2. im Anhang).

Bei der Nennung von Eigenschaften, die aus den Stimuli generiert wurden, und die nicht lediglich die dargebotenen Informationen wiedergeben und somit bereits Urteilscharakter haben, zeigt sich, dass die meisten Nennungen von den Probanden der Filmgruppe (Mittelwert $M = 4,87$) gemacht wurden, während in der Gruppe Text mit Bild die wenigsten generierten Eigenschaften und Merkmale (Mittelwert $M = 2,95$) genannt wurden (Ergebnisdarstellung siehe Punkt 3. im Anhang).

In der Kategorie Zusammenfassungen wurden wiederum in der Filmgruppe mit einem Mittelwert $M = 5,87$ die meisten Items, in der Gruppe Text mit Bild mit einem Mittelwert $M = 4,19$ die wenigsten Items reproduziert (Ergebnisdarstellung siehe Punkt 4. im Anhang).

Die hier aufgeführten Mittelwertsunterschiede sind – bis auf den oben genannten signifikanten Unterschied zwischen den Gruppen Text und Text mit Bild für die reproduzierten Originalinformationen - zwar nicht signifikant auf dem 5 % Niveau, weisen jedoch alle in die gleiche Richtung: die Reproduktionsrate von Items, die originalgetreu ist oder sinngemäß der dargebotenen Stimulusinformation entspricht, ist in der Gruppe der Stimulusvariante Text mit Bild am höchsten, und zugleich werden in dieser Gruppe die wenigsten Eigenschaftszuschreibungen und die wenigsten zusammengefassten Inhalte genannt. Die Ergebnisse der Stimulusvariante Film tendieren dazu genau entgegengesetzt: hier werden die wenigsten Originalinformationen und synonymen Informationen wiedergegeben, wogegen die meisten generierten Eigenschaften und zusammengefasste Verhaltensbeschreibungen genannt werden. Die Ergebnisse der Stimulusvariante Text liegen in allen Analysekatoren zwischen denen der Stimuli Text mit Bild und Film.

Signifikante Gruppenunterschiede auf dem 5 % Niveau ergeben sich für die Anzahl an fälschlicherweise wiedergegebenen Items zwischen den Gruppen Text und Film ($F = 5,445$; $p < .05$) und zwischen den Gruppen Text mit Bild und Film ($F = 9,784$; $p < .05$): hier wurden von den Probanden der Filmgruppe die wenigsten Intrusionen (Mittelwert $M = 0,33$) genannt, während die Probanden der Gruppe Text mit Bild die meisten Intrusionen (Mittelwert $M = 1,43$) wiedergab. Die Textgruppe lag mit einem Mittelwert $M = 1,19$ für fälschlicherweise genannte Items wiederum zwischen den beiden anderen Stimulusvariationen. Alle Ergebnisse zu den genannten Intrusionen finden sich detailliert im Anhang unter Punkt 5.

10.1.2. Ergebnisse zur Beurteilung der Geschworenen

Bei der statistischen Auswertung der freien Beurteilung zu den dargebotenen Personen wurden die Anzahl der genannten Eigenschaftszuschreibungen und Persönlichkeitsmerkmale für jeden Geschworenen, die Anzahl der genannten Verhaltensbeschreibungen aus der Stimulusinformation und die Anzahl der genutzten Dimensionen im Sinne des Fünf-Faktoren-Modells der Persönlichkeit (vgl. Norman, 1963; Henss, 1998) unter Anwendung der einfaktoriellen Varianzanalyse (ANOVA) auf Unterschiede zwischen den drei Stimulusvarianten untersucht.

Für alle Stimulusvariationen wurden unter der Instruktion der freien Beurteilung der Geschworenen von den Probanden deutlich mehr erschlossene bzw. generierte Eigenschaften und Personmerkmale (Mittelwert $M = 4,21$) genannt als Verhaltensweisen (Mittelwert $M = 1,73$), die eher Reproduktionen der Stimulusinformation widerspiegeln.

Punkt 11. im Anhang enthält detailliert die Mittelwerte und Standardabweichungen für die genannten Eigenschaftszuschreibungen in Abhängigkeit der dargebotenen Stimulusinformation. Hierbei zeigte sich zwischen den Gruppen Text und Film ($F=11,918$; $p<.05$) und zwischen den Gruppen Text und Text mit Bild ($F=4,622$; $p<.05$) signifikante Unterschiede auf dem 5 % Niveau. Die wenigsten Eigenschaftszuschreibungen und Nennungen für Persönlichkeitsmerkmale erfolgten für die Stimulusvariante Text (Mittelwert $M = 3,65$), die meisten Eigenschaftszuschreibungen wurden von Probanden der Filmgruppe vorgenommen (Mittelwert $M = 5,04$). Mit einem Mittelwert $M = 4,34$ lagen die Ergebnisse der Bedingung Text mit Bild zwischen den Bedingungen Text und Film.

Im Anhang unter Punkt 12. sind die Mittelwerte und Standardabweichungen für die genannten Verhaltensweisen der Geschworenen in Abhängigkeit der Stimulusvariation dargestellt. Die Gruppenunterschiede zwischen den Stimulusvariationen Text und Film ($F=15,003$; $p<.05$) und Text mit Bild und Film ($F=22,405$; $p<.05$) sind auf dem 5 % Niveau hoch signifikant. Die meisten Verhaltensbeschreibungen wurden von den Probanden der Filmgruppe (Mittelwert $M = 2,57$) vorgenommen, während sich die Anzahl der Verhaltensbeschreibungen der Stimulusgruppen Text (Mittelwert $M = 1,59$) und Text mit Bild (Mittelwert $M = 1,31$) kaum unterscheidet. Für die Anzahl der genutzten Dimensionen ergaben sich bei der statistischen Auswertung keine signifikanten Unterschiede (siehe Anhang Punkt 13.).

Bei der qualitativen Inhaltsanalyse wurden die Unterschiede hinsichtlich der für das Urteil herangezogenen Kategorien, der verwendeten Urteilkategorien und Bewertungen zu Sympathie und Antipathie vorgenommen.

Hinsichtlich der für die Urteilsbildung herangezogenen Kategorien zeigten sich die meisten Unterschiede zwischen den Stimulusvarianten Text / Text mit Bild und Film. Bei der Stimulusvariante Film werden für die Urteilsbildung visuelle Informationen und akustische Informationen genutzt, während für die Variante Text mit Bild visuelle Informationen nur zu einem verschwindend geringen Anteil genutzt wurden. So wurden für die Bedingung Text mit Bild insgesamt 6 Nennungen zum äußeren Erscheinungsbild der zu beurteilenden Personen vorgenommen, die sich auf das Alter und den Gesichtsausdruck bezogen. Bei der Bedingung Film wurden hingegen insgesamt 134 Äußerungen zum Erscheinungsbild der zu beurteilenden Geschworenen gemacht, die sich auf Geschlecht, Alter, Kleidung, Gestik, Gesichtsausdruck, Sitzhaltung und Sitzplatz bezogen. Die Nutzung von akustischen Informationen in der Bedingung Film wurde durch 44 Nennungen zum Gebrauch der Stimme, die sich auf die Lautstärke und die Geschwindigkeit der Sprache, die Artikulation, die Modulation und Lachen bezogen, dargestellt.

Hinsichtlich der Verwendung der Urteilkategorien zeigten sich die folgenden Unterschiede. In den Bedingungen Text und Text mit Bild wurde für drei der fünf zu beurteilenden Geschworenen als Eigenschaftskategorie mit den häufigsten oder den zweithäufigsten Nennungen (zwischen 6 und 10 Nennungen pro Geschworener und pro Stimulus) die Eigenschaft „vorschnell verurteilend“ verwendet. Diese Eigenschaftskategorie, die in den Bedingungen Text und Text mit Bild so häufig verwendet wurde, wird in der Stimulusvariante Film nicht ein einziges Mal – für keinen der Geschworenen – genannt.

Einschätzungen, bei denen die Geschworenen einer globalen übergeordneten Kategorie zugeordnet wurden (z.B. Mitläufer, mittelbürgerlicher Familienvater), wurden von Probanden aller drei Stimulusvariationen vorgenommen, jedoch mit unterschiedlicher Häufigkeit. Für die Bedingung Text wurden mit 4 Nennungen die wenigsten globalen Kategorisierungen für die zu beurteilenden Personen verwendet, während für die Bedingungen Text mit Bild und Film die Häufigkeit mit je 13 Nennungen gleich hoch war. Nennungen von Eigenschaften und Bewertungen, die sehr global sind und zu denen die dargebotenen Stimuli keine Informationen geben (z.B. unzufrieden mit eigenem Leben), wurden ebenfalls in allen drei

Stimulusbedingungen mit unterschiedlicher Häufigkeit vorgenommen. Die Textgruppe nannte lediglich zwei dieser weitreichenden Eigenschaften, die Gruppe Text mit Bild verwandte vier und die Filmgruppe acht globale Eigenschaften.

Nur für die Stimulusvariation Film – und hier nur in der Gruppe der Novizen – fanden sich viele weitere Bewertungen und Äußerungen, die sehr verallgemeinert wurden und die zum Teil nicht mehr mit dem Inhalt der dargebotenen Stimulusinformation in unmittelbarem Zusammenhang stehen. So wurden beispielsweise Mutmaßungen über Intentionen der Geschworenen dargestellt oder Verhaltensvorhersagen für einzelne Geschworene gemacht. Für die Stimulusbedingung Text mit Bild wurden insgesamt 7 Äußerungen zur Sympathie / Antipathie vorgenommen, in den beiden anderen Stimulusvariationen wurden keine Angaben zu diesen Attributen geäußert.

Des Weiteren fand sich als Ergebnis der qualitativen Inhaltsanalyse, dass die Einschätzungen und Bewertungen für den Stimulus Text für die einzelnen Geschworenen in die gleiche Richtung wiesen. Beispielsweise wurde einer der Geschworenen von mehreren Probanden als „entscheidungsscheu“, „zurückhaltend“, „nicht dominant“ oder „verunsichert“ bewertet. Für die Stimulusvariante Text mit Bild findet sich in allen Beurteilungen zu den einzelnen Geschworenen eine größere Bandbreite hinsichtlich genannter Eigenschaften, beispielsweise werden hier für den gleichen Geschworenen „lässig“, „objektiv“, „überfordert“ und „gefühlvoll“ gleich häufig genannt. Die größten konträren Bewertungen innerhalb eines Stimulus hinsichtlich der Einschätzung der Geschworenen ergaben sich in der Stimulusgruppe Film. So wurde z.B. von vier Probanden der Filmgruppe bei dem gleichen Geschworenen eingeschätzt, dass er eine feststehende Meinung habe, die gleiche Anzahl an Probanden der Filmgruppe urteilte jedoch, dass er keine eigene Meinung habe, weiterhin wurde ein Geschworener gleich oft als sicher und als unsicher beschrieben.

Die statistische Auswertung des Ratings ergab keine signifikanten Unterschiede zwischen den drei verschiedenen Stimulusdarbietungen. Bei der Untersuchung der Ergebnisse aus dem Rating fanden sich hingegen große individuelle Unterschiede. Die individuellen Unterschiede lassen sich im Umgang der Probanden bei der Einschätzung der Geschworenen auf der vorgegebenen Ratingskala aufdecken. Die Versuchspersonen scheinen der Bewertung der Geschworenen hinsichtlich vorgegebener Eigenschaften auf einer Ratingskala eine Art eigenen Maßstab zugrunde zu legen, der sich in der Nutzung der Ausprägungen niederschlägt.

So nutzten einige Probanden – unabhängig vom dargebotenen Stimulusmaterial – vor allem die Extremausprägungen 1 und 9, andere nutzten die Ausprägungen am unteren und oberen Ende überhaupt nicht und bewerteten vorwiegend mit mittleren Ausprägungen zwischen 3 und 7, wiederum andere Probanden nutzten alle möglichen Differenzierungen, weitere Versuchspersonen nutzten hauptsächlich die Ausprägungen 2 und 8 und die mittlere Ausprägung 5. Diese scheinbar individuelle Art des Umgangs mit den Bewertungen von Eigenschaften auf einer Ratingskala findet sich für die einzelnen Versuchspersonen für alle 5 zu beurteilenden Geschworenen wieder.

10.1.3. Ergebnisse zur subjektiven Urteilssicherheit

Zur Untersuchung der subjektiven Urteilssicherheit wurden die auf dem Kontinuum einer Skala zwischen „0“ und „1“ von den Versuchspersonen angegebenen Werte, wie sicher sie sich bei der Beurteilung der einzelnen Geschworenen gefühlt haben, in einen prozentualen Wert umgerechnet und zwischen den Gruppen auf Unterschiede untersucht. Hierbei ergab sich zwischen den Gruppen Text und Film ($p < .05$) und zwischen den Gruppen Text mit Bild und Film ($p < .05$) signifikante Mittelwertsunterschiede. Die höchste subjektive Sicherheit (Mittelwert $M = 73,7\%$) bei der Beurteilung der Geschworenen wurde von den Probanden der Filmgruppe empfunden, während die Probanden der Gruppen Text (Mittelwert $M = 56,9\%$) und Text mit Bild (Mittelwert $M = 57,9\%$) ihr subjektives Sicherheitsempfinden bei der Beurteilung der einzelnen Geschworenen annähernd gleich einschätzten.

10.1.4. Hypothesenbezug und Zusammenfassung der Ergebnisse hinsichtlich der Variation der Stimulusinformation

Für die Variation der Stimulusinformation wurden Gruppenunterschiede hinsichtlich der Reproduktionsleistung bei einer Gedächtnisabfrage zu den Personinformationen, hinsichtlich der Beurteilung der Personen und hinsichtlich des subjektiven Sicherheitsgefühls der Probanden bei der Urteilsbildung erwartet.

1.

Für die Reproduktion von Gedächtnisinhalten zeigten sich in der Bedingung Text mit Bild lediglich Tendenzen, dass mehr Informationen originalgetreu und sinngemäß wiedergegeben wurden als in der Bedingung Text und in der Bedingung Film. Gleichzeitig wurden in der Bedingung Text mit Bild bei der Wiedergabe von Personinformationen die wenigsten generierten Eigenschaften und Verallgemeinerungen sowie Zusammenfassungen genannt.

2.

Wie postuliert, wurden für die Beurteilung der Geschworenen in der Bedingung Text mit Bild signifikant mehr Eigenschaften und Persönlichkeitsmerkmale genannt als in der Bedingung Text, was darauf hinweist, dass die Probanden in dieser Stimulusvariation das „mehr“ an Informationen für die Urteilsbildung nutzten. Hinweise auf eine Nutzung der visuellen Informationen des Bildes geben vor allem die genannten Einschätzungen, bei denen die zu beurteilenden Personen weitreichenden globalen Kategorien zugeordnet wurden und die diesen Personen zugeschriebenen Eigenschaften - die sehr global sind und zu denen die dargebotenen Stimuli keine Informationen geben - , die in der Bedingung Text mit Bild mit insgesamt 17 Nennungen deutlich mehr sind als in der Textbedingung mit insgesamt 6 globalen Einschätzungen.

3.

Die Hypothese zur Wirkung eines zusätzlich vorgelegten Fotos zu einem Text auf die subjektiv empfundene Urteilssicherheit konnte nur zum Teil bestätigt werden. Die subjektive Urteilssicherheit in der Bedingung Text mit Bild war erwartungsgemäß niedriger als in der Filmbedingung. Der erwartete Unterschied zwischen den subjektiven Urteilssicherheiten für die Bedingungen Text und Text mit Bild konnte nicht nachgewiesen werden.

4.

Für die Bedingung Film wurde für die Reproduktion von Gedächtnisinhalten eine weniger gute Behaltensleistung für originalgetreue Informationen als in der Bedingung Text mit Bild postuliert. Es wurde erwartet, dass in dieser Stimulusvariation vermehrt zusammengefasste Inhalte der dargebotenen Informationen genannt werden. Signifikante Unterschiede konnten hierfür nicht nachgewiesen werden. Es zeigten sich lediglich Tendenzen in dieser Richtung.

5.

Die für die Beurteilung der Personen postulierte weitere Nennung von Eigenschaften und Persönlichkeitsmerkmalen, die durch Nutzung von Informationen, die aus den zusätzlich dargebotenen Modalitäten Bewegungen der Personen und akustische Informationen resultieren, konnte quantitativ und qualitativ nachgewiesen werden. Es zeigte sich hinsichtlich der Anzahl der Eigenschaftszuschreibungen, dass in der Filmgruppe signifikant die meisten generierten Persönlichkeitsmerkmale genannt wurden. Die qualitative Auswertung zeigt, dass die durch diese Stimulusvariante dargebotenen zusätzlichen visuellen und akustischen Informationen für die Urteilsbildung genutzt wurden.

Ein deutlicher Unterschied findet sich in der Verwendung einer Beurteilungskategorie: für die Varianten Text und Text mit Bild wurde für drei Geschworene als häufigste bzw. zweithäufigste Eigenschaftszuschreibung das Attribut „vorschnell verurteilend“ verwendet, während dieses Attribut bei der Filmbedingung keinem der Geschworenen zugeschrieben wurde. Für die Unterschiedlichkeit der verwendeten Beurteilungskategorien findet sich ein weiteres Indiz: in der Bedingung Film wurden mit insgesamt 21 Nennungen im Vergleich zu den beiden anderen Stimulusvarianten (6 Nennungen für Text und 17 Nennungen für Text mit Bild) die meisten globalen Kategorisierungen und Eigenschaftszuschreibungen für die zu beurteilenden Personen vorgenommen. In der Filmgruppe – und hier nur für die Gruppe der Novizen – fanden sich auch weitere Bewertungen und Äußerungen, die weit über die Informationen der Stimulusdarbietung hinausreichen, die nicht in die gewählten Analysekatoren eingeordnet werden konnten, denen aber aus Sicht des Autors Beachtung geschenkt werden sollte. So wurden beispielsweise Mutmaßungen über Intentionen der zu beurteilenden Personen gemacht und mögliches zukünftiges Verhalten vorhergesagt.

6.

Hinsichtlich der Einschätzung der subjektiven Urteilssicherheit konnte die Hypothese bestätigt werden, dass diese in der Bedingung Film am höchsten ausgeprägt ist.

10.2. Ergebnisse hinsichtlich der Variation der Beurteilererfahrung

10.2.1. Ergebnisse zur Reproduktion von Gedächtnisinhalten

Die statistische Auswertung der Ergebnisse zur Reproduktion von Gedächtnisinhalten zur Untersuchung von Unterschieden zwischen den Gruppen mit unterschiedlicher Beurteilererfahrung erfolgte durch Anwendung der einfaktoriellen Varianzanalyse (ANOVA) für die gewählten Analysekategorien Originalinformation, synonyme Information, Eigenschaftsnennungen, Zusammenfassungen und fälschlicherweise wiedergegebene Intrusionen.

Punkt 6. im Anhang stellt die Mittelwerte und Standardabweichungen für die reproduzierten Originalinformationen in Abhängigkeit der Beurteilererfahrung dar. Die meisten Informationen wurden von der Trainingsgruppe (Mittelwert $M = 5,22$) originalgetreu wiedergegeben, die wenigsten von der Expertengruppe (Mittelwert $M = 2,05$). Die Gruppe der Novizen lag mit einem Mittelwert $M = 3,69$ für richtig wiedergegebene Originalinformationen zwischen den Gruppen der Trainierten und der Experten. Für die Gruppe der Trainierten und die Gruppe der Experten ist der genannte Mittelwertsunterschied mit $p < .05$ auf dem 5 % Niveau signifikant.

Die Mittelwertsunterschiede und Standardabweichungen für die sinngemäße Wiedergabe von Originalinformationen für die einzelnen Beurteilergruppen sind in Punkt 7. im Anhang dargestellt. Die Gruppe der Experten nannte in dieser Kategorie die weitaus wenigsten Items (Mittelwert $M = 7,79$). Die Trainierten reproduzierten die meisten synonymen Informationen (Mittelwert $M = 33,28$). Die Gruppe der Novizen lag mit der Anzahl der Reproduktionen für sinngemäße Informationen mit einem Mittelwert $M = 24,46$ zwischen den Experten und den Trainierten. Alle Gruppenunterschiede für die synonyme Nennung von Originalinformationen sind auf dem 5 % Niveau signifikant. Statistiktabelle 7. im Anhang gibt die Auswertung zur Signifikanz für die Gruppenunterschiede für die Reproduktion sinngemäßer Informationen wieder.

Die Mittelwertsunterschiede für die Nennung von generierten Eigenschaften aus den Stimulusinformationen, die nicht nur eine Reproduktion der Originalinformationen darstellen, sind in Punkt 8. im Anhang zu finden. Die meisten Persönlichkeitsmerkmale und

Eigenschaften wurden von der Novizengruppe (Mittelwert $M = 5,73$) aufgezählt, die wenigsten von den Trainierten (Mittelwert $M = 2,11$). Die Gruppe der Experten lag mit einer mittleren genannten Anzahl $M = 3,21$ zwischen den Novizen und Trainierten. Die genannten Unterschiede zwischen den einzelnen Gruppen sind jedoch nicht signifikant.

Bei der Nennung von zusammengefassten Inhalten finden sich zwischen den Bedingungen Novizen und Experten ($p < .05$) und zwischen den Bedingungen Trainierte und Experten ($p < .05$) signifikante Gruppenunterschiede. Alle Ergebnisse zu den zusammengefassten Informationen sind unter Punkt 9. im Anhang dargestellt. Die Trainierten nannten die wenigsten zusammengefassten Items (Mittelwert $M = 3,17$), die Experten nahmen die meisten Zusammenfassungen (Mittelwert $M = 7,26$) für die dargebotenen Informationen vor, die Novizen lagen mit einem Mittelwert $M = 4,42$ zwischen diesen Gruppen.

Für die falsch genannten Items ergaben sich keine signifikanten Mittelwertsunterschiede zwischen den Gruppen. Hier nannten die Experten die wenigsten (Mittelwert $M = 0,79$), die Trainierten die meisten (Mittelwert $M = 1,39$) und die Novizen (Mittelwert $M = 1,04$) eine mittlere Anzahl an Intrusionen. Die detaillierte Ergebnisdarstellung findet sich unter Punkt 10. im Anhang.

10.2.2. Ergebnisse zur Beurteilung der Geschworenen

Zur statistischen Auswertung der freien Beurteilung der Geschworenen wurden die Gruppenunterschiede zwischen den Probanden mit unterschiedlicher Beurteilererfahrung mittels einfaktorieller Varianzanalyse (ANOVA) hinsichtlich der für die zu beurteilenden Personen genannten Eigenschaftsattribute, hinsichtlich der genannten Verhaltensbeschreibungen aus der Stimulusinformation und hinsichtlich der für die Eigenschaftszuschreibungen genutzten Anzahl an Dimensionen im Sinne des Fünf-Faktoren-Modells der Persönlichkeit (vgl. Norman, 1963; Henss, 1998) untersucht.

Bei der freien Beurteilung zeigten sich bei der Nennung von Eigenschaftsattributen für die repräsentierten Personen zwischen allen Gruppen signifikante Mittelwertsunterschiede, die in Punkt 14. im Anhang dargestellt sind. Die meisten Persönlichkeitsmerkmale für die zu beurteilenden Geschworenen wurden von der Gruppe der Novizen (Mittelwert $M = 5,15$)

genannt. Die wenigsten Eigenschaftsattribute wurden von den Experten (Mittelwert $M = 3,30$) zur Beurteilung herangezogen. Die Trainierten lagen mit einem Mittelwert $M = 3,99$ für die Eigenschaftszuschreibung zwischen den beiden Novizen und den Experten.

Für die Nennung von Verhaltensbeschreibungen ergaben sich signifikante Unterschiede zwischen den Gruppen Novizen und Experten ($p < .05$) und Trainierte und Experten ($p = .05$). Die Experten nannten auch hier die wenigsten Items (Mittelwert $M = 1,20$), während die Zahl der Verhaltensnennungen bei den Novizen (Mittelwert $M = 2,05$) und den Trainierten (Mittelwert $M = 2,03$) gleich hoch ist. Die Ergebnisse gibt Punkt 15. im Anhang wider.

Für die Anzahl der Dimensionen, die bei der Zuschreibung von Eigenschaften bei der Beurteilung berücksichtigt wurden, ergibt sich ein signifikanter Mittelwertsunterschied für die Novizen, die Attribute für die meisten Dimensionen nannten (Mittelwert $M = 2,35$) und die Experten, die die wenigsten Dimensionen berücksichtigten (Mittelwert $M = 1,93$). In Punkt 16. im Anhang finden sich die Ergebnisse zu den bei der Beurteilung der Geschworenen verwandten Eigenschaftsdimensionen.

Bei der qualitativen Auswertung der Ergebnisse zur Variation der Beurteilererfahrung wurden die Verwendung der für die Urteilsbildung herangezogenen Kategorien, die verwendeten Urteilstkategorien, Bewertungen zu Sympathie/Antipathie untersucht.

Unabhängig der eben genannten Aspekte soll ein Befund genannt werden, der in der Interpretation der Ergebnisse nicht übergangen werden sollte. In der Expertengruppe – jedoch nur für den Stimulus Text – nahmen insgesamt 6 Experten – also mehr als die Hälfte der Probanden in dieser Gruppe – für einen, zwei oder sogar drei Geschworene keine verbale Beurteilung vor und führten für diese kein Rating durch. In der Gruppe der Novizen waren insgesamt drei Probanden, die – nur für den Stimulus Film – jeweils einen Geschworenen nicht beurteilten.

Hinsichtlich der Nutzung der für eine Beurteilung herangezogenen Kategorien konnten Unterschiede zwischen den Gruppen mit unterschiedlicher Beurteilererfahrung gefunden werden. So wurden von den Trainierten mit 80 Nennungen die meisten Ausführungen zum äußeren Erscheinungsbild der Geschworenen gemacht, die Experten nannten 35 äußere Merkmale und die Novizen nannten nur 25. Ähnlich, wenn auch nicht in dieser Stärke zeigten sich die Gruppenunterschiede bei der Nutzung von sprachlichen Kategorien. Die Trainierten

fürten 18, die Experten 14 und die Novizen 12 Äußerungen zum Gebrauch der Stimme, die sich auf Lautstärke, Geschwindigkeit, Artikulation, Modulation und Lachen bezogen, an. Bei der Verwendung von Urteilkategorien fanden sich folgende Unterschiede. Einschätzungen, die die zu beurteilenden Personen einer globalen übergeordneten Kategorie zuordneten (z.B. Mitläufer, Spielertyp), wurden am häufigsten von Probanden der Novizengruppe vorgenommen. Die Novizen führten insgesamt 17 solcher Kategorisierungen an, während die Trainierten mit 6 Nennungen die wenigsten Kategorisierungen vornahmen. Die Experten lagen mit 7 Nennungen für derartige globale Kategorisierungen leicht über dem Wert der Trainierten. Die Nennung von Eigenschaften, die sehr global sind, und zu denen die dargebotenen Stimuli keine Informationen geben, erfolgte in den drei Beurteilergruppen ebenfalls mit unterschiedlicher Häufigkeit: die Novizen nannten 6, die Experten 5 und die Trainierten nur 3 dieser weit reichenden Eigenschaftsattribute. Im Zusammenhang mit der Nennung von Eigenschaften und Kategorien, die weit über Einschätzungsmöglichkeiten, die durch den Inhalt der Stimulusinformation gegeben würden, hinausgehen steht auch der Befund, dass die Novizen – und nur bei der Stimulusvariante Film – viele weitere Verallgemeinerungen und Bewertungen vornahmen. So wurde beispielsweise Verhalten für die Geschworenen außerhalb der Gerichtsverhandlung vorhergesagt oder Mutmaßungen über Intentionen der Geschworenen dargelegt.

Bei der Beurteilung der Geschworenen wurden insgesamt 7 Äußerungen zur Sympathie / Antipathie vorgenommen, die sich auf die Novizengruppe mit 5 Äußerungen und auf die Expertengruppe mit 2 Äußerungen aufteilen. Von den Trainierten wurden zur Sympathie und Antipathie in der Beurteilung keine Angaben gemacht.

Die statistische Auswertung des Ratings brachte keine signifikanten Unterschiede.

10.2.3. Ergebnisse hinsichtlich der Urteilssicherheit

Zur Untersuchung der subjektiv empfundenen Sicherheit bei der Beurteilung der Geschworenen wurden die auf dem Kontinuum einer Skala zwischen „0“ und „1“ von den Probanden angegebenen Werte in einen prozentualen Wert umgerechnet und zwischen den Gruppen mit unterschiedlicher Beurteilererfahrung auf Mittelwertsunterschiede untersucht. Hierbei zeigte sich zwischen den Gruppen Novizen und Experten ($p < .05$) und zwischen Trainierten und Experten ($p < .05$) ein hoch signifikanter Mittelwertsunterschied auf dem 5 % Niveau. Für

die Novizen und die Trainierten wurde der Mittelwertsunterschied zwar nicht signifikant, er tendiert jedoch mit Mittelwerten $M = 67,3 \%$ für die Novizen und $M = 61,4 \%$ für die Trainierten in die Richtung, dass die Trainierten ihre Sicherheit bei der Beurteilung der Geschworenen niedriger einschätzten als die Novizen. Die Experten gaben mit einem Wert $M = 52,0 \%$ ihre subjektive Sicherheit bei der Einschätzung der dargebotenen Personen am niedrigsten an.

10.2.4. Hypothesenbezug und Zusammenfassung der Ergebnisse hinsichtlich der Variation der Beurteilererfahrung

Für die Gruppen Novizen, Trainierte und Experten, die sich hinsichtlich ihrer Ausbildung und ihrer Erfahrung in Bezug auf die Beurteilung von Personen unterscheiden, wurden Unterschiede hinsichtlich der Gedächtnisleistung für dargebotene Personinformationen, der Beurteilung der dargebotenen Personen und des subjektiven Sicherheitsgefühls der Probanden bei der Urteilsbildung erwartet.

1.

Es wurde erwartet, dass sich die Gruppe der Trainierten von den Novizen in der Hinsicht unterscheidet, dass sie bei der Reproduktion von gegebenen Personinformationen mehr Originalinformationen erinnern und dass sie in einer Gedächtnisabfrage weniger Zusammenfassungen und weniger generierte Eigenschaften nennt.

Die Trainierten reproduzierten signifikant die meisten synonymen Informationen, sowohl im Vergleich zu den Novizen als auch zu den Experten. Die Gruppe der Trainierten wies von den drei Beurteilergruppen die besten Leistungen für die Wiedergabe von Originalinformationen auf, jedoch war hier nur der Unterschied zwischen den Trainierten und den Experten signifikant. Die Trainierten nannten in der Gedächtnisabfrage die wenigsten generierten Persönlichkeitsmerkmale und Eigenschaften der dargebotenen Personen und sie fassten die dargebotenen Informationen auch am wenigsten zusammen.

2.

Die statistische Auswertung der Ergebnisse zur freien Beurteilung der Geschworenen erbrachte den Befund, dass die Trainierten signifikant weniger Eigenschaftszuschreibungen als die Novizen vornehmen.

Die qualitative Auswertung der Protokolle bestätigt die Hypothese, dass die Trainierten bei der Beurteilung von Personen mehr Details der Originalinformationen berücksichtigen. So notierten die Trainierten im Vergleich zu den Novizen und den Experten die meisten Beschreibungen zum äußeren Erscheinungsbild der zu beurteilenden Probanden und achteten auch am meisten auf Kategorien, die sich auf die Sprache bezogen. Hypothesenkonform ist auch der Befund, dass die Trainierten die wenigsten globalen und weitreichenden Eigenschaftszuschreibungen für die zu beurteilenden Personen vornehmen. Bei der Beurteilung der Personen nutzen sie am wenigsten Kategorisierungen, zu denen die dargebotenen Stimuli keine Informationen geben. Die Trainierten geben zu keiner der zu beurteilenden Person eine Einschätzung zu Sympathie oder Antipathie.

3.

Für die Einschätzung der subjektiv empfundenen Sicherheit bei der Beurteilung der einzelnen Geschworenen konnte die Hypothese, dass die Trainierten ihre Urteilssicherheit niedriger einschätzen als die Novizen, nicht durch signifikante Befunde bestätigt werden. Es finden sich lediglich Tendenzen hinsichtlich der vorliegenden Mittelwertsunterschiede, dass die Trainierten ihre Urteilssicherheit nicht so hoch wie die Novizen einschätzen.

4.

Für die Gruppe der Experten wurde erwartet, dass sie in einer Gedächtnisabfrage im Vergleich zu den Novizen weniger Originalinformationen reproduzieren, jedoch mehr verallgemeinerte und zusammengefasste Inhalte nennen. Diese Hypothese konnte durch die Ergebnisse bestätigt werden. Die Experten gaben die wenigsten Originalinformationen wieder. Ebenso reproduzierten sie signifikant die wenigsten synonymen Informationen, sowohl im Vergleich zu den Novizen als auch zu den Experten. Für die Nennung zusammengefasster Items ergaben sich wie erwartet signifikante Unterschiede zwischen den Novizen und den Experten und zwischen den Trainierten und den Experten. Die Experten nahmen die meisten Verallgemeinerungen und Zusammenfassungen der dargebotenen Inhalte vor.

5.

Die statistische Auswertung der Ergebnisse zur freien Beurteilung der Geschworenen erbrachte ebenfalls den erwarteten Befund, dass die Experten die wenigsten Eigenschaftsattribute für die zu beurteilenden Personen verwendeten und dass sie die

wenigsten Verhaltensbeschreibungen nannten. Auch die Anzahl der genutzten Eigenschaftsdimensionen ist im Vergleich zu Novizen und Trainierten am niedrigsten. Die qualitative Auswertung der Ergebnisse weist entsprechend der Hypothese darauf hin, dass die Experten stärker als die Novizen, jedoch nicht so stark wie die Trainierten Informationen nutzen, die sich auf das äußere Erscheinungsbild der zu beurteilenden Personen oder deren sprachliche Informationen beziehen. Die Einschätzungen, die sehr global sind, wurden von den Experten weit weniger vorgenommen als von den Novizen, jedoch lagen sie noch knapp über der von den Trainierten genannten Anzahl.

6.

Hinsichtlich der subjektiv empfundenen Urteilssicherheit bei der Beurteilung der einzelnen Geschworenen konnte die Hypothese nur teilweise bestätigt werden. Die subjektiv empfundene Urteilssicherheit der Expertengruppe lag hypothesenkonform unter der der Novizen. Die Erwartung, dass die Trainierten ihre Urteilssicherheit noch niedriger einschätzen, als die Experten, konnte nicht bestätigt werden. Stattdessen zeigten sich signifikante Unterschiede sowohl zwischen Novizen und Experten und als auch zwischen Trainierten und Experten für die empfundene Sicherheit bei der Beurteilung der Geschworenen.

11. Diskussion der Ergebnisse

11. 1. Diskussion hinsichtlich der Variation der Stimulusinformation

Ausgangspunkt der vorliegenden Untersuchung war die Frage, wie sich unterschiedliche Stimuli, die identische Personinformation darbieten, auf die Beurteilung der Personen auswirken. Hierfür wurde als Stimulusmaterial zum einen ein Textausschnitt, aus einem Bühnenstück, der ein Gespräch zwischen fünf Geschworenen wiedergibt, zum zweiten der identische Textausschnitt, zu dem jeweils ein Bild des Geschworenen vorgelegt wurde und zum dritten ein Filmausschnitt, der dem Textinhalt identisch ist, ausgewählt. So hatten alle Probanden hinsichtlich des Gespräches zwischen den Geschworenen die gleichen Informationen. Eine Gruppe hatte zusätzlich ein Foto von jedem der Geschworenen, die dritte Gruppe konnte die Informationen nicht nur lesen, sondern das Gespräch und die Personen beobachten.

Hinsichtlich der Reproduktion der dargebotenen Personinformation wurde deutlich, dass das Ziel einer in etwa gleichen Ausgangsbasis der Gedächtnisinhalte für die unterschiedlichen Stimuli Text und Film erreicht werden konnte. In der Stimulusvariante Text mit Bild ergab sich im Vergleich zur Textgruppe ein signifikanter Unterschied für die Erinnerung an Originalinformationen, des Weiteren wurden hier die meisten synonymen Informationen erinnert. Die Vorlage eines Fotos von einer Person, von der sonst nur schriftliche Informationen vorliegen, scheint sich also als zusätzliche Modalität (Engelkamp, 1991) bzw. wie postuliert als eine Art Bildanker (Baltes et al., 2002) auszuwirken und führt zu besseren Erinnerungsleistungen für die dargebotenen schriftlichen Informationen. Betrachtet man diesen Befund unter praktischen Gesichtspunkten, so ergibt sich, dass, wenn eine Erinnerungsleistung über eine Person gefordert ist, die Verwendung eines Fotos einer Person durchaus bessere Ergebnisse hinsichtlich der Erinnerungsgüte bringen kann. Nützlich wäre dies beispielsweise bei der Auswertung von Gruppendiskussionen in Assessment Centern, wenn im Auswertungsgespräch der Assessment Center Beobachter ein Foto der jeweiligen Bewerber vorliegen würde und dies die Erinnerungsleistung der Beobachter für das gezeigte Verhalten des Bewerbers im AC verbessern würde und die anschließende Beurteilung somit auf eine solidere Grundlage gestellt werden würde.

Unter der Instruktion, die Gedächtnisinhalte wiederzugeben, wurden von den Probanden der Filmgruppe die meisten Zusammenfassungen und Verallgemeinerungen des Inhalts vorgenommen und die meisten Beurteilungen für die handelnden Personen formuliert. Die Beobachtung eines Filmausschnittes in Echtzeit, der aufgrund der Realitätsnähe zu Situationen im täglichen Leben ähnliche Verarbeitungsprozesse wie bei Beobachtungen im Alltag evozieren sollte, scheint auch unter Experimentalbedingungen ein so hoch automatisierter Prozess zu sein, dass, auch wenn eine andere Instruktion gegeben wurde (hier Wiedergabe von Gedächtnisinhalten), unwillkürlich Urteile über die handelnden Personen gebildet werden, die sich dann in einer Gedächtnisabfrage niederschlagen. Das gehäufte Aufzählen von Zusammenfassungen und Verallgemeinerungen des dargebotenen Inhalts und das Nennen von Einschätzungen und Bewertungen der Personen bereits in der Gedächtnisabfrage ist Ausdruck der großen Komplexitätsreduktion, die bei der Beobachtung komplexer sozialer Situationen vorgenommen werden muss, um die quasi unendliche Informationsmenge mit den begrenzten zur Verfügung stehenden kognitiven Ressourcen verarbeiten zu können.

Gestützt auf die Ergebnisse zur Gedächtnisreproduktion, dass die Textgruppe und die Filmgruppe ähnlich viele Items reproduzierten, wäre als eine mögliche praktische Anwendung zu nennen, dass ein Film ebenso wie ein Lehrbuchtext geeignet ist, Lerninhalte zu vermitteln. Hierbei kann es erforderlich sein, den Film mehrfach darzubieten (siehe Voruntersuchungen). Bei Lehrfilmen für den Schulbereich wird dies bereits häufig umgesetzt: wichtige Schlüsselszenen und Erklärungen oder Inhalte, die von den Schülern unbedingt eingepreßt werden sollten, werden wiederholt und oftmals erläutert der Moderator ein zweites bzw. drittes Mal, worauf es gerade hier ankommt.

Für das Nennen von Intrusionen ergaben sich signifikante Unterschiede zwischen den Gruppen der verschiedenen Stimulusvariationen. Dass die wenigsten falschen Erinnerungen von der Filmgruppe genannt wurden, könnte daran liegen, dass der als Stimulusmaterial gewählte Filmausschnitt realitätsnah und detailgetreu in Echtzeit dargeboten wurde, so dass es für die Beobachter zum einen nur schwer möglich und zum anderen nicht notwendig sein dürfte, konstruktive Prozesse bei der Verarbeitung auszulösen. Bei der Textverarbeitung spielt der Aufbau eines mentalen Modells, wie im theoretischen Teil der Arbeit in Abschnitt „4.1. Stimulusinformation Text“ ausführlich erläutert, eine sehr große Rolle und konstruktive Prozesse sind für den Aufbau eines Situationsmodells erforderlich. Dass hier, wenn für den Aufbau des mentalen Modells konstruktive Prozesse ablaufen, diese auch in einer Gedächtnisabfrage als wahrgenommene Inhalte wiedergegeben werden, scheint nicht ungewöhnlich (vgl. Walther, 1997). Die höchste Intrusionsrate in der Variation Text mit Bild dürfte der zusätzlichen Vorlage eines Fotos geschuldet sein. Bereits das kurze Betrachten des Fotos einer Person versetzt den Betrachter in die Lage, ein differenziertes Urteil über diese Person abzugeben (Henss, 1998, Jendraczyk, 1991). Die konstruktiven Prozesse, die durch das Betrachten des Bildes ausgelöst werden, sind so ausgeprägt, dass der Betrachter in der Lage ist, allein aufgrund des Fotos zu entscheiden, ob er die betrachtete Person sympathisch oder unsympathisch findet und ob er diese Person gern als Kollege/Kollegin haben würde oder nicht (Henss, 1998), er kann sich also z.B. vorstellen, wie es wäre, wenn er mit der völlig fremden Fotoperson zusammen arbeiten sollte. Dass sich derartige konstruktive Prozesse durch eine erhöhte Intrusionsrate bei der Wiedergabe von Gedächtnisinhalten niederschlagen, ist durch das so genannte „Source Monitoring“ – das Verwechseln von Quellen erklärbar (Johnson, Hashtroudi & Lindsay, 1993). Ausgangspunkt des Source-Monitoring Ansatzes ist, dass unterschiedliche Informationsverarbeitungsmodi verschiedenartige Repräsentationen erzeugen, die durch ihre Hinweisreize dazu führen, dass die erinnerten Informationen der

jeweiligen Quelle richtig zugeordnet werden können. Die richtige Zuordnung der erinnerten Information ist jedoch nicht immer korrekt (Johnson et al., 1993, Walther, 1997). Die Versuchspersonen glauben also, dass zumindest ein kleiner Teil der Erinnerung, den sie in konstruktiven Prozessen aufgebaut haben, der Originalinformation entstammt und nennen daher vermehrt Items, die nicht dem Originalmaterial zugeordnet werden können, was sich in der höheren Intrusionsrate bei den Probanden der Gruppe Text mit Bild niederschlägt.

Für die verschiedenen Stimulusvariationen zeigten sich für die Beurteilung der Geschworenen erhebliche Unterschiede. Obwohl alle Probanden hinsichtlich des Gesprächsinhaltes der zu beurteilenden Personen identische Informationen erhielten, erfolgte die Einschätzung der Geschworenen in großem Maße unterschiedlich. Als ein wesentlicher Befund der vorliegenden Untersuchung kann herausgestellt werden, dass, wenn semantisch identische Informationen von mehreren Personen in unterschiedlichen Medien präsentiert werden, diese unterschiedliche Präsentation zu unterschiedlichen Beurteilungsergebnissen führt. Die Informationsquellen, die über die verbale Information von Gesprächsinhalten hinausgehen, werden von Beurteilern aktiv genutzt. Ein Urteil über eine Person wird qualitativ und quantitativ stark beeinflusst von dem Medium, über welches die Darbietung der Personeninformation erfolgt. Für die Urteilsbildung über Personen ist von Bedeutung, ob eine Information ausschließlich schriftlich vorgelegt wird, ob der Urteiler ein Foto von der zu beurteilenden Person erhält oder ob er die Person tatsächlich agieren sieht. Die Beurteilung einer Person ist demzufolge nicht nur abhängig von der über die Person vorliegenden semantischen Information, sondern gleichermaßen von der Art und Weise, wie diese Informationen zum Beurteiler gelangen.

Wird Beurteilern nur ein Text mit Informationen zu Personen vorgelegt, sind sie durchaus in der Lage die beteiligten Personen zu beurteilen. Für den Stimulus Text orientieren sich die Urteiler stark an den vorgelegten Stimulusinformationen und die Urteile weisen hinsichtlich der Einschätzung der zu beurteilenden Personen eine gewisse Homogenität auf. Bei einer freien Beurteilung aufgrund eines Textes neigen Urteiler wenig dazu, Kategorien oder Dimensionen einzuschätzen, zu denen der Text keine Aussagen trifft. Ein weiteres Indiz dafür ist die Tatsache, dass nur für den Stimulus Text – und hier nur Probanden aus der Expertengruppe – einzelne Geschworene nicht beurteilt wurden.

Dies ändert sich massiv, wenn zum gleichen Text ein Foto der handelnden Person vorgelegt wird. Zum einen wurden für alle Geschworenen Bewertungen abgegeben. Zum anderen steigt die Anzahl der Persönlichkeitsmerkmale für die zu beurteilenden Personen in einer freien Bewertung gegenüber der Anzahl generierter Eigenschaften beim Stimulus Text signifikant an. In einer freien Beurteilung werden Einschätzungen zu Persönlichkeitseigenschaften oder Kategorien gegeben, zu denen die vorgelegten Informationen keinerlei Angaben enthalten. Allein die Vorlage eines Fotos scheint die Urteiler bei einer Beurteilung „mutiger“ zu machen, in dem Sinne, dass sie den zu beurteilenden Personen Eigenschaften zuschreiben und sie Kategorien zuordnen, die anhand der vorliegenden Informationen in keiner Weise nachvollziehbar sind. Diese Tatsache spiegelt sich auch in dem Umstand wieder, dass die Homogenität des Urteils über die einzelnen Geschworenen deutlich abnimmt und dass hier auch sich widersprechende Eigenschaften genannt werden. Interessant ist auch der Befund, dass nur für diese Stimulusinformation Nennungen zu Sympathie / Antipathie für die zu beurteilenden Geschworenen gemacht wurden.

Nähert sich die gegebene Stimulusinformation noch mehr Realbedingungen an, in dem ein Filmausschnitt wiedergegeben wird, ändern sich die Urteilsergebnisse im Vergleich zu den beiden anderen Stimulusinformationen erheblich. Eine semantisch gleiche Information über dieselbe Person führt durch das Hinzufügen weiterer Modalitäten - die visuellen Informationen, die das Äußere und die Bewegungen der zu beurteilenden Personen widerspiegeln und die akustischen Informationen (Stimmlage, Lautstärke, Sprechgeschwindigkeit, Artikulation, Modulation, Atmung) - zu Urteilsunterschieden. Zum einen konnte nachgewiesen werden, dass bei Darbietung eines Filmausschnittes die wenigsten Informationen originalgetreu oder synonym wiedergegeben werden und dass hier die meisten zusammengefassten Items erinnert werden, was der erwarteten Komplexitätsreduktion aufgrund der großen Informationsmenge dieser Stimulusvariante entspricht. Für die im Filmausschnitt präsentierten Personen wurden die meisten Eigenschaften und Persönlichkeitsfaktoren genannt. In dieser Bedingung, die Realbedingungen recht nahe kommt, beziehen die Urteiler die Informationen, die zusätzlich zum semantischen Inhalt gegeben wurden, als Grundlage mit ein, was dazu führen kann, dass mehrere Urteiler die gleiche Person unterschiedlich beurteilen.

Die praktische Relevanz dieser Befunde und dessen Bedeutung soll nachfolgend anhand einiger Beispiele erläutert werden. Die Art und Weise der Darbietung einer Information über

Personen – auch wenn sie semantisch dieselben Inhalte wiedergibt – hat einen nachhaltigen Einfluss auf die Gedächtnisinhalte des Beurteilers über die Person und bewirkt eine unterschiedliche Beurteilung von Personen. Gerade in Bereichen, in denen ein Urteil über eine Person sehr präzise sein sollte und Abweichungen wenig tolerierbar sind, ist es wichtig, sich über die unterschiedliche Wirkung verschiedener Informationsarten bewusst zu sein. Das Lesen eines schriftlichen Protokolls eines Gesprächs oder beispielsweise einer Gerichtsverhandlung wird zu anderen Ergebnissen bei der Beurteilung der sprechenden Personen führen, als wenn zusätzliche Informationen, z.B. ein Foto vorliegen würde und nochmals anders wird die Beurteilung ausfallen, wenn man bei dem Gespräch als Beobachter (beispielsweise als Schöffe in einer Gerichtsverhandlung) anwesend war.

Die praktischen Bereiche, in denen es um die Einschätzung von Personen geht, sind vielfältig. Wie oft findet sich eine Beurteilung, beispielsweise im Gesundheitsbereich (z.B. Einschätzung der Pflegestufe) oder im beruflichen Kontext (z.B. berufliche Rehabilitation nach Unfall), die lediglich aufgrund „Aktenlage“ verfasst wurde. Die Realitätsnähe und die Qualität einer solchen Beurteilung darf gerade, wenn man die vorliegenden Ergebnisse betrachtet, stark angezweifelt werden. Allein ein Foto zusätzlich zu schriftlichen Informationen beeinflusst die Beurteilung bereits in starkem Maße. Eine direkte Beobachtung der zu begutachtenden Person wird immer zu anderen Eindrücken führen, als das Lesen von unzähligen Berichten, Protokollen oder Stellungnahmen über die Person.

Auch die Qualität der Ausbildung für die Bereiche, in denen Menschen tagtäglich darauf angewiesen sind, andere Personen zu beurteilen, könnte von den Ergebnissen profitieren, wenn ihnen in dem Maße Rechnung getragen wird, dass Ausbildung oder Trainings zu Beurteilungen mit Material durchgeführt werden, das die Bedingungen der realen Lebenswelt möglichst gut repräsentiert. Für Supervisionen im Moderatoren- oder Therapiebereich bietet sich beispielsweise an, Videos aufzunehmen und auszuwerten. Allein die Protokollierung würde viele Informationen nicht berücksichtigen.

Ebenso sollte aus Sicht des Autors die Ausbildung von Sicherheitskräften zur Beurteilung von Personen so realitätsnah wie möglich erfolgen, damit etwa das Abschätzen von Gefahren, die von Personen ausgehen können, praktisch geübt wird. Allein das Lesen einer Beschreibung einer Person in wenigen Sätzen, woraus dann die weiteren Schritte abgeleitet und (schriftlich) dargelegt werden sollen, erscheint wenig sinnvoll, gerade wenn man bedenkt, innerhalb welcher

kurzer Zeit ein Sicherheitsbeamter im Ernstfall eine Entscheidung über das weitere Handeln treffen muss. Durch Darstellung von Fehlentscheidungen in den öffentlichen Medien wird dann bewusst, wie sensibel gerade dieser Bereich des Einschätzens von Menschen ist und wie wichtig daher gerade im Bereich der Urteilsbildung eine gute Ausbildung ist.

Darüber hinaus spiegelt sich die Relevanz der Ergebnisse für den rein privaten Bereich wider. Moderne Partnerbörsen im Internet haben großen Zulauf, da die Hürden des Aufeinanderzugehens durch das vorerst schriftliche Kommunizieren per Mail relativ niedrig sind. Lernt man sich dann richtig kennen und sieht sich das erste Mal, kann der vermeintliche Traumpartner, der so schöne Mails schreiben konnte, plötzlich zur großen Enttäuschung werden, weil man ihn sich so überhaupt nicht vorgestellt hat. Eine persönliche Begegnung – und sei sie auch nur kurz – führt eben zu ganz anderen Eindrücken als ein langer Schriftwechsel.

Die große Eindruckswirkung eines zusätzlichen Bildes einer Person zu einem Text wird am Beispiel der Analyse einer Bewerbungsmappe besonders deutlich. Selbst wenn man sich vornimmt, dem Bild keine große Bedeutung beizumessen, wird der Bildeindruck sofort und unwillkürlich bestimmte Bewertungen hervorrufen. Schuler und Berger (1979) konnten zeigen, dass attraktivere Bewerber tatsächlich besser Einstellungschancen hatten. Eine Analyse der Bewerbungsunterlagen, die als messtechnische Qualitätskriterien Objektivität, Reliabilität und Validität aufweisen sollte, ist in der täglichen Arbeit der Personalabteilungen eher die Ausnahme (Kanning, 2002). Die von Kanning (2002) genannte Empfehlung für den Umgang mit Bewerbungsunterlagen, zumindest in den Branchen, in denen das äußere Erscheinungsbild nicht von essentieller Bedeutung ist (z.B. Pressereferent, Modelbranche), dass vor der inhaltlichen Begutachtung die Lichtbilder entfernt werden sollten, um einer Verzerrung der Urteilsbildung, die allein dem Foto geschuldet ist, vorzubeugen, kann durch die hier gefundenen Ergebnisse nur unterstützt werden.

Die Einschätzung der Geschworenen auf den Ratingskalen brachte keine signifikanten Unterschiede in Abhängigkeit des dargebotenen Stimulusmaterials. Vielmehr schien es, als ob jede Versuchsperson eine Art eigenen Maßstab bei der Beurteilung nutzte. Als mögliche Erklärung dieses Befundes werden die unterschiedlich ausgewählten Vergleichsstandards der Beurteiler (Mussweiler et al., 2004b) herangezogen. Ein und dasselbe Verhalten einer Person kann aufgrund des gewählten Vergleichsstandards von unterschiedlichen Beurteilern

verschieden eingeschätzt werden und so zu einer anderen Beurteilung der Zielperson führen (Mussweiler et al., 2004b).

Für die Einschätzung der subjektiven Sicherheit bei der Beurteilung der Geschworenen wurde bewusst eine Skala verwendet, die keine fest vorgeschriebenen Abstufungen enthielt, sondern an deren Enden nur die Extremausprägungen „überhaupt nicht sicher“ und „absolut sicher“ festgeschrieben waren. In der Ratingskala zur Einschätzung der Geschworenen sollte jeder Versuchsteilnehmer angeben, wie seiner Meinung nach die genannten Merkmale ausgeprägt sind. Unter anderem befanden sich Dimensionen und Merkmalskategorien dabei, die, objektiv betrachtet, anhand des vorgelegten Stimulusmaterials nicht für jeden der Geschworenen eingeschätzt werden konnten. Trotzdem finden sich in der Einschätzung der subjektiv empfundenen Urteilssicherheit Sicherheitsangaben mit Mittelwerten zwischen $M = 57\%$ und $M = 74\%$. Damit finden sich auch in dieser Untersuchung Belege, die die Befunde von Hoffrage (1993) stützen.

Die Beurteiler fühlten sich in der Gruppe, die den Stimulus Film dargeboten bekommen hatte, am sichersten in ihrem Urteil, obwohl sich dort in der freien Beurteilung der Geschworenen die größten inhaltlichen Unterschiede fanden. Es scheint so, dass die Sicherheit in der Beurteilung für die Stimulusvariante Film subjektiv am höchsten empfunden wird, weil die Beurteiler sich aufgrund der Realitätsnähe des Mediums und dem damit verbundenen bekannten (weil im Alltag ständig praktizierten) Beurteilungsprozess sicher sind, die Beurteilung „richtig“ gemacht zu haben. In Aus- und Weiterbildungsangeboten für professionelle Beurteiler empfiehlt sich eine Sensibilisierung für die Thematik des übersteigerten Sicherheitsgefühles bei Beurteilungen vorzunehmen.

11.2. Diskussion hinsichtlich der Variation der Beurteilererfahrung

Ausgangspunkt des zweiten Teils der empirischen Untersuchung dieser Arbeit war die Frage, ob und gegebenenfalls wie sich zum einen ein Training in der Personenbeurteilung und zum anderen eine Expertise in der Personenbeurteilung auswirken. Hierzu wurden nach der Darbietung von Personinformationen die Gedächtnisinhalte, die Beurteilungen und die Einschätzung der subjektiv empfundenen Urteilssicherheit innerhalb der einzelnen Gruppen ausgewertet.

Die Ergebnisse aus den durchgeführten Experimenten zeigen, dass sich sowohl ein Training als auch eine langjährige Erfahrung in der Beurteilung von Personen auswirken. Die Wirkung eines Trainings zur Personenbeurteilung findet im Gegensatz zu einer langjährigen Berufserfahrung in der Personenbeurteilung einen anderen Niederschlag. Ein durchgeführtes Training in der Personenbeurteilung mit Personen, die keine Ausbildung in der Beurteilung von Personen haben und die im beruflichen Kontext nicht häufig Beurteilungen vornehmen müssen, wirkt spezifisch. Das heißt, die Trainierten verändern ihre Beobachtungs- und Beurteilungsstrategien genau in den Bereichen, die im Training thematisiert und geübt wurden.

Die Wirkungen des Trainings sind an den Ergebnissen gut zu demonstrieren. Die Sensibilisierung der Trainingsteilnehmer für den Unterschied zwischen Beobachtung und Bewertung spiegelt sich einerseits in den guten Ergebnissen zur Gedächtnisreproduktion wider, wo die Trainierten die besten Leistungen für die Reproduktion von Originalinformationen und für synonyme Informationen erbrachten und gleichzeitig die wenigsten Informationen zusammenfassten oder bereits Bewertungen generierten. Selbst in der freien Beurteilung der Geschworenen findet sich, dass das Training die Versuchspersonen derart sensibilisierte, dass die Trainierten vorsichtiger bei der Zuschreibung von Eigenschaften vorgingen und die wenigsten Eigenschaften und Persönlichkeitsmerkmale generierten. Dass die Trainierten am wenigsten Eigenschaftskategorien benutzten, zu denen die dargebotenen Stimuli keine Informationen enthielten, spricht ebenfalls für die Wirkung des Trainings. Stattdessen wurden von den Trainierten auch unter der Aufgabenstellung „freie Beurteilung“ viele wahrgenommene Beobachtungen genannt, die keine Bewertung enthalten (z.B. Angaben zum äußeren Erscheinungsbild). Ebenso zeigen die Ergebnisse, dass in der Gruppe der Trainierten keinerlei Angaben zu Sympathie/Antipathie gemacht wurden, dass die Trainingsteilnehmer verinnerlicht haben, dass eine Einschätzung über das persönliche Empfinden für diese Person (Sympathie/Antipathie) mit einer objektiven Beurteilung nicht im Einklang steht.

Wenn als „bessere Beurteilung“ eine Veränderung der Personenbeurteilung in dem Sinne verstanden wird, dass der Beurteiler zuerst ein Hauptaugenmerk auf Beobachtungen legt und erst anschließend Bewertungen des Wahrgenommenen vornimmt (z.B. Kanning, 1999), so kann man von einer Qualitätssteigerung der Personenbeurteilung durch ein vorheriges intensives Training sprechen. Die Teilnehmer haben in den Bereichen, in denen das Training

Inhalt vermittelte und die praktisch anhand von Beispielen geübt wurden, messbare Veränderungen im Vergleich zu Novizen erreicht. Diese Ergebnisse sprechen für die Anwendung von Trainings im Bereich der Personenbeurteilung. Wie lange ein solcher Trainingseffekt anhält, ist in dieser Arbeit nicht untersucht worden. Dies könnte ein möglicher Ansatzpunkt für weitere Arbeiten sein.

Die Hypothese, dass die Trainierten ihre subjektiv empfundene Urteilssicherheit niedriger einschätzen würden als Novizen, konnte nicht durch signifikante Unterschiede bestätigt werden. Ein Bewusstsein beim Beurteiler dafür zu entwickeln, dass eine Personbeurteilung aufgrund eines kleinen dargebotenen Ausschnittes niemals umfassend und objektiv vorgenommen werden kann, scheint eine herausfordernde Aufgabe zu sein, die in einem einzigen Training nicht erfüllt werden kann.

Die Experten, die sich von den beiden anderen Kontrastgruppen zum einen dadurch unterscheiden, dass sie eine formalisierte Ausbildung absolviert haben, in denen auch Inhalte zum Thema Personenbeurteilung vermittelt wurden, und zum anderen eine langjährige Berufspraxis haben, in der die Beurteilung von Personen alltägliche Aufgabe ist, differieren im Vergleich zu Novizen und Trainierten auch hinsichtlich der Beurteilung von Personen. Zum einen lässt sich den Ergebnissen entnehmen, dass die Experten bereits bei ihrer Beobachtung sehr stark die Komplexität der Informationen reduzieren, was sich z.B. darin widerspiegelt, dass in der Gedächtnisabfrage kaum Originalinformationen und wenige synonyme Informationen reproduziert werden, dass sie aber viele Zusammenfassungen und Verallgemeinerungen nennen.

Bei der Bewertung beziehen die Experten, wie die Trainierten, jedoch nicht so stark, beobachtete Informationen, wie äußere Merkmale, mit in die Beurteilung der Geschworenen ein. Der Befund, dass die Experten bei der freien Beurteilung der Geschworenen die wenigsten Eigenschaften nennen, ist möglicherweise dadurch erklärbar, dass die Experten ihren Fokus besonders auf Merkmalsdimensionen lenken, die den Kategorien, die sie in ihrer täglichen beruflichen Praxis zur Beurteilung benötigen, sehr ähnlich sind und dass andere Kategorien nicht zusätzlich in die Bewertung einbezogen werden. Dieses Ergebnis würde somit auch die postulierte Annahme stützen, dass sich bei den hier untersuchten Experten eine auf den Kontext bezogene Expertise herausgebildet hat (Gruber, 1994), die dazu führt, dass sie Beurteilungen über Personen anhand der in ihrer Domäne vorherrschenden Kategorien

vornehmen und dass sie in anderen Kontexten wenig flexibel reagieren (vgl. Holyak, 1991). Die Befunde, dass die Experten weit weniger Einschätzungen, die sehr weit über die dargebotenen Informationen hinausgehen, als die Novizen vornahmen, sprechen dafür, dass sie sich der begrenzten Möglichkeiten einer umfassenden Einschätzung von Personen anhand weniger Informationen durchaus bewusst sind. Die Tatsache, dass ein Teil der Experten in der Stimulusvariation, die die wenigsten Informationen enthielt (nur Text), für einige der Geschworenen keine Bewertung vorgenommen hat, erhärten ebenfalls die Vermutung, dass die Expertengruppe ein recht gut ausgebildetes Bewusstsein für die Begrenztheit einer objektiven Beurteilung anhand geringer Informationen hat.

In die gleiche Richtung weisen auch die Ergebnisse hinsichtlich der subjektiv empfundenen Urteilssicherheit. Der Befund, dass die subjektive Urteilssicherheit der Experten signifikant unter der der Novizen und der der Trainierten lag, deutet ebenfalls darauf hin, dass die Experten über eine realistischere Einstellung darüber verfügen, wie gut bzw. wie schlecht es für sie möglich ist, einen Menschen objektiv zu beurteilen. Es ist jedoch auch möglich, dass die Unsicherheit der Experten bei der Beurteilung der Geschworenen auf die geringe Informationsmenge des dargebotenen Stimulusmaterials zurückzuführen ist. Im beruflichen Kontext haben sie die Möglichkeit, für eine Einschätzung weitere Informationen (entweder innerhalb der Unterrichtsstunde oder über einen längeren Zeitraum, etwa vor dem Verbalurteil auf dem Zeugnis oder Meinungen von anderen Kollegen zu einem Schüler) zu sammeln. Doch auch wenn dies der Fall gewesen sein sollte, dass die Experten ihre Urteilssicherheit aufgrund der geringen Informationsmenge des dargebotenen Stimulusmaterials so niedrig einschätzten, deutet es auf eine gute Fähigkeit der Experten hin, dass sie selbst realistisch einschätzen können, dass diese Informationsmenge nicht ausreicht, um Menschen umfassend einzuschätzen. Gerade wenn man in Betracht zieht, dass die Experten durch Honorierung ihres Expertenstatus für die Untersuchung gewonnen wurden, wäre es durchaus denkbar gewesen, dass sich dies auch in einer übersteigerten subjektiven Sicherheit niederschlägt. Insofern spiegelt sich in dem doch unerwarteten Ergebnis, dass die subjektive Urteilssicherheit der Experten signifikant niedriger im Vergleich zu den Trainierten, die gerade für diese Thematik in einem Training sensibilisiert wurden, ist, eine Seite der ausgeprägten Expertise in der Personenbeurteilung wider. Das Training allein genügte nicht, um das Sicherheitsgefühl der Trainierten bei der Beurteilung so zu beeinflussen, dass sie das Bewusstsein für die Schwierigkeit einer objektiven Beurteilung erlangt hätten.

Im Zuge einer kritischen Interpretation der Ergebnisse sollen nun noch Faktoren, die einen nicht kontrollierbaren Einfluss auf die Untersuchungsergebnisse haben könnten, erörtert werden. Die Experten unterscheiden hinsichtlich des Alters und des beruflichen Status von den beiden anderen Kontrastgruppen. Wenn man in einer Art Querschnittsstudie auf Versuchspersonen zurückgreift, die ihr Personenbeurteilungsvermögen über viele Jahre hinweg trainiert und entwickelt haben und andererseits auf Probanden, die in diesem Bereich keine oder wenige professionelle Erfahrung haben, dann ist der unkontrollierte Einfluss weiterer Größen, wie soziale Kompetenz, Intelligenz und anderer Persönlichkeitsfaktoren praktisch kaum auszuschließen. Die Expertengruppe hat zudem ihre Expertise „unkontrolliert“ – also außerhalb des Experimentes erworben. Auch innerhalb der Expertengruppe gibt es relevante Altersunterschiede und eine unterschiedliche Anzahl von absolvierten Dienstjahren im Schuldienst. Lern- und erfahrungsabhängige oder dispositionale Verursachungsfaktoren (Begabung, Talent) für den Erwerb der Expertise können nicht gemessen werden. Beim Erwerb der Expertise hat jede einzelne Versuchsperson ihre eigene Entwicklung durchlaufen und eigene Erfahrungen gemacht.

Als Experten wurden für die durchgeführte Untersuchung Pädagogen im Schuldienst mit mehreren Jahren Berufserfahrung ausgewählt. Es ist denkbar, dass andere Menschen, die langjährige Erfahrungen in der professionellen Personenbeurteilung gesammelt haben, wie beispielsweise Richter, in der experimentellen Situation zu anderen Einschätzungen kommen. Die im beruflichen Kontext erworbenen Fähigkeiten in der Personenbeurteilung entstammen bei Menschen in anderen Berufen völlig anderen Arbeitsinhalten, so dass ihre Entwicklung der Expertise auf diesem Gebiet notwendigerweise anders verlaufen ist. Als Ergebnis wären bei anderen Probanden andere Befunde denkbar. Trotz dieser Faktoren, die die Ergebnisse der Untersuchung unkontrolliert beeinflussen können, überwiegen die Vorteile beim Vergleich zwischen Experten und Novizen (Gruber, 1994), da die Ergebnisse einen starken Bezug zum täglichen Leben haben und über die Experimentalsituation hinaus Anwendung finden können. Möglichen Einfluss auf die Untersuchungsergebnisse könnte auch der Umstand haben, dass die Experten am Experiment nach einem für sie meist anstrengenden Arbeitstag teilnahmen. Es ist nicht auszuschließen, dass die Leistungsfähigkeit beispielsweise für die Reproduktion der Gedächtnisinhalte höher gewesen wäre, wenn die Untersuchung in den Morgenstunden stattgefunden hätte. Dies war jedoch praktisch nicht realisierbar.

12. Zusammenfassung und Ausblick

In der vorliegenden Arbeit konnte gezeigt werden, dass semantisch identische Personinformationen, die durch verschiedene Medien dargeboten werden, zu unterschiedlichen Ergebnissen sowohl hinsichtlich der Gedächtnisinhalte über die Personen als auch hinsichtlich der Beurteilung dieser Personen führen. Gerade vor dem Hintergrund, dass in der Vergangenheit eine große Reihe von Untersuchungen, in denen es um Personenbeschreibungen geht, mit Versuchsmaterial durchgeführt wurde, welches sich auf die Vorgabe von Eigenschaftsbegriffen und/oder kurze Verhaltensbeschreibungen in Form einer Liste beschränkte (Krolak-Schwerdt, 2000), erscheint die Kritik an einer mangelnden ökologischen Validität der Ergebnisse, deren Umsetzung in der Praxis schwierig ist (Kanning, 1999; Höft, 2005), gerechtfertigt.

Die aus den Untersuchungen zur Variation des Stimulusmaterials gewonnenen Ergebnisse erlauben die Schlussfolgerung, dass es für die Gewinnung von validen Forschungsergebnissen auf dem Gebiet der Personbeurteilung unerlässlich ist, eine gezielte Auswahl des Stimulusmaterials vorzunehmen.

Für eine verbesserte ökologische Validität von Untersuchungsergebnissen gilt es, auch kleine Variationen im Stimulusmaterial, die bislang wenig Beachtung fanden, beispielsweise, ob das dargebotene Stimulusmaterial ein Foto enthält oder nicht, zu berücksichtigen. Wenn Aussagen über Zusammenhänge für Beobachtungen oder Bewertungen über den experimentellen Rahmen hinaus getroffen werden sollen, sollten die Untersuchungen so wirklichkeitsnah wie möglich gestaltet werden. Dies betrifft nicht nur den semantischen Inhalt der Stimulusinformation, sondern auch die Auswahl des Mediums und dessen Eigenschaften. Bei der Ausgestaltung von Assessment Centern wird bei einer professionellen Personalauswahl diesem Umstand bereits Rechnung getragen: die von den Bewerbern zu erledigenden Aufgaben orientieren sich stark an den Anforderungen des zukünftigen Arbeitsplatzes. Jedoch ist das Assessment Center nur ein sehr kleiner Ausschnitt, in dem Personenbeurteilung im professionellen Kontext häufig angewandt wird. Das Problem der Personenbeurteilung ist sowohl im täglichen Leben und im beruflichen Kontext allgegenwärtig und könnte hier wohl auch in vielen Bereichen eine stärkere Beachtung finden, sei es in der Ausbildung oder in der täglichen Berufspraxis. In diesem Zusammenhang erscheinen weitere Untersuchungen, z.B. zur Wirkung von Personenbeurteilung bei einer teilnehmenden Beobachtung interessant. Hier sind sicher nochmals andere Wirkungen bei der

Wiedergabe von Gedächtnisinhalten und bei der Beurteilung von Personen zu erwarten, wenn man zum Beispiel an Zeugen von Unfällen oder Opfer von Straftaten denkt, die in dieser Situation ihren Fokus auf andere Aspekte lenken werden (z.B. „Was ist jetzt nach dem Unfall zu tun? Muss ich Rettungskräfte verständigen?“ etc.). Eine wesentliche Rolle in diesem Feld der Personenbeurteilung wird sicherlich den Emotionen der involvierten Personen zuzuschreiben sein, deren Untersuchung lohnenswert erscheint.

Im zweiten experimentellen Teil ging es um die Untersuchung der Frage, wie sich ein Training bzw. eine langjährige Expertise auf die Beurteilung von Personen auswirken. Anhand der Ergebnisse lassen sich sowohl die Wirkung eines Trainings als auch die Wirkung von Expertise in diesem Bereich verdeutlichen. Die nach Höft & Funke (2001) empfohlene Trainingspraxis für Assessment Center-Beobachter, ein Beurteilerfehlertraining mit einem Verhaltensbeobachtungstraining zu kombinieren, wurde hier in dem mit den Versuchspersonen durchgeführten Training zur Personenbeurteilung angewandt.

Die Wirkungen des Trainings zeigen sich spezifisch in den Bereichen, in denen die Versuchspersonen trainiert wurden. So konnte gezeigt werden, dass die Trainierten durch die erfolgte Sensibilisierung für die Thematik die Trennung zwischen Beobachtung und Bewertung in der Gedächtnisabfrage am besten umsetzten. Die Beobachtungsakkuratheit kann demzufolge auch bei Probanden, die keine spezielle Ausbildung oder längere Vorerfahrung in der Personenbeurteilung haben, durch ein Training messbar verbessert werden. Wie lange diese Wirkungen eines einmaligen Trainings anhalten und zu einer längerfristigen Verbesserung des Beobachtungsverhaltens führen, wurde in der vorliegenden Arbeit nicht untersucht. Für einen Langzeiterfolg werden sicher mehrere Trainings in zeitlichen Abständen notwendig sein, da auch für langjährige erfahrene Assessment Center-Beobachter wiederholte Schulungen empfohlen werden (Kanning, 2002; Schuler, 2001), um diese erneut für mögliche Beurteilungsfehler zu sensibilisieren.

In wiederholten Trainings könnten die Trainingsteilnehmer dann auch weiter mit der Problematik der subjektiven Urteilssicherheit konfrontiert werden, da die experimentellen Ergebnisse dieser Arbeit belegen, dass ein einmaliges Training nicht genügt, um das Bewusstsein der Beurteiler dafür zu schärfen, wie schwierig es ist, eine objektive Beurteilung anhand weniger Informationen über eine Person vorzunehmen.

Über den Rahmen dieser Arbeit hinaus wäre es wünschenswert, die Trainingseffekte in der Personenbeurteilung weiter zu untersuchen. Von Bedeutung ist beispielsweise die Frage, ob und wie ein Training, welches mit Experten durchgeführt wird, wirkt und welche Wechselwirkungen hierbei zustande kommen. Für die professionelle Personenbeurteilung im beruflichen Kontext wäre es lohnenswert, die Wirkungen von Trainings näher zu untersuchen, damit vor allem Berufsgruppen, für die die Beurteilung von Personen alltägliche Praxis ist, davon profitieren könnten. Wenn ein Training bzw. mehrere Trainingssequenzen Veränderungen im Sinne einer Objektivierung der Beurteilung von Personen bewirken, sollte beispielsweise die Aufnahme solcher Trainings in Curricula von Berufsausbildungen bzw. Berufs begleitende Weiterbildungen forciert werden.

Literaturangaben

- Albrecht, J.E. & O'Brien, E.J. (1993). Updating a mental model: Maintaining both local and global coherence. In *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, S. 1061-1070.
- Alley, T.R. (1988). *Social and applied aspects of perceiving faces*. Hillsdale, NJ: Erlbaum.
- Allport, G.W. (1949). *Persönlichkeit. Struktur, Entwicklung und Erfassung der menschlichen Eigenart*, Stuttgart: Klett [Original: 1937].
- Allport, G.W. & Odbert, H.S. (1936). Trait-names. A psychol-lexical study. *Psychological Monographs*, 47, 1, Whole number 211.
- Anderson, J.R. & Bower, G.H. (1973). *Human associative memory*. Washington, D.C.: Winston & Sons.
- Archer, D., Iritani, B., Kimes, D.D. & Barrios, M. (1983). Face-ism: Five studies of sex differences in facial prominence. *Journal of Personality and Social Psychology*, 45, S. 725-735.
- Arkes, H.R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, 110, S. 486-498.
- Aronson, D.E., & Akamatsu, T.J. (1981). Validation of a Q-sort task to assess MMPI Skills. *Journal of Clinical Psychology*, 37, S. 831-836.
- Aronson, E., Wilson, T.D. & Akert, R.M. (2002). *Social Psychology* (4th edition). London: Prentice-Hall.
- Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.) *The psychology of learning and motivation: Advances in research and theory 2*. New York: Academic Press.
- Backhaus, K. (2004) *Persönlichkeit als Forschungsgegenstand der Psychologie*. Psyreon - Psychological Research Online 1.
- Baltes, P. B., Glück, J., & Kunzmann, U. (2002). Wisdom: Its structure and function in regulating successful lifespan development. In C. R. Snyder & S. J. Lopez (Eds.), *The handbook of positive psychology*, S. 327-347. New York: Oxford University Press
- Bargh, J.A. (1984). Automatic and conscious processing of social information. In R.S. Wyer Jr. & T.K. Srull (Hrsg.). *Handbook of Social Cognition*, Bd. 3, S. 3-43, Hillsdale: NJ Erlbaum.
- Barnett, S.M. & Koslowski, B. (2002). Adaptive expertise: Effects of experience and the level of theoretical understanding it generates. In *Thinking and Reasoning*, 2002, 8 (4), S. 237 – 267.

Barrows, H.A., Norman, G.R., Neufeld, V.R. & Feightner, J. (1982). The Clinical reasoning of randomly selected physicians in general medical practice. *Clinical Investigative Medicine*, 5, S. 49-55.

Baumgarten, F. (1933). *Die Charaktereigenschaften. Beiträge zur Charakter- und Persönlichkeitsforschung* (Nr. 1) Bern: A. Francke.

Berliner, D.C. (1987a). Der Experte im Lehrerberuf: Forschungsstrategien und Ergebnisse. *Unterrichtswissenschaft*, 15, S. 295-305.

Berliner, D.C. (1987b). Ways of thinking about students and classroom by more and less experienced teachers. In J. Calderhead (Ed.). *Exploring teacher thinking*, S. 60-83.

Berliner, D.C. (1990). Characteristics of experts in the pedagogical domain. Paper presented at the International Symposium "Research on Effective and Responsible Teaching", Fribourg.

Berliner, D.C. & Carter, K.J. (1986). Differences in processing classroom information by expert and novice teachers. Paper presented at the Conference of the International Study Association on Teachers Thinking (ISATT), Leuven.

Berliner, D.C., Stein, P., Sabers, D., Clarridge, P., Cushing, K. & Pinnegar, S. (1988). Implications of research on pedagogical expertise and experience for mathematics teaching. In D. Grouws & T. Cooney (Eds.), *Perspectives on research on effective mathematics teaching*, S. 67-95.

Beyer, R. (1987). Psychologische Untersuchung zur Textverarbeitung unter besonderer Berücksichtigung des Modells von Kintsch & van Dijk (1978). *Zeitschrift für Psychologie mit Zeitschrift für angewandte Psychologie*, Suppl. 8.

Billig, M. (1985). Prejudice, categorization and particularization. From a perceptual to a rhetorical approach. *European Journal of Social Psychology*, 15, S. 79-103.

Bierhoff, H.-W. (2000). *Sozialpsychologie. Ein Lehrbuch*. Stuttgart: Kohlhammer.

Borgatta, E.F. (1964). The structure of personality characteristics. *Behavioral Science*, 9, S. 8-17.

Borkenau, P. & Liebler, A. (1992). The cross-modal consistency of personality: Inferring strangers traits form visual or acoustic information. *Journal of Research in Personality*, 26, 183-204.

Borkenau, P. & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology*, 65, S. 546-553.

Borkenau, P. & Liebler, A. (1994). The factor structure of trait ratings depends on the extent of information available to the judges. *European Review of Applied Psychology*, 44, S. 3-7.

Bower, G.H. (1998). An associative theory of implicit and explicit memory. In Conway, M.A., Gathercole, S.E. & Cornoldi, C.,(Eds.) *Theories of memory*, Vol. 2, S. 25-60, East Sussex: Psychology Press Ltd.

Brewer, M.B. (1988). A dual process model of impression formation. In T.K. Srull & R.S. Wyer (Hrsg.) *Advances in social cognition*. S. 1-36. Hillsdale, NJ: Erlbaum.

Brewer, M.B. & Feinstein, A.S. (1999). Dual processes in the cognitive representation of persons and social categories. In S. Chaiken & Y. Trope (Hrsg.) *Dual process theories in social psychology*. S. 255-270. New York: Guilford.

Brewer, W.F. & Dupree, D.A. (1983). Use of plan schemata in the recall and recognition of goal-directed actions. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 9, S. 117-129.

Bromme, R. (1992). *Der Lehrer als Experte: zur Psychologie des professionellen Wissens*, Bern.

Brown, A.L. & Scott, M.S. (1971). Recognition memory for pictures in preschool children. *Journal of Experimental Child Psychology*, 11, S. 401-412.

Bruner, J.S. (1951). Personality dynamics and the process of perceiving. In R.R. Blake & G.V. Ramsey (Eds.), *Perception, an approach to personality*, S. 121-147.

Bruner, J.S. & Postman, L. (1951). An approach to social perception. In W. Dennis & R. Lippitt (Eds.), *Current trends in social psychology*, S. 71-118.

Calderhead, J. (1981). A psychological approach to research on teachers' classroom decision making. *British Educational Research Journal*, 7, 51-57.

Calderhead, J. (1983). Research into teachers' and student teachers' cognitions: Exploring the nature of classroom practice. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Carter, K., Sabers, D., Cushing, K., Pinnegar, P. & Berliner, D. (1987). Processing and using information about students: A study of expert, novice and postulant teachers. *Teaching and Teacher Education*, 3, S. 147-157.

Chaiken, S. & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology*, 2, 641-653.

Charness, N. (1981). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance*, 7, S. 467-476

Chase, W.G. & Ericsson, K.A. (1981). Skilled memory. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (S. 141-189). Hillsdale, NJ: Erlbaum.

Chase, W.G. & Ericsson, K.A. (1982). Skill and working memory, In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, S. 1-58). New York: Academic Press.

Chase, W.G. & Simon, H.A. (1973a). Perception in chess. *Cognitive Psychology*, 5, S. 55-81.

Chase, W.G. & Simon, H.a. (1973b). The mind's eye in chess. In Chase, W.G. (ed.), Visual information processing, S. 215-281, New. York: Academic Press.

Chi, M.T.H. (1978). Knowledge structures and memory development. In Siegler, R.S. (Hrsg.), Children's thinking: What develops? (S. 73-96) Hillsdale, NJ: Erlbaum.

Chi, M.T.H. , Feltovich, P.J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, S. 121-152.

Chi, M.T. H., Glaser, R. & Farr, M. J. (Eds.). (1988). The nature of expertise. Hillsdale, NJ: Erlbaum

Chi, M.T.H., Glaser, R. & Rees, E. (1982). Expertise in Problem Solving. In Sternberg, R.J. (ed.), *Advances in the psychology of human intelligence*, Vol. I, S. 7-75. Hillsdale, NJ: Erlbaum.

Chiesi, H.L., Spilich, G.J. & Voss, J.F. (1979). Acquisition of domainrelated information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 18, S. 257-273.

Clarke, Vanessa, J. & Lamberts, Koen (1997). Strategy Shifts and Expertise in Solving Transformation Rule Problems. In *Thinking and Reasoning*, 3 (4), S. 271-290.

Costa, P.T. & McCrae, R.R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, 38, S. 668-678.

Costa, P.T. & McCrae, R.R. (1994). Stability and change in personality from adolescence through adulthood. In C.F. Halverson, G.A. Kohnstamm & R.P. Martin (Eds.) *The developing structure of temperament and personality from infancy to adulthood*, S. 139-150.

Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, S. 671-684.

Davis, M. (1979). *Understanding body movement. An annotated bibliography*. New York: Arno Press.

Dawson, V.L., Zeitz, C.M & Wright, J.C. (1989). Expert-novice differences in person perception: Evidence of experts' sensitivities to the organization of behaviour. *Social cognition*, 7, S. 1-30.

De Groot, A.D. (1965). *Thought and choice in chess*. The Hague: Mouton.

De Saint-Exupéry, A. (1943). *Le Petit Prince*. Titel der Französischen Originalausgabe.

De Saint-Exupéry, A. (1943). *Der Kleine Prinz*. Ins Deutsche Übertragen von Grete und Josef Leitgeb. Düsseldorf, Karl Rauch Verlag, verlegt 1993.

- De Vega, M. (1995). Backward updating of mental models during continuous reading of narratives. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, S. 373-385.
- Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of personality and social psychology*, 56, S. 5-18.
- Digman, J.M. & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology*, 50, 116-123.
- Dirscherl, B., Goß, S. & Hollweck, C. (2002). Face-ism. Der Einfluss der Größe des Gesichtsanteils bei photographischen Personendarstellungen bei der Eindrucksbildung. Unveröffentlichter Bericht an der Universität Regensburg.
- Dörner, D. (1997). Über die Gefahren und die Überflüssigkeit der Annahme eines „propositionalen“ Gedächtnisses. In Lürer, G. & Lass, U. (Hrsg.). *Erinnern und Behalten: Wege zur Erforschung des menschlichen Gedächtnisses*. S. 172-198. Göttingen: Vandenhoeck und Ruprecht.
- Dörner, D., Schaub, H., Stäudel, T. & Strohschneider, S. (1988). Ein System zur Handlungsregulation oder die Interaktion von Emotion, Kognition und Motivation. *Sprache & Kognition*, 7, S 217 - 232.
- Duncan, J. (1987). Attention and reading wholes and parts in shape recognition – a tutorial review. In Coltheart (Ed.). *Attention and performance Xii: The psychology of reading*, S. 39 – 61, NJ: Hillsdale, Erlbaum.
- Dunning, D. & Hayes, A.F. (1996). Evidence for Egocentric Comparison in Social Judgment. In *Journal of Personality and Social Psychology*, Vol. 71, Nr. 2, S. 213-229.
- Dutke, S. (1998). Zur Konstruktion von Sachverstandsrepräsentation. In *Zeitschrift für experimentelle Psychologie*, S. 42-59.
- Edeler, B. (1992). Strukturierung von Gedächtnisrepräsentationen über Personen bei der sozialen Urteilsbildung. Unveröffentlichte Habilitation, Friedrich-Schiller-Universität, Jena.
- Elo, A.E. (1978). *The rating of chessplayers, past and present*. New York: Arco.
- Elstein, A.S., Shulman, L.S. & Sprafka, S.A. (1978). *Medical Problem solving: An analysis of clinical reasoning*. Cambridge, Mass.: Harvard University Press.
- Engelkamp, J. (1973). *Semantische Struktur und die Verarbeitung von Sätzen*. Bern: Huber.
- Engelkamp, J. (1991). *Das menschliche Gedächtnis*. Göttingen: Hogrefe.
- Engelkamp, J. (1998). Gedächtnis für Bilder. In K.Sachs-Hombach & K.Rehkämper (Hrsg.). *Bild – Bildwahrnehmung - Bildverarbeitung*. Wiesbaden, Deutscher Universitätsverlag, S. 227-242.
- Engelkamp, J. & Pechmann, T. (Hrsg.) (1993). *Mentale Repräsentation*. Bern: Huber.

- Englich, B. & Mussweiler, T. (2001). Sentencing under Uncertainty: Anchoring Effects in the Courtroom. In *Journal of Applied Social Psychology*, 2001, 31, 7 S. 1535-1551.
- Erdfelder, E. (2003). Das Gedächtnis des Augenzeugen. In *Report Psychologie* (28) 7/8, S. 434- 443.
- Ericsson, K.A. (1985). Memory skill. *Canadian Journal of Psychology*, 39, S. 188-231.
- Ericsson, K.A. & Crutcher, R.J. (1990). The nature of exceptional performance. In Baltes, P.B. Featherman, D.L. & Ferner, R.M. (eds.), *Life-span development and behaviour* (Vol. 10, S. 187-217). Hillsdale, NJ: Erlbaum.
- Ericsson, K.A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, S. 211-245.
- Fiedler, K., Koch, S., Haag, S. Immesberger, P. & Ullrich, M. (1996). Stereotypes without biases: A cognitive-environmental approach.
- Fiedler, K. & Plessner, H. (2004). Induktives Schließen: Umgang mit Wahrscheinlichkeiten. Verfügbar über: http://www.psychologie.uni-heidelberg.de/ae/allg/enzykl_denken/Enz_05_Induktion.pdf [Datum des Zugriffs: 01.06.2005]
- Fiedler, K. (1997). Die Verarbeitung sozialer Informationen für Urteilsbildung und Entscheidungen. In: Strobe, W, Hewstone, M & Stephenson, G.M (Hrsg.) *Sozialpsychologie*.
- Fiske, S.(1993). Social cognition and social perception. *Annual Review of Psychology*, 44, S. 155-194.
- Fiske, S.T., Lin, M. & Neuberg, S.L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Hrsg.) *Dual process theories in social psychology*. S. 231-254. New York: Guilford.
- Fiske, S.T. & Neuberg, S.L. (1990). A continuum model of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M.P. Zanna: *Advances in Experimental Social Psychology*. S. 1-74. New York: Academic Press.
- Fiske, S.T. & Taylor, S.E. (1991). *Social cognition*. New York: McGraw-Hill.
- Fletcher, C.R. (1994). Levels of representation in memory for discourse. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*, S. 589-607.
- Foertsch, J. & Gernsbacher, M.A. (1994). In search of complete comprehension: Getting „minimalists“ to work. *Discourse Processes*, 18, S. 271-296.
- Frensch, P. & Sternberg, R.J. (1989). Expertise and intelligent thinking: When is it worse to know better? In Sternberg, R.S. (ed.). *Advances in the psychology of human intelligence*, 5 (S. 157-188). Hillsdale, NJ: Erlbaum.
- Frey, P. W. & Adesman, P. (1976). Recall memory for visually presented chess positions. *Memory and Cognition*, 4, 541-547.

Frey, S. (1999a). Neue Wege in der Kommunikationsforschung. In: D.Ganten, E.Meyer-Galow, H. Ropers, H. Scheich, H.Schwarz, K. Urban und E. Truscheit (Hrsg.) Gene, Neurone, Qubits & Co. Unsere Welten der Information. Verhandlungen der Gesellschaft Deutscher Naturforscher und Ärzte. 120. Versammlung, 19.-22. September 1998 in Berlin. Stuttgart: Hirzel.

Frey, S. (1999b). Die Macht des Bildes. Der Einfluß der nonverbalen Kommunikation auf Kultur und Politik. Bern, Göttingen: Verlag Hans Huber.

Frey, S. & Bente, G. (1989). Mikroanalyse medienvermittelter Informationsprozesse. Zur Anwendung zeitreihen-basierter Notationsprinzipien auf die Untersuchung von Fernsehnachrichten. In: M. Kaase und W. Schulz (Hrsg.) Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 30/1989: Massenkommunikation. Theorien, Methoden, Befunde. Opladen: Westdeutscher Verlag, S. 508-526.

Frey, S., Hirsbrunner, H.P., Pool, J. & Daw, W. (1981). Das Berner System zur Untersuchung nonverbaler Interaktion: I. Die Erhebung des Rohdatenprotokolls. In P.Winkler (Hrsg.): Methoden der Analyse von Face-to-Face-Situationen. Stuttgart: Metzler, S. 203-236.

Frey, S. Hirsbrunner, H.P., Florin, A., Daw, W. & Crawford, R. (1983). A unified approach to the investigation of nonverbal and verbal behaviour in communication research. In W. Doise and S. Moscovici (eds.). Current issues in European social psychology. Cambridge: Cambridge University Press, S. 99-143.

Funder, D.C. & Colvin, C.R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, 55, S. 149-158.

Galton, F. (1884). Measurement of character. *Fortnightly Review*, 36, S. 179-185.

Garb, H.N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review*, 4, S. 641-655.

Garb, H.N. (1986). The appropriateness of confidence ratings in clinical judgment. *Journal of Clinical Psychology*, 42, S. 190-197.

Garb, H.N. (1989). Clinical Judgment, Clinical Training, and Professional Experience. In *Psychological Bulletin* 105, No. 3, S. 387-396.

Genero, N. & Cantor, N. (1987). Exemplar prototypes and clinical diagnosis: Toward a cognitive economy. *Journal of Social and Clinical Psychology*, 5, S. 59-78.

Gernsbacher, A.M., Goldsmith, H.H. & Robertson, R.R.W.(1992). Do readers mentally represent characters' emotional states? In *Cognition and Emotion*, 6, S. 89-111.

Gilbert, D.T. (1989). Thinking lightly about others: Automatic components of the social inference process. In: J. S. Uleman & J.A. Bargh (Eds.) *Unintended thought*. New York: the Guilford Press.

- Glenberg, A.M., Meyer, M. & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. In *Journal of Memory and Language*, 26, S. 69-83.
- Goldberg, L.R. (1981). Language and individual differences: The search for universals in the personality lexicons. In L. Wheeler (Ed.) *Review of Personality and Social Psychology*, Vol. 2, S. 141-166.
- Goldberg, L.R. (1982). From ace to zombie: Some explorations in the language of personality. In C.D. Spielberger & J.N. Butcher (Eds.). *Advances in Personality Assessment*. Vol. 1, S. 203-234.
- Goldberg, L.R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59, S. 1216-1229.
- Goldberg, L.R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4, S. 26-42.
- Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48, S. 26-34.
- Goldin, S.E. (1978). Effects of orienting tasks on recognition of chess positions. *American Journal of Psychology*, 91, S. 659-671.
- Gordon, R.A. & Anderson, R.S. (1995). Perceptions of race-stereotypic and race-nonstereotypic crimes: The impact of response-time instructions on attributions and judgments. *Basic and Applied Social Psychology*, 16, 455-470.
- Grabowski, J. (1991). *Der propositionale Ansatz der Textverständlichkeit: Kohärenz, Interessanztheit und Behalten*. Münster: Aschendorffsche Verlagsbuchhandlung.
- Grebstein, L. (1963). Relative accuracy of actuarial prediction, experienced clinicians, and graduate students in a clinical judgment task. *Journal of Consulting Psychology*, 37, S. 127-132.
- Greenberg, J. & Baron, R.A. (2000). *Behavior in organizations* (7th edition). London: Prentice-Hall.
- Griswold, P.M. & Dana, R.H. (1970). Feedback and experience effects on psychological reports and predictions of behaviour. *Journal of Clinical Psychology*, 26, S. 439-442.
- Gruber, H. & Ziegler, A. (1990). Comparing different kinds of expertise: Chess problem solving in problemists and tournament players. *The German Journal of Psychology*, 14, 115-116.
- Gruber, H. & Mandl, H. (1992). *Begabung und Expertise*. In Hany, E.A. & Nickel, H. (Hrsg.), *Begabung und Hochbegabung* (S. 59-73). Bern, Göttingen: Huber.
- Gruber, Hans (1994). *Expertise. Modelle und empirische Untersuchungen*. Opladen: Westdeutscher Verlag.

- Hacker, W. (1992). Expertenkönnen – Erkennen und Vermitteln. Göttingen, Stuttgart: Verlag für Angewandte Psychologie.
- Hamilton, D.L. & Sherman, J. (1994). Stereotypes. In R.S. Wyer Jr. & T.K. Srull (Hrsg.) Handbook of Social Cognition. S. 1-68, Hillsdale, NJ: Erlbaum.
- Harnad, S. (Ed.) (1987). Categorical perception. The groundwork of cognition. Cambridge, Mass.: Cambridge University Press.
- Harvey, L.O.Jr. (1992). The Critical Operating Characteristic and the Evaluation of Expert Judgment. In Organizational Behavior and Human Decision Processes 53, S. 229 – 251.
- Hastie, R. & Park, B. (1986). The relationship between memory and judgement depends on whether the judgement task is memory-based or on-line. Psychological Review, 93, S. 258-268.
- Hastie, R. & Pennington, N. (1989). Notes on the distinction between memory-based versus on-line judgment. In J.N. Bassili (Ed.), On-line cognition in person perception, S. 1-17. Hillsdale, NJ: Erlbaum.
- Hatano, G. (1982). Cognitive consequences of practice in culture specific procedural skills. The Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 4 (1), S. 15-18.
- Hatano, G. (1988). Social and motivational bases for mathematical understanding. In G.B. Saxe & M. Gerhart (Eds.), Children's mathematics, 55-70, San Francisco: Jossey-Bass.
- Heider, F. (1958). The psychology of interpersonal relations. New York.
- Helmholtz, H. (1867). Handbuch der Physiologischen Optik. In G. Karsten (Hrsg.). Allgemeine Enzyklopädie der Physik, Band 9, Leipzig: Voss.
- Henss, R. (1995). Das Fünf-Faktoren-Modell der Persönlichkeit bei der Beurteilung von Gesichtern. Report Psychologie, 20, 28-39.
- Henss, R. (1998). Gesicht und Persönlichkeitseindruck. Hogrefe: Göttingen.
- Herr, P.M. (1986). Consequences of priming: Judgment and behavior. Journal of Personality and Social Psychology, 51, S. 1106-1115.
- Hertel, G. & Bless, H. (2000). „On-line“ und erinnerungsgestützte Urteilsbildung: Auslösefaktoren und empirische Unterscheidungsmöglichkeiten. In Psychologische Rundschau, 51 (1), S. 19-28.
- Hertwig, R. (1993). Frequency-validity-effect und hindsight-bias: Unterschiedliche Phänomene – gleiche Prozesse? In W. Hell, K. Fiedler & G.Gigerenzer, (Hrsg.), Kognitive Täuschungen, S. 39-71.
- Hiebsch, H. (1986). Interpersonelle Wahrnehmung und Urteilsbildung – Psychologische Grundlagen der Beurteilung von Menschen, Berlin.

- Hinton, P.R. (1995). *The psychology of interpersonal perception*. London: Routledge.
- Hoffrage, U. (1993). Die Illusion der Sicherheit bei Entscheidungen unter Unsicherheit. In W. Hell, K. Fiedler & G. Gigerenzer (Hrsg.). *Kognitive Täuschungen*, S. 73-97. Heidelberg: Spektrum.
- Höft, S. (2005). Assessment Center-Forschung und –Praxis: Zwei getrennte Welten? In *Zeitschrift für Personalpsychologie*, 4 (4), S. 147-150. Göttingen: Hogrefe.
- Höft, S. & Funke, U. (2001). Simulationsorientierte Verfahren der Personalauswahl. In Schuler, H. (Hrsg.). *Lehrbuch der Personalpsychologie.*, S. 135 – 173. Göttingen: Hogrefe.
- Hogan, R. (1991). Personality and personality measurement. In M.D. Dunnette & L.M. Hough (Eds.) *Handbook of industrial and organizational psychology*, S. 873-919.
- Holding, D.H. (1979). The evaluation of chess positions. *Simulation & Games*, 10, 207-221.
- Holding, D.H. (1985). *The psychology of chess skill*. Hillsdale, NJ: Erlbaum.
- Holding, D.H. (1992). Theories of chess skill. *Psychological Research*, 54, 10-16.
- Holding, D.H. & Pfau, H.D. (1985). Thinking ahead in chess. *American Journal of Psychology*, 98, 271-282.
- Holyak, K.J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K. Anders Ericsson & J. Smith (Eds.) *Toward a general theory of expertise: Prospects and limits*, 301 – 335, New York: Cambridge University Press.
- Intraub, H. (1979). The role of implicit naming in pictorial encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 5, S. 78-87.
- Jendraczyk, M. (1991). “Snap Judgments” in der Personwahrnehmung. Diplomarbeit im Nebenfach Psychologie. Universität Duisburg.
- Johnson, M. K. & Raye, C. L. (1981a). Reality monitoring. *Psychological Review*, 88, 67-85.
- Johnson, M. K., Hashtroudi, S., & Sindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28.
- Johnson, P.E. (1982). Cognitive Models of Medical Problem Solvers. In Connelly, D.P. et al (eds.), *Clinical Decisions and Laboratory Use*, S. 39-51.
- Johnson, P. E., Duran, A.S., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P.J. & Swanson, D.B. (1981b). Expertise and error in diagnostic reasoning. *Cognitive Science* 5, S. 235-283.
- Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inferences, and consciousness*. Cambridge: Cambridge University Press.

- Joseph, G.-M. & Patel, V.L. (1986). Specificity of Expertise in Clinical Reasoning. In Cognitive Science Society, The Eighth Annual Conference of the Cognitive Science Society (S. 331-345). Hillsdale, NJ: Erlbaum
- Joseph, G.-M. & Patel, V.L. (1987). Problem Representation and Hypotheses Generation in Diagnostic Reasoning. In Cognitive Science Society, The Ninth Annual Conference of the Cognitive Science Society, s. 243-257, Hillsdale, NJ: Erlbaum.
- Jung, A. (1989). Gedächtnis für Handlungen bei verbalen, visuellen und motorischen Einkodierungen – eine empirische Untersuchung. Unveröffentl. Ms. Saarbrücken (Dipl.-Arbeit).
- Kamphaus, G. (1980). Veränderung der Personbeurteilung als Trainingseffekt. Vergleich der angewandten Gruppendynamik mit lernpsychologisch orientierten Verfahren. In Hochschulsammlung Philosophie. Psychologie (Bd. 7). Freiburg: Hochschulverlag.
- Kanning, U.P. (1999). Die Psychologie der Personenbeurteilung. Göttingen: Hogrefe.
- Kanning, U.P. (2002). Tipps für die Anwendung nicht-standardisierter Methoden. In Kanning, U.P. & Holling, H. (Hrsg.). Handbuch personaldiagnostischer Instrumente. S. 493 – 544. Göttingen: Hogrefe.
- Kanning, U.P., Hofer, S., Schulze Willbrenning, B. (2004). Professionelle Personenbeurteilung. Ein Trainingsmanual. Hogrefe Göttingen.
- Kanning, U.P. & Leisten, K. (2004). Selektive Verarbeitung nonverbaler Informationen durch AC-Beobachter. Wissenschaftlicher Vortrag auf dem 44. Kongress Deutscher Psychologen in Göttingen.
- Kant, I. (1980). Anthropologie in pragmatischer Hinsicht. Hrsg. Von Karl Vorländer. Hamburg: Meiner.
- Kelley, B.W. (1973). The process of causal attribution. In American Psychologist, 28, S. 107-128.
- Kempter, G. (1998). Das Bild vom Anderen. Skriptanimation als Methode zur Untersuchung spontaner Attributionsprozesse. Habilitationsschrift an der Universität Duisburg.
- Kendell, R.E. (1973). Psychiatric diagnoses. A study of how they are made. British Journal of Psychiatry, 122, S. 437-445.
- Kenny, D.A. (1991). A general model of consensus and accuracy in interpersonal perception: Acquaintance and the big five. Psychological Bulletin, 116, S. 245-258.
- Kenny, D.A. (1994). Interpersonal perception. A social relations analysis. New York: Guilford Press.
- Kihlstrom, J.F. (1980). Posthypnotic amnesia for recently learned materials. Interactions with episodic and semantic memory. Cognitive Psychology, S. 227-251.
- Kintsch, W. (1974). The representation of meaning in memory. New York: Wiley.

- Kintsch, W. & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, 5, S. 257-274.
- Klages, L. (1926). *Die Grundlagen der Charakterkunde*. Leipzig: Johann Ambrosius Barth.
- Kolodner, J.L. (1983). Towards an understanding of the role of experience in the evolution from novice to expert. *International Journal of Man-Machine Studies*, 19, S. 497-518.
- Krause, C. (2000). *Posthypnotische Amnesie für therapeutische Geschichten*. Dissertation an der Eberhard-Karls-Universität Tübingen.
- Krems, Josef F. (1994). *Wissensbasierte Urteilsbildung: diagnostisches Problemlösen durch Experten und Expertensysteme*. Bern.
- Krems, J. & Bachmaier, M. (1991). Hypothesenbildung und Strategieauswahl in Abhängigkeit vom Expertisegrad. *Zeitschrift für experimentelle und angewandte Psychologie*, 38, 394-410.
- Krolak-Schwerdt, S. (2000). *Kategorien des Personengedächtnisses. Einfluß von Stimulusmerkmalen und Verarbeitungszielen auf ihre Aktivierung*. Habilitationsschrift, Saarbrücken: Universität des Saarlandes.
- Krolak-Schwerdt, S. & le Coutre, C. (2002). Ist das Gedächtnis für klinisch-diagnostische Personmerkmale abhängig vom Verarbeitungsziel? In *Psychologische Beiträge Band 44*, 2002, S. 428 – 446.
- Krolak-Schwerdt, S. & Wintermantel, M. (2002). Einfluss der Textkohärenz auf das Behalten von Personbeschreibungen. Vortrag auf dem 43. Kongress der Deutschen Gesellschaft für Psychologie. Humboldt Universität Berlin.
- Krolak-Schwerdt, S. & Wintermantel, M. (2004). Verarbeitung von personbeschreibender Information als Textverstehen. In *Psychologische Rundschau 55 (1)*, S. 11-18.
- Kuipers, B. & Kassirer, J.P. (1984). Causal reasoning in medicine: analysis of a protocol. *Cognitive Science*, 8, S. 363-385.
- Laine, T., Hyötyniemi, H. & Saarilouma, P. (1998). Foundations of simulative theorizing. In H. Prade (Ed.) *The Proceedings of the 13th European Conference on Artificial Intelligence*. EXAI 98. London.
- Larkin, J. H. (1983). The role of problem representation in physics. In Gentner, D. & Stevens, A. (eds.). *Mental models*, S. 75-98. Hillsdale, N.J.: Erlbaum.
- Larkin J.H., McDermott, J., Simon, D.P. & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.
- Lavater, J.C. (1775-1778). *Physiognomische Fragmente zur Beförderung der Menschenkenntnis und Menschenliebe*. Vier Bände. Leipzig: Weidmann Erben & Reich.

- Leibrich, J. & White, K.G. (1983). Information stored when expecting recall or recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9, S. 440-455.
- Levenberg, S.B. (1975). Professional training, psychodiagnostic skill and Kinetic Family Drawings. *Journal of Personality Assessment*, 39, S. 389-393.
- Lichtenberg, G.C. (1778/1991). Goettinger Taschen=Calender vom Jahr 1778. Faksimileausgabe 1991. Mainz: Dieterich'sche Verlagsbuchhandlung.
- Lichtenberg, G.C. (1972). Schriften und Briefe. Dritter Band. Aufsätze; Entwürfe, Gedichte; Erklärungen der Hogarthischen Kupferstiche. Herausgegeben von Wolfgang Promies. München: Carl Hanser Verlag.
- Lichtenberg, G.C. (1980). Schriften und Briefe. Erster Band. Sudelbücher I. Herausgegeben von Wolfgang Promies. 3. Auflage. München: Carl Hanser Verlag.
- Lilli, W. (1984). Die Hypothesentheorie der sozialen Wahrnehmung. In D. Frey & M. Irle (Hrsg.) *Theorien der Sozialpsychologie, Band I, Kognitive Theorien*, S. 19-46.
- Livesley, W.J. & Bromley, D.B. (1973). *Person perception in childhood and adolescence*. London: Wiley.
- Loftus, G.R. & Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, S. 435-411.
- Lorenz, K. & Leyhausen, P. (1968). *Antriebe tierischen und menschlichen Verhaltens*. München: Piper.
- Macrae, C.N., Hewstone, M. & Griffiths, R.J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, 23, S. 77-87.
- Madigan, S. (1983) Picture memory. In J.C. Yuille (Ed.), *Imagery, memory and cognition*. Hillsdale, NJ: Erlbaum.
- Mandl, H., Gruber, H. & Renkl, A. (1993). Kontextualisierung von Expertise. In Mandl, H., Dreher, M. & Kornadt, H.-J. (Hrsg.), *Entwicklung und Denken im kulturellen Kontext*. (S. 203-229). Göttingen: Hogrefe.
- Mandl, H. & Spada, H. (1988). *Wissenspsychologie*. München: Psychologie Verlags Union.
- Mayring, P. (2000). Qualitative Inhaltsanalyse. *Forum Qualitative Sozialforschung / Forum Qualitative Social Research [On-line Journal]*, 1 (2). Verfügbar über: <http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-d.htm> [Datum des Zugriffs: 01.06.2005].
- McArthur, L.Z. & Baron, R. (1983). Toward an ecological theory of social perception. *Psychological review*, 90, S. 225-238.
- McConnell, A.R., Sherman, S.J. & Hamilton, D.L. (1994). On-line and memory-based aspects of individual and group target judgements. *Journal of Personality and Social Psychology*, 67, S. 173-185.

- McCrae, R.R. (1989). Why I advocate the five-factor model: Joint analyses of the NEO-PI with other instruments. In D.M. Buss & N. Cantor (Eds.). *Personality psychology. Recent trends and emerging directions*, S. 237-245.
- McCrae, R.R. & Costa, P.T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, S. 81-90.
- McCrae, R.R. & Costa, P.T. (1990). *Personality in adulthood*. New York: Guilford Press.
- Mehle, T. (1982). Hypothesis generation in an automobile malfunction inference task. *Acta Psychologica*, 52, S. 87-106.
- Metz, C. (2001). In: Monaco, J.: *Film verstehen – Kunst, Technik, Sprache, Geschichte und Theorie des Films und der neuen Medien*. Reinbeck bei Hamburg, Rowohlt Taschenbuch Verlag.
- Miller, G.A., (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, S. 81-96.
- Monaco, J. (2001). *Film verstehen – Kunst, Technik, Sprache, Geschichte und Theorie des Films und der neuen Medien*. Reinbeck bei Hamburg, Rowohlt Taschenbuch Verlag.
- Morine, G & Vallance, E. (1975). Special study B: A study of teacher and pupil perceptions of classroom interaction (Technical Report Nr. 75-11-6). San Francisco.
- Morrow, D.G., Bower, G.H. & Greenspan, S.L. (1989). Updating situation models during narrative comprehension. In *Journal of Memory and Language*, 28, S. 292-312.
- Morrow, D.G. (1994). Spatial models created from text. In van Oostendorp, H. & Zwaan, R.A. (Eds.) *Naturalistic text comprehension*, S. 57-78.
- Moser, D.V. (1992). Does memory affect judgment? Self-generated versus recall measures. *Journal of Personality and Social Psychology*, 62, S. 555-563.
- Murphy, G.L. & Wright, J.C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 10, S. 144-155.
- Mussweiler, T. (2003a). Comparison Processes in Social Judgment: Mechanisms and Consequences. In *Psychological Review*, Vol. 110, Nr. 3, S. 472-489.
- Mussweiler, T. (2003b). “Everything is relative”: Comparison processes in social judgment. In *European Journal of Social Psychology* 33, S. 719-733.
- Mussweiler, T., Rüter, K. & Epstude, K. (2004a). The man who wasn't there: Subliminal social comparison standards influence self-evaluation. In *Journal of Experimental Social Psychology* 40, S. 689-696.
- Mussweiler, T., Rüter, K. & Epstude, K. (2004b). The Ups and Downs of Social Comparison: Mechanisms of Assimilation and Contrast. In *Journal of Personality and Social Psychology* Vol. 87, Nr. 6, S. 832-844.

- Muzzin, L.J., Norman, G.R., Jacoby, L.L., Feightner, J.V., Tugwell, P. & Guvett, G.H. (1982). Manifestations of expertise in recall of clinical protocols. In Papers of the Assoc. of American Medical Colleges (S. 163-168), Washington, D.C.: AAMC.
- Nieding, G. & Ohler, P. (1999). Der Einfluss von Protagonisten-Zielstrukturen auf räumliche mentale Modelle beim narrativen Textverstehen von Kindern. In Zeitschrift für Sprache und Kognition, 18.Jg. Heft 3/4, S. 146-158.
- Niermeyer, R. (1999). Beobachterkompetenz. In W. Jochmann (Hrsg.), Innovationen im Assessment-Center (S. 157-179). Stuttgart: Schäffer-Poeschel.
- Norman, W.T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structures in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66, 574-583
- Norman, W.T. (1967). 2.800 personality trait descriptors: Normative operating characteristics for a university population. Unveröff. Manuskript. University of Michigan.
- Norman, G., Jacoby, L., Feightner, J. & Campbell, E. (1979). Clinical Experience and the structure of memory. In Papers of the Association of American Medical Colleges, S. 214-218), Washington, D.C.: AAMC.
- Obermann, C. (2002). Assessment Center: Entwicklung, Durchführung, Trends. Wiesbaden: Gabler.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 29, S. 261-265.
- Ostrom, T.M., Pryor, J.B. & Simpson, D.D. (1981). The organization of social information. In E.T. Higgins, C.P. Herman & M.P. Zanna (Hrsg.). *Social Cognition: The Ontario Symposium*, S. 3-38, Hillsdale, NJ: Erlbaum.
- Oswald, M.E. (1993). Hypothesentesten: Suche und Verarbeitung hypothesenkonformer und hypothesenkonträrer Informationen. In W. Hell, K. Fiedler & G. Gigerenzer (Hrsg.) *Kognitive Täuschungen*, S. 189-211, Heidelberg: Spektrum.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehard & Winston.
- Pashler, H. & Carrier, M. (1996). Overview of human memory. In Bjork, E.L. & Bjork, R.A. (Eds.) *Memory*, S. 3-23, San Diego: Academic Press.
- Perrig, W.J. (1988). *Vorstellungen und Gedächtnis*. Heidelberg: Springer.
- Pohl, R.F. & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational behaviour and Human Decision Processes*, 67, S. 49-58.
- Posner, M.I. (1988). Introduction: What is to be an expert? In M.T.H. Chi, R. Glaser & M.J. Farr (Eds.), *The nature of expertise* (pp. xxix- xxxvi). Hillsdale, NJ: Erlbaum

- Postman, L. (1951). Toward a general theory of cognition. In J.H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads*, S. 242-272.
- Pöhls, U. (1989) *Politik im Fernsehen: Zur Personalisierung in den Fernsehnachrichten der USA, Frankreichs und der Bundesrepublik Deutschland*. Inaugural-Dissertation zur Erlangung des Doktorgrades der Philosophischen Fakultät der Heinrich-Heine-Universität Düsseldorf.
- Preckel, D. & Schüpbach, H. (2005). Zusammenhänge zwischen rezeptiver Selbstdarstellungskompetenz und Leistung im Assessment Center. In *Zeitschrift für Personalpsychologie* 4(4), S. 151-158, Göttingen: Hogrefe.
- Priest, A.G. & Lindsay, R.O. (1992). New light on novice-expert differences in physics problem solving. *British Journal of Psychology*, 83, S. 389-405.
- Pryor, J.B. & Ostrom, T.M. (1981). The cognitive organization of social information: A converging operations approach. *Journal of Personality and Social Psychology*, 41, S. 628-641.
- Püschel, A. (2005). Absicherung von Entscheidungen im Assessment Center durch Einführung von Begriff und Quantifizierung der Messunsicherheit. In *Zeitschrift für Personalpsychologie* 4 (4), S. 181-186.
- Putz-Osterloh, W. (1987). Gibt es Experten für komplexe Probleme? *Zeitschrift für Psychologie*, 195, S. 63-84.
- Putz-Osterloh, W. (1988). Wissen und Problemlösen. In Mandl, H. & Spada, H. (Hrsg.), *Wissenspsychologie*, S. 247-263.
- Ratcliff, R. & McKoon, G. (1978). Priming the intern recognition. Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior* 17, S. 403-418.
- Rinck, M. & Bower, G.H. (1995). Anaphora resolution and the focus of attention in situation models. In *Journal of Memory and Language*, 34, S. 110-131.
- Rinck, M. (2000). Situationsmodelle und das Verstehen von Erzähltexten: Befunde und Probleme. In *Psychologische Rundschau*, 51, (3), S. 115-122.
- Roediger, H.L. & Weldon, M.S. (1987). Reversing the picture superiority effect. In M. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes*. New York: Springer.
- Rose, R. (1954). *Die zwölf Geschworenen (Twelve angry Men)*. Berlin: Felix Bloch Erben.
- Saariluoma, P. (1995). *Chess players` thinking*. London.
- Saariluoma, P. & Laine, T. (2001). Novice construction of chess memory. In *Scandinavian Journal of Psychology*, 2001, 42, S. 137 – 146.

Schmidt, H. & Boshuizen, H. (1990). Effects of activation of prior knowledge on the recall of a clinical case. Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

Schmidt, H., Boshuizen, H. & Hobus, P. (1988). Transitory stages in the development of medical expertise: The intermediate effect in clinical case representation studies. Paper presented at the 10th Annual Conference of the Cognitive Science Society, Montreal, Canada.

Schneider, W. & Körkel, J. (1989). The knowledge base and text recall: Evidence from a short-term longitudinal study. In *Contemporary Educational Psychology*, 14, S. 382-393.

Schneider, W., Körkel, J. & Weinert, F.E. (1990). Expert Knowledge, General Abilities, and Text Processing. In Schneider, W. & Weinert, F.E. (eds.) *Interactions among Aptitudes, Strategies, and Knowledge in Cognitive Performance*, S. 235-251, New York: Springer.

Schuler, H. (2001). *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe.

Schuler, H. & Berger, W. (1979). Physische Attraktivität als Determinante von Beurteilung und Einstellungsempfehlung. *Psychologie und Praxis*, 2, S. 59-70.

Schulz, W. (1975). Wirkungsqualitäten verschiedener Medien. Experimentelle Untersuchungen über die Vermittlung von konnotativer Bedeutung durch unterschiedliche Formen medialer Darstellung. *Rundfunk und Fernsehen*, 23, S. 57-72.

Schulz, W. (1982). Ausblick am Ende des Holzweges. Eine Übersicht über die Ansätze der neuen Wirkungsforschung. *Publizistik*, 27, S. 49-73.

Seitz, G. (1979). Film als Rezeptionsform von Literatur. Zum Problem der Verfilmungen von Thomas Manns Erzählungen „Tonio Kröger“, „Wälsungenblut“ und „Der Tod in Venedig“. München: tuduv-Verlagsgesellschaft.

Showers, C. & Cantor, N. (1985). Social cognition: A look at motivated strategies. *Annual Review of Psychology*, 26, S. 275-305.

Simon, D. P. & Simon, H.A. (1978). Individual differences in solving physics problems. In Siegler, R.S. (ed.), *Children's thinking: What develops* (pp. 325-348). Hillsdale, NJ: Erlbaum

Sines, L.K. (1959). The relative contribution of four kinds of data to accuracy in personality assessment. *Journal of Consulting Psychology*, 23, 483-492.

Singer, M. (1994). Discourse inference processes. In Gernsbacher, M.A. (Ed) *Handbook of psycholinguistics*, S. 479-515.

Smith, E. R. & Mackie, D.M. (2000). *Social psychology* (2nd edition) Philadelphia: Psychology Press.

Snyder, M. & Swann, W.B.(1978). Hypothesis-testing strategies in social interaction. *Journal of Personality and Social Psychology*, 36, 1202-1212.

Snyder, M. & Uranowitz, S.2. (1978). Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology*, 36, S. 941-950.

Spiro, R.J. (1980). Accommodative reconstruction in prose recall. *Journal of Verbal Learning and Verbal Behavior*, 19, S. 84-95.

Spiro, R.J., Feltovich, P.J., Jacobson, M.J. & Coulson, R.L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in illstructured domain. *Educational Technology*, 31, 24-33

Srull, T.K. & Wyer, R.S. (1989). Person memory and judgement. *Psychological Review*, 96, S. 58-83.

Stahlberg, D. & Frey, D. (1990). Einstellungen I: Struktur, Messung und Funktionen. In W. Stroebe, M. Hewstone, J.-P. Codol & G.M. Stephenson (Hrsg.) *Sozialpsychologie*. Berlin: Springer-Verlag, S. 144-170.

Standing, L. (1973). Learning 10.000 pictures. *Quarterly Journal of Experimental Psychology*, 25, S. 207-222.

Strack, F. (1988). Social Cognition: Sozialpsychologie innerhalb des Paradigmas der Informationsverarbeitung. *Psychologische Rundschau*, 39, S. 72-82.

Strack, F. (2000). Urteilsbildung im sozialen Kontext. In *Zeitschrift für Psychologie*, 208, S. 226-241.

Trabasso, T. & v.d.Broek, P. (1985). Causal thinking and the representation of narrative events. In *Journal of Memory and Language*, 24, S. 612-630.

Tulving, E. (1972). Episodic and semantic memory. In Tulving, E. & Donaldson, W. (Eds.) *Organization of memory*. S. 381 – 403, New York, Academic Press.

Tupes, E.C. & Christal, R.E. (1961/1992). Recurrent personality factors based on trait ratings. *Journal of Personality*, 60, S. 225-251.

Tversky, A. & Kahnemann, D. (1974). *Judgment under Uncertainty: Heuristics and Biases*. American Association for the Advancement of Science. Washington. S. 1124-1131.

Van den Broek, P. (1994). Comprehension and memory of narrative texts. Inferences and coherence. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*, S. 539-588.

Van Dijk, T.A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York.

Vassend, O. & Skrandal, A. (1995). Factor analytic studies of the NEO Personality Inventory and the five-factor model: The problem of high structural complexity and conceptual indeterminacy. *Personality and Individual Differences*, 19, S. 135-147.

Vessey, I. (1985). Expertise in debugging computer systems: A process analysis. *International Journal of Man-Machine Studies*, 23, S. 459-494.

Voss, J. F., Fincher-Kiefer, R.H., Green, T.R. & Post, T.A. (1986). Individual Differences in Performance: the Contrastive Approach to Knowledge. In Sternberg, R.J. (ed.) *Advances in the Psychology of Human Intelligence*, 3, S. 297-334. Hillsdale, NJ: Erlbaum.

- Waara, H. (2003). Literarisches Erzählen in Text und Film. Unveröffentlichtes Skript an der Örebro Universität – Humanistiska institutionen.
- Walther, E. (1997). Der Einfluß von Wissensstrukturen auf konstruktive Prozesse der sozialen Urteilsbildung. *Psychologia Universalis*, Band 7.
- Wanderer, Z.W. (1969). Validity of clinical judgments based on human figure drawings. *Journal of Consulting and Clinical Psychology*, 33, S.143-150.
- Watson, C.G. (1967). Relationship of distortion to DAP diagnostic accuracy among psychologists at three levels of Sophistication. *Journal of Consulting Psychology*, 53, S. 95-108.
- Waxner, P. (1976). Nonverbal cues for depth of depression: Set versus no set. *Journal of Consulting and Clinical Psychology*, 44, 493.
- Weiss, H.J. (1982). Die Wahlkampfberichterstattung und –kommentierung von Fernsehen und Tagespresse zum Bundestagswahlkampf 1980. *Media Perspektiven*, S. 263-275.
- Werner, P.D., Rose, T.L., & Yesavage, J.A. (1983). Reliability, accuracy, and decision-making strategy in clinical predictions of imminent dangerousness. *Journal of Consulting and Clinical Psychology*, 51, S. 815-825.
- Werth, L. (2004). *Psychologie für die Wirtschaft. Grundlagen und Anwendungen*. Heidelberg: Spektrum Akademischer Verlag.
- Wessells, M.G. (1994). *Kognitive Psychologie*. München: Ernst Reinhardt.
- Wiggins, J.S. & Trapnell, P.D. (1997). Personality Structure. The Return of the Big Five. In *Handbook of Personality Psychology*, S. 737-765.
- Winterhoff-Spurk, P. (1986). *Fernsehen: Psychologische Befunde zur Medienwirkung*. Bern: Huber.
- Wintermantel, M. (2002). Verstehen von Personbeschreibungen. In G.Rickheit, Th. Herrmann & W. Deutsch, (Hrsg.), *Handbuch Psycholinguistik*.
- Wintermantel, M. & Christmann, U. (1983). Person description: Some empirical findings concerning the production and reproduction of a specific text type. In G. Rickheit & M. Bock (Eds.) *Psycholinguistic studies in language processing*, S. 137-151.
- Wintermantel, M. & Krolak-Schwerdt, S. (2002). Eindrucksbildung aus Personbeschreibungen: Der Einfluss der syntaktischen Kohärenz auf die soziale Kognition. In *Zeitschrift für Sozialpsychologie*, 33, S. 45-64.
- Wiseman, S. & Neisser, U. (1974). Perceptual organization as a determinant of visual recognition memory. *American Journal of Psychology*, 87, S. 675-681.
- Woehr, D.J. & Huffcutt, A.I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67, S. 189 – 205.

- Wontorra, H. & Krems, J. (1989). Acquisition of strategies depending on the opponents' competence level. *Archiv für Psychologie*, 141, 113-126.
- Wyer, R.S. & Srull, T.K. (1986). Human cognition in its social context. *Psychology review*, 93, S. 322-359.
- Yekovich, F. R., Walker, C.H., Ogle, L.T. & Thompson, M.A. (1990). The influence of domain knowledge on inferencing in low-aptitude individuals. In A.C. Graesser & G.H. Bower (Eds.) *The psychology of learning and motivation* Vol. 25, S. 175-196.
- Young, R. K., Kennedy, A.H., Newhouse, A., Browne, P. & Thiessen, D. (1993). The effects of names on perceptions of intelligence popularity, and competence. *Journal of Applied Social Psychology*, 23, S. 1770-1988.
- Zajchowski, R., & Martin, J. (1993). Differences in the problem solving of stronger and weaker novices in physics: knowledge, strategies, or knowledge structure? *Journal of Research in Science Teaching*, 20, S. 459-470.
- Zebrowitz, L.A. (1996). Impression formation. In A.S. Manstead & M. Hewstone (Hrsg.). *The Blackwell encyclopedia of social psychology*. S. 309-314. Oxford: Blackwell.
- Zebrowitz, L.A. (1997). *Reading faces: Window to the soul?* Boulder, CO: West Press.
- Zuckermann, M. & Kieffer, S.C. (1994). Race differences in face-ism: Does facial prominence imply dominance? *Journal of Personality and Social Psychology*, 66, S. 86-92.
- Zwaan, R.A., Langston, M.C. & Graesser, A.C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6, S. 292-297.
- Zwaan, R.A. & Radvansky, G.A. (1998). Situation Models in Language Comprehension and Memory. In *Psychological Bulletin*, 123, (2), S. 162-185.
- Zwaan, R.A. & van Oostendorp, H. (1993). Do readers construct spatial representations in naturalistic story comprehension? *Discourse Processes*, 16, S. 125-143.

Anhang

Ergebnisse der Voruntersuchungen

1. Voruntersuchung

Ergebnisse zur Untersuchung, nach welcher Anzahl der Darbietung des Filmausschnittes die Gedächtnisinhalte zwischen den Stimulusvariationen Text und Film etwa gleich groß sind

Tabelle 1 enthält die Anzahl der mittleren reproduzierten Informationen in Abhängigkeit des Stimulusmaterials Text bzw. Film und der Anzahl der Darbietungen des Filmausschnittes.

Tabelle 1:

Mittelwerte der Reproduktionen	Text	Anzahl der Darbietung des Filmausschnittes				
		1	2	3	4	5
Originalinformationen	2,5	1,2	1,4	2,6	2,8	2,9
Synonyme Informationen	20,8	3,4	7,8	18,6	22,2	24,3
generierte Eigenschaften	4,2	1,1	1,2	4,5	4,8	4,8
Zusammenfassungen	4,8	1,3	2,2	5,2	5,6	5,8

2. Voruntersuchung

Ergebnisse zur Generierung der Eigenschaftsitems für die Ratingskala

Eigenschaftszuschreibung

Anzahl der Nennung

Geschworener 1

sachlich	12
korrekt	11
ruhig	10
angespannt	9
verantwortungsbewusst	7
durchsetzungsfähig	7

Geschworener 3

autoritär	14
aggressiv	13
verbittert	12
voreingenommen	10
ungerecht	9
respektlos	8

Geschworener 7

egoistisch	12
desinteressiert	11
überzeugend	9
intolerant	8
oberflächlich	7
humorvoll	7

Geschworener 8

gerecht	18
intelligent	13
besonnen	12
offen	11
herzlich	11
clever	10

Geschworener 10

unbeherrscht	11
aufbrausend	10
unhöflich	9
provozierend	7
ungeduldig	7
selbstbewusst	6

Vorinformation

Die Vorinformation wurde den Versuchspersonen vor der Darbietung des Stimulusmaterials als Tonbandansage dargeboten.

„Meine Herren Geschworenen,

damit ist dieser komplizierte Fall abgeschlossen. Der vorsätzliche Mord ist die schwerste Anklage, die von unseren Schwurgerichten erhoben werden kann.

Sie haben die Zeugenaussagen gehört. Die Paragraphen des Gesetzes, die auf diesen Fall zutreffen, wurden ihnen erläutert.

Es ist jetzt ihre Pflicht, Wahres vom Unwahren zu trennen. Ein Leben wurde ausgelöscht, über ein zweites sollen sie entscheiden.

Wenn sie einen begründeten Zweifel an der Schuld des Angeklagten hegen, einen wirklich begründeten Zweifel, dann muss ihr Urteilsspruch lauten „Nicht schuldig.“

Wenn jedoch kein begründeter Zweifel besteht, so müssen sie mit ruhigem Gewissen den Anklagten schuldig sprechen.

Aber wie sie auch entscheiden mögen, ihr Urteil muss einstimmig sein.

Für den Fall, dass sie den Angeklagten für schuldig befinden, wird das Gericht ein Gnadengesuch nicht befürworten. Dann ist das gefällte Todesurteil unabänderlich.

Somit tragen sie eine schwere Verantwortung. Ich danke den Geschworenen.“

(Originalton wurde dem Film „Die 12 Geschworenen“ von Reginald Rose, 1957 entnommen.)

Instruktionen

Instruktion für Stimulus Text und Text mit Bild vor der Textbearbeitung

Instruktion:

Bitte lesen Sie den vorgelegten Text aufmerksam durch.

Lesen Sie den Text so oft und intensiv, bis Sie das Verhalten der handelnden Personen möglichst detailgetreu wiedergeben können.

Bitte hören Sie erst dann mit der Textbearbeitung auf, wenn Sie das Verhalten den einzelnen Geschworenen-Nummern zuordnen können.

Instruktion für Stimulus Text und Text mit Bild nach der Textbearbeitung zur Reproduktion der Gedächtnisinhalte

Instruktion:

Bitte notieren Sie nun alles aus dem Textabschnitt, woran Sie sich noch erinnern können.

Instruktion für Stimulus Film vor der Darbietung des Filmausschnittes

Instruktion:

Bitte schauen Sie sich den folgenden Filmausschnitt aufmerksam an. Achten Sie dabei vor allem auf die handelnden Personen.

**Instruktion für Stimulus Film nach der Darbietung des Filmausschnittes zur
Reproduktion der Gedächtnisinhalte**

Instruktion:

**Bitte notieren Sie nun alles aus dem Filmausschnitt, woran Sie sich noch erinnern
können.**

Instruktion für freie Beurteilung für alle Stimulusvariationen

Instruktion:

Beurteilen Sie jetzt bitte jeden Geschworenen und formulieren Sie Ihre Einschätzung der Charaktereigenschaften so detailliert wie möglich.

Instruktion für alle Stimulusvariationen zur Einschätzung der Geschworenen anhand der vorgelegten Ratingskala

Instruktion:

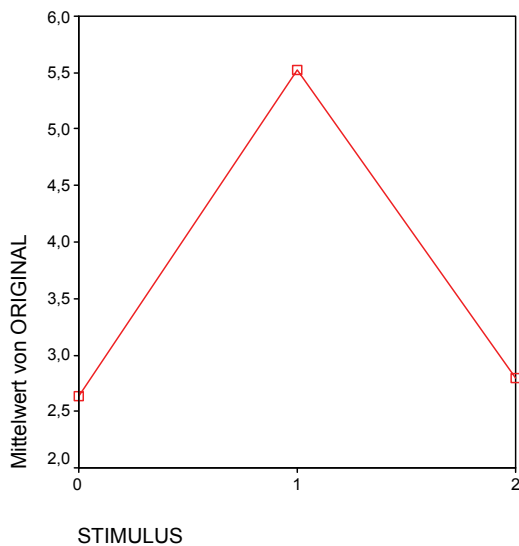
Bitte markieren Sie auf der vorliegenden Skala wie Ihrer Meinung nach das jeweils genannte Merkmal für jede Person ausgeprägt ist.

(Hierbei gibt es keine richtigen und keine falschen Antworten. Es zählt ihr persönlicher Eindruck.)

Darstellung der Ergebnisse zur Reproduktion der Gedächtnisinhalte

1. Darstellung der Mittelwerte und Standardabweichungen der reproduzierten Originalinformationen pro Versuchsperson bei Variation der Stimulusinformation

Grafik 1 - Darstellung der Mittelwerte



Legende zur Grafik 1

Abszisse: Anzahl der mittleren reproduzierten Originalinformationen
 Ordinate: Medium der dargebotenen Information
 0 ... Stimulus Text
 1 ... Stimulus Text mit Bild
 2 ... Stimulus Film

Ergebnistabelle 1 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

ORIGINAL								
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
0	32	2,63	3,963	,763	1,06	4,20	0	17
1	30	5,52	5,627	1,228	2,96	8,09	0	18
2	31	2,80	2,678	,691	1,32	4,28	0	7
Gesamt	93	3,63	4,498	,567	2,50	4,77	0	18

Legende zu Ergebnistabelle 1

0 ... Stimulusvariation Text
 1 ... Stimulusvariation Text mit Bild
 2 ... Stimulusvariation Film
 N... Gruppengröße

Statistiktable 1 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Stimulus Text und Film hinsichtlich der reproduzierten Originalinformationen

ANOVA

ORIGINAL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,280	1	,280	,022	,883
Within Groups	508,696	40	12,717		
Total	508,976	41			

Gruppenvergleich zwischen Stimulus Text und Text mit Bild hinsichtlich der reproduzierten Originalinformationen

ANOVA

ORIGINAL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	98,945	1	98,945	4,370	,042
Within Groups	1041,534	46	22,642		
Total	1140,479	47			

Gruppenvergleich zwischen Stimulus Text mit Bild und Film hinsichtlich der reproduzierten Originalinformationen

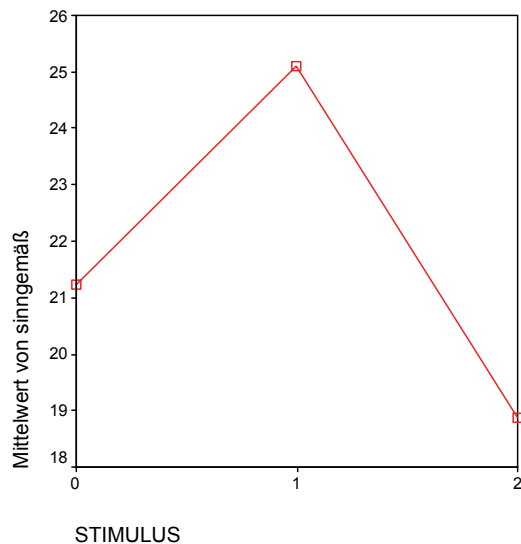
ANOVA

ORIGINAL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	64,917	1	64,917	3,009	,092
Within Groups	733,638	34	21,578		
Total	798,556	35			

2. Darstellung der Mittelwerte und Standardabweichungen der sinngemäß reproduzierten Informationen pro Versuchsperson bei Variation der Stimulusinformation

Grafik 2 – Darstellung der Mittelwerte



Legende zur Grafik 2

Abszisse: Anzahl der mittleren reproduzierten sinngemäßen Informationen

Ordinate: Medium der dargebotenen Information

0 ... Stimulus Text

1 ... Stimulus Text mit Bild

2 ... Stimulus Film

Ergebnistabelle 2 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

sinngemäß								
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
0	32	21,22	15,624	3,007	15,04	27,40	3	51
1	30	25,10	16,226	3,541	17,71	32,48	5	53
2	31	18,87	11,957	3,087	12,24	25,49	2	46
Gesamt	93	21,95	15,016	1,892	18,17	25,73	2	53

Legende zu Ergebnistabelle 2

0 ... Stimulusvariation Text

1 ... Stimulusvariation Text mit Bild

2 ... Stimulusvariation Film

N... Gruppengröße

Statistiktable 2 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Stimulus Text und Film hinsichtlich der sinngemäß reproduzierten Informationen

ANOVA

sinngemäß

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	53,505	1	53,505	,256	,615
Within Groups	8348,400	40	208,710		
Total	8401,905	41			

Gruppenvergleich zwischen Stimulus Text und Text mit Bild hinsichtlich der sinngemäß reproduzierten Informationen

ANOVA

sinngemäß

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	177,190	1	177,190	,702	,406
Within Groups	11612,476	46	252,445		
Total	11789,667	47			

Gruppenvergleich zwischen Stimulus Text mit Bild und Film hinsichtlich der sinngemäß reproduzierten Informationen

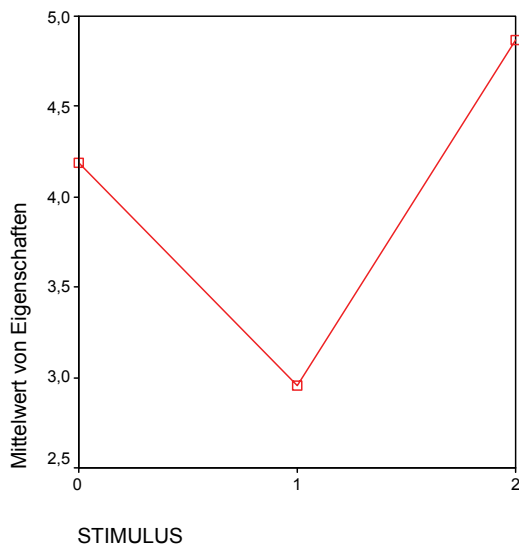
ANOVA

sinngemäß

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	339,457	1	339,457	1,588	,216
Within Groups	7267,543	34	213,751		
Total	7607,000	35			

3. Darstellung der Mittelwerte und Standardabweichungen der aus den Informationen generierten Eigenschaften pro Versuchsperson bei Variation der Stimulusinformation

Grafik 3 – Darstellung der Mittelwerte



Legende zur Grafik 3

Abszisse: Anzahl der mittleren aus den dargebotenen Informationen generierten Eigenschaften

Ordinate: Medium der dargebotenen Information

0 ... Stimulus Text

1 ... Stimulus Text mit Bild

2 ... Stimulus Film

Ergebnistabelle 3 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Eigenschaften								
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
0	32	4,19	6,451	1,242	1,63	6,74	0	31
1	30	2,95	3,324	,725	1,44	4,47	0	12
2	31	4,87	5,780	1,492	1,67	8,07	0	23
Gesamt	93	3,94	5,397	,680	2,58	5,30	0	31

Legende zu Ergebnistabelle 3

0 ... Stimulusvariation Text

1 ... Stimulusvariation Text mit Bild

2 ... Stimulusvariation Film

N... Gruppengröße

Statistiktable 3 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Stimulus Text und Film hinsichtlich der aus den dargebotenen Informationen generierten Eigenschaften

ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4,478	1	4,478	,116	,736
Within Groups	1549,807	40	38,745		
Total	1554,286	41			

Gruppenvergleich zwischen Stimulus Text und Text mit Bild hinsichtlich der aus den dargebotenen Informationen generierten Eigenschaften

ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	17,953	1	17,953	,634	,430
Within Groups	1303,026	46	28,327		
Total	1320,979	47			

Gruppenvergleich zwischen Stimulus Text mit Bild und Film hinsichtlich der aus den dargebotenen Informationen generierten Eigenschaften

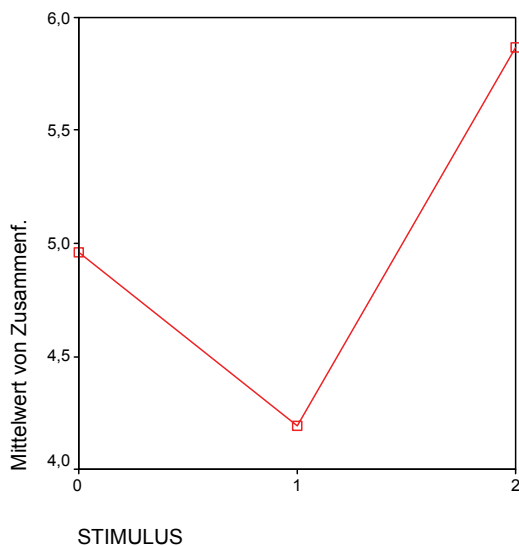
ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	32,064	1	32,064	1,583	,217
Within Groups	688,686	34	20,255		
Total	720,750	35			

4. Darstellung der Mittelwerte und Standardabweichungen der reproduzierten zusammengefassten Informationen pro Versuchsperson bei Variation der Stimulusinformation

Grafik 4 – Darstellung der Mittelwerte



Legende zur Grafik 4

Abszisse: Anzahl der mittleren reproduzierten zusammengefassten Informationen

Ordinate: Medium der dargebotenen Information

0 ... Stimulus Text

1 ... Stimulus Text mit Bild

2 ... Stimulus Film

Ergebnistabelle 4 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Zusammenf.

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
0	32	4,96	3,848	,741	3,44	6,49	0	13
1	30	4,19	3,669	,801	2,52	5,86	0	12
2	31	5,87	3,907	1,009	3,70	8,03	1	12
Gesamt	93	4,92	3,794	,478	3,97	5,88	0	13

Legende zu Ergebnistabelle 4

0 ... Stimulusvariation Text

2 ... Stimulusvariation Film

1 ... Stimulusvariation Text mit Bild

N...Gruppengröße

Statistiktable 4 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Stimulus Text und Film hinsichtlich der reproduzierten zusammengefassten Informationen

ANOVA

Zusammenf.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7,875	1	7,875	,526	,472
Within Groups	598,696	40	14,967		
Total	606,571	41			

Gruppenvergleich zwischen Stimulus Text und Text mit Bild hinsichtlich der reproduzierten zusammengefassten Informationen

ANOVA

Zusammenf.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7,049	1	7,049	,496	,485
Within Groups	654,201	46	14,222		
Total	661,250	47			

Gruppenvergleich zwischen Stimulus Text mit Bild und Film hinsichtlich der reproduzierten zusammengefassten Informationen

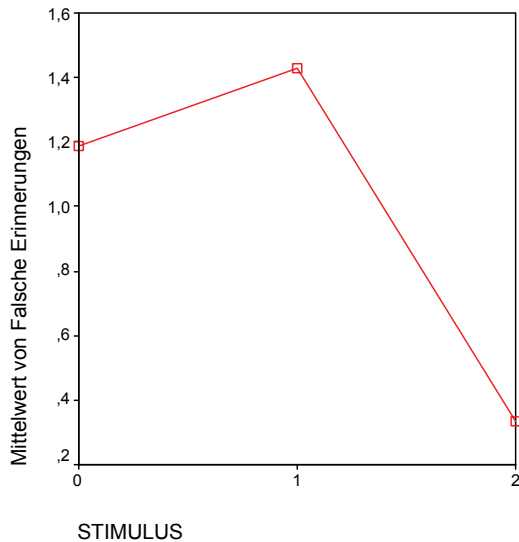
ANOVA

Zusammenf.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	24,584	1	24,584	1,731	,197
Within Groups	482,971	34	14,205		
Total	507,556	35			

5. Darstellung der Mittelwerte und Standardabweichungen der genannten Intrusionen pro Versuchsperson bei Variation der Stimulusinformation

Grafik 5 – Darstellung der Mittelwerte



Legende zur Grafik 5

Abszisse: Anzahl der mittleren genannten Intrusionen

Ordinate: Medium der dargebotenen Information

0 ... Stimulus Text

1 ... Stimulus Text mit Bild

2 ... Stimulus Film

Ergebnistabelle 5 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Falsche Erinnerungen								
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
0	32	1,19	1,360	,262	,65	1,72	0	6
1	30	1,43	1,287	,281	,84	2,01	0	5
2	31	,33	,488	,126	,06	,60	0	1
Gesamt	93	1,06	1,243	,157	,75	1,38	0	6

Legende zu Ergebnistabelle 5

0 ... Stimulusvariation Text

1 ... Stimulusvariation Text mit Bild

2 ... Stimulusvariation Film

N... Gruppengröße

Statistiktable 5 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Stimulus Text und Film hinsichtlich der genannten Intrusionen

ANOVA

Falsche Erinnerungen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6,997	1	6,997	5,445	,025
Within Groups	51,407	40	1,285		
Total	58,405	41			

Gruppenvergleich zwischen Stimulus Text und Text mit Bild hinsichtlich der genannten Intrusionen

ANOVA

Falsche Erinnerungen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,700	1	,700	,396	,532
Within Groups	81,217	46	1,766		
Total	81,917	47			

Gruppenvergleich zwischen Stimulus Text mit Bild und Film hinsichtlich der genannten Intrusionen

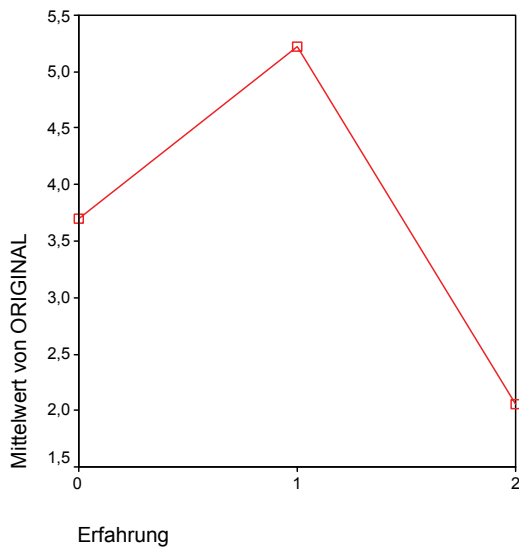
ANOVA

Falsche Erinnerungen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10,496	1	10,496	9,784	,004
Within Groups	36,476	34	1,073		
Total	46,972	35			

6. Darstellung der Mittelwerte und Standardabweichungen der reproduzierten Originalinformationen pro Versuchsperson bei Variation der Beurteilererfahrung

Grafik 6 - Darstellung der Mittelwerte



Legende zur Grafik 6

Abszisse: Anzahl der mittleren reproduzierten Originalinformationen

Ordinate: Kontrastgruppe mit variiertes Beurteilererfahrung

0 ... Novizen

1 ... Trainierte

2 ... Experten

Ergebnistabelle 6 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

ORIGINAL									
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum	
					Untergrenze	Obergrenze			
0	30	3,69	4,352	,854	1,93	5,45	0	18	
1	32	5,22	5,826	1,373	2,32	8,12	0	17	
2	31	2,05	2,483	,570	,86	3,25	0	9	
Gesamt	93	3,63	4,498	,567	2,50	4,77	0	18	

Legende zu Ergebnistabelle 6

0 ... Novizen

1 ... Trainierte

2 ... Experten

N... Gruppengröße

Statistiktable 6 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Novizen und Experten hinsichtlich der reproduzierten Originalinformationen

ANOVA

ORIGINAL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	29,514	1	29,514	2,171	,148
Within Groups	584,486	43	13,593		
Total	614,000	44			

Gruppenvergleich zwischen Novizen und Trainierten hinsichtlich der reproduzierten Originalinformationen

ANOVA

ORIGINAL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	24,896	1	24,896	,995	,324
Within Groups	1050,650	42	25,015		
Total	1075,545	43			

Gruppenvergleich zwischen Trainierten und Experten hinsichtlich der reproduzierten Originalinformationen

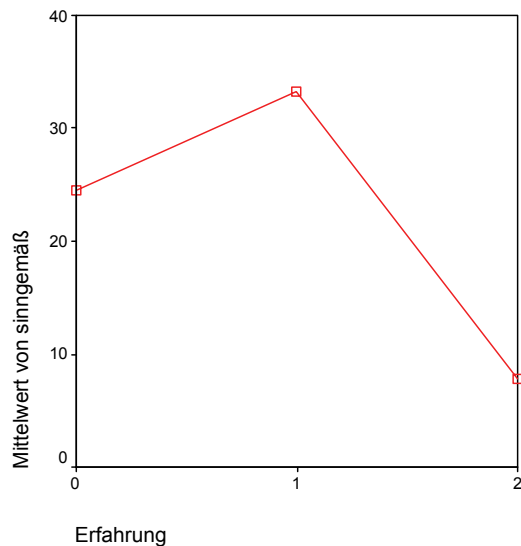
ANOVA

ORIGINAL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	92,860	1	92,860	4,724	,037
Within Groups	688,058	35	19,659		
Total	780,919	36			

7. Darstellung der Mittelwerte und Standardabweichungen der sinngemäß reproduzierten Informationen pro Versuchsperson bei Variation der Beurteilererfahrung

Grafik 7 - Darstellung der Mittelwerte



Legende zur Grafik 7

Abszisse: Anzahl der mittleren sinngemäß reproduzierten Informationen

Ordinate: Kontrastgruppe mit variiertem Beurteilererfahrung

0 ... Novizen

1 ... Trainierte

2 ... Experten

Ergebnistabelle 7 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

sinngemäß									
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum	
					Untergrenze	Obergrenze			
0	30	24,46	13,915	2,729	18,84	30,08	2	53	
1	32	33,28	11,955	2,818	27,33	39,22	13	51	
2	31	7,79	4,860	1,115	5,45	10,13	3	19	
Gesamt	93	21,95	15,016	1,892	18,17	25,73	2	53	

Legende zu Ergebnistabelle 7

0 ... Novizen

1 ... Trainierte

2 ... Experten

N... Gruppengröße

Statistiktable 7 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Novizen und Experten hinsichtlich der sinngemäß reproduzierten Informationen

ANOVA

sinngemäß

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3051,358	1	3051,358	24,918	,000
Within Groups	5265,619	43	122,456		
Total	8316,978	44			

Gruppenvergleich zwischen Novizen und Trainierten hinsichtlich der sinngemäß reproduzierten Informationen

ANOVA

sinngemäß

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	826,723	1	826,723	4,776	,034
Within Groups	7270,073	42	173,097		
Total	8096,795	43			

Gruppenvergleich zwischen Trainierten und Experten hinsichtlich der sinngemäß reproduzierten Informationen

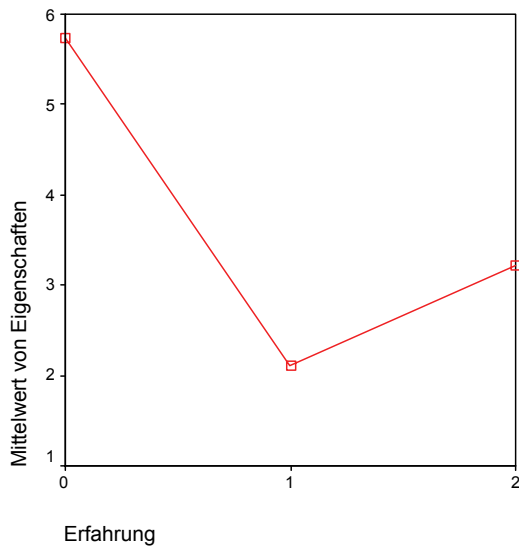
ANOVA

sinngemäß

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6004,907	1	6004,907	73,621	,000
Within Groups	2854,769	35	81,565		
Total	8859,676	36			

8. Darstellung der Mittelwerte und Standardabweichungen der aus den Informationen generierten Eigenschaften pro Versuchsperson bei Variation der Beurteilererfahrung

Grafik 8 – Darstellung der Mittelwerte



Legende zur Grafik 8

Abszisse: Anzahl der mittleren aus den dargebotenen Informationen generierten Eigenschaften

Ordinate: Kontrastgruppe mit variiertem Beurteilererfahrung

0 ... Novizen

1 ... Trainierte

2 ... Experten

Ergebnistabelle 8 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Eigenschaften								
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
0	30	5,73	7,214	1,415	2,82	8,64	0	31
1	32	2,11	3,740	,882	,25	3,97	0	12
2	31	3,21	2,507	,575	2,00	4,42	0	8
Gesamt	93	3,94	5,397	,680	2,58	5,30	0	31

Legende zu Ergebnistabelle 8

0 ... Novizen

1 ... Trainierte

2 ... Experten

N... Gruppengröße

Statistiktable 8 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Novizen und Experten hinsichtlich der aus den dargebotenen Informationen generierten Eigenschaften

ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	69,727	1	69,727	2,120	,153
Within Groups	1414,273	43	32,890		
Total	1484,000	44			

Gruppenvergleich zwischen Novizen und Trainierten hinsichtlich der aus den dargebotenen Informationen generieren Eigenschaften

ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	139,357	1	139,357	3,803	,058
Within Groups	1538,893	42	36,640		
Total	1678,250	43			

Gruppenvergleich zwischen Trainierten und Experten hinsichtlich der aus den dargebotenen Informationen generierten Eigenschaften

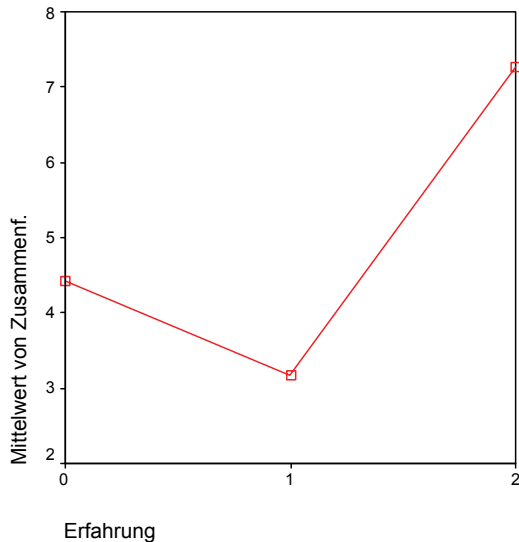
ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11,172	1	11,172	1,114	,298
Within Groups	350,936	35	10,027		
Total	362,108	36			

9. Darstellung der Mittelwerte und Standardabweichungen der reproduzierten zusammengefassten Informationen pro Versuchsperson bei Variation der Beurteilererfahrung

Grafik 9 – Darstellung der Mittelwerte



Legende zur Grafik 9

Abszisse: Anzahl der mittleren reproduzierten zusammengefassten Informationen

Ordinate: Kontrastgruppe mit variiertem Beurteilererfahrung

0 ... Novizen

1 ... Trainierte

2 ... Experten

Ergebnistabelle 9 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Zusammenf.								
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
0	30	4,42	3,828	,751	2,88	5,97	0	12
1	32	3,17	3,276	,772	1,54	4,80	0	12
2	31	7,26	3,142	,721	5,75	8,78	2	13
Gesamt	93	4,92	3,794	,478	3,97	5,88	0	13

Legende zu Ergebnistabelle 9

0 ... Novizen

1 ... Trainierte

2 ... Experten

N... Gruppengröße

Statistiktable 9 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Novizen und Experten hinsichtlich der reproduzierten zusammengefassten Informationen

ANOVA

Zusammenf.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	88,547	1	88,547	6,999	,011
Within Groups	544,030	43	12,652		
Total	632,578	44			

Gruppenvergleich zwischen Novizen und Trainierten hinsichtlich der reproduzierten zusammengefassten Informationen

ANOVA

Zusammenf.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	16,790	1	16,790	1,285	,263
Within Groups	548,846	42	13,068		
Total	565,636	43			

Gruppenvergleich zwischen Trainierten und Experten hinsichtlich der reproduzierten zusammengefassten Informationen

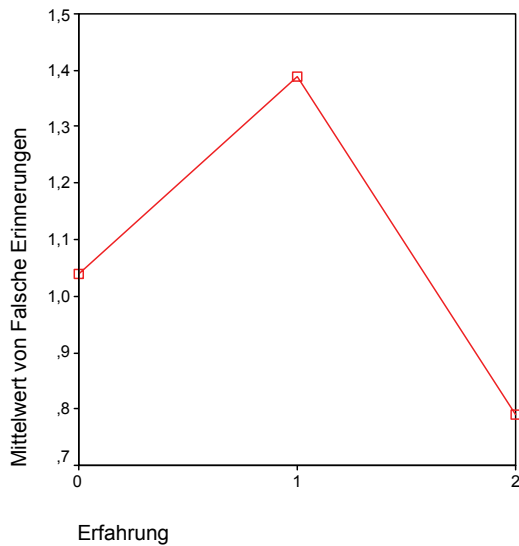
ANOVA

Zusammenf.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	155,113	1	155,113	15,073	,000
Within Groups	360,184	35	10,291		
Total	515,297	36			

10. Darstellung der Mittelwerte und Standardabweichungen der genannten Intrusionen pro Versuchsperson bei Variation der Beurteilererfahrung

Grafik 10 – Darstellung der Mittelwerte



Legende zur Grafik 10

Abszisse: Anzahl der mittleren genannten Intrusionen
 Ordinate: Kontrastgruppe mit variiertem Beurteilererfahrung
 0 ... Novizen
 1 ... Trainierte
 2 ... Experten

Ergebnistabelle 10 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Falsche Erinnerungen									
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum	
					Untergrenze	Obergrenze			
0	30	1,04	,958	,188	,65	1,43	0	3	
1	32	1,39	1,685	,397	,55	2,23	0	6	
2	31	,79	1,084	,249	,27	1,31	0	4	
Gesamt	93	1,06	1,243	,157	,75	1,38	0	6	

Legende zu Ergebnistabelle 10

0 ... Novizen
 1 ... Trainierte
 2 ... Experten
 N... Gruppengröße

Statistiktable 10 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Novizen und Experten hinsichtlich der genannten Intrusionen

ANOVA

Falsche Erinnerungen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,681	1	,681	,663	,420
Within Groups	44,119	43	1,026		
Total	44,800	44			

Gruppenvergleich zwischen Novizen und Trainierten hinsichtlich der genannten Intrusionen

ANOVA

Falsche Erinnerungen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1,306	1	1,306	,770	,385
Within Groups	71,239	42	1,696		
Total	72,545	43			

Gruppenvergleich zwischen Trainierten und Experten hinsichtlich der genannten Intrusionen

ANOVA

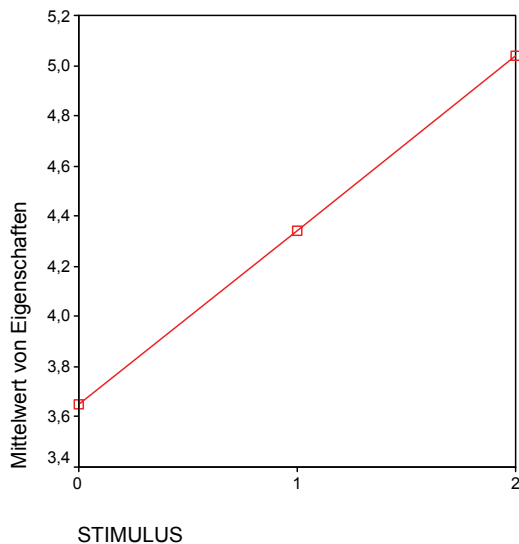
Falsche Erinnerungen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3,321	1	3,321	1,674	,204
Within Groups	69,436	35	1,984		
Total	72,757	36			

Darstellung der Ergebnisse zur Beurteilung der Geschworenen

11. Darstellung der Mittelwerte und Standardabweichungen der zur Beurteilung der Geschworenen genannten Eigenschaften pro Versuchsperson bei Variation der Stimulusinformation

Grafik 11 - Darstellung der Mittelwerte



Legende zur Grafik 11

Abszisse: Anzahl der mittleren genannten Eigenschaften pro zu beurteilende Person

Ordinate: Medium der dargebotenen Information

0 ... Stimulus Text

1 ... Stimulus Text mit Bild

2 ... Stimulus Film

Ergebnistabelle 11 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Eigenschaften								
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
0	160	3,65	2,224	,191	3,27	4,03	0	10
1	150	4,34	2,756	,269	3,81	4,88	0	13
2	155	5,04	3,600	,416	4,21	5,87	0	18
Gesamt	465	4,21	2,825	,159	3,90	4,53	0	18

Legende zu Ergebnistabelle 11

0 ... Stimulusvariation Text

1 ... Stimulusvariation Text mit Bild

2 ... Stimulusvariation Film

N... Gruppengröße

Statistiktable 11 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Stimulus Text und Film hinsichtlich der genannten Eigenschaften

ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	92,907	1	92,907	11,918	,001
Within Groups	1621,517	208	7,796		
Total	1714,424	209			

Gruppenvergleich zwischen Stimulus Text und Text mit Bild hinsichtlich der genannten Eigenschaften

ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	28,202	1	28,202	4,622	,033
Within Groups	1452,294	238	6,102		
Total	1480,496	239			

Gruppenvergleich zwischen Stimulus Text mit Bild und Film hinsichtlich der genannten Eigenschaften

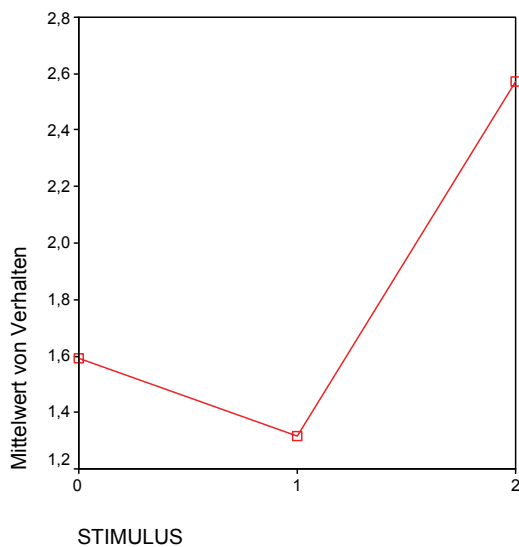
ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	21,263	1	21,263	2,165	,143
Within Groups	1748,537	178	9,823		
Total	1769,800	179			

12. Darstellung der Mittelwerte und Standardabweichungen der zur Beurteilung der Geschworenen genannten Verhaltensweisen pro Versuchsperson bei Variation der Stimulusinformation

Grafik 12 - Darstellung der Mittelwerte



Legende zur Grafik 12

Abszisse: Anzahl der mittleren genannten Verhaltensweisen pro zu beurteilende Person

Ordinate: Medium der dargebotenen Information

0 ... Stimulus Text

1 ... Stimulus Text mit Bild

2 ... Stimulus Film

Ergebnistabelle 12 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Eigenschaften									
	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum	
					Untergrenze	Obergrenze			
0	160	3,65	2,224	,191	3,27	4,03	0	10	
1	150	4,34	2,756	,269	3,81	4,88	0	13	
2	155	5,04	3,600	,416	4,21	5,87	0	18	
Gesamt	465	4,21	2,825	,159	3,90	4,53	0	18	

Legende zu Ergebnistabelle 12

0 ... Stimulusvariation Text

1 ... Stimulusvariation Text mit Bild

2 ... Stimulusvariation Film

N... Gruppengröße

Statistiktable 12 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Stimulus Text und Film hinsichtlich der genannten Verhaltensweisen

ANOVA

Verhalten

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	46,375	1	46,375	15,003	,000
Within Groups	642,939	208	3,091		
Total	689,314	209			

Gruppenvergleich zwischen Stimulus Text und Text mit Bild hinsichtlich der genannten Verhaltensweisen

ANOVA

Verhalten

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4,575	1	4,575	1,940	,165
Within Groups	561,221	238	2,358		
Total	565,796	239			

Gruppenvergleich zwischen Stimulus Text mit Bild und Film hinsichtlich der genannten Verhaltensweisen

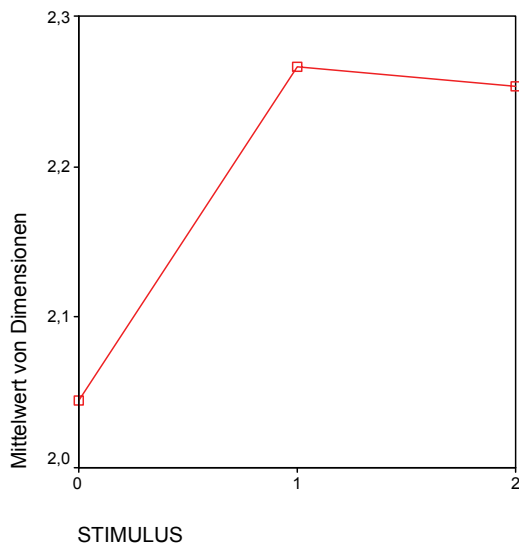
ANOVA

Verhalten

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	69,353	1	69,353	22,405	,000
Within Groups	550,975	178	3,095		
Total	620,328	179			

13. Darstellung der Mittelwerte und Standardabweichungen der genannten Eigenschaftsdimensionen pro Versuchsperson bei Variation der Stimulusinformation

Grafik 13 - Darstellung der Mittelwerte



Legende zur Grafik 13

Abszisse: Anzahl der mittleren genannten Eigenschaftsdimensionen pro zu beurteilende Person

Ordinate: Medium der dargebotenen Information

0 ... Stimulus Text

1 ... Stimulus Text mit Bild

2 ... Stimulus Film

Ergebnistabelle 13 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Dimensionen	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
					0	160		
1	150	2,27	1,031	,101	2,07	2,47	0	5
2	155	2,25	1,264	,146	1,96	2,54	0	5
Gesamt	465	2,17	1,100	,062	2,05	2,29	0	5

Legende zu Ergebnistabelle 13

0 ... Stimulusvariation

Text

1 ... Stimulusvariation

Text mit Bild

2 ... Stimulusvariation

Film

N... Gruppengröße

Statistiktable 13 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Stimulus Text und Film hinsichtlich der genannten Eigenschaftsdimensionen

ANOVA

Dimensionen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2,104	1	2,104	1,646	,201
Within Groups	265,920	208	1,278		
Total	268,024	209			

Gruppenvergleich zwischen Stimulus Text und Text mit Bild hinsichtlich der genannten Eigenschaftsdimensionen

ANOVA

Dimensionen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2,917	1	2,917	2,688	,102
Within Groups	258,267	238	1,085		
Total	261,183	239			

Gruppenvergleich zwischen Stimulus Text mit Bild und Film hinsichtlich der genannten Eigenschaftsdimensionen

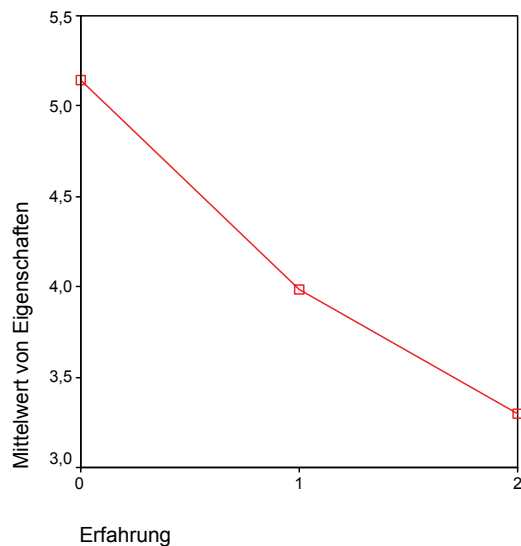
ANOVA

Dimensionen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7,778E-03	1	7,778E-03	,006	,938
Within Groups	228,720	178	1,285		
Total	228,728	179			

14. Darstellung der Mittelwerte und Standardabweichungen der zur Beurteilung der Geschworenen genannten Eigenschaften pro Versuchsperson bei Variation der Beurteilererfahrung

Grafik 14 - Darstellung der Mittelwerte



Legende zur Grafik 14

Abszisse: Anzahl der mittleren genannten Eigenschaften
 Ordinate: Kontrastgruppe mit variiertem Beurteilererfahrung
 0 ... Novizen
 1 ... Trainierte
 2 ... Experten

Ergebnistabelle 14 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Eigenschaften		N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
						Untergrenze	Obergrenze		
0		150	5,15	3,262	,286	4,58	5,71	0	18
1		160	3,99	1,960	,234	3,52	4,45	1	9
2		155	3,30	2,399	,224	2,85	3,74	0	9
Gesamt		465	4,21	2,825	,159	3,90	4,53	0	18

Legende zu Ergebnistabelle 14

0 ... Novizen
 1 ... Trainierte
 2 ... Experten
 N... Gruppengröße

Statistiktable 14 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Novizen und Experten hinsichtlich der genannten Eigenschaften

ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	208,956	1	208,956	25,035	,000
Within Groups	2028,171	243	8,346		
Total	2237,127	244			

Gruppenvergleich zwischen Novizen und Trainierten hinsichtlich der genannten Eigenschaften

ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	61,271	1	61,271	7,410	,007
Within Groups	1637,209	198	8,269		
Total	1698,480	199			

Gruppenvergleich zwischen Trainierten und Experten hinsichtlich der genannten Eigenschaften

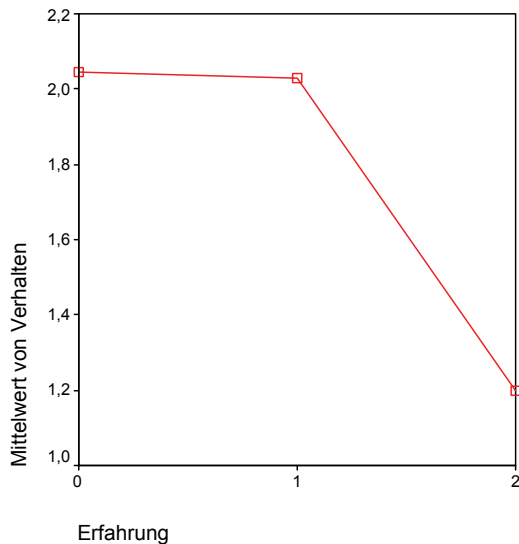
ANOVA

Eigenschaften

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	20,721	1	20,721	4,117	,044
Within Groups	920,934	183	5,032		
Total	941,654	184			

15. Darstellung der Mittelwerte und Standardabweichungen der zur Beurteilung der Geschworenen genannten Verhaltensweisen pro Versuchsperson bei Variation der Beurteilererfahrung

Grafik 15 - Darstellung der Mittelwerte



Legende zur Grafik 15

Abszisse: Anzahl der mittleren genannten Verhaltensweisen

Ordinate: Kontrastgruppe mit variiertem Beurteilererfahrung

0 ... Novizen

1 ... Trainierte

2 ... Experten

Ergebnistabelle 15 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Verhalten	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
					0	150		
1	160	2,03	1,865	,223	1,58	2,47	0	8
2	155	1,20	1,639	,153	,90	1,50	0	8
Gesamt	465	1,73	1,741	,098	1,54	1,93	0	8

Legende zu Ergebnistabelle 15

0 ... Novizen

1 ... Trainierte

2 ... Experten

N... Gruppengröße

Statistiktable 15 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Novizen und Experten hinsichtlich der genannten Verhaltensweisen

ANOVA

Verhalten

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	43,689	1	43,689	16,083	,000
Within Groups	660,123	243	2,717		
Total	703,812	244			

Gruppenvergleich zwischen Novizen und Trainierten hinsichtlich der genannten Verhaltensweisen

ANOVA

Verhalten

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1,407E-02	1	1,407E-02	,005	,945
Within Groups	593,666	198	2,998		
Total	593,680	199			

Gruppenvergleich zwischen Trainierten und Experten hinsichtlich der genannten Verhaltensweisen

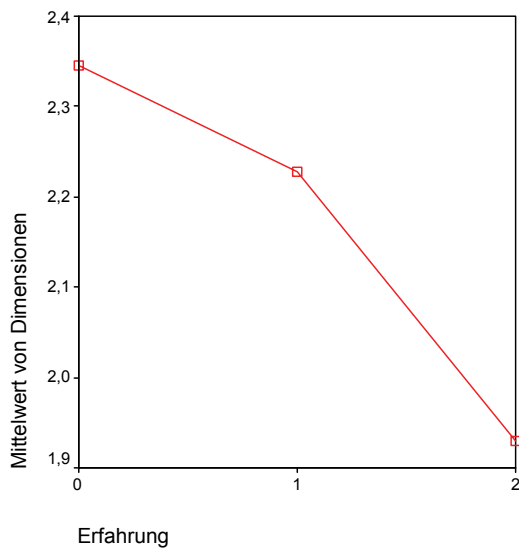
ANOVA

Verhalten

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	29,873	1	29,873	10,006	,002
Within Groups	546,343	183	2,985		
Total	576,216	184			

16. Darstellung der Mittelwerte und Standardabweichungen der genannten Eigenschaftsdimensionen pro Versuchsperson bei Variation der Beurteilererfahrung

Grafik 16 - Darstellung der Mittelwerte



Legende zur Grafik 16

Abszisse: Anzahl der mittleren genannten Eigenschaftsdimensionen pro zu beurteilende Person

Ordinate: Kontrastgruppe mit variiertem Beurteilererfahrung

0 ... Novizen

1 ... Trainierte

2 ... Experten

Ergebnistabelle 16 - Deskriptive Statistik der Mittelwerte und Standardabweichungen

ONEWAY deskriptive Statistiken

Verhalten	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
					0	150		
1	160	2,03	1,865	,223	1,58	2,47	0	8
2	155	1,20	1,639	,153	,90	1,50	0	8
Gesamt	465	1,73	1,741	,098	1,54	1,93	0	8

Legende zu Ergebnistabelle 16

0 ... Novizen

1 ... Trainierte

2 ... Experten

N... Gruppengröße

Statistiktable 16 - Inferenzstatistische Auswertung

Gruppenvergleich zwischen Novizen und Experten hinsichtlich der genannten Eigenschaftsdimensionen

ANOVA

Dimensionen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10,546	1	10,546	8,139	,005
Within Groups	314,867	243	1,296		
Total	325,412	244			

Gruppenvergleich zwischen Novizen und Trainierten hinsichtlich der genannten Eigenschaftsdimensionen

ANOVA

Dimensionen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,629	1	,629	,605	,437
Within Groups	205,766	198	1,039		
Total	206,395	199			

Gruppenvergleich zwischen Trainierten und Experten hinsichtlich der genannten Eigenschaftsdimensionen

ANOVA

Dimensionen

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3,868	1	3,868	3,250	,073
Within Groups	217,786	183	1,190		
Total	221,654	184			

Ehrenwörtliche Erklärung

Ich erkläre hiermit,

dass mir die geltende Promotionsordnung bekannt ist;

dass ich die vorliegende Dissertation selbständig angefertigt habe, insbesondere die Hilfe eines Promotionsberaters nicht in Anspruch genommen habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe;

dass Frau Karen Möhrle und Frau Sylke Schönburg bei der inhaltsanalytischen Auswertung der Rohdaten unentgeltlich mitgewirkt haben;

dass darüber hinaus Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen;

dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe; und

dass ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule oder anderen Fakultät als Dissertation eingereicht habe.

Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Leipzig, 02.04.2006