

Laienpsychologische Persönlichkeitsbeurteilung:
Übereinstimmung von Fremd- und Selbsteinschätzungen
am Beispiel eines Depressionsinventars

Diplomarbeit

vorgelegt von

Andreas Hübner

an der Universität Koblenz-Landau

Fachbereich 8 – Psychologie

am 20. Juli 2006

Betreuer und Erstgutachter:

Prof. Dr. Manfred Schmitt

Zweitgutachterin:

Dr. Christine Altstötter-Gleich

Inhaltsverzeichnis

Zusammenfassung	6
1. Einleitung	7
2. Laienpsychologische Persönlichkeitsbeurteilung	10
2.1 Judgment	10
2.1.1 Allgegenwart und Wichtigkeit der Persönlichkeitsbeurteilung	10
2.1.2 Accuracy: Wie kann Urteilspräzision sinnvoll bestimmt werden?.....	12
2.1.3 Agreement: Wie kann Urteilsübereinstimmung quantifiziert werden?	17
2.1.4 Gibt es präzise Persönlichkeitsbeurteilungen und wovon hängen sie ab?	21
2.1.5 Kurzer Vergleich von Selbst- und Fremdbeurteilungen	27
2.2 Judgmental Process	33
2.2.1 Das Realistic Accuracy Model (RAM) und seine Teilprozesse	33
2.2.1.1 Relevance: Bedeutsamkeit schaffen.....	34
2.2.1.2 Availability: Zugang erhalten.....	35
2.2.1.3 Detection: Hinweise aufnehmen	35
2.2.1.4 Utilization: Hinweise nutzen	36
2.2.2 Theoretische Verortung von RAM.....	36
2.2.3 Fazit.....	37
2.3 Moderators of judgmental Accuracy	38
2.3.1 Taxonomie nach RAM	38
2.3.2 The good Judge: Merkmale des Beurteilers.....	38
2.3.2.1 Theoretische Erwägungen	38
2.3.2.2 Empirische Ergebnisse.....	40
2.3.3 The good Target: Merkmale des Beurteilten	41
2.3.3.1 Theoretische Erwägungen	41
2.3.3.2 Empirische Ergebnisse.....	42
2.3.4 Good Information: Merkmale der geteilten Hinweisbasis	44
2.3.4.1 Theoretische Erwägungen	44
2.3.4.2 Empirische Ergebnisse.....	44
2.3.5 The good Trait: Merkmale der zu beurteilenden Eigenschaft	47
2.3.5.1 Theoretische Erwägungen	47
2.3.5.2 Empirische Ergebnisse.....	47
3. Fragestellung und Hypothesen	49
3.1 Herleitung der bearbeiteten Fragestellung	49
3.1.1 Judge: Offenheit, Empathie, Perspektivübernahmetendenz, Alter	50
3.1.2 Target: Privatheit, Konsistenz, Selbst- und Fremdtäuschungstendenz.....	51
3.1.3 Information: Acquaintanceship	52
3.1.4 Trait: Beobachtbarkeit	52
3.2. Hypothesen.....	53

4.	Methodenteil	55
4.1	Verwendete Erhebungsinstrumente.....	55
4.1.1	Vergleichsskala: Das BDI-V	55
4.1.2	Empathie / Perspektivübernahme: Skala von Maes, Schmitt und Schmal	58
4.1.3	Offenheit des Wertesystems: Skala O6 des NEO-PI-R	59
4.1.4	Soziale Erwünschtheit: Balanced Inventory of Desirable Responding	60
4.1.5	Privatheit: Skala N der deutschen Ausgabe des 16 PF-R.....	62
4.1.6	Acquaintanceship: Aspekte des Bekanntschaftsgrades und Umgangs	63
4.1.7	Beobachtbarkeit: gemittelte Augenscheinvalidität	64
4.2	Durchführung der Datenerhebung	65
4.3	Stichprobenbeschreibung	65
4.4	Auswertungsstrategien.....	66
4.4.1	Zwei Ebenen der Analyse: Personen-Level und Item-Level	66
4.4.2	Moderatorvariablenansatz	67
4.4.3	Operationalisierung der Urteilsübereinstimmung	69
4.4.3.1	Zur Problematik einfacher Globalmaße.....	69
4.4.3.2	Auswahl spezifischer Maße.....	71
4.4.4	Zusammenfassung Auswertungsansatz	74
5.	Ergebnisteil	75
5.1	Meßgüte des BDI-V.....	75
5.2	Allgemeine Urteilsübereinstimmung (Hypothese 1)	76
5.2.1	Personen-Level Ergebnisse.....	76
5.2.2	Item-Level Ergebnisse	77
5.2.3	Fazit allgemeine Urteilsübereinstimmung	78
5.3	Judge: Moderatoren der Fremdbeurteiler (Hypothesen 2-5).....	79
5.3.1	Personen-Level Ergebnisse.....	79
5.3.2	Item-Level Ergebnisse	80
5.4	Target: Moderatoren der Selbstbeurteiler (Hypothesen 6-9)	82
5.4.1	Personen-Level Ergebnisse.....	82
5.4.2	Item-Level Ergebnisse	84
5.5	Information: Moderator Acquaintanceship (Hypothesen 10/11).....	86
5.5.1	Bekannschaftsdauer der Urteilerpaare (quantitativer Aspekt)	86
5.5.2	Treffensfrequenz der Urteilerpaare (quantitativer Aspekt)	87
5.5.3	Bekanntschaftsgrad der Urteilerpaare (qualitativer Aspekt)	87
5.5.4	Liierte vs. nicht-liierte Urteilerpaare (qualitativer Aspekt)	89
5.5.5	Fazit Acquaintanceship	90
5.6	Trait: Moderator Beobachtbarkeit (Hypothese 12).....	91
5.6.1	Erfassung der Observability.....	91
5.6.2	Zusammenhang mit Urteilsübereinstimmung	92

6.	Diskussion	93
6.1	Zusammenfassung und Interpretation der Ergebnisse	93
6.2	Kritische Bewertung, Grenzen vorliegender Arbeit	97
6.3	Ausblick	102
7.	Literatur	104
8.	Anhang	112
Anhang A:	Hauptinstruktion Fragebögen	113
Anhang B:	Fragebögen Judge.....	114
	Umschlag, Instruktion Judge.....	114
	BDI-V (Fremdeinschätzungsform).....	115
	Empathie / Perspektivübernahme	116
	Offenheit des Wertesystems	117
	Bekanntschaftsgrad, Demographie	118
Anhang C:	Fragebögen Target.....	119
	Umschlag, Instruktion Target.....	119
	BDI-V (Selbsteinschätzungsform)	120
	Selbst- und Fremdtäuschung.....	121
	Privatheit	122
	Bekanntschaftsgrad, Demographie	123
Anhang D:	Screenshots Onlinefragebogen Beobachtbarkeit	124
Anhang E:	Übereinstimmungskoeffizienten.....	130
Anhang F:	Weitere Ergebnistabellen	131

Abbildungsverzeichnis

Abbildung 1:	Das Johari-Fenster der verschiedenen Bereiche der Persönlichkeit	28
Abbildung 2:	RAM und die getrennten Domänen der Persönlichkeits- und Sozialpsychologie..	37

Tabellenverzeichnis

Tabellen im Hauptteil der Arbeit:

Tabelle 1:	Voraussetzungen eines gelingenden Urteilsprozesses laut RAM.....	33
Tabelle 2:	Felder von Moderatoren der Urteilsgüte und assoziierte RAM-Kernprozesse	38
Tabelle 3:	Wortlaut der Items des BDI-V	56
Tabelle 4:	Interne Konsistenzen der BDI-V Selbst- und Fremdeinschätzungsskalen.....	75
Tabelle 5:	Trennschärfen und Reliabilitäten der Items des BDI-V (Selbst- und Fremdeinschätzungen)	75
Tabelle 6:	Verteilungswerte der dyadischen Übereinstimmungsmaße	76
Tabelle 7:	Korrelationen der dyadischen Übereinstimmungsmaße untereinander	76

Tabelle 8:	Korrelationen der Selbst- und Fremdeinschätzungen aller BDI-V Items.....	77
Tabelle 9:	Verteilungswerte der potentiellen Moderatoren seitens des Judges	79
Tabelle 10:	Korrelationen der potentiellen Moderatoren seitens des Judges mit den Kennwerten der dyadischen Urteilsübereinstimmung	79
Tabelle 11:	Ergebnisse der Regressionsanalysen bzgl. der Moderatoren des Judges: Zusätzliche Varianzaufklärung, i.e. Veränderung von R^2 durch Eingabe des jeweiligen Interaktionsterms	81
Tabelle 12:	Verteilungswerte der potentiellen Moderatoren seitens des Targets	82
Tabelle 13:	Korrelationen der potentiellen Moderatoren seitens des Targets mit den Kennwerten der dyadischen Urteilsübereinstimmung	82
Tabelle 14:	Abweichung der BDI-V Selbst- und Fremdeinschätzungen (Gesamtscore).....	83
Tabelle 15:	Korrelation und Semipartialkorrelation der Selbsttäuschungstendenz mit der absoluten Abweichung der BDI-V Selbst- und Fremdeinschätzungen.....	84
Tabelle 16:	Ergebnisse der Regressionsanalysen bzgl. der Moderatoren des Targets: Zusätzliche Varianzaufklärung, i.e. Veränderung von R^2 durch Eingabe des jeweiligen Interaktionsterms	85
Tabelle 17:	Korrelationen der kontinuierlichen Variable Bekanntschaftsdauer mit den Kennwerten der dyadischen Urteilsübereinstimmung	86
Tabelle 18:	Schema der Erzeugung eines Gesamtwertes des Bekanntschaftsgrades durch die Kombination der beiden Einzelangaben von Judge und Target	87
Tabelle 19:	Häufigkeitswerte der kategorialen Variable Bekanntschaftsgrad.....	88
Tabelle 20:	Korrelationen der kategorialen Variable Bekanntschaftsgrad mit den Kennwerten der dyadischen Urteilsübereinstimmung	88
Tabelle 21:	Korrelationen der Selbst- und Fremdeinschätzungen aller BDI-V-Items (Rohwerte, personenzentrierte Abweichungswerte), Subgruppen liiert ja/nein ..	89
Tabelle 22:	Absteigend angeordnete Observability-Werte aller BDI-V Items: Gesamtwert und Normalform, Abweichung der positiv umgepolten Form, Korrelation zwischen beiden Formen	92
Tabelle 23:	Moderatorwirkungen auf Personen-Level: Signifikante Korrelationen der potentiellen Moderatoren mit den Maßen der dyadischen Urteilsübereinstimmung.....	94

Tabellen im Anhang F:

Tabelle F1:	Semipartialkorrelationen (ohne Bekanntschaftsgrad) der potentiellen Moderatoren seitens des Judges mit den Kennwerten der dyadischen Urteilsübereinstimmung	131
Tabelle F2:	Semipartialkorrelationen (ohne Bekanntschaftsgrad) der potentiellen Moderatoren seitens des Targets mit den Kennwerten der dyadischen Urteilsübereinstimmung	131
Tabelle F3:	Mittelwerte der dyadischen Urteilsübereinstimmung, Subgruppen liierte und nicht-liierte Urteilerpaare	131
Tabelle F4:	Mittelwerte der dyadischen Urteilsübereinstimmung, Subgruppen Bekanntschaftsdauer 0 bis 1, 1 bis 6, 6 bis 21 und mehr als 21 Jahre	132
Tabelle F5:	Fisher-z-Werte der Selbst-Fremd-Korrelationen aller BDI-V-Items, Subgruppen Bekanntschaftsdauer 0 bis 1, 1 bis 6, 6 bis 21 und mehr als 21 Jahre	132
Tabelle F6:	Korrelationen (Spearman's Rho) der kategorialen Variable Treffensfrequenz mit den Kennwerten der dyadischen Urteilsübereinstimmung	133
Tabelle F7:	Absteigend angeordnete Agreement-Werte aller BDI-V-Items und deren Observability: Fisher-z-Werte, Observability-Scores und jeweilige Rangplätze ..	133

Zusammenfassung

Die vorliegende Studie dient der weiteren Validierung des BDI-V, der neuen Kurzfassung des Beck-Depressions-Inventars (SCHMITT & MAES, 2000). Dazu wurden an einer heterogenen Gelegenheitsstichprobe paarweise Selbst- und Fremdeinschätzungen auf dem BDI-V erhoben und anschließend auf Personen- und Item-Level miteinander verglichen. Die Konvergenz der Urteile war befriedigend, i.e. die Selbstbericht-Fremdbericht-Korrelationen erreichten eine für Persönlichkeitsskalen typische Höhe (Personen-Level: mittlere dyadische Profilkorrelation über alle 20 Items von $r=.46$, Item-Level: Korrelation der aggregierten BDI-V Gesamtscores über alle 177 Urteilerpaare von $r=.52$, mittlere Korrelation der BDI-V Items über alle 177 Urteilerpaare von $r=.36$). Das Hauptanliegen der Untersuchung wurde damit erreicht, die Angemessenheit des BDI-V zur Erfassung depressiver Zustände konnte erneut bestätigt werden.

Darüber hinaus wurde der Versuch unternommen, verschiedene Moderatorvariablen empirisch zu belegen, die mit dem urteiler- bzw. itemspezifischen Übereinstimmungsgrad zusammenhängen sollten. Die zu testenden Moderatoren wurden aus dem Realistic Accuracy Model von David C. Funder abgeleitet, das den Prozeß der laienpsychologischen Persönlichkeitsbeurteilung beschreibt und gut zur Strukturierung des Gegenstandsbereichs geeignet ist (FUNDER, 1995, 1999).

Der empirische Nachweis eines mittleren Einflusses des Bekanntschaftsgrades der Selbst- und Fremdbeurteiler ($r=.29$) und eines schwächeren Einflusses bewußter und unbewußter sozialer Erwünschtheitstendenzen der Selbstbeurteiler (jeweils $r=.16$) auf die dyadische Urteilsübereinstimmung gelang. Die anderen erfaßten personenbezogenen Variablen (Empathie, Perspektivübernahme und Offenheit der Fremdeinschätzer, Privatheit und Konsistenz der Selbsteinschätzer) zeigten dagegen in der Analyse keine erwartungsgemäße Wirkung. In einer zusätzlichen Erhebung wurde die Beobachtbarkeit der einzelnen depressiven Verhaltens- bzw. Erlebensweisen bewertet, die in den Items des BDI-V thematisiert werden. Es konnte ein moderater Zusammenhang zwischen dieser unabhängig eingeschätzten Fremdbeobachtbarkeit der Items und ihrer tatsächlichen itemspezifischen Übereinstimmung belegt werden ($r=.20$).

Die Untersuchungsfragen werden in den breiteren Rahmen der Debatte über die Qualität menschlicher Urteilsgüte gestellt. Die erzielten erwartungsgemäßen Ergebnisse werden kritisch eingeordnet, außerdem werden mögliche Gründe für die Nichtwirksamkeit der anderen erhobenen potentiellen Moderatoren diskutiert.

1. Einleitung

No person can understand any other person completely because no human being shares directly the motives, thoughts, and feelings of another. The only self to which we have immediate access is our own. Knowledge of other people comes to us directly and in fragments. At best we catch glimpses of one another. Yet we try ardently to bridge the chasm between mind and mind, for our happiness and survival depend on correct judgments of persons (ALLPORT, 1937, S.497).

Jeder normale Mensch hat Vorstellungen davon, was für Persönlichkeitseigenschaften andere Menschen haben und wie man ihre Gesamtpersönlichkeit charakterisieren kann. Man schätzt ab, wie diese anderen in bestimmten Situationen wohl reagieren werden, um seine eigenen Handlungen zumindest teilweise danach auszurichten. Jeder versucht mehr oder weniger bewußt das zu verstehen, was er selber niemals einnehmen kann: die Perspektive des anderen. Und jeder hat eigene Theorien bezüglich des Wesens und Charakters anderer Menschen.

Die wissenschaftliche Psychologie bezeichnet diese Vorgänge in Abgrenzung zu ihrer eigenen systematischen Eigenschaftsbeschreibung etwas abfällig als "laienpsychologische Persönlichkeitsbeurteilungen" (*engl.* lay personality judgments). Die Existenz solcher Beurteilungen wurde niemals bezweifelt, aber es wurden andere gewichtige Fragen aufgeworfen: Sind Persönlichkeitseinschätzungen von einer philosophischen Ebene aus betrachtet überhaupt angemessen oder führen sie epistemologisch in die Irre, weil es in Wirklichkeit gar keine stabilen Persönlichkeitseigenschaften gibt? Sind solche Beurteilungen wenigstens auf einer pragmatischen Ebene funktional? Wie kann man ihre Güte ermitteln? Und wovon hängt die Urteilsgüte dann im konkreten Einzelfall ab?

Natürlich kann die Beurteilung der Persönlichkeit anderer Menschen niemals in einem absoluten Sinne vollkommen korrekt sein, also exhaustiv, da sie immer auf notgedrungen begrenzten Eindrücken von außen beruht. Viele Menschen empfinden dennoch, daß sie zumindest in ihrem engsten Umfeld andere Personen haben, die sie recht gut kennen und recht gut einschätzen können. Aber läßt sich dieser subjektive Eindruck von Urteilsgüte auch objektiv belegen? Sind die alltäglichen Persönlichkeitsbeurteilungen des echten Lebens meistens tendenziell richtig oder prinzipiell immer verzerrt? Anders ausgedrückt: Wie gut ist die menschliche Beurteilungsfähigkeit ausgebildet?

Ausgangspunkt der vorliegenden Arbeit war die weitere Validierung einer von SCHMITT UND MAES (2000) neu entwickelten Kurzfassung eines bekannten Depressionsinventars über den Vergleich von Selbst- und Fremdeinschätzungen miteinander bekannter Laien auf diesem Instrument. Der angestrebte Nachweis einer guten durchschnittlichen Konvergenz beider Urteile soll neben der Validität und Reliabilität der Skala in einer breiteren Perspektive auch die Reliabilität der menschlichen Beurteiler belegen, i.e. eine durchschnittlich vorhandene gewisse Güte der laienpsychologischen Persönlichkeitsbeurteilung. Daneben ist es das Ziel dieser Untersuchung, bestimmte theoretisch postulierte Bedingungen von und Einflußfaktoren auf Urteilsgüte empirisch zu überprüfen, indem versucht wird, Moderatoren der Urteilsübereinstimmung zu belegen.

In der kognitiven Sozialpsychologie gab und gibt es starke Tendenzen, menschliche Urteilsfähigkeit unter den Generalverdacht der Befangenheit, Fehlerbehaftetheit, Verzerrtheit und Unangemessenheit zu stellen. Damit spiegelt sie letztlich einen Hang zur Defizitorientierung und Pathologisierung wider, der leider in der gesamten wissenschaftlichen Psychologie weit verbreitet ist (HASELTON & FUNDER, in press).

Erst in der jüngsten Zeit ist eine Gegenbewegung entstanden, die sogenannte "positive Psychologie". Statt problemzentriert zu sein, betont sie die 'guten' Aspekte des menschlichen Lebens und Miteinanders und untersucht Entwicklungs- und Coping-Möglichkeiten, menschliche Stärken und Ressourcen und bietet damit eine neue und wichtige Sichtweise (z.B. SELIGMAN & CSIKSZENTMIHALYI, 2000, SNYDER & LOPEZ, 2002). In diesem Ansatz wird dem Laien eine deutlich größere Kompetenz zugestanden und somit auch die Qualität und Güte der alltäglichen Persönlichkeitsbeurteilung optimistischer gesehen (ASPINWALL & STAUDIGER, 2003).

Und auch aus einer evolutionspsychologischen Perspektive erscheint es unlogisch, daß die Menschheit bei ihrer langen Sozialisierung auf dem Weg aus der Steinzeit in die technisierten und hochkomplex organisierten Gesellschaften der Moderne die so wichtige Sozialtechnik der gegenseitigen Einschätzung nicht adäquat erlernt haben sollte. Viele der theoretisch postulierten und im Laborversuch bestätigten Urteilsverzerrungsfaktoren der kognitiven Sozialpsychologie könnten im echten Leben durchaus funktional sein und zu sinnvollen Persönlichkeitsbeurteilungen beitragen (HASELTON & BUSS, 2003), moderne Fehlermanagement-Theorien integrieren daher bestimmte Formen der Urteilsverzerrung in ihre Kosten-Nutzen-Modelle (HASELTON & NETTLE, 2006).

Im folgenden Theorieteil wird zunächst in den Forschungsbereich Personality Judgment eingeführt, indem die Allgegenwart und Relevanz der laienpsychologischen Persönlichkeitsbeurteilung hervorgehoben wird (vgl. Abschnitt 2.1.1). Danach wird beschrieben, welche allgemeinen philosophischen Problematiken die Suche nach Kriterien der Urteilsgröße aufwirft (vgl. Abschnitt 2.1.2) und welche methodischen Problematiken bei der Operationalisierung eines dieser Kriterien, der Urteilsübereinstimmung, überwunden werden müssen (vgl. Abschnitt 2.1.3). Der Beschreibung dieser formalen Aspekte des Forschungsbereichs wird einiger Platz eingeräumt, da eine mangelhafte Methodik vieler früherer Untersuchungen ein wesentlicher Grund dafür ist, daß es in der Psychologie bis zum heutigen Tag erstaunlich wenig gesicherte Erkenntnisse über die näheren Bedingungen gelingender Beurteilungsprozesse gibt.

Der nächste Abschnitt widmet sich der langjährigen Debatte innerhalb der wissenschaftlichen Psychologie, ob Menschen überhaupt dazu in der Lage sind, zutreffende Persönlichkeitsbeurteilungen zu fällen und welche Eigenschaften einen Menschen zu einem guten Urteiler machen (vgl. Abschnitt 2.1.4). Der allgemeine Theorieteil endet mit dem Vergleich von Selbst- und Fremdbeurteilungen (vgl. Abschnitt 2.1.5).

Im speziellen Theorieteil wird dann ein neueres Modell von David C. Funder vorgestellt, das den Prozeß der Persönlichkeitsbeurteilung beschreibt (vgl. Abschnitt 2.2). Anhand der Kategorien des Modells werden theoretisch ableitbare Einflußfaktoren herausgearbeitet, die die Güte von Beurteilungsprozessen determinieren sollten, und mit bisher vorliegenden empirischen Ergebnissen kontrastiert (vgl. Abschnitt 2.3). Aus den dargestellten theoretischen Erwägungen und empirischen Resultaten werden zuletzt die konkreten Fragestellungen bzw. Hypothesen der vorliegenden Untersuchung abgeleitet (vgl. Abschnitte 3.1 und 3.2).

Im folgenden Methodenteil werden die verwendeten Befragungsinstrumente und die Durchführung der Datenerhebung beschrieben sowie die Auswertungsstrategien der Studie (vgl. Abschnitt 4), bevor die Ergebnisse in Tabellenform aufgeführt und detailliert beschrieben werden (vgl. Abschnitt 5), zusammengefaßt und interpretiert (vgl. Abschnitt 6.1) und zuletzt kritisch eingeordnet (vgl. Abschnitt 6.2).

Die Arbeit schließt mit einem kurzen Ausblick, der Anregungen für weitere Forschungsvorhaben zum Thema Persönlichkeitsbeurteilung im Allgemeinen und zur näheren Analyse der Items des BDI-V im Besonderen gibt (vgl. Abschnitt 6.3).

2. Laienpsychologische Persönlichkeitsbeurteilung

2.1 Judgment

2.1.1 Allgegenwart und Wichtigkeit der Persönlichkeitsbeurteilung

Wir fällen in unserem Alltag ständig Urteile und bewerten sowohl Menschen als auch soziale Situationen. Die Beurteilung von Personen kann in einem privaten oder beruflichen Umfeld (beispielsweise als Lehrer, Vorgesetzter, Juror o.ä.) erfolgen und sich einerseits auf Personen und Gruppen beziehen, mit denen wir unmittelbar in Kontakt treten, andererseits auch auf solche, die wir nur indirekt kennen (KANNING, 1999).

Menschen bewerten in ihrem eigenen Umfeld einzelne Angehörige, Freunde, Bekannte, Kollegen, Nachbarn und ihnen unbekannte Passanten oder Wartende. Bewertet werden aber auch die Familie, die Clique, die Nachbarschaft oder die Mitarbeiter im Betrieb als Ganzes. Wenn Urteile über die Persönlichkeit der Bundeskanzlerin, des Papstes oder der Nationalspieler gefällt werden, also über Menschen, die der Urteilende vermutlich niemals getroffen hat und treffen wird, beruhen diese Urteile nur auf weitergegebenen Fremdeindrücken und sind notgedrungen oberflächlicher und störungsanfälliger als Beurteilungen, die auf persönlichem Kontakt basieren. Noch abstrakter ist die Beurteilung von Gruppen, die eine große Zahl von Individuen zusammenfassen, von denen der Urteiler die allermeisten natürlich auch nicht kennt, beispielsweise die Einschätzung der Persönlichkeit 'der Skinheads' oder 'der Börsenyuppies'. In solchen Fällen bezieht sich das Urteil gar nicht mehr primär auf konkrete Menschen, sondern auf virtuelle Prototypen, welche die subjektiv relevanten Persönlichkeitseigenschaften der Mitglieder solcher Gruppen stellvertretend repräsentieren und aufgrund von Erfahrungen und Informationen durch Lernprozesse gebildet wurden (ANDERSON, 1996).

Trotz unterschiedlicher Ausgangsvoraussetzungen und Informationsgrundlagen bei der Beurteilung direkt oder nur indirekt bekannter realer oder gar nicht existenter virtueller Menschen laufen die Urteilsvorgänge selber offenbar ähnlich ab und scheinen ähnlichen Einschränkungen zu unterliegen (JONES, 1990). Der Untersuchungsgegenstand der vorliegenden Arbeit ist jedoch die laienpsychologische Persönlichkeitsbeurteilung konkreter Menschen, die auf direktem, persönlichem Umgang im Lebensumfeld beruht. Alle weiteren Ausführungen beziehen sich daher auf diese Art der Beurteilung, die unmittelbar mit dem Leben der Individuen zusammenhängt, aus ihm heraus entsteht und in ihm unmittelbare, direkte Konsequenzen für die Urteilenden hat.

Menschen fällen nicht nur häufig Urteile über ihre Mitmenschen, sie sind als soziale Wesen auch darauf angewiesen dies zu tun. In der ständigen Interaktion mit der Umwelt gilt es, die Passung an diese zu optimieren, um negative Konsequenzen zu minimieren. Dazu werden Verhaltensprognosen der anderen benötigt und daher gibt es keine Alternative zum Bewerten: "Wir können nicht einfach versuchen, die Beurteilung anderer Menschen ... zu vermeiden, da sie einen zentralen Bestandteil unserer eigenen Verhaltenssteuerung darstellt" (KANNING, 1999, S. 17).

Innerhalb konkreter Situationen werden konkrete Verhaltensvorhersagen benötigt. Diese können zwar sehr allgemein gehalten bleiben, müssen jedoch konkret in dem Sinne sein, daß sie für die aktuelle Situation gültig sind. Konkrete Vorhersagen werden aus unmittelbaren Erfahrungen mit der betreffenden Person in ähnlichen Situationen abgeleitet, aber auch aus allgemeineren Vermutungen über ihre Charaktereigenschaften und Persönlichkeitszüge. Im Wechselspiel von Induktion und Deduktion werden solche generellen Persönlichkeitseigenschaften als übergeordnete, abstrakte Beschreibungen aus konkreten Erfahrungen abgeleitet und dienen dann wieder zur theoriegeleiteten Prognose konkreten Verhaltens: "Personality is important *because* of the actions it affects, and we can only know about personality by *observing* what an individual does" (FUNDER, 1999, S.126, Hervorhebung im Originaltext).

Die laienpsychologische Persönlichkeitsbeurteilung wird als häufiges wie auch wichtiges Phänomen des Lebens automatisch zum Untersuchungsgegenstand der wissenschaftlichen Psychologie. Die erste zentrale Frage im Forschungsfeld ergibt sich dabei quasi von selbst: Wie kann man die Präzision von Persönlichkeitsbeurteilungen bestimmen, um anschließend Vergleiche zwischen ihnen zu ermöglichen?

Auf Basis eines Kriteriums der Urteilsgüte kann man dann untersuchen, wie es um die Urteilspräzision des Alltags bestellt ist, ob es dort überhaupt korrekte Urteile gibt, wenn ja wann und ob Persönlichkeitseigenschaften besser von den betroffenen Menschen selbst oder von Fremdbeobachtern eingeschätzt werden können. Weiterhin sollte es Ziel der Psychologie sein, Bedingungen und Einflußfaktoren von Persönlichkeitsbeurteilungen empirisch zu identifizieren und einen theoretischen Rahmen für die Beschreibung gefundener Determinanten und Moderatoren der Urteilspräzision zu schaffen. Vor der Diskussion dieser Themen widmen sich die folgenden Abschnitte 2.1.2 und 2.1.3 aber zunächst den Problemen, die die Definition und Operationalisierung eines Gütekriteriums der Persönlichkeitsbeurteilung aufwirft.

2.1.2 Accuracy: Wie kann Urteilspräzision sinnvoll bestimmt werden?

A natural avenue of inquiry into the domain of person perception would be to identify those who are good, accurate judges of others and to find out what makes them so. One does not have to travel very far along this avenue, however, before realizing that there are many potholes, barriers, and detour signs. The biggest problem of all is deciding what we mean by a 'good, accurate judge' - accurate with respect to what? (JONES, 1990, S. 20).

Um den Grad an Exaktheit, Präzision und Treffgenauigkeit (*engl.* accuracy) einer Persönlichkeitsbeurteilung zu bewerten, braucht man ein adäquates Kriterium der Urteilsgröße, dessen Definition einige philosophische und methodische Komplikationen aufwirft (vgl. FUNDER, 1999, 2004). Die methodischen Aspekte werden in Abschnitt 2.1.3 behandelt. Auf die philosophischen Erwägungen, die v.a. erkenntnistheoretischer Natur sind und das Verhältnis von Wahrnehmung zu Wirklichkeit berühren und die Angemessenheit stabiler subjektiver Persönlichkeitszuschreibungen, kann in der vorliegenden Arbeit aus Platzgründen leider nur kurz eingegangen werden. Ausführlichere Darstellungen bieten z.B. FUNDER (1999, S. 27ff) oder KRUGLANSKI (1989).

Im Wesentlichen müssen zwei Vorannahmen akzeptiert werden, wenn man Kriterien für die Präzision der laienpsychologischen Persönlichkeitsbeurteilung aufstellen möchte: Zunächst die Existenz einer subjektunabhängigen Realität und die Möglichkeit der Annäherung an sie, in dieser Realität wiederum die Existenz stabiler Persönlichkeitseigenschaften. Beide Aussagen wurden zeitweilig innerhalb der Psychologie bzw. Sozialwissenschaften in Frage gestellt (SCHÖNPFLUG, 2000).

Im Sinne des radikalen Konstruktivismus gibt es keine objektive Realität, die dem menschlichen Erkenntnisvermögen zugänglich wäre, sondern nur verschiedene subjektive Konstruktionen der Welt, welche die Sichtweisen der jeweiligen Personen bestimmen (GLASERSFELD, 1998). Die Suche nach einem Kriterium für Urteilspräzision macht vor einem solchen Hintergrund keinen Sinn: Alle Persönlichkeitsbeurteilungen wären prinzipiell gleich 'wahr', da sie alle Bestandteil einer existierenden individuellen Konstruktion von Wirklichkeit sind. Im Ansatz des radikalen Konstruktivismus läßt sich zwar leicht erklären, warum menschliche Wahrnehmung und darauf aufbauende Interpretationen und Urteile nur 'relativ' sind, jedoch nicht ohne weiteres, warum diese Urteile nicht vollständig beliebig sein und keinerlei Einschränkungen unterliegen sollten.

Die Gegenposition vertritt der hypothetische Realismus: Es wird angenommen, daß es eine reale Welt mit gewissen Strukturen gibt, die teilweise oder sogar vollständig erkennbar sind. Das Wissen über diese reale Welt behält zwar immer mutmaßlichen Charakter, hat aber als richtig zu gelten, solange es funktional ist, i.e. Erklärungswert für das Individuum hat und keine Widersprüche erzeugt (CHALMERS, 2001). Auch der moderate, kognitive Konstruktivismus postuliert, daß die Welt zumindest in Teilen subjektunabhängig existiert, wenn auch niemals unverfälscht erfahrbar. Eine Annäherung an die Realität sieht er in der Schnittmenge verschiedener individueller, subjektiver Repräsentationen der Umwelt. Die Unzuverlässigkeit von Wahrnehmung und Erinnerung und die Verzerrungstendenzen, denen sie unterliegen, werden im Sozialkonstruktivismus stärker betont, als das im hypothetischen Realismus der Fall ist (NÜSE, 1995).

Die Annahme zu akzeptieren, daß es eine Form von äußerer Realität gibt, der man sich nähern kann, heißt jedoch noch nicht, daß der Begriff der Persönlichkeitseigenschaft eine sinnvolle Beschreibungskategorie für diese Realität ist. Der Situationismus hält solche Eigenschaften (abgesehen von Intelligenz) für Fiktionen der Alltagspsychologie, die durch die transssituationale Variabilität und mangelnde Konsistenz des Verhaltens widerlegt seien (MISCHEL, 1968). Persönlichkeitsmerkmale im Sinne relativ stabiler, konstanter Verhaltens- und Erlebensdispositionen (*engl.* traits) existieren demnach überhaupt nicht. Unter solchen Voraussetzungen würde die Suche nach einem Kriterium für Urteilspräzision ebenfalls keinen Sinn machen: Alle Persönlichkeitsbeurteilungen wären prinzipiell gleich 'unwahr', da sie nur fälschlich internale Attribuierungen von Verhalten im Rahmen falscher impliziter Persönlichkeitstheorien der Beurteiler wären. Die Debatte zwischen Situationisten und Anhängern der Trait-Theorien, die Verhalten als Funktion in der Person liegender Faktoren beschreiben (z.B. GUILFORD, 1959, CATTELL, 1964) wurde in den siebziger Jahren des letzten Jahrhunderts sehr kontrovers geführt und entzog der Accuracy-Forschung lange den Boden (FUNDER, 1999).

Als Kompromiß mündete der Situationismus später in Synthese aus Person- und Situations-Einfluß auf das Verhalten in einen Interaktionismus und hat sich in mehrere Unterarme verzweigt, die den situativen Anteil jeweils unterschiedlich stark gewichten und das Zusammenspiel von situativen und personalen Faktoren unterschiedlich modellieren (KLEITER, 1987). In seiner reinen Form wird der Situationismus mittlerweile kaum noch vertreten, bleibt aber nach wie vor einflußreich (FUNDER, 2006).

After years of debate, most social scientists agree that personality exists - people have enduring personality traits that are useful in predicting their behavior. The only thing shocking now about this is how long it took to arrive at this conclusion (HASLTON & FUNDER, in press, S.13).

Natürlich deckt sich dieses Erkenntnis mit der Alltagspsychologie: Bei einer lexikalischen Analyse der englischen Sprache fanden ALLPORT UND ODBERT (1936) insgesamt 17953 verschiedene Wörter zur Bezeichnung allgemeiner und spezifischer Persönlichkeitseigenschaften, was die überragende Wichtigkeit von Traitkonzepten für die Laienbeschreibung anderer Menschen und der sozialen Umwelt belegt.

Werden die beiden Annahmen akzeptiert, daß man sich irgendeiner Form von äußerer Realität annähern kann und daß Persönlichkeitseigenschaften dabei sinnvolle Beschreibungskategorien sind, kann untersucht werden, wie zutreffend Beurteilungen solcher Eigenschaften ausfallen und wodurch sie beeinflußt werden. Für die wissenschaftliche Analyse laienpsychologischer Persönlichkeitsbeurteilungen gibt es mindestens vier verschiedene Kriterien zur Evaluation der Urteilsgüte (KRUGLANSKI, 1989):

1. Consistency (Übereinstimmungsgrad mehrerer 'Urteilsfacetten' in einem Urteil)
2. Adaptive Value (Anpassungswert des Urteils für das urteilende Individuum)
3. Correspondence (Übereinstimmungsgrad des Urteils mit einem Außenkriterium)
4. Consensus (Übereinstimmungsgrad mehrerer Urteile mehrerer Beurteiler)

Vor- und Nachteile dieser Gütekriterien werden im folgenden kurz besprochen.

Consistency: Die 'Stimmigkeit' eines Urteils als Kriterium seiner Güte heranzuziehen ist ein anschauliches Konzept und unter den Bedingungen des echten Lebens sicherlich weit verbreitet. Die Methode ist jedoch hochinterpretativ und für die systematische Erforschung laienpsychologischer Urteilsgüte weniger gut geeignet (FUNDER, 1987).

Adaptive Value: Unabhängig von 'Wahrheit' oder Genauigkeit kann man Urteilsgüte streng utilitaristisch definieren und mit dem Anpassungswert des Urteils gleichsetzen, also seinem Nutzen in der sozialen Interaktion für das Individuum. SWANN (1984) schlug vor, eine Beurteilung als pragmatisch korrekt (*engl.* *practical accurate*) anzusehen, wenn der Beurteilte die Erwartungen des Beurteilers erfüllt. Die Operationalisierung wirft jedoch in der Praxis einige Probleme auf und mündet letztlich wieder in einem Spezialfall der Definition über ein Außenkriterium (KRUGLANSKI, 1989).

Correspondence: Urteilsgüte kann durch den Grad der inhaltlichen Übereinstimmung des Urteils mit einem Außenkriterium bestimmt werden. Dessen Auswahl muß sorgfältig begründet werden, da es die Anforderung maximaler Objektivität erfüllen muß.

In manchen Studien werden sogenannte "Expertenurteile" von Psychiatern, Psychologen oder Pädagogen über bestimmte Persönlichkeitseigenschaften als Außenkriterien genutzt, um die Güte laienpsychologischer Persönlichkeitsbeurteilungen zu ermitteln (vgl. FUNDER, 1987). HASTIE UND RASINSKI bemerken nicht zu Unrecht: "This means that a subject's 'innaccurate judgment' boils down to a simple competition between the subject's judgment and the researcher's judgment" (1988, S. 197). Eine spezielle Klasse von Außenkriterien ist besser geeignet: Bezieht sich das Persönlichkeitsurteil auf einen Trait, der unmittelbar und eindeutig mit spezifischen Verhaltensweisen verbunden ist oder wird es direkt in Form einer Verhaltensprognose abgegeben, können empirisch erhobene Verhaltensdaten benutzt werden (vgl. FUNDER & SNEED, 1993). Der Vergleich mehrerer aus Persönlichkeitseinschätzungen abgeleiteter Verhaltensvorhersagen verschiedener Urteiler anhand ihrer Übereinstimmung mit tatsächlichem Verhalten ist eine reliable Methode, um die Güte der einzelnen Persönlichkeitsbeurteilungen zu bestimmen (z.B. KOLAR, FUNDER & COLVIN, 1996, SPAIN, EATON & FUNDER, 2000).

Consensus: Stehen keine Verhaltensdaten zur Verfügung, kann Urteilsgüte durch die Konvergenz mehrerer Beurteiler operationalisiert werden. Die Nutzung der Urteilsübereinstimmung (*engl.* interjudge agreement) von Selbsteinschätzern und Fremdeinschätzern oder von mehreren Fremdeinschätzern als Beleg für Urteilspräzision ist eine der Standardmethoden des Forschungszweiges geworden (vgl. FUNDER, 1999, HOFSTEE, 1994, KOLAR ET AL., 1996, KRUGLANSKI, 1989). Obwohl sich diese Methode bewährt hat und auch in der vorliegenden Untersuchung Consensus als Gütekriterium genutzt wird, darf eine Übereinstimmung von Persönlichkeitsbeurteilungen nicht einfach mit deren Korrektheit gleich gesetzt werden, aus theoretischer Sicht sollte Interjudge Agreement vielmehr als negatives Ausschlußkriterium dienen:

Although consensus does not positively establish accuracy (after all, collective fallacies have been known to occur), a lack of consensus may indirectly imply someone's inaccuracy. In other words, consensus may constitute a necessary but not a sufficient condition for accuracy (KRUGLANSKI, 1989, S. 396).

Fazit: Natürlich gibt es keine Quelle einer hundertprozentig akkuraten und zutreffenden Beschreibung der 'wahren' Persönlichkeitseigenschaften einer Person, die als Vergleichskriterium zur Ermittlung der Urteilsgüte anderer Persönlichkeitseinschätzungen dienen könnte. Urteilspräzision ist daher nicht im Sinne einer klassischen Korrespondenztheorie der Wahrheit definierbar, sondern nur im Rahmen eines pragmatischen oder konsenstheoretischen Wahrheitsbegriffs.

Je breiter ein solcher Konsens ermittelt wird, desto zuverlässiger fällt die Beurteilung der Urteilspräzision aus: "The assessment of the accuracy ... may require a multimethod approach that incorporates multiple indexes of accuracy based on different sources of information" (Vogt & Colvin, 2005, S. 267). Ein sinnvoller und angemessener Weg zur Evaluation der Urteilsgüte von Persönlichkeitsbeurteilungen liegt in der Nutzung von Consensus (i.e. dem Übereinstimmungsgrad verschiedener Beurteilungen) *und* Correspondence (i.e. der Zutreffensgenauigkeit von Verhaltensprognosen) als kombiniertem Kriterium (z.B. BORKENAU & LIEBLER, 1993, FUNDER & COLVIN, 1988, 1991, KOLAR ET AL., 1996, LEVESQUE & KENNY, 1993, SPAIN ET AL., 2000).

Das Problem bei diesem kombinierten Ansatz liegt in dem hohen Aufwand einer reliablen Verhaltensmessung in der realen Lebensumwelt (BORTZ & DÖRING, 1995, S. 240ff). Aus pragmatischen Gründen wird daher dennoch oftmals Interjudge Agreement als alleiniges Kriterium genutzt (vgl. FUNDER, 1999, KRUGLANSKI, 1989). Gleiches gilt auch für die vorliegende Studie: Das Erfassen unabhängiger Verhaltensdaten an einer größeren Stichprobe hätte den Rahmen einer Diplomarbeit gesprengt, deshalb wird Urteilspräzision ausschließlich über Urteilsübereinstimmung operationalisiert.

Wie laienpsychologische Persönlichkeitsbeurteilungen erfaßt werden müssen, um ihre Übereinstimmung anschließend sinnvoll beziffern zu können und welche methodischen Herausforderungen vor der Berechnung eines solchen Übereinstimmungswertes liegen, wird im folgenden Abschnitt 2.1.3 beschrieben.

2.1.3 Agreement: Wie kann Urteilsübereinstimmung quantifiziert werden?

Die laienpsychologische Persönlichkeitsbeurteilung im echten Leben verläuft ungeordnet, unformatiert und auf diversen Ebenen gleichzeitig ab, da sich solche Beurteilungen oft auf mehrere verschiedene Eigenschaften einer Person beziehen und oft in Situationen zustande kommen, in denen verschiedenste kognitive Anforderungen simultan zu bewältigen sind, u.U. auch die gleichzeitige Beurteilung mehrerer Personen (JONES, 1990). Viele Urteile werden daher unbewußt oder halbbewußt gefällt und die meisten Menschen hätten vermutlich Schwierigkeiten, die Gesamtheit ihrer bestehenden Urteile über die Persönlichkeit eines anderen klar und präzise zu verbalisieren.

Entsprechend ist die Einschätzung, ob und inwieweit sich eigene Urteile mit denen anderer Menschen decken, eine höchst subjektive Interpretation. Es gibt Belege dafür, daß viele Menschen die Verbreitung ihrer eigenen Einstellungen, Überzeugungen und Verhaltensweisen (und damit die Übereinstimmung mit anderen) überschätzen, was als False Consensus Effect bezeichnet wird (ROSS, GREENE & HOUSE, 1977). Andererseits umgeben sich Menschen oft mit Personen aus dem gleichen Milieu, die ähnliche Verhaltensweisen und Meinungen haben, daher könnten übertrieben erscheinende Ähnlichkeitseinschätzungen bezogen auf die unrepräsentative 'Stichprobe' des Bekanntenkreises durchaus zutreffend sein. False Consensus wurde in der Sozialpsychologie lange als abgesicherte Tatsache angesehen, obwohl er v.a. im Experiment zu Tage tritt und vielleicht nur ein der spezifischen Labormethodik geschuldeter Artefakt ist und kein grundlegendes, stabiles Phänomen des echten Lebens (DAWES & MULFORD, 1996).

Unabhängig von der Übereinstimmung mit anderen evaluieren Menschen die Angemessenheit ihrer Persönlichkeitseinschätzungen ohnehin mehr nach deren 'Stimmigkeit', i.e. nach der Konsistenz ihrer verschiedenen Ansichten über eine Person und nach der Plausibilität dieses Gesamtbildes unter dem Gesichtspunkt ihrer eigenen, individuellen impliziten Persönlichkeitstheorien (KRUGLANSKI, 1989). Generell gilt, daß es im täglichen Leben keinen Grund gibt, an der Richtigkeit von Persönlichkeitsbeurteilungen zu zweifeln, solange sie soziale Situationen kohärent und für das Individuum befriedigend interpretierbar machen. Erst unerwartete und nicht mehr ohne weiteres erklärbar erlebte Ereignisse und Beobachtungen führen zu einer Erweiterung und Veränderung bestehender Urteile (FESTINGER, 1957).

Die wissenschaftliche Psychologie braucht natürlich 'harte' Kriterien zur Bestimmung des Ausmaßes an Urteilsübereinstimmung, daher stellt sich zunächst das Problem, wie laienpsychologische Persönlichkeitsbeurteilungen überhaupt sinnvoll erfaßt werden können. Optimal wäre ein holistischer Ansatz, bei dem man sinnverstehende, interpretative Beschreibungen am Einzelfall entwickelt, direkt aus dem sozialen Feld heraus und unter Berücksichtigung aller sozialen Rahmenbedingungen (HOLZKAMP, 1983). Mit qualitativen Ansätzen ist es aber nur unter extremen Aufwand möglich, systematische Zusammenhänge innerhalb größerer Populationen zu analysieren (LAMNEK, 2005).

Um die Urteile einer großen Personenzahl systematisch vergleichen zu können, müssen deren Einschätzungen in einer systematisierten, numerisch umcodierbaren Form gemessen werden. Der mit diesem Vereinfachungsschritt einhergehende Informationsverlust muß notgedrungen akzeptiert werden, birgt aber große Gefahren: Wenn die Beurteilungen nicht reliabel erfaßt werden, können später weder Unterschiede zwischen verschiedenen Urteilen noch Zusammenhänge mit anderen Merkmalen statistisch zuverlässig nachgewiesen werden, weil die Messwerte von Zufallsfaktoren abhängen (BORTZ & DÖRING, 1995, S.181ff). Mangelt es der Erfassung der Urteile dagegen an Validität, sind zwar zuverlässige Berechnungen möglich, aus diesen können jedoch keine sinnvollen Aussagen abgeleitet werden, weil die ursprünglich gemessenen Werte keine inhaltliche Treffgenauigkeit besitzen (BORTZ & DÖRING, 1995, S. 185ff). Reliabilität und Validität bei der Messung sind daher die Mindestvoraussetzungen für die adäquate Quantifizierung der Übereinstimmung von Persönlichkeitsbeurteilungen.

Die Frage des Übereinstimmungsmaßes hängt dann stark mit dem Skalenniveau (Definition des Begriffs bei BORTZ, 1999, S. 20) der erfaßten Werte zusammen. Es gibt zwei Hauptmethoden, um Einschätzungen bestimmter Aspekte und Dimensionen von Persönlichkeit so zu messen, daß sie nichtinterpretativ in Zahlen umsetzbar sind: Die Zuordnung zu Kategorien erzeugt nominal- oder ordinalskalierte Meßwerte, während Einschätzungen auf Ratingskalen als intervallskaliert angenommen werden. Im ersten Fall muß entschieden werden, ob eine bestimmte Ausprägung eines Merkmals vorliegt bzw. welche genau von mehreren, im zweiten wird die Intensität eines Persönlichkeitsmerkmals graduell abgestuft eingeschätzt (BORTZ & DÖRING, 1995). Durch die Einordnung diverser einzelner Merkmale auf einer Ratingskala entsteht ein intervallskaliertes Antwortprofil, das eine komplexe Persönlichkeitsbeurteilung differenziert abbildet.

Bei echt nominalskalierten Kategoriensystemen wiegen sämtliche Nichtübereinstimmungen prinzipiell gleich schwer. Bei Ratingskalen sind Abweichungen weniger gravierend, da "das zugrundeliegende Merkmal ... meist als kontinuierlich angesehen werden kann. Bei Urteilen auf kontinuierlichen Skalen liegt es in der Natur der Sache, daß exakte Übereinstimmung einen Ausnahmefall darstellt" (WIRTZ & CASPAR, 2002, S. 34). Dennoch wurde die Übereinstimmung intervallskalierter Antwortprofile bis in die fünfziger Jahre des letzten Jahrhunderts fast ausschließlich über die mittlere quadratische Distanz (zwischen den Urteilern über alle Items) operationalisiert, einen Globalindex, in den jede Abweichung eingeht (vgl. FUNDER, 1995, 1999, KRUGLANSKI, 1989).

CRONBACH (1955) konnte mathematisch aufzeigen, daß in einem solchen Wert vier Varianzquellen vereinigt sind (konstanter Fehler, Variabilitätsfehler, stereotype Genauigkeit und differentielle Genauigkeit), von denen zwei urteilsobjektunabhängige, idiosynkratische Bewertungs- bzw. Skalennutzungstendenzen der Urteiler widerspiegeln, die den Index massiv beeinflussen können (vgl. Abschnitt 4.4.3.1). Seine vielbeachtete Kritik gipfelte in der Warnung: "Investigators run much risk of giving psychological interpretation to mathematical artifacts when they use measures which combine the components [of interjudge agreement]" (CRONBACH, 1955, S. 177). Als Folge seiner berechtigten Einwände kam die Erforschung von Urteilsübereinstimmungen und ihren Voraussetzungen lange Zeit zum Erliegen (vgl. Abschnitt 2.1.4).

KENNY (1994) entwickelte aus Cronbachs Ansatz eine komplexe Berechnungsmethode auf korrelativer Basis, das "Social Relations Model", mit der die Übereinstimmung zwischen verschiedenen Persönlichkeitsbeurteilungen in ihre Varianzkomponenten zerlegt werden kann, die dann einzeln analysierbar sind. Der Aufwand für die Anwendung ist verhältnismäßig hoch, da dyadische Daten aus einem Ringversuch (*engl.* round robin design) benötigt werden, in dem jede Person jede andere beurteilt. Die Interpretation der Ergebnisse ist anspruchsvoll, da keine Agreement-Kennwerte und keine individuellen Accuracy-Kennwerte für Urteiler und Beurteilte berechnet werden (FUNDER, 1999, S. 97f). Das Social Relations Model wurde und wird dennoch vielfach eingesetzt (KENNY, 2005). Liegen lediglich paarweise Beurteilungen vor, bei denen ein Beurteiler immer genau eine Zielperson einschätzt und diese Einschätzung dann mit genau einer Selbsteinschätzung verglichen wird, ist keine vollständige Komponentenanalyse im Sinne Cronbachs möglich (BERNIERI, ZUCKERMAN, KOESTNER & ROSENTHAL, 1994).

Doch auch innerhalb eines solchen vereinfachten Untersuchungsparadigmas gibt es Möglichkeiten, mit verschiedenen Indizes jeweils verschiedene der gewollten und ungewollten Komponenten mathematisch zu kontrollieren. Im Abschnitt 4.4.3.2 werden einige dieser Möglichkeiten im Zusammenhang der Auswertungsstrategien der Arbeit näher erläutert. Dort wird dann auch noch einmal ausführlicher auf die inhaltliche Bedeutung der vier Cronbachschen Übereinstimmungskomponenten eingegangen.

An dieser Stelle soll zunächst nur verdeutlicht werden, daß der Berechnung eines Wertes der Urteilsübereinstimmung eine Klärung vorangehen muß, welche Aspekte der Übereinstimmung man überhaupt quantifizieren möchte und welche Anteile man gleichzeitig mathematisch kontrollieren will. SCHMITT (1989) nennt drei ähnlichkeitsrelevante Informationen intervallskaliertter Antwortprofile: Profilhöhe (Mittelwert des Urteilers über alle Items, quasi die Urteilsstrenge), Profilstreuung (mittlere Abweichung der Items vom Urteilsmittelwert, quasi die Urteilsbreite) und Profilverlaufsgestalt (Antwortmuster über alle Items, quasi die spezifische Ausdifferenzierung des Urteils).

Eine intervallskalierte Persönlichkeitseinschätzung kann im Vergleich zu anderen Beurteilungen unverzerrt sein, aber gleichzeitig unpräzise (*engl.* unbiased but inaccurate): Eine Person wird zutreffend als sozial aufgeschlossen eingeschätzt, ohne daß vernünftig differenziert wird zwischen den aktuellen Attributen, die bei der Person mehr oder weniger typisch für die Gesamteigenschaft sind. Der Fall entspricht einer hohen Profilhöhenähnlichkeit bei niedriger Profilverlaufsähnlichkeit. Umgekehrt kann eine Persönlichkeitseinschätzung im Vergleich zu anderen Beurteilungen verzerrt sein, aber dennoch präzise (*engl.* biased but accurate): Die relativen Ausprägungen der einzelnen Attribute der Gesamteigenschaft, also ihre Stellung zueinander, werden richtig eingeschätzt, aber insgesamt zu milde oder zu streng gesehen, z.B. aufgrund von Sympathie oder Antipathie, Vorurteilen oder selbstwertdienlichen Urteilsverzerrungen. Dieser Fall entspricht einer niedrigen Profilhöhenähnlichkeit bei hoher Profilverlaufsähnlichkeit.

Korrelative Maße der Ähnlichkeit von Antwortprofilen quantifizieren die Kovarianz der Variablen unabhängig von Mittelwert und Varianz der beiden Profile, andere Maße berücksichtigen nur die absoluten Höhen der Variablen und vernachlässigen die Verlaufsgestalt (SCHMITT, 1989). Am Ende hängt die Wahl des adäquaten Koeffizienten immer von den Untersuchungsfragen und der Datenstruktur ab (WIRTZ & CASPAR, 2002).

2.1.4 Gibt es präzise Persönlichkeitsbeurteilungen und wovon hängen sie ab?

Die Psychologie hat sich mit der Beantwortung dieser beiden auf den ersten Blick simpel erscheinenden Fragen erstaunlich schwer getan: Die Voraussetzungen präziser laienpsychologischer Persönlichkeitsbeurteilungen unter Alltagsbedingungen sind erst in den letzten Jahren theoretisch beschrieben worden und nur eine Handvoll neuerer, methodisch einwandfreier empirischer Studien hat bisher mögliche Moderatoren der Urteilsgüte solcher Beurteilungen getestet (vgl. FUNDER, 1999, 2004). Ein Grund dafür liegt in der wechselvollen Geschichte des Forschungszweiges.

In den dreißiger Jahren des letzten Jahrhunderts zweifelte die entstehende Persönlichkeitspsychologie nicht an, daß es präzise und korrekte Persönlichkeitsbeurteilungen im Alltag gibt bzw. prinzipiell geben kann, warf aber die Frage auf, wann Einschätzungen tatsächlich korrekt sind und was für Menschen zur Abgabe präziser Urteile in der Lage sind (z.B. ALLPORT, 1937, ESTES, 1938). Bis in die fünfziger Jahre hinein wurden viele Studien durchgeführt, um Merkmale guter Beurteiler ausfindig zu machen: "Searching for the good judge of personality" (FUNDER & DOBROTH, 1987). Die Ergebnisse waren jedoch insgesamt enttäuschend, es wurden nur schwache und inkonsistente Zusammenhänge der Urteilsübereinstimmung mit Eigenschaften der Beurteiler gefunden, außerdem fiel das Agreement generell niedrig bis schlecht aus. Einen Überblick dieser ersten Forschungswelle gibt TAFT (1955).

Dann veröffentlichte Lee J. Cronbach einen vielbeachteten Artikel, in dem er nachweisen konnte, daß ein Großteil dieser Studien methodische Ungenauigkeiten enthielt, die zu einer Verschätzung der Accuracy führen konnten (CRONBACH, 1955, vgl. auch Abschnitt 2.1.3). Das Echo war so nachhaltig, daß der klassische Forschungsansatz zum Thema Personality Judgment, die Untersuchung der inhaltlichen Übereinstimmung verschiedener Urteile, aufgegeben wurde. Und das, obwohl Cronbach eigentlich detaillierte Vorschläge unterbreitet hatte, wie Urteilsübereinstimmung in Zukunft angemessener zu quantifizieren sei (GAGE & CRONBACH, 1955).

Cronbach's intention may have been to improve accuracy research rather than to shut it down, but the latter is very nearly what happened. Both the complex statistical adjustments recommended in a precomputer era and the large amount of data gathering that Cronbach argued was required apparently proved daunting to many investigators (FUNDER, 1995, S. 654).

Statt dessen begann nun im Zeitgeist der aufkommenden kognitiven Psychologie die Beschäftigung mit dem Prozeß des Urteilens. Der Fokus des Interesses lag jetzt auf häufig auftretenden Fehlern der Informationsverarbeitung (*engl.* biases) und gängigen kognitiven Vereinfachungsstrategien, sogenannten Heuristiken (z.B. KAHNEMAN & TVERSKY, 1973, NISBETT & ROSS, 1980). Die alte Forschungsfrage 'Wovon hängt Urteils-genauigkeit ab?' wurde in der prozessualen Perspektive verdrängt durch die Frage 'Wie entsteht Urteilsungenauigkeit?'. KRUEGER UND FUNDER kommentieren:

"[It] challenged the axiomatic status of rationality in economic theories of choice, and social psychologists soon saw the potential of this new paradigm for the study of social cognition" (2004, S. 324).

Die Persönlichkeitspsychologie entwickelte weiterhin Fragebögen zur Meßung von Persönlichkeitseigenschaften in Form klar definierter psychologischer Konstrukte und prüfte diese auch mit Übereinstimmungswerten auf Reliabilität, ihre Ziele waren aber die möglichst exakte klinische Messung der Ausprägung der Merkmale, die Untersuchung der Unterschiede zwischen einzelnen Personen und deren Auswirkungen auf das Verhalten (ASENDORPF, 1999). Die Frage, ob und auf welche Art präzise und korrekte Persönlichkeitsbeurteilungen im Alltag zustande kommen, fiel in die Domäne der mittlerweile entstandenen Sozialpsychologie, aber hier vergaß man über der Beschäftigung mit Fehlern, Fehlinterpretationen, Vereinfachungen und Verzerrungsfaktoren kognitiver Prozesse im Labor die Determinanten und Moderatoren gelingender Urteilsprozesse im echten Leben (FUNDER, 1987). Ab den sechziger Jahren spielte Accuracy als eigenständiges Konzept de facto keine Rolle mehr, eine Verbesserung der Urteils-genauigkeit sollte quasi *en passant* durch die Reduktion typischer Urteilsfehler (wie z.B. Halo Effect, Overconfidence, Hindsight Effect, False Consensus) erreicht werden. Die Angemessenheit dieser Strategie ist heute widerlegt: Beurteilerfehler-Maße und Accuracy sind nahezu unkorreliert, niedrigere Bias-Werte dürfen auf keinen Fall als indirekter Beleg für Urteilsgüte gewertet werden (MURPHY & BALZER, 1989).

In den Termini des damals populären Informationsverarbeitungsansatzes wurde das Phänomen Persönlichkeitsbeurteilung auch begrifflich reduziert, es hieß jetzt "Personenwahrnehmung" (*engl.* person perception). Die erforschten Wahrnehmungen bezogen sich aber gar nicht auf reale Menschen, sondern auf Persönlichkeitssplitter virtueller Personen im Labor und waren in Wirklichkeit 'inhaltsfrei', der Forschung war die soziale Substanz des Vorgangs verloren gegangen (FUNDER, 1995).

Ab den siebziger Jahren diskreditierte in der Persönlichkeitspsychologie zusätzlich der Einfluß des Situationismus das Traitkonzept, indem er bezweifelte, daß es Persönlichkeitseigenschaften überhaupt gibt (vgl. Abschnitt 2.1.2), Ross (1977) postulierte in einer sehr einflußreichen Arbeit die systematische Unterschätzung des Einflusses der Situation auf das Verhalten und nannte das "fundamentaler Attributionsfehler". Langsam aber sicher hatte sich die Meinung ausgebreitet, daß menschliche Urteilsfähigkeit generell limitiert und laienpsychologische Persönlichkeitsbeurteilungen prinzipiell falsch bzw. unangemessen seien (vgl. KRUEGER & FUNDER, 2004).

However, accuracy was never the point of this kind of research on person perception, ... nearly all of it used experimentally manipulated, artificial social stimuli that were ideally suited for testing models of some of the fine cognitive processes involved in person perception. None of it said very much about accuracy one way or the other (FUNDER, 1995, S. 654).

Erst in den achtziger Jahren des letzten Jahrhunderts begann in den USA wieder eine intensivere Auseinandersetzung mit dem Thema 'Urteilsgüte im Alltag'. Dem Fehlerparadigma (*engl.* error paradigm) wurde ein Präzisionsparadigma (*engl.* accuracy paradigm) entgegengesetzt, das einen Fokus auf die Beschreibung gelingender Urteilsprozesse und zutreffender Persönlichkeitsbeurteilungen sowie ihrer Voraussetzungen beinhaltet (vgl. FUNDER, 1983, 1987). Es wurde betont, daß laienpsychologische Persönlichkeitsbeurteilungen in natürlichen Umgebungen oftmals zuträfen und daß die vielfach behaupteten Biases vor allem der Laborsituation geschuldet seien (z.B. MCARTHUR & BARON, 1983, SWANN, 1984). Verlangt wurde deshalb, Persönlichkeitsbeurteilungen wo immer möglich nicht experimentell anhand der Bewertung abstrakten Versuchsmaterials sondern quasi-experimentell anhand der Bewertung echter Menschen aus echten Umgebungen zu erforschen und dabei Kriterien der Urteilsgüte zu nutzen, die aus diesen sozialen Umwelten abgeleitet sind: Consensus verschiedener Urteile und Correspondence mit Verhaltenskriterien (z.B. FUNDER, 1995, 1999, FUNDER & WEST, 1993, KENNY, 1994, KRUGLANSKI, 1989).

Ähnliche Forderungen nach einer ökologischen Perspektive bei der Erforschung sozialer Wahrnehmung und Persönlichkeitsbeurteilung hatte vor langer Zeit bereits BRUNSWIK (1956) erhoben. Seine Vorschläge und Anregungen waren damals aber im allgemeinen Schockzustand nach Cronbachs massiver Kritik untergegangen, jetzt wurde er als geistiger Urvater des Präzisionsparadigmas angesehen (FUNDER, 1995).

Durch die Nutzung von Korrespondenzkriterien zur Bestimmung der Urteilsgüte statt der Kohärenzkriterien (Abweichung von einem normativen Modell) des Fehlerparadigmas und durch komplexere, angemessenere Ansätze zur Quantifizierung der Urteilsübereinstimmung ergab sich ein deutlich positiveres Bild der Qualität menschlichen Urteilsvermögens im allgemeinen und laienpsychologischer Persönlichkeitsbeurteilungen im besonderen (vgl. HAMMOND, 1996).

KENRICK UND FUNDER (1988) antworteten in einem Überblicksartikel zur Situationismus-Personalismus-Debatte auf die radikale These, daß Persönlichkeitsbeschreibungen ohne weiteren Realitätsbezug nur losgelöst im Auge des Betrachters existierten:

Many Studies ... have shown that the impressions others have of your personality agree to an impressive extent both with each other and with your impressions of yourself. So although personality judgments may or may not often be wrong, they are not typically solipsistic (S. 25).

In der Tat hatte eine Reihe neuerer Untersuchungen belegt, daß es bei laienpsychologischen Persönlichkeitsbeurteilungen ein großes Maß an Übereinstimmung gibt und daß aus diesen Beurteilungen auch valide Verhaltensprognosen abgeleitet werden können (z.B. CHEEK, 1982, FUNDER, 1980, 1983, MOSKOWITZ & SCHWARZ, 1982). Die Beantwortung der Frage, ob es präzise Persönlichkeitsbeurteilungen im Alltag gibt, hatte damit endlich eine reliable empirische Basis. Mittlerweile wurden diese positiven Ergebnisse so oft bestätigt, daß heute die vorsichtig formulierte Aussage als gesichert gelten kann, daß es zumindest guten Bekannten zumindest prinzipiell möglich ist, zutreffende Persönlichkeitsbeurteilungen abzugeben, aus denen einigermaßen zutreffende Verhaltensvorhersagen abgeleitet werden können (FUNDER, 1999). Diese These klingt zunächst sehr zurückhaltend, stellt jedoch eine fundamentale Abkehr vom Fehlerparadigma dar.

Seit den neunziger Jahren wurden zunehmend Studien durchgeführt, die zeigen konnten, daß selbst auf sehr geringer Informationsbasis beruhende Beurteilungen tendenziell zutreffend sein können: Gänzlich unbekannte Personen können aufgrund begrenzter Informationen überzufällig korrekt beurteilt werden (BORKENAU & LIEBLER, 1993, WATSON, 1989), Traits können schon aus sehr kurzen Beobachtungen überzufällig korrekt abgeleitet werden (AMBADY, BERNIERI & RICHESON, 2000), gleiches gilt für Persönlichkeitsstörungen (OLTMANN, FRIEDMAN, FIEDLER & TURKHEIMER, 2004) und sexuelle Orientierungen (AMBADY, HALLAHAN & CONNER, 1999).

Die Behauptung einer generellen Fehlerbehaftetheit menschlicher Urteilsprozesse wird arg in Frage gestellt durch die erstaunlichen Leistungen, zu denen Beurteiler in der Lage sind. Beispielsweise korrelieren studentische Einschätzungen amerikanischer Dozenten vor dem Semester aufgrund nur halbminütiger, stummer Videosequenzen signifikant mit den Evaluations-Scores, welche die Dozenten am Ende des Semesters erhalten (AMBADY & ROSENTHAL, 1993). Und sogar aus einem kurzen Blick in das Schlafzimmer oder auf den Schreibtisch im Büro können valide Aussagen über Persönlichkeitseigenschaften abgeleitet werden (GOSLING, KO, MANNARELLI & MORRIS, 2002).

Das Fehlerparadigma war und ist unglaublich einflußreich und in seiner Wirkung auf die gesamte Sozialpsychologie kaum zu überschätzen (HASELTON & FUNDER, in press). Trotz vielfältiger Evidenzen für gelingende Urteilsvorgänge wird es weiterhin prominent vertreten (z.B. GILOVICH, GRIFFIN & KAHNEMAN, 2002, MYERS, 2004). Gleichzeitig werden aber immer stärkere Forderungen erhoben, die Erforschung sozialer Urteilsvorgänge nicht nur defizitorientiert und pathologisierend zu gestalten: "A more balanced social psychology would yield not only a more positive view of human nature, but also an improved understanding of the bases of good behavior and accurate judgment, coherent explanations of occasional lapses, and theoretically-grounded suggestions for improvement" (KRUEGER & FUNDER, 2004, S. 314). Das Präzisionsparadigma ist somit auch im Kontext der seit Anfang des neuen Jahrhunderts aufkommenden "positiven Psychologie" zu sehen, welche die 'guten' Aspekte des menschlichen Lebens und Miteinanders untersuchen und menschliche Stärken, Ressourcen, Entwicklungs- und Coping-Möglichkeiten aufzeigen will (z.B. SELIGMAN & CSIKSZENTMIHALYI, 2000, SNYDER & LOPEZ, 2002).

Die Frage, ob es zutreffende laienpsychologische Persönlichkeitsbeurteilungen überhaupt gibt bzw. ob laienpsychologische Persönlichkeitsbeurteilungen des Alltagslebens typischerweise eher zutreffend sind, wird also seit einiger Zeit wieder bejaht. Aufgrund der langen Blockierung des Forschungszweiges und der unzulänglichen Methodik vor Cronbach gab es aber tatsächlich Ende der achtziger Jahre, fünf Dekaden nach Gordon Allports ersten Schriften zum Thema Personality Judgment (ALLPORT, 1937) immer noch kaum Erkenntnisse darüber, welche Faktoren die Güte von Beurteilungsprozessen moderieren (FUNDER, 1987) und somit auch keine Antwort auf den zweiten Teil der Eingangsfrage dieses Abschnitts, 'Wovon hängen präzise Urteile ab?'

Erst in den letzten fünfzehn Jahren wurden eine Reihe potentieller Moderatoren der Urteilsübereinstimmung und -präzision empirisch überprüft und dabei erste Ergebnisse gefunden (vgl. FUNDER, 1999, 2004). David C. Funder schuf mit seinem Realistic Accuracy Model, kurz RAM, das prozessuale Perspektive und soziale Sichtweise bei der Beschreibung der Persönlichkeitsbeurteilung vereint, einen Bezugsrahmen, der die theoretische Einordnung von Moderatoren in Hauptkategorien und dadurch eine systematischere Suche ermöglicht (FUNDER, 1995, 1999). Dennoch gilt bis zum heutigen Tag, daß die wissenschaftliche Psychologie alles in allem nur wenige abgesicherte Erkenntnisse über die Voraussetzungen präziser, zutreffender laienpsychologischer Persönlichkeitseinschätzungen hat, da bisher nur wenige Moderatoren stabil und replizierbar nachgewiesen werden konnten (vgl. FUNDER, 1999, 2004).

Das Realistic Accuracy Model und seine Teilprozesse werden in Abschnitt 2.2 detailliert vorgestellt, danach werden in Abschnitt 2.3 auf Basis der Kategorien des Modells theoretisch ableitbare Moderatoren diskutiert und dazu bisher vorliegende empirische Ergebnisse zu Moderatoren der Urteilspräzision bzw. -übereinstimmung referiert, aus denen sich schließlich die konkreten Fragestellungen der vorliegenden Untersuchung ableiten (Abschnitt 3.1 und 3.2).

Vorher werden aber im folgenden Abschnitt 2.1.5 zunächst noch einmal die grundsätzlichen Unterschiede von Selbst- und Fremdeinschätzungen und die potentiellen Vor- und Nachteile der beiden Perspektiven kursorisch besprochen. Dazu werden einige unterschiedliche Präferierungen der wissenschaftlichen Psychologie im historischen Verlauf vorgestellt, ein einfaches Modell zur Strukturierung des Gegensatzes Selbst- und Fremdsicht auf Persönlichkeit und zuletzt empirische Resultate.

2.1.5 Kurzer Vergleich von Selbst- und Fremdbeurteilungen

Das (Fremd-)Einschätzen der Persönlichkeit anderer Menschen ist ein zentraler Bestandteil der menschlichen Verhaltenssteuerung (vgl. Abschnitt 2.1.1). Eine realistische und angemessene (Selbst-)Einschätzung der eigenen Persönlichkeit und ihrer Attribute ist ein wichtiger Aspekt mentaler Gesundheit (JAHODA, 1958, ROGERS, 1961). Beide Beurteilungsarten sind demnach von großer Bedeutung für das Individuum. Aber welche der Sichtweisen ist reliabler? Und wie unterscheiden sie sich?

Historische Notizen:

Als Wilhelm Wundt im Jahre 1879 in Leipzig das erste psychologische Labor errichtete, war die meistgenutzte Untersuchungsmethode die sogenannte Introspektion. Dabei berichteten geschulte Beobachter unter kontrollierten Bedingungen über ihre Bewußtseinsinhalte. Selbstbeobachtung und darauf beruhende Selbsteinschätzungen wurden von den ersten Psychologen eindeutig präferiert. Man hielt sie für den direkten Zugang zum individuellen Erleben der Wahrnehmung, des Denkens und Fühlens (Lück, 2002).

Ab den zwanziger Jahren des letzten Jahrhunderts kam der Behaviorismus auf, der Introspektion als unwissenschaftlich ablehnte und sich nur noch mit der Messung des äußeren, beobachtbaren Verhaltens befaßte. Innere, geistige Vorgänge laufen nach behavioristischer Diktion in einer prinzipiell uneinsehbaren "black box" ab (WATSON, 1976). Damit waren Fremdeinschätzung und mehr noch Fremdbeobachtung die Methoden der Wahl geworden. Persönlichkeit galt nur als Summe ausgebildeter Angewohnheiten, die Innensicht darauf war konzeptuell irrelevant.

Da viele Phänomene des Lebens ohne die Beschreibung innerer Prozesse und Strukturen einfach nicht adäquat erfaßt werden können, entstand ab den sechziger Jahren die kognitive Psychologie, in der wieder die Wahrnehmungen und Geisteseinhalte der Menschen zum Thema wurden. Beeinflußt vom Informationsverarbeitungsansatz wurden logisches Denken, Entscheidungsfindung, Aufmerksamkeitssteuerung, Sprachproduktion und andere rationale Teile des Bewußtseins erforscht (ANDERSON, 1996). Darüber hinaus beeinflusste die "kognitive Wende" jedoch alle Bereiche der Psychologie. Selbstauskünfte wurden wieder als Datenquellen akzeptiert, ohne den Fehler zu begehen, im Umkehrschluß erneut auf Fremdb Berichte zu verzichten. Beide Sichten sollten sich nun ergänzen und keine von beiden wurde als 'wahrer' angenommen (SCHÖNPFLUG, 2000).

In der Persönlichkeitspsychologie ist die Erfassung von Persönlichkeitseigenschaften mittels Selbstreport-Fragebögen eine bis heute weit verbreitete und allgemein als reliabel akzeptierte Methode. Fremdeinschätzungen sind prinzipiell ebenfalls relevant und kommen auch in persönlichkeitspsychologischen Untersuchungen zum Einsatz, aus pragmatischen Gründen allerdings deutlich seltener (HOFSTEE, 1994).

In der Sozialpsychologie sieht man Selbsteinschätzungen meist als verzerrt und fehlerbehaftet an: "Many theorists are less than sanguine about the ability of people to perceive their behavior objectively" (GOSLING, JOHN, CRAIK & ROBINS, 1998, S. 1339, vgl. auch Abschnitt 2.1.4). Dieses Mißtrauen bedeutet jedoch keine Favorisierung der Fremdperspektive, vielmehr wird Laienpsychologen die Fähigkeit zur Fällung objektiver Urteile per se abgesprochen, egal auf wen sie sich beziehen (vgl. FUNDER, 1999).

Theoretische Erwägungen:

Die Organisationspsychologen LUFT UND INGHAM (1955) entwickelten ein einfaches grafisches Schema der Persönlichkeitsbereiche, die der Selbst- bzw. Fremdbeobachtung zugänglich sind, das sogenannte Johari-Fenster.

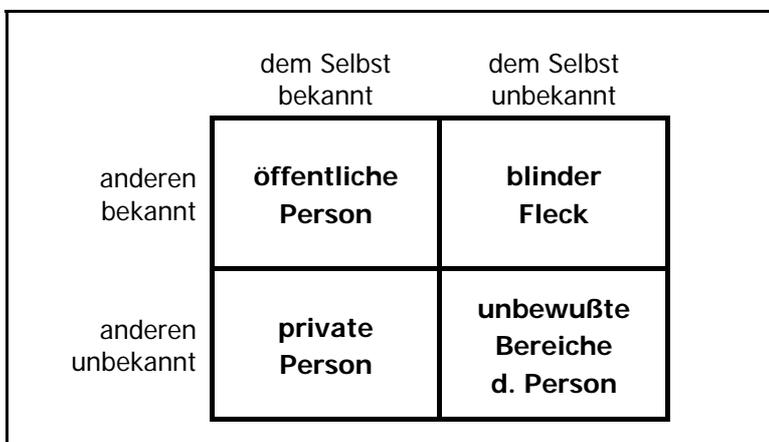


Abbildung 1: Das Johari-Fenster der verschiedenen Bereiche der Persönlichkeit

Das Johari-Fenster umfaßt alle Bereiche der Persönlichkeit, auch unbewußte, die niemals direkt zugänglich und beobachtbar sind. Zwischen privater und öffentlicher Person verläuft keine statische Demarkationslinie, sondern eine fließende Grenze. Sie variiert in Abhängigkeit vom Setting und vom Vertrautheitsgrad der jeweiligen Interaktionspartner in diesem Setting. In verschiedenen Settings werden unterschiedliche Aspekte der Persönlichkeit akzentuiert und dadurch zur öffentlichen Person gemacht.

Das Ausmaß des blinden Flecks differiert interindividuell, abhängig vom Grad der Selbstreflexivität, bleibt aber prinzipiell immer veränderbar. Im Bereich der Organisationspsychologie wird beispielsweise versucht, durch 360-Grad-Feedbacks und ähnliche Ansätze den blinden Fleck zu minimieren, indem Führungskräfte Rückmeldungen über die Mitarbeiter-Einschätzungen ihrer Persönlichkeit und ihres Auftretens erhalten (Brinkmann, 1998). Die Güte laienpsychologischer Persönlichkeitsbeurteilungen entscheidet sich im Spannungsfeld der drei Bereiche öffentliche Person, private Person und blinder Fleck: Welche Perspektive günstiger ist, spezifische Persönlichkeitszüge korrekt zu beurteilen, hängt von den relativen Ausmaßen der privaten Person und des blinden Flecks zueinander ab. Das Johari-Fenster verdeutlicht damit, daß der Vertrautheitsgrad von Judge und Target eine wichtige Determinante der Urteilsqualität ist.

Selbsturteile haben gegenüber Fremdurteilen natürlich den Vorteil der größeren Datenbasis: "We know how we act across different settings, we have a unique perspective on our private experiences, and in general we have access to more information about ourselves than any observer could possibly have" (KOLAR ET AL., 1996, S. 313). Man könnte daher vermuten, daß Selbsturteile meistens präziser und besser ausfallen als Fremdurteile. Auf das Johari-Fenster übertragen hieße das, das Gewicht des blinden Flecks würde aufgewogen durch die Kenntnis der privaten Person und aller Facetten der eigenen öffentlichen Person über alle Situationen hinweg.

Aber auch die Gegenthese, daß zumindest enge Bekannte besser in der Lage sind, die Persönlichkeit einer anderen Person zu beurteilen als diese Person selbst, läßt sich begründen. Aus der Attributionsforschung ist bekannt, daß Menschen dazu tendieren, den Handlungen anderer Personen innere Ursachen zuzuschreiben (dispositionelle Attributionstendenz), während sie ihr eigenes Verhalten eher mit äußeren Faktoren erklären (situative Attributionstendenz). JONES UND NISBETT (1971) erklären diese komplementären Neigungen mit den unterschiedlichen Perspektiven von Handelndem und Beobachter. Das Beobachtungsfeld eines Handelnden ist die Umgebung, auf deren Analyse er Aufmerksamkeit richten muß, so daß er seinen eigenen Anteil an der Situation und die Persönlichkeitszüge, die sich in seinem Handeln ausdrücken, nicht immer adäquat wahrnehmen kann. Ein Beobachter dagegen kogniziert den Handelnden als 'Figur' vor einem kontextuellen Hintergrund. Automatisch wird er mit den anderen Handelnden des Beobachtungsfeldes verglichen (FESTINGER, 1954).

Je weniger sich eine Person ihres eigenen Verhaltens und Charakters bewußt ist, desto größer ist ihr blinder Fleck. Ab einer gewissen Ausdehnung werden Fremdeinschätzungen der Persönlichkeit potentiell reliabler als die Selbstbeurteilungen des Individuums. Die Schwierigkeit gleichzeitigen Erlebens und Beobachtens und die prinzipiell vom Individuum weggerichtete Sichtweise begünstigen die Entstehung solcher blinden Flecken. KOLAR ET AL. (1996) bezeichnen dies anschaulich als "Fish-and-Water Effect":

[It's] difficult to detect one`s own stable behavioral tendencies on roughly the same reason that fish are said to find it difficult to detect water (S. 330).

Neben der grundsätzlichen Diskrepanz der Perspektiven, die natürlich für alle Menschen gilt, bedrohen interindividuell unterschiedlich stark ausgeprägte Verhaltenstendenzen die Güte von Selbstbeurteilungen. Die bewußte Verfälschung von Selbstberichten aufgrund sozialer Erwünschtheit ist ein wohlbekanntes und in der testtheoretischen Literatur breit diskutiertes Phänomen (AMELANG & ZIELINSKI, 2002). Außerdem kann es zu einer unbewußten Verzerrung der Selbstwahrnehmung und -beurteilung aufgrund selbstwertdienlicher motivationaler Prozesse kommen (STEELE, 1988).

Beispielsweise benutzen Menschen positiv besetzte Persönlichkeitseigenschaften typischerweise eher zur Beschreibung ihres eigenen Charakters als dem anderer (BROWN, 1986) und halten sich selbst für gelassener und gefestigter, als andere sie sehen (FARAH & ATOUM, 2002). Das Motiv der Selbsterhöhung unterstützt die Aufrechterhaltung des Selbstwertgefühls. Andererseits wurde auch ein Motiv der realistischen Selbstbewertung postuliert: Demnach sind Menschen bestrebt, eine realitätsnahe und akkurate Einschätzung ihrer Eigenschaften zu gewinnen, und zwar unabhängig von möglichen positiven oder negativen Implikationen dieser Urteile für das Selbst (FESTINGER, 1954).

Empirische Ergebnisse:

Die objektive Beurteilung, ob in einem konkreten Fall eine Fremd- oder eine Selbsteinschätzung der Persönlichkeit zutreffender ist, wirft wieder das Problem eines angemessenen Kriteriums der Urteilspräzision auf. Wie in Abschnitt 2.1.2 erläutert wurde, kann man die Güte eines aus einer Fremd- und einer Selbsteinschätzung bestehenden Persönlichkeitsurteils pragmatisch durch den Übereinstimmungsgrad der beiden Einschätzungen operationalisieren. Dieser Wert läßt aber natürlich keine Aussage über die Accuracy der beiden Einzelbeurteilungen zu.

Die relative Güte einzelner Beurteilungen kann nur bestimmt werden, wenn mehr als zwei Einschätzungen der jeweils gleichen Zielperson erfaßt werden. Der Consensus der Urteile mehrerer Fremdeinschätzer ist dann typischerweise geringfügig größer als der von Selbst- und Fremdeinschätzern (z.B. JOHN & ROBINS, 1993, 1994, KENNY, 1994), was eine leichte Überlegenheit der Fremdperspektive andeutet. Selbsteinschätzungen bezüglich hoch angesehener Eigenschaften wie Freundlichkeit, Aufgeschlossenheit oder Hilfsbereitschaft fallen im Schnitt etwas positiver aus als Fremdeinschätzungen, ein Beleg für selbstwertdienliche Urteilsverzerrungen (z.B. COLVIN, BLOCK & FUNDER, 1995). Dieser Effekt schwankt jedoch interindividuell sehr stark, da er durch Persönlichkeitseigenschaften der Selbsteinschätzer moderiert wird: "Individual's substantial differences in self-perception bias suggest, that the self-enhancement tendency should not be treated as a general law of social perception" (GOSLING ET AL., 1998, S. 1344).

Zur Bestimmung der absoluten Güte einzelner Persönlichkeitsbeurteilungen müssen Verhaltensvorhersagen abgeleitet und mit unabhängigen Verhaltenskriterien verglichen werden. KENNY (1994) referiert einige Untersuchungen dieser Art: Bei der Vorhersage bestimmter beruflicher Schritte und des Erfolges in einer Psychotherapie waren Selbsteinschätzungen realistischer, ebenso bei Prognosen bezüglich Schlaflosigkeit oder Suchtmittelgebrauch, dagegen trafen Fremdeinschätzungen besser zu, wenn es um betriebswirtschaftliche Erfolge, Dauer von Partnerschaften oder zukünftige Delinquenz ging. Die zusammengetragenen Ergebnisse sind divergent und wirken zufällig. In den meisten Studien wurden lediglich Auftretenswahrscheinlichkeiten eng umrissener Verhaltensweisen eingeschätzt, keine umfassenderen Persönlichkeitseigenschaften.

Verhaltensprognosen werden aus Persönlichkeitseigenschaften abgeleitet, die wiederum aus der Beobachtung zurückliegenden bzw. aktuellen Verhaltens erschlossen werden: "Behavioral acts constitute the building blocks of interpersonal perception and the basis for inferences about personality traits" (GOSLING ET AL., 1998, S. 1337). Schon bei der Beobachtung des Verhaltens kann es zu selbstwertdienlichen Wahrnehmungsverzerrungen kommen. JOHN UND ROBINS (1994) gingen der Frage der 'Wahrnehmungsgüte' nach. Nach einer Gruppendiskussion erfaßten sie die retrospektiven Einschätzungen aller Teilnehmer bezüglich der Auftretenshäufigkeit bestimmter psychologisch relevanter Verhaltensweisen. Diese wurden dann mit simultan erzeugten, standardisierten und aggregierten Beobachtungsscores verglichen.

Dabei zeigte sich, daß die Teilnehmer die Häufigkeit der Beiträge der anderen Personen korrekter einschätzten als die ihrer eigenen. Auf der Beobachtungsebene scheint also die Präzision der Fremdperspektive höher zu sein als die der Selbstperspektive. Aber wie sieht es mit der Urteilebene aus?

KOLAR ET AL. (1996) fanden klare Hinweise darauf, daß Fremdbeurteilungen durchaus zutreffender sein können als Selbsteinschätzungen. Sie verglichen die Fähigkeit von gut miteinander bekannten Selbst- und Fremdeinschätzern, psychologisch bedeutsames Verhalten (ängstliches Auftreten, Enthusiasmus im Gespräch, Offenheit, humorige Gesprächsführung etc.) vorherzusagen. Die Personenbeurteilungen und Verhaltensprognosen wurden mit einer detaillierten psychometrischen Skala erfaßt, zu jedem Selbsturteil kamen jeweils zwei Fremdurteile von Freunden der Person. Die Verhaltenskriterien wurden nach einem festen Auswertungsschlüssel aus längeren, auf Video aufgenommenen Interaktionssequenzen codiert. Es zeigte sich, daß ein einzelnes Fremdurteil im Vergleich zur Selbsteinschätzung etwa die gleiche bis leicht bessere prognostische Validität besaß und daß zwei aggregierte Fremdurteile eine deutlich bessere Vorhersagegenauigkeit hatten als das Selbsturteil. In einer ähnlich aufwendigen Studie replizierten SPAIN ET AL. (2001) die Ergebnisse und belegten ebenfalls eine Überlegenheit der Fremdperspektive bei der Abgabe von Verhaltensprognosen.

The most reliable source of information about an individual's personality is to be found not in his or her self-judgments, but in the consensus of the judgments of the community of his or her peers. This difference in reliability would only increase, of course, as more informants ... were recruited (KOLAR ET AL., 1996, S. 330).

Fazit: Aus den verschiedenen theoretischen Ansätzen heraus gibt es Gründe, die für die Überlegenheit der Fremdeinschätzung sprechen (Beobachterperspektive, emotionaler Abstand, Unverzerrtheit), aber auch Gründe, welche die umgekehrte These einer Präferenz der Selbsteinschätzung unterstützen (situativ breitere Informationsbasis, Kenntnis innerer Zustände). Die empirischen Ergebnisse sprechen für ein leichte Überlegenheit der Fremdperspektive, deren Urteilsgüte tendenziell etwas höher zu sein scheint. Voraussetzung ist natürlich, daß der Fremdbeurteiler die zu beurteilende Person gut kennt. Die Präzision von Fremdeinschätzungen kann durch Aggregation der Einschätzungen mehrerer Beurteiler potentiell weiter erhöht werden. Dies entspricht der sozialkonstruktivistischen Sichtweise, wonach sich der Grad der Annäherung an die objektive Realität im Grad der Überschneidung individueller Sichtweisen manifestiert.

2.2 Judgmental Process

2.2.1 Das Realistic Accuracy Model (RAM) und seine Teilprozesse

RAM prescribes that a portrayal of the actual psychological attributes of the target person be sought in combination with a wide range of information and the accuracy of a judgment of the target's personality be evaluated in terms of its congruence with this portrayal (FUNDER, 1995, S. 656f).

David C. Funder's sozialkognitives Realistic Accuracy Model (FUNDER, 1995, 1999) beschreibt den Prozeß der laienpsychologischen Persönlichkeitsbeurteilung einer Zielperson (*engl.* target) durch einen Beurteiler (*engl.* judge) über die Zerlegung in dabei ablaufende essentielle Kernprozesse, deren Gelingen die Urteilsgröße determiniert. Statt nach Fehlerquellen und Verzerrungsfaktoren zu suchen, ist RAM positiv orientiert und will Möglichkeiten und Leistungsvermögen gelingender, präziser Urteilsprozesse und deren Voraussetzungen systematisch beschreibbar machen. Die Vorannahmen des Modells sind: Persönlichkeitseigenschaften im Sinne stabiler Traits existieren, Menschen fällen manchmal Urteile über die Eigenschaften anderer Menschen und diese Urteile sind manchmal zutreffend (FUNDER, 1999, S. 118f, vgl. auch Abschnitt 2.1.4).

Aus Gründen der Übersichtlichkeit geht das Modell vereinfachend von einer 'one-trait one-cue' Situation aus, also dem Schließen auf genau eine Persönlichkeitseigenschaft aus genau einem dazu vorliegenden Hinweis. In Wirklichkeit setzen sich viele Beobachtungen und Erfahrungen zu einem mosaikartigen Bild zusammen, das wir uns von den Persönlichkeitseigenschaften anderer machen. Tatsächliches Urteilen im echten Leben ist als Integration vieler einzelner Urteilsprozesse im Sinne von RAM zu verstehen.

Ein einzelner Urteilsprozeß besteht laut RAM aus vier Teil- oder Kernprozessen, die in Tabelle 1 zunächst überblicksartig dargestellt sind und dann in den folgenden Unterabschnitten detaillierter beschrieben werden.

Tabelle 1: Voraussetzungen eines gelingenden Urteilsprozesses laut RAM

Teilprozeß	Art	betroffene Personen
Relevance: Bedeutsamkeit schaffen	situativ	Target
Availability: Zugang erhalten	situativ	Target und Judge
Detection: Hinweise aufnehmen	perzeptiv	Judge
Utilization: Hinweise nutzen	kognitiv	Judge

2.2.1.1 Relevance: Bedeutsamkeit schaffen

Zunächst muß eine Zielperson, um bezüglich einer bestimmten Persönlichkeitseigenschaft beurteilbar zu sein, Informationen abgeben, die für diese Eigenschaft relevant sind. Das geschieht typischerweise in Form von Verhalten, welches in irgendeiner Verbindung zu der Eigenschaft steht. Bestimmte Traits können sich in bestimmten Situationen aber auch in 'Nichtverhalten' ausdrücken, beispielsweise Feigheit oder Opportunismus. Neben Verhalten können natürlich auch verbale Informationen, also das, was eine Zielperson sagt, und andere, subtilere Informationen wie Körperhaltung, Gesichtsausdruck, Stimmlage, Kleidungsstil und Ähnliches auf Persönlichkeitseigenschaften hinweisen (FUNDER & SNEED, 1993).

Der Punkt der Bedeutsamkeit ist sehr wichtig und stellt sich in der Realität deutlich komplizierter dar als im Modell. Einerseits kann ein gegebenes Verhalten von mehreren Persönlichkeitseigenschaften beeinflusst sein, u.U. sogar von sehr vielen verschiedenen, so daß die Verbindung unklar ist und schwierig zu erkennen (MCCRAE & COSTA, 1995). Andererseits kann eine Eigenschaft, die ein gegebenes Verhalten verursacht, manchmal auch erst nach wiederholten Beobachtungen über einen längeren Zeitraum deutlich werden. Beispielsweise kann hinter einem Geschenk Großzügigkeit stecken, es kann aber auch taktische, manipulative Gründe haben oder einfach nur der Konvention geschuldet sein und mit gar keiner spezifischen Persönlichkeitseigenschaft zusammenhängen (außer vielleicht Konventionalität). Die Großzügigkeit einer Person wird meistens nicht in einem singulären Verhaltensakt, sondern erst in einem wiederkehrenden Verhaltensmuster erkennbar (FUNDER, 1991). Andere Persönlichkeitseigenschaften drücken sich wiederum nur in zugespitzten, eher extremen und damit selteneren Situationen aus, z.B. Mut oder Streßresistenz (FUNDER, 1999).

Zudem kann Relevance auch durch situative Zwänge verhindert werden, welche die Verhaltensvariabilität einschränken. Verhalten hängt dann eher von Determinanten der Situation ab als von Persönlichkeitseigenschaften. SNYDER UND ICKES (1985) sprechen in dem Zusammenhang von 'schwachen' und 'starken' Situationen: Starke Situationen sind vorstrukturierte, in denen Menschen weniger Möglichkeiten des Persönlichkeitsausdrucks haben. So hat ein Judge, der das Target nur im Kontext niederer Erwerbsarbeit kennt, sicherlich eine schlechtere Basis für ein Urteil, als ein Beurteiler, der auch in der Freizeit, also in schwachen Situationen, mit der Zielperson Kontakt hat.

2.2.1.2 Availability: Zugang erhalten

Wenn relevantes Verhalten und andere Hinweise seitens eines Targets emittiert werden, müssen sie auch einem Judge zugänglich sein, damit eine Beurteilung stattfinden kann. Verschiedene Verhaltensweisen einer Person werden verschiedenen Menschen in verschiedenen Situationen zugänglich, daher fällen sie auch unterschiedliche Urteile über diese Person. Je mehr verschiedene Situationen ein Beurteiler über die Zeit beobachten kann, desto präziser wird sein Urteil ausfallen (FUNDER, 1999).

Beurteilungsgenauigkeit hängt demnach eng mit dem Grad der Bekanntschaft (*engl. acquaintanceship*) zusammen. Diese Erkenntnis deckt sich natürlich mit dem Alltagswissen: Personen, die man länger und besser kennt, kann man auch besser beurteilen. Bekanntschaft hat quantitative und qualitative Aspekte. Ein Beurteiler, der eine Zielperson schon seit fünf Jahren kennt, hat vermutlich eine bessere Urteilsgrundlage als einer, der die Person erst seit fünf Wochen kennt. Von zwei Beurteilern, die beide die Zielperson seit fünf Jahren kennen, hat vermutlich einer, der mit der Person schon eine Weltreise gemacht hat, eine bessere (weil situativ breitere) Grundlage als einer, der das Target nur aus dem nachbarschaftlichen Gespräch über den Gartenzaun kennt.

2.2.1.3 Detection: Hinweise aufnehmen

Wenn relevante Hinweise emittiert werden und einem Judge prinzipiell zugänglich sind, muß er sie auch wahrnehmen bzw. müssen die Stimuli in irgendeiner Form vom Nervensystem des Urteilers erfaßt werden. Dieser Vorgang muß nicht unbedingt bewußt ablaufen, ein Judge kann auch sublimale Details und Hinweise aufnehmen, die seine Urteile beeinflussen, ohne daß ihm das klar ist (FUNDER, 1999).

Ob Detection stattfindet, hängt mit dem aktuellen Zustand des Judges, der Art der Hinweise und einer Interaktion zwischen Judge und Informationen zusammen. Seitens des Beurteilers wird die Informationsaufnahme z.B. durch Unaufmerksamkeit, Ablenktheit, Erregtheit oder Desinteresse gefährdet. Seitens der Hinweise spielt die Salienz der Informationen eine Rolle: dramatische, überraschende, lebendige oder auch nur laute Ereignisse werden eher wahrgenommen als unspektakuläre, regelmäßig auftretende und leise. Schließlich neigen Beurteiler dazu, bei anderen Personen gerade solche Informationen wahrzunehmen, die sich auf Eigenschaften zu beziehen scheinen, die für ihr eigenes Selbstkonzept zentral sind (SEDIKIDES & SKOWRONSKI, 1993).

2.2.1.4 Utilization: Hinweise nutzen

Die Abgabe relevanter Informationen durch das Target, deren Zugänglichkeit für und Aufnahme durch den Judge reichen noch nicht aus, ein präzises Urteil zu gewährleisten: Natürlich muß der Beurteiler diese Informationen dann auch adäquat nutzen und aus den vorliegenden Hinweisen sinnvoll auf die Existenz bzw. Ausprägung spezifischer Persönlichkeitszüge schließen. Urteilsprozesse erfordern kognitive Informationsverarbeitungsressourcen, die begrenzt sind. Bei Überlastung bzw. Vollauslastung des Systems (*engl.* cognitive load) aufgrund beliebiger anderer situativer Anforderungen kann es Interferenzen geben, welche die Qualität des Urteilsprozesses senken (FUNDER, 1999). Allgemein gilt: Je umfassender eine Beurteilungsaufgabe, desto größer die Verzerrung der Informationen durch kognitive Vereinfachungsprozesse im Dienste der Komplexitätsreduktion (GAUGLER & THORNTON, 1989).

Die Nutzung verschiedener Informationen für die Persönlichkeitsbeurteilung setzt bestimmte kognitive Strategien und allgemeines Wissen über Persönlichkeitseigenschaften und ihren Ausdruck in Verhalten voraus, also implizite Persönlichkeitstheorien seitens des Urteilers. Diese werden im Laufe des Lebens erlernt. Eine optimale Informationsnutzung wird durch Biases und starre, vorurteilshafte Schemata von Persönlichkeit gefährdet. Andererseits sind nicht alle Stereotypen falsch und können manchmal auch effizient sein und im Sinne einer automatisierten Situationsanalyse zu validen Urteilen führen (BRODT & ROSS, 1998).

2.2.2 Theoretische Verortung von RAM

Social psychology has concerned itself with how cues are detected and used in judgment without addressing the degree to which these cues are actually diagnostic of the traits they are used to judge under realistic circumstances. Its approach has been almost exclusively experimental, in which cues are manipulated and serve as the independent variable. In a parallel fashion, personality psychology has concerned itself with the behaviors that tend to be emitted by people who are characterized by certain traits, but has not addressed the manner in which laypersons might use these behaviors when inferring these traits. Its approach has been almost exclusively correlational, in which subjects' preexisting personality traits serve as the independent variable and their behavior serves as the dependent variable. Notice how the dependent variable of personality psychology becomes the independent variable of social psychology in RAM's conceptualization (FUNDER, 1995, S.658).

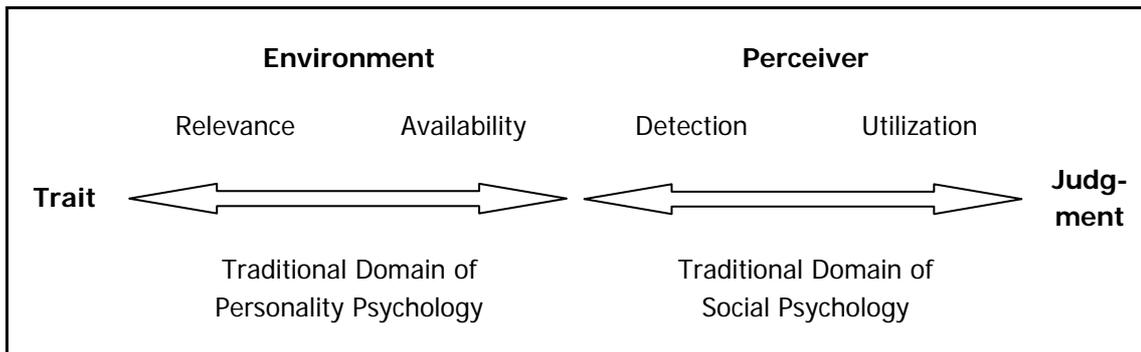


Abbildung 2: RAM und die getrennten Domänen der Persönlichkeits- und Sozialpsychologie

2.2.3 Fazit

Anders als frühere kognitive Ansätze zur Personenwahrnehmung bezieht das Realistic Accuracy Model neben Prozessen im Kopf des Judges auch Eigenschaften des Targets und den Informationsaustausch zwischen beiden in der geteilten sozialen Umgebung mit ein. In dieser umfassenden Perspektive wird deutlich, daß präzises Urteilen ein komplexer, schwieriger Prozeß ist. Sein Gelingen hängt vom Gelingen mehrerer Kernprozesse ab, die wiederum von sehr vielen verschiedenen Faktoren beeinflusst werden können. Daraus folgt, daß die Zahl potentieller Moderatoren der Urteilsgüte als sehr groß angesehen werden muß.

Das Modell gibt dem Phänomen Judgment eine gewisse Ordnung, da Einflußfaktoren der Urteilsgüte ihre Effekte in einem oder mehreren der Teilprozesse des Modells entfalten müssen. Dadurch wird es einerseits möglich, empirisch gefundene Moderatoren von Urteilsvorgängen sinnvoll einzuordnen, andererseits können theoriegeleitet neue, bisher noch nicht getestete Moderatoren erschlossen und anschließend empirisch überprüft werden. Außerdem könnten mit Hilfe des Modells Interaktionen von Moderatoren besser erklärt werden. Zur Zeit gibt es noch zu wenig empirische Ergebnisse und kaum bestätigte Hauptwirkungen von Moderatoren, die Beschreibung und Analyse von Interaktionswirkungen ist jedoch eines der Ziele im Forschungsfeld (FUNDER, 1999).

Im Anschluß werden im folgenden Abschnitt 2.3 einige konkrete Überlegungen bezüglich potentieller Moderatoren erörtert, die aus RAM ableitbar sind, außerdem werden bisher vorliegende empirische Ergebnisse zu Moderatoren der Urteilsgüte bzw. -übereinstimmung gesichtet. Aus den Ausführungen des Abschnitts 2.3 ergeben sich die konkreten Fragestellungen der vorliegenden Untersuchung, die in Abschnitt 3.1 dargestellt und dann in Abschnitt 3.2 in Form von Hypothesen zusammengefaßt werden.

2.3 Moderators of judgmental Accuracy

2.3.1 Taxonomie nach RAM

Empirisch gefundene und theoretisch abgeleitete Moderatorvariablen, die den Grad der Genauigkeit bzw. Übereinstimmung von Persönlichkeitsbeurteilungen beeinflussen oder beeinflussen könnten, werden von RAM eingeteilt in Eigenschaften des Urteilers, des Beurteilten, der von Urteiler und Beurteiltem geteilten Hinweisbasis und der zu beurteilenden Persönlichkeitseigenschaft.

Tabelle 2: Felder von Moderatoren der Urteilsgüte und assoziierte RAM-Kernprozesse

Moderatorenfeld	Kernprozesse
Urteiler / Judge	Detection, Utilization
Beurteilter / Target	Relevance, Availability
Hinweisbasis / Information	Relevance, Availability
Persönlichkeitseigenschaft / Trait	Relevance, Availability

In den folgenden Abschnitten 2.3.2 bis 2.3.5 werden diese vier Moderatorenfelder nacheinander betrachtet. Dabei werden jeweils zunächst grundsätzliche Erwägungen zu den RAM-Prozessen angestellt und theoriebasiert potentielle Moderatoren der Urteilsgüte abgeleitet und danach, soweit vorliegend, empirische Ergebnisse zu bisher getesteten Moderatoren diskutiert.

2.3.2 The good Judge: Merkmale des Beurteilers

2.3.2.1 Theoretische Erwägungen

Interindividuelle Unterschiede der Beurteilungsfähigkeit müssen laut RAM mit den Prozessen Detection und Utilization zusammenhängen. Die zur adäquaten, unverzerrten Wahrnehmung und Nutzung persönlichkeitsrelevanter Informationen nötigen Voraussetzungen eines Beurteilers sind laut FUNDER (1995, 1999):

Motivation: Konkret die aus der Wichtigkeit des Urteils abgeleitete Motivation angemessen zu urteilen, allgemein das Nichtvorliegen eines die Informationsverarbeitung verzerrenden motivationalen Stils wie Feindseligkeit oder Narzißmus.

Knowledge: Implizites und explizites Wissen über Persönlichkeitseigenschaften und ihren Ausdruck in Verhalten und anderer Form.

Ability: Allgemeine perzeptive und kognitive Fähigkeiten (Intelligenz) und kognitive und attributive Komplexität.

Im Bereich der motivationalen Voraussetzungen sind Empathie, Respekt für andere, Neugierde, Extraversion und geistige Offenheit Beispiele für potentiell förderliche Züge in der Beurteilerpersönlichkeit und Rechthaberei, Dogmatismus, Humorlosigkeit, Narzißmus und Neurotizität Beispiele für hinderliche (FUNDER, 1999). Knowledge in Form von implizitem und explizitem Wissen beruht auf Erfahrung und Lernprozessen, daher kommt das Lebensalter des Beurteilers als Moderator in Betracht, auf jeden Fall das Ausmaß seiner sozialen Erfahrung, das wiederum auch mit Extraversion zusammenhängen sollte, da extravertierte Menschen vermutlich mehr Gelegenheiten haben, intensivere soziale Erfahrungen zu machen. Ein zumindest schwacher Zusammenhang von Extraversion und Beurteilungsfähigkeit wäre also möglich (FUNDER, 1995). Insgesamt dürfte das Ausmaß der sozialen Erfahrung jedoch schwierig zu erfassen sein. Abilities wie Intelligenz, kognitive Komplexität und ähnliche allgemeinere Konstrukte geistiger Fähigkeiten können hingegen gut, wenn auch aufwendig gemessen werden (AMELANG & ZIELINSKI, 2002).

Leider steht ein spezifischeres Konstrukt noch aus, das Knowledge und Ability vereint und so etwas wie 'sowohl wissensbasiert-rationale, logisch-analytische als auch empathiebasiert-emotionale, rollenübernehmende und -nachvollziehende Beurteilungskompetenz' als komplexe Persönlichkeitseigenschaft beschreibt und v.A. operationalisiert. Ein wichtiger Teil so eines Konstruktes wäre die Fähigkeit zur sozialen Perspektivübernahme (*engl.* perspective taking).

FUNDER UND HARRIS (1986) beschrieben die zum Abgeben präziser Beurteilungen nötige Persönlichkeitseigenschaft als "sozialen Scharfsinn" (*engl.* social acuity): "The ability and inclination to perceive the psychological state of others and act accordingly" (S. 530). Sie erhoben an einer größeren Stichprobe mit unterschiedlichen Meßmethoden verschiedene Aspekte dieses hypothetischen Konstrukts und konnten zeigen, daß die unabhängig gemessenen Eigenschaften "Self-Monitoring", "Empathy", "Attributional Complexity" und "Nonverbal Sensitivity" hoch korrelieren, außerdem gibt es einen starken Zusammenhang des berechneten Social Acuity Gesamtwertes mit "Interpersonal Effectiveness", die per Fremd- und Selbsteinschätzung ermittelt wurde. FUNDER UND HARRIS (1986) sahen darin deutliche Belege für Reliabilität und konvergente Validität des Konstrukts. Der Ansatz wurde aber leider nicht weiter entwickelt und Social Acuity bisher nicht als Moderator getestet.

2.3.2.2 Empirische Ergebnisse

Die generelle Existenz von Beurteilungsfähigkeit als Persönlichkeitseigenschaft konnte durch eine Reihe von Studien nachwiesen werden, welche die Konsistenz interindividueller Unterschiede der Urteilsgüte von Beurteilern über mehrere Beurteilte und verschiedene Trait-Domänen hinweg belegten (z.B. KOLAR, 1995, MARANGONI, GARCIA, ICKES & TENG, 1995). Keinen Zusammenhang gibt es allerdings zwischen der von den Judges selbst eingeschätzten Beurteilungsfähigkeit und ihrer tatsächlich erzielten Urteilsgüte (MARANGONI ET AL., 1995, SWANN & GILL, 1997).

TAFT (1955) zog in seinem Überblicksartikel der ersten Accuracy-Forschungswelle der dreißiger bis fünfziger Jahre das Fazit: "The best judge of personality tended to be intelligent, socially skilled, and psychologically well adjusted" (S. 20). Die Ergebnisse der damaligen Zeit sind aufgrund der methodischen Vorbehalte von CRONBACH (1955) natürlich vorsichtig zu interpretieren, aber die Aussage ist sehr allgemein gehalten und deckt sich mit neueren Forschungsergebnissen. In KOLARS (1995) aufwendiger Studie, bei der diverse Persönlichkeitseigenschaften der Judges erhoben wurden, charakterisierten sich die guten männlichen Beurteiler der Stichprobe als extravertiert, angenehm, gut organisiert und nicht-ängstlich, von Freunden wurden sie als extravertiert, interessant und freundlich beschrieben. Die guten weiblichen Beurteiler bezeichneten sich selbst als sozial aufmerksam, breit interessiert, unkonventionell und nicht-konservativ, die Fremdbeschreibungen ergaben bei ihnen kein klares Bild. Alles in allem scheinen Frauen im Durchschnitt etwas bessere Beurteilungen abzugeben als Männer (z.B. AMBADY, HALLAHAN & ROSENTHAL, 1995, BERNIERI ET AL., 1994, VOGT & COLVIN, 2003).

Motivation: FLINK UND PARK (1991) konnten belegen, daß die Urteilsgüte besser wird, wenn die Beurteiler glauben, daß an dem Urteil wichtige soziale Konsequenzen hängen. Personen mit hoher interpersoneller Orientierung fällen tendenziell bessere Urteile (VOGT & COLVIN, 2003). Für den positiven Zusammenhang von sozialer Urteilskompetenz und Extraversion gibt es ebenfalls einige empirische Belege (z.B. AKERT & PANTER, 1988, LESTER, 1991). Ebenso für den negativen Zusammenhang von Narzißmus und Accuracy (JOHN & ROBINS, 1994). Personen, die sehr viel kognitive Energie für Impression Management aufwenden, fällen schlechtere Urteile (PATTERSON, 1994).

Ability: Der Intelligenzquotient wurde bisher als Moderator kaum geprüft, mindestens eine Studie belegt jedoch einen Zusammenhang zwischen dem IQ und der Accuracy von Verhaltensbeurteilungen (HAVENSTEIN & ALEXANDER, 1991).

2.3.3 The good Target: Merkmale des Beurteilten

2.3.3.1 Theoretische Erwägungen

There are 'open' personalities about whom all judges agree exceptionally well; there are 'enigmatic' personalities about whom they agree hardly at all. Who are these people? (ALLPORT, 1937, S. 443).

Interindividuelle Unterschiede in der Beurteilbarkeit verschiedener Targets müssen laut RAM mit den Prozessen Relevance und Availability zusammenhängen.

Relevance: Die Bedeutsamkeit der durch ihr Verhalten abgegebenen Informationen sollte höher sein bei Menschen, deren Verhalten konsistent und deren Persönlichkeit stabil und kohärent organisiert ist (BEM & ALLEN, 1974). Da das wesentliche Bestandteile der Definition des Konstruktes der "psychologischen Anpassung" sind (*engl.* psychological adjustment, *Def.* correspondence between private and public self, role variability, and behavior predictability), müssten hochangepasste Menschen besser beurteilbar sein. Im Umkehrschluß sollten psychisch zerrissene, disorganisierte und inkohärente Personen schwieriger einzuschätzen sein, da ihr Verhalten nicht valide und somit nicht-diagnostisch ist (COLVIN, 1993a). Und auch für neurotische und extrem impulsive Menschen gilt, daß ihr Verhalten von außen tendenziell undurchschaubar und unvorhersagbar bleibt (FUNDER, 1999).

Während niedrige psychologische Anpassung und die damit verbundene niedrigere Konsistenz und Kohärenz von Persönlichkeit und Verhalten sicherlich kein von den Betroffenen erwünschter und gezielt herbeigeführter Zustand ist, beschreibt das Konzept des "Self-Monitoring" (SNYDER, 1987) letztlich das Ausmaß intentionaler Inkonsistenz: Hoch selbstüberwachende Personen sind um Anerkennung bemüht und gleichzeitig hochsensibel für soziale Hinweisreize. Daher kontrollieren sie ihr eigenes Verhalten fast permanent und passen es selbst an subtilste Veränderungen in der Umgebung an. Die Beurteilung ihrer 'wahren' Persönlichkeit wird dadurch erschwert.

Narzißtische Menschen versuchen ebenfalls, ein übertrieben positives und gefälliges Bild ihrer selbst zu vermitteln. Generell sollten Menschen, die zu Täuschung und Betrug neigen, schwieriger zu beurteilen sein, da jede Form unaufrichtigen Verhaltens per definitionem uninformativ bezüglich der wirklichen Einstellung einer Person ist. Bei stärkeren Ausprägungen dürfte jedoch über einen gewissen Zeitraum das Muster betrügerischen Verhaltens an sich und damit die Eigenschaft des notorischen Lügneriums selbst ganz gut beurteilbar sein (FUNDER, 1999).

BAUMEISTER UND TICES (1988) Metatrait-Hypothese, nach der die Urteilsgüte von Fremdeinschätzungen von Persönlichkeitsmerkmalen davon abhängt, ob das Target die Eigenschaftsdimension überhaupt besitzt, kann eines gewissen Charmes nicht entbehren. Sie fußt auf der idiographischen Annahme von BEM UND ALLEN (1974), daß die zur Beschreibung der Persönlichkeit angemessenen und nötigen Traitdimensionen interindividuell differieren. Sollen Eigenschaften beurteilt werden, die für die Persönlichkeit eines Individuums keine Rolle spielen und in Wirklichkeit in keiner spezifischen Richtung ausgeprägt sind, resultieren natürlich Zufallseinschätzungen. Die theoretische Konzeption des Targetmoderators Metatrait als quantifizierbare Typikalität und Zentralität einer Eigenschaft für die Beschreibung der Persönlichkeit eines Individuums ist interessant, seine vorgeschlagene Operationalisierung jedoch enttäuschend: Die Messung soll über das Streumaß der Selbstauskunft erfolgen. "High variance signifies that the person responded erratically, variably, or inconsistently to the different items, suggesting that he or she is untraited on the dimension measured by the scale" (BAUMEISTER & TICE, 1988, S. 581). Damit wird dann wieder nur die individuelle Konsistenz der zu beurteilenden Persönlichkeitseigenschaft gemessen.

Availability: Extravertierte Menschen müssten besser einschätzbar sein, da sie mehr und lebendigeres Verhalten zeigen als Introvertierte und der Umwelt damit in quantitativer Hinsicht einfach mehr Hinweise über ihre Persönlichkeit geben. Daneben hat Zugänglichkeit auch qualitative Aspekte: Das Ausmaß, in dem Menschen Emotionen zeigen und bereit sind, über Gefühle, Gedanken, Probleme und alles andere, was sie bewegt zu sprechen, hängt ab vom Grad an Privatheit (*engl.* privacy) bzw. Selbstoffenbarung und Schüchternheit des Targets (FUNDER, 1995).

2.3.3.2 Empirische Ergebnisse

Die generelle Existenz von Beurteilbarkeit als Persönlichkeitseigenschaft konnte durch eine Reihe von Studien nachwiesen werden, welche die Konsistenz interindividueller Unterschiede der Güte von Urteilen bezüglich verschiedener Targets über mehrere Judges und wiederholte Messungen hinweg belegten (Metaanalyse bei ZUCKERMAN, BERNIERI, KOESTNER & ROSENTHAL, 1989). COLVIN (1993b) fand in einer Längsschnittanalyse über einen Zeitraum von fünf Jahren eine große Rangreihen-Stabilität der Beurteilbarkeit, und das sogar in der von vielen persönlichen, psychologisch bedeutsamen Umbrüchen geprägten Altersspanne von 18 bis 23 Lebensjahren.

Ein Überblicksartikel zum Moderatorvariablenansatz in der Persönlichkeitspsychologie (ZUCKERMAN, KOESTNER, DEBOY, GARCIA, MARESCA & SARTORIUS, 1988) nennt als die am häufigsten empirisch geprüften Moderatoren der Urteilsübereinstimmung auf Seiten des Targets die individuelle Konsistenz des zu beurteilenden Traits und assoziierten Verhaltens über Zeit und Situationen (*engl.* consistency) und die selbstberichtete Wichtigkeit und Bedeutung (*engl.* importance) des Traits für das Selbstbild der Person. Weder Consistency noch Importance zeigten einen signifikanten Gesamteffekt in der Metaanalyse, beide wurden aber zumindest in einigen der referierten Einzelstudien belegt. In ihrer eigenen Untersuchung mit fast 500 Versuchspersonen fanden ZUCKERMAN ET AL. (1988) eine signifikante Verbindung von Importance und Consistency mit dem Ausmaß der Urteilsübereinstimmung, wobei der Effekt der selbstberichteten Wichtigkeit etwas stärker ausfiel als der Effekt der Konsistenz.

CHAPLIN (1991) sichtete Studien, in denen Selbstüberwachung, Selbstbewußtheit, Selbstreflexivität und sozialkommunikative Kompetenz als mögliche Indikatoren der Beurteilbarkeit eines Targets getestet wurden. Im Endeffekt hing keines der gemessenen Konstrukte signifikant mit Accuracy zusammen.

Dafür korrelierte das unabhängig erfaßte Ausmaß der psychischen Anpassung einer Targetperson signifikant mit ihrer Beurteilbarkeit, operationalisiert durch die Intraklassenkorrelation mehrerer Fremdurteile (COLVIN, 1993a). Auch die Extravertiertheit eines Targets hat einen empirisch meßbaren Einfluß auf die Beurteilbarkeit seiner Persönlichkeit (BORKENAU & LIEBLER, 1992). Wie nah eine Person mit Ihren Werten auf den Big Five Faktordimensionen am Bevölkerungsdurchschnitt liegt, was COLVIN (1993a) als Grad der Üblichkeit ihrer Persönlichkeitsstruktur (*engl.* commonness) interpretiert, scheint jedoch keine Rolle für ihre Beurteilbarkeit zu spielen.

2.3.4 Good Information: Merkmale der geteilten Hinweisbasis

2.3.4.1 Theoretische Erwägungen

Das Ausmaß verfügbarer Hinweise auf Persönlichkeitseigenschaften kommt laut RAM in den Prozessen Availability und Relevance zum Tragen. Die Zugänglichkeit zu einer angemessen breiten und aussagekräftigen Informationsbasis ist eine entscheidende Grundlage für ein gelingendes Urteil. Die Hinweisbasis besteht aus allem, was eine Person tut und/oder sagt. Sie ist umso größer, je mehr Zeit Menschen miteinander verbringen (quantitativer Aspekt), aber auch je intensiver und variabler die gemeinsamen Situationen sind (qualitativer Aspekt), daher sollte Urteilsgüte durch den Grad der Bekanntschaft zwischen Target und Judge moderiert werden (COLVIN & FUNDER, 1991).

Eine Liaison, Partnerschaft oder Ehe ist die engste und intimste Beziehung, die Menschen zueinander eingehen können und bietet eine besonders gute und breite Urteilsgrundlage. Beziehungspartner sollten sich daher besonders gut einschätzen können (MCCRAE, STONE, FAGAN & COSTA, 1998). Die Beobachtbarkeit einzelner Traits (vgl. Abschnitt 2.3.5) sollte innerhalb einer Liebesbeziehung eine geringere Rolle spielen, da es ab einem gewissen, sehr engen Grad an Bekanntschaft vermutlich kaum noch unbeobachtbare Persönlichkeitseigenschaften gibt.

Die Postulation eines Bekanntschaftseffekts ist naheliegend und wenig überraschend, sein praktischer Nachweis jedoch nicht ohne Tücken: "The variable is rather crude, certainly as operationalized to date and perhaps even in principle" (FUNDER, 1995, S. 662).

2.3.4.2 Empirische Ergebnisse

Es gibt diverse Belege dafür, daß schon auf einer äußerst geringen Informationsgrundlage Persönlichkeitsbeurteilungen möglich sind, deren Präzision deutlich über dem Zufallsniveau liegt (z.B. BORKENAU & LIEBLER, 1992, 1993, LEVESQUE & KENNY, 1993, WATSON, 1989). Wird das Informationslevel einer solchen "Zero-Acquaintance" Situation experimentell leicht variiert, zeigt sich bereits ein Zusammenhang zwischen der Quantität und Qualität der verfügbaren Informationen und der erzielten Urteilsgüte (z.B. AMBADY ET AL., 2000, BORKENAU, MAUER, RIEMANN, SPINATH, & ANGLEITNER, 2004, LETZRING, WELLS & FUNDER, in press). Selbst wenn der Unterschied nur darin besteht, ob die Videoaufnahme der Interaktion einer unbekannte Targetperson für die Dauer von fünf bis zehn oder von 20 bis 25 Minuten gezeigt wird, erhöht sich die Urteilsübereinstimmung unter mehreren Fremdeinschätzern (BLACKMAN & FUNDER, 1998).

Natürlich nimmt die Urteilsübereinstimmung zwischen Selbst- und Fremdeinschätzungen und zwischen mehreren Fremdurteilen über eine Person noch deutlich stärker zu, wenn man statt Quasi-Fremden (*engl.* relative strangers) enge Bekannte der Zielperson befragt. Einige frühere Studien hatten in diesem Punkt noch kontraintuitive Resultate ergeben (z.B. GORMLY & EDELBERG, 1974, STELMACHERS & MCHUGH, 1964, zit. n. PAUNONEN, 1989). Mittlerweile ist der Acquaintanceship-Effekt jedoch vielfach repliziert und breit dokumentiert worden (z.B. COLVIN & FUNDER, 1991, FUNDER & COLVIN, 1988, PAULHUS & BRUCE, 1992, PAULHUS & REYNOLDS, 1995, PAUNONEN, 1989). Es konnte untermauert werden, daß der Effekt tatsächlich auf ein höheres Informationsniveau zurückzuführen ist und nicht auf Assumed Similarity (CRONBACH, 1955) und False Consensus (ROSS, GREENE & HOUSE, 1977): Enge Bekannte können sich besser einschätzen, auch wenn sie sich nicht ähnlicher sind als beliebige Nichtbekannte, außerdem hängt die Übereinstimmung der Selbstbeschreibung eines Judges mit seinen abgegebenen Fremdeinschätzungen nicht mit deren Präzision zusammen (FUNDER, KOLAR & BLACKMAN, 1995).

Die meisten Studien, die den Acquaintanceship-Effekt stützen, belegten einen Einfluß des Bekanntschaftsgrades indirekt durch den Vergleich der mittleren Urteilspräzision von Subgruppen, die nach abgestuftem Kenntnisgrad gebildet wurden. Acquaintanceship wird so zu einer diskreten bzw. kategorialen Variable. Die Operationalisierung von Acquaintanceship als kontinuierliche Variable ergab dagegen bisher kaum brauchbare Resultate (PAULHUS & BRUCE, 1992, WATSON, HUBBARD & WIESE, 2000).

MCCRAE (1994) untersuchte Fremdeinschätzungen einer großen Stichprobe Erwachsener und fand keinen direkten linearen Zusammenhang der Urteilsgüte mit einzelnen, isolierten Variablen wie z.B. Dauer der Bekanntschaft, Frequenz der Treffen, wahrgenommener Ähnlichkeit oder Sympathie füreinander. SWANN UND GILL (1997) fanden ebenfalls keinen direkten Zusammenhang zwischen Bekanntschaftsdauer und Urteilspräzision, lediglich das Vertrauen in die Güte der eigenen Urteile scheint mit der Zeit zu steigen. BERNIERI ET AL. (1994) fanden an einer Stichprobe studentischer Zimmergenossen keinen Einfluß der Bekanntschaftsdauer auf die Güte der gegenseitigen Persönlichkeitseinschätzungen, ebensowenig PARK, KRAUS UND RYAN (1997) in einer ähnlich gelagerten Studie mit mehreren Meßzeitpunkten. PAUNONEN (1989) belegte einen Zusammenhang zwischen der Selbsteinschätzung der Bekanntschaftsenge und Urteilsübereinstimmung, fand jedoch keinen Einfluß der Bekanntschaftsdauer.

KENNY (1994) kommt zu dem Schluß, daß die Accuracy der gegenseitigen Einschätzungen zu Beginn einer Bekanntschaft / Freundschaft zunächst stark und schnell zunimmt, danach langsamer steigt und sich dann ab einem bestimmten Punkt kaum noch verändert. SULLIVAN UND CHAPDELAINE (2001) konnten diese Vermutung empirisch untermauern: In ihrer Längsschnittstudie mit vier Meßzeitpunkten, verteilt über sechs Monate, nahm der Konsensus bei der Bewertung der Eigenschaften anderer Personen unter den Mitgliedern mehrerer neu gegründeter studentischer Arbeitsgruppen zu Beginn deutlich stärker zu als gegen Ende des Semester. Zur exakten Beschreibung der Steigung und Form der Acquaintanceshipfunktion von Urteilerdyaden wären aber sorgfältig kontrollierte Längsschnittstudien über einen erheblich längeren Zeitraum nötig, beginnend mit dem allerersten Aufeinandertreffen von Target und Judge, um die Entwicklung von Bekanntschaft und die dadurch bewirkte Veränderung von Urteilsgenauigkeit über die Zeit präzise zu dokumentieren - ein nahezu unumsetzbares Unterfangen (FUNDER, 2004).

Abgesehen von kontinuierlichen und künstlich diskretisierten Acquaintanceship-Variablen gibt es auch echt kategorialen Variablen, die auf qualitativ grundsätzlich unterschiedliche Formen gegenseitiger Kenntnis und Verbindung verweisen. Die Auswirkungen des Beziehungsstatus (z.B. liiert ja / nein, Familienmitglied ja / nein) auf die Güte von Persönlichkeitseinschätzungen sollten einfacher nachzuweisen sein.

Besonders groß ist die Nähe und damit die geteilte Informationsbasis innerhalb von Familien, deshalb ist es nicht überraschend, daß die Urteilsübereinstimmung unter Familienmitgliedern umfassender ausfällt als in anderen Gruppen (BRANJE, VAN AKEN, VAN LIESHOUT & MATHIJSSSEN, 2003). Innerhalb von Liebesbeziehungen kommt es zwar zur Überschätzung von Gemeinsamkeiten, aber das meßbar hohe Niveau dieses Biases senkt nicht die Beurteilungsgüte (KENNY & ACITELLI, 2001). Es konnte gezeigt werden, daß die Urteilsübereinstimmung unter Ehepartnern im Schnitt höher ausfällt als unter Freunden und Bekannten, die empirische Identifikation von Moderatoren, die Unterschiede der Urteilsgüte innerhalb der Gruppe der verheirateten Paare erklären könnten, gelang jedoch nicht (MCCRAE ET AL., 1998).

2.3.5 The good Trait: Merkmale der zu beurteilenden Eigenschaft

2.3.5.1 Theoretische Erwägungen

Unterschiede in der Beurteilbarkeit verschiedener Persönlichkeitseigenschaften müssen laut RAM mit Unterschieden in den Prozessen Availability und Relevance zusammenhängen. Die Zugänglichkeit oder Beobachtbarkeit (*engl.* observability) eines Traits hängt einerseits an der schieren Sichtbarkeit (*engl.* visibility) assoziierter Verhaltensweisen. Beispielsweise haben Redseligkeit, Dominanzstreben, Schüchternheit oder Aggressivität unmittelbare und damit leichter erkennbare Auswirkungen als beispielsweise Geduld, mathematisches Talent, Kunstgeschmack oder handwerkliches Geschick.

Andererseits hängt die Beobachtbarkeit einer Eigenschaft auch mit ihrem Ansehen (*engl.* favorability) zusammen, und zwar im Sinne der bewußten und gezielten Unterdrückung sozial unerwünschter Verhaltensweisen, die dadurch natürlich auch schwieriger zu beurteilen sind. Beispielsweise drückt sich Geselligkeit in einer hohen Frequenz emotional warmer und positiver sozialer Interaktionen aus und kann im Allgemeinen gut beobachtet werden, während sich Unehrlichkeit z.B. in Lügen, Betrügereien und Diebstählen äußert, die aus naheliegenden Gründen weitestgehend verdeckt praktiziert werden, also unbeobachtbar. Auch Persönlichkeitszüge wie Neid oder Mißgunst drücken sich oft eher indirekt und ungewollt aus, da den meisten Personen bewußt ist, das solche Eigenschaften von anderen zumeist abgelehnt werden.

Selten auftretende Verhaltensweisen sind natürlich immer schwieriger zu beobachten, erzeugen davon abgesehen aber auch automatisch eine geringere statistische Übereinstimmung: "If an act has a low base rate of occurrence, then observers are more likely to miss it over the course of an interaction. Moreover, on psychometric grounds, low base-rate acts will have less variance across targets, which will tend to reduce correlations between observers" (GOSLING ET AL., 1998, S. 1339).

2.3.5.2 Empirische Ergebnisse

Einige Befunde aus dem Bereich der psychologischen Forschung zum Thema Interjudge Agreement zeigen, das die Urteilsübereinstimmung bei beobachtbareren Eigenschaften (*engl.* more observable traits) tatsächlich größer ist, und zwar sowohl die Übereinstimmung zwischen Selbst- und Fremdurteilen als auch die zwischen mehreren Fremdurteilen über die gleiche Person (z.B. FUNDER & COLVIN, 1988, JOHN & ROBINS, 1993, 1994, PAUNONEN, 1989, ZUCKERMAN ET AL., 1989).

In anderen, älteren Studien konnte allerdings kein Einfluß der Observability auf das Agreement als unabhängiger Haupteffekt nachgewiesen werden (z.B. FUNDER, 1980, KAMMANN, SMITH, MARTIN & MCQUEEN, 1984). Die Ursache dieser Diskrepanz könnte in der diffizilen Ermittlung eines objektiven Wertes für die Beobachtbarkeit einer Persönlichkeitseigenschaft bzw. einer konkreten Verhaltens- und Erlebensweise liegen, da bei der Messung schon geringfügige sprachliche Änderungen der Items zu stark veränderten Observability-Werten führen können (WATSON ET AL., 2000).

Außerdem beeinflußt der Ausprägungsgrad einer Eigenschaft ihre Beobachtbarkeit, so daß es zu schwierig zu analysierenden Interaktionswirkungen kommen kann (FUNDER & COLVIN, 1991). Der Zusammenhang kann inhaltlich-qualitative Gründe haben, da sich die 'positiven' und 'negativen' Pole eines bipolaren Konstruktes im Spontanausdruck unterscheiden und dadurch in unterschiedlichem Maße beobachtbar sein können (PAUNONEN, 1989), resultiert aber auch aus der simplen Tatsache, daß häufiger auftretende Verhaltensweisen wahrscheinlicher wahrgenommen und bewertet werden können.

FUNDER & DOBROTH (1987) operationalisierten die Beurteilbarkeit verschiedener Persönlichkeitseigenschaften sehr aufwendig über sieben einzelne Einschätzungen: Imaginability of (dis)conforming Behaviors, Frequency of Occasions allowing (dis)conforming Behaviors, Number of behavioral Instances required to (dis)conform Trait und Easiness to judge. Alle Werte korrelierten mittel bis hoch mit der Accuracy der Einschätzung der Traits (die als Intraklassenkorrelation über eine Selbst- und zwei Fremdeinschätzungen ermittelt worden war) und wirkten somit als Moderatoren der Urteilsübereinstimmung. Die ebenfalls miterhobene Favorability der Persönlichkeitseigenschaften hatte keinen Einfluß auf ihre Beurteilungsgüte, hing interessanterweise aber positiv mit der geschätzten Easiness to Judge zusammen. In anderen Studien ergab sich dagegen eine positiv-lineare Beziehung zwischen der eingeschätzten Attraktivität einer Eigenschaft und dem Ausmaß des Interjudge Agreement bei ihrer Beurteilung, i.e. die Übereinstimmung bei der Bewertung attraktiver Eigenschaften war größer (FUNDER & COLVIN, 1988, HAYES & DUNNING, 1997).

Der Einfluß der Beobachtbarkeit einer Eigenschaft auf die durchschnittliche Güte der Fremdbeurteilung dieser Eigenschaft ist nicht konstant, sondern interagiert mit dem Bekanntheitsgrad: Die Observability zu beurteilender Traits ist bei kaum bis gering miteinander bekannten Personen eine stärkere Determinante der Urteilsübereinstimmung als bei gut bis sehr gut bekannten Personen (PAUNONEN, 1989, WATSON, 1989).

3. Fragestellung und Hypothesen

3.1 Herleitung der bearbeiteten Fragestellung

Ziel der vorliegenden Untersuchung ist es zunächst, einen gewissen, passablen Übereinstimmungsgrad von Selbst- und Fremdeinschätzungen bei der laienpsychologischen Beurteilung der Häufigkeit depressiver Symptome nachzuweisen. Damit soll einerseits weitere Evidenz für die Sichtweise der positiven Psychologie geschaffen werden, welche die menschliche Fähigkeit zur Persönlichkeitsbeurteilung für insgesamt gut entwickelt hält (vgl. Abschnitt 2.1.4), andererseits sollen dadurch weitere Belege für die Validität des Erfassungsinstruments der Beurteilungen, des BDI-V, gesammelt werden.

Die interessante Frage, ob Selbsteinschätzungen oder Fremdeinschätzungen der Ausprägung bzw. Häufigkeit depressiver Symptome tendenziell zutreffender sind, kann in dieser Untersuchung leider nicht geklärt werden, da der Aufwand der Erfassung der dazu nötigen unabhängigen Verhaltenskriterien den Rahmen einer Diplomarbeit gesprengt hätte. Über den Vergleich der durchschnittlichen Selbst- und Fremdeinschätzungen soll jedoch geprüft werden, ob es eine generelle Abweichung der Selbstbeurteiler in Richtung einer weniger depressiven Einschätzung gibt, was ein Hinweis auf einen generell wirksamen selbstwertdienlichen Bias wäre (vgl. Abschnitt 2.1.5).

Darüber hinaus soll der Einfluß mehrerer Persönlichkeits- und Umgebungsvariablen auf die Urteilsübereinstimmung getestet werden, die in der vorliegenden Arbeit aus pragmatischen Gründen trotz gewisser theoretischer Vorbehalte (vgl. Abschnitt 2.1.2) als direktes Kriterium der Urteilsgüte genutzt wird. Die Auswahl der Variablen orientiert sich an den vier Feldern von Moderatoren der Urteilspräzision, die das Realistic Accuracy Model beschreibt. Für den Judge (im konkreten Fall: den BDI-V Fremdbeurteiler), das Target (im konkreten Fall: den BDI-V Selbstbeurteiler), die von ihnen geteilte Informationsbasis und den Trait (im konkreten Fall: die einzelnen Verhaltens- und Erlebensweisen, die in den Items des BDI-V thematisiert werden, vgl. Abschnitt 4.1.1) wurden potentielle Moderatoren ausgewählt, die aus den Anforderungen der Teilprozesse von RAM ableitbar sind, zu denen jedoch nach Möglichkeit noch keine empirischen Ergebnisse vorliegen sollten.

In den folgenden Abschnitten 3.1.1 bis 3.1.4 wird die konkrete Auswahl dieser zu testenden Moderatorvariablen näher erläutert. Danach werden im Abschnitt 3.2 noch einmal alle Untersuchungsfragen als Hypothesen zusammengefaßt dargestellt.

3.1.1 Judge: Offenheit, Empathie, Perspektivübernahmetendenz, Alter

Urteilsqualität hängt laut RAM auf Seiten des Beurteilers von Motivation, Knowledge und Ability ab (vgl. Abschnitt 2.3.2.1). Für die vorliegende Untersuchung wurden aus dem Bereich der Motivation Offenheit und Empathie und aus dem Überschneidungsbereich von Motivation und Ability die Tendenz zur Perspektivübernahme als zu testende Konstrukte ausgewählt. Zu allen scheint es bisher keine empirischen Ergebnisse zu geben, aus RAM läßt sich jedoch ableiten, daß hohe Ausprägungen aller drei Eigenschaften jeweils mit höherer Urteilsgüte einhergehen sollten (vgl. Abschnitt 2.3.2.2).

Offenheit des Denkens sollte im Gegensatz zu rigiden, dogmatischen Denkmustern dazu motivieren, anderen Menschen vorurteilsfrei, aufgeschlossen und mit Interesse zu begegnen und sie weniger stereotyp zu beurteilen.

Empathie und Perspektiv- bzw. Rollenübernahme sind verwandte Vorgänge, lassen sich aber konzeptuell klar voneinander abgrenzen. Perspektivübernahme ist eine kognitive Leistung, bei der neben der Sichtweise des anderen zwar auch dessen Gefühle 'verstanden' werden, jedoch nicht selbst empfunden. Empathie basiert auf einer gelungenen Rollenübernahme, vollzieht aber auch die emotionale Lage des anderen nach, so daß ähnliche Emotionen erlebt werden. Und das auch dann, wenn die eigene emotionale Reaktion in der selben Situation eigentlich anders ausfallen würde. Empathie hat insofern kognitive und emotionale Elemente als notwendige, alleine nicht hinreichende Bedingungen und ist eine affektive Leistung (SCHMITT, 1982). Beide Konzepte gehören im vorliegenden Zusammenhang zunächst einmal zu den Abilities. Erfasst wird jedoch nicht die schiere Fähigkeit zur Perspektivübernahme und darauf folgender Empathie, sondern das habituelle Ausmaß, in dem die Beurteiler empathisch empfinden und ihre Neigung, die Perspektive zu wechseln. Als Neigungen und Verhaltenstendenzen fallen sie dann in den Bereich der Motivation. Beide sollten realistischere und detailliertere Urteile fördern und damit die Accuracy erhöhen.

Für den schwierig zu erfassenden Bereich der Knowledge wird stellvertretend der Einfluß des Alters des Judges untersucht. Mit dem Alter eines Beurteilers sollte das Ausmaß seiner sozialen Erfahrung zusammenhängen, damit wiederum die Komplexität seiner impliziten Persönlichkeitstheorien. Der über diesen Umweg erfaßte Einfluß der Knowledge auf die Urteilsgüte dürfte kaum noch meßbar sein, tendenziell sollte das Alter jedoch positiv mit Urteilspräzision zusammenhängen.

3.1.2 Target: Privatheit, Konsistenz, Selbst- und Fremdtäuschungstendenz

Urteilsqualität hängt laut RAM auf Seiten des Beurteilten von Relevance und Availability ab (vgl. Abschnitt 2.3.3.1). Für die vorliegende Untersuchung wurde aus dem Bereich der Availability die Privatheit des Targets als zu testendes Konstrukt ausgewählt. Als Gegenpol von Extraversion sollte sie mit schlechterer Beurteilbarkeit und damit geringerer Urteilsübereinstimmung einhergehen, da Personen mit hoher Privatheit wenig von sich offenbaren und verschlossen sind. Der positive Zusammenhang von Extraversion und Accuracy wurde bereits empirisch belegt, der negative Zusammenhang von Privacy und Urteilsgüte bisher jedoch noch nicht (vgl. Abschnitt 2.3.3.2).

Für den Bereich Relevance wurde die Streuung der Selbstauskunft als potentieller Moderator der Urteilsübereinstimmung ausgewählt. Interindividuelle Unterschiede der intraindividuellen Itemvarianz reflektieren, wie einheitlich und geschlossen der Ausprägungsgrad depressiver Symptome jeweils ist. BAUMEISTER UND TICE (1988) nannten diese Eigenschaft Metatrait, hier soll sie etwas weniger hochtrabend als Konsistenz der Selbstauskunft bezeichnet werden. Hohe Konsistenz sollte mit besserer Beurteilbarkeit und damit höherer Urteilsübereinstimmung einhergehen (vgl. Abschnitt 2.3.3.2).

In der vorliegenden Studie gibt der Beurteilte gleichzeitig auch eine Selbsteinschätzung ab. Daher wirken sich auch Eigenschaften, die die Güte von Selbstbeurteilungen beeinflussen, auf die Urteilsübereinstimmung aus. Die Verzerrung von Selbstberichten aufgrund sozialer Erwünschtheit wurde in der Psychologie vielfach dokumentiert. Deshalb werden zwei Komponenten der sozialen Erwünschtheitstendenzen des Targets erfaßt: Die unbewußte Selbsttäuschung und die bewußte Fremdtäuschung.

Hohe Selbsttäuschungswerte des Targets sollten die Urteilsübereinstimmung senken, da solche Personen sich selbst in einem durchgehend günstigeren, i.e. weniger depressiven Licht sehen als Fremdbeobachter. Je höher die Selbsttäuschungswerte ausfallen, desto stärker sollte die positive Abweichung der Selbsteinschätzung vom Fremdurteil sein. Hohe Fremdtäuschungswerte des Targets sollten ebenfalls die Urteilsübereinstimmung senken, da die Verstellungstendenzen solcher Personen dazu führen sollten, daß depressive Verhaltensweisen weniger offen ausagiert werden, sondern eher verborgen bleiben und daher schlechter zu beurteilen sind. Gleichzeitig verfälschen Fremdtäuscher die Werte der Selbstauskunft, da sie diese eher an antizipierten Normen und sozialen Erwartungen ausrichten, als an ihren tatsächlichen Eigenschaften.

3.1.3 Information: Acquaintanceship

Urteilsqualität hängt laut RAM maßgeblich von Relevance und Availability persönlichkeitsbezogener Hinweise ab (vgl. Abschnitt 2.3.4.1). Je enger eine Bekanntschaft ist, desto mehr relevante Informationen sind einem Beurteiler zugänglich. Deshalb wurden einige Items generiert, die qualitative und quantitative Aspekte des Bekanntschaftsgrades messen sollen. Ein linearer Einfluß quantitativer Variablen wie Bekanntschaftsdauer und Begegnungshäufigkeit auf die Urteilsübereinstimmung konnte bisher noch nicht belegt werden (vgl. Abschnitt 2.3.4.2), dennoch werden diese Variablen erneut getestet. Klare Ergebnisse zeigten verschiedene kategoriale Variablen des Bekanntschaftsgrades (vgl. Abschnitt 2.3.4.2), über deren Ausprägungen z.B. Subgruppen definiert und auf Unterschiede der Urteilsübereinstimmung untersucht werden können.

In der vorliegenden Untersuchung werden eine Globaleinschätzung des Bekanntschaftsgrades kategorial erfaßt sowie einige dichotome Variablen, die qualitative Aspekte der Verbindung beschreiben. Judge und Target sollen beispielsweise angeben, ob sie zusammen "sehr intensive Erlebnisse" hatten oder ob sie sich in "privaten Situationen" sehen. Weiterhin wird erfaßt, ob das Urteilerpaar liiert ist. Beziehungspartner sollten sich besonders gut kennen und daher besser einschätzen können. Diese These wurde bisher jedoch erst von einer Studie empirisch bestätigt (vgl. Abschnitt 2.3.4.2).

3.1.4 Trait: Beobachtbarkeit

Urteilsqualität hängt laut RAM auch von Relevance und Availability der zu beurteilenden Persönlichkeitseigenschaft und mit ihr assoziierter Verhaltensweisen ab (vgl. Abschnitt 2.3.5.1). Bei der Beurteilung der Big Five Persönlichkeitsdimensionen zeigen sich zwischen den einzelnen Domänen konstante Unterschiede der Urteilsübereinstimmung, die mit der Observability und Favorability der Eigenschaften korrespondieren (vgl. Abschnitt 2.3.5.2). In der vorliegenden Studie beziehen sich natürlich alle Fragen des Vergleichsinstruments BDI-V auf die gleiche Persönlichkeitsdimension, nämlich Eigenschaftsdepression. Da die Items aber eine Reihe von Verhaltens- und Erlebensweisen beschreiben, deren 'Innerlichkeit' differiert (vgl. Abschnitt 4.1.1, insb. Tabelle 3), scheint es dennoch theoretisch angemessen, die Beobachtbarkeit der BDI-V Items getrennt einschätzen zu lassen und die Korrespondenz dieser Werte mit der itemspezifischen Urteilsübereinstimmung zu untersuchen.

3.2. Hypothesen

Hypothese 1:

allgemeine Urteilsübereinstimmung

Es wird vermutet, daß die menschliche Urteils- und Beurteilungsfähigkeit generell gut entwickelt ist, die BDI-V Selbst- und Fremdbeurteiler sollten daher eine gute durchschnittliche Übereinstimmung erreichen.

Hypothese 2:

Empathie der Judges

Es wird vermutet, daß hohe habituelle Empathie für präzise Urteilsprozesse förderlich ist, hohe Empathiewerte seitens der Fremdbeurteiler sollten daher mit höherer Urteilsübereinstimmung einhergehen.

Hypothese 3:

Perspektivübernahme der Judges

Es wird vermutet, daß eine stärkere Neigung zur sozialen Perspektivübernahme für präzise Urteilsprozesse förderlich ist, hohe Perspektivübernahmewerte seitens der Fremdbeurteiler sollten daher mit höherer Urteilsübereinstimmung einhergehen.

Hypothese 4:

Offenheit der Judges

Es wird vermutet, daß eine größere Offenheit des Normen- und Wertesystems für präzise, vorurteilsfreie Urteilsprozesse förderlich ist, hohe Offenheitswerte seitens der Fremdbeurteiler sollten daher mit höherer Urteilsübereinstimmung einhergehen.

Hypothese 5:

Alter der Judges

Es wird ein schwacher positiver Zusammenhang des Alters der Fremdbeurteiler mit der Urteilsübereinstimmung vermutet, da die Angemessenheit der zur präzisen Persönlichkeitsbeurteilung nötigen impliziten Persönlichkeitstheorien mit dem Ausmaß sozialer Erfahrung zusammenhängen sollte und damit indirekt auch mit dem Lebensalter.

Hypothese 6:

Fremdtäuschung der Targets

Es wird vermutet, daß eine hohe Tendenz zur sozialen Erwünschtheit zur bewußten Verzerrung von Selbstberichten führt, hohe Fremdtäuschungswerte seitens der Selbstbeurteiler sollten daher mit niedrigerer Urteilsübereinstimmung einhergehen.

Hypothese 7: *Privatheit der Targets*

Es wird vermutet, daß verschlossene, zurückhaltende Menschen schwieriger zu beurteilen sind, hohe Privatheitswerte seitens der Selbstbeurteiler sollten daher mit niedrigerer Urteilsübereinstimmung einhergehen.

Hypothese 8: *Konsistenz der Targets*

Es wird vermutet, daß Menschen schwieriger zu beurteilen sind, deren Ausprägung depressiver Symptome inkonsistent und atypisch ist, höhere Varianz der Selbsteinschätzung sollte daher mit niedrigerer Urteilsübereinstimmung einhergehen.

Hypothese 9: *Selbsttäuschung der Targets*

Es wird vermutet, daß die Selbsteinschätzungen durch selbstwertdienliche Urteilsverzerrungen im Schnitt etwas positiver ausfallen als die Fremdeinschätzungen (genereller self-serving Bias), das Ausmaß der Abweichung sollte mit dem Ausmaß der Selbsttäuschungstendenz der Targets zusammenhängen (spezifischer self-serving Bias).

Hypothese 10: *Bekanntschaftsgrad*

Es wird vermutet, daß die Urteilsübereinstimmung mit zunehmender Enge der Bekanntschaft von Selbst- und Fremdbeurteiler zunimmt. Verschiedene quantitative und qualitative Aspekte des Bekanntschaftsgrades sollten daher jeweils positiv mit dem Grad der Urteilsübereinstimmung zusammenhängen.

Hypothese 11: *lierte vs. nicht-lierte Urteilerpaare*

Die engste Bekanntschaft ist in einer Liebesbeziehung realisiert. Es wird daher insbesondere vermutet, daß die Urteilsübereinstimmung im Durchschnitt höher ausfällt wenn Selbst- und Fremdbeurteiler liiert sind, als wenn das nicht der Fall ist.

Hypothese 12: *Beobachtbarkeit und Beurteilbarkeit spezif. Verhaltensweisen*

Es wird vermutet, daß die Fremdbeurteilbarkeit verschiedener Verhaltens- und Erlebensweisen unterschiedlich ist, daher sollte sich die durchschnittliche Urteilsübereinstimmung auf den einzelnen Items des BDI-V unterscheiden. Diese Unterschiede sollten mit der getrennt erhobenen Beobachtbarkeit der Items zusammenhängen.

4. Methodenteil

4.1 Verwendete Erhebungsinstrumente

4.1.1 Vergleichsskala: Das BDI-V

Die deutsche Kurzform des Beck-Depressions-Inventars (SCHMITT & MAES, 2000) wurde im Rahmen eines DFG-geförderten Längsschnittprojektes zu den psychologischen Folgen der Wiedervereinigung (Abschlußbericht von SCHMITT & MAES, 2001) entwickelt und eingesetzt. Das BDI ist ein weit verbreitetes Depressionsmaß, seine Reliabilität ist hoch, seine Validität gilt als belegt (RICHTER, 1991). Es erschien jedoch hinsichtlich seiner Nützlichkeit für breite epidemiologische Screenings noch optimierbar.

In der zugrundeliegenden Originalversion BDI-O (BECK & STEER, 1987, deutsche Fassung HAUZINGER, BAILER, WORALL & KELLER, 1994) werden, orientiert am DSM-III, insgesamt 21 Hauptsymptome der klinischen Depression abgefragt. Die Intensitätsskalierung erfolgt über je vier Aussagen zunehmender Schwierigkeit, deren zutreffendste angekreuzt werden soll. Dieses Format ist umständlich und unökonomisch, da das Lesen der insgesamt 84 Items einen hohen zeitlichen und kognitiven Aufwand erfordert. SCHMITT UND MAES (2000) behielten alle Symptome bis auf den wenig trennscharfen Gewichtsverlust bei, operationalisierten sie aber über jeweils nur eine Aussage. Die Symptomstärke wird nun über eine sechsstufige Häufigkeitsskala mit numerischer Stufenverankerung und Verbalverankerung der Extremwerte (nie/0-1-2-3-4-5/fast immer) ermittelt¹. Damit umfasst das BDI-V nur noch 20 Items und kann effizienter bearbeitet werden. Beide Fassungen des BDI messen überwiegend die persistente Eigenschaftsdepressivität, die im Gegensatz zur instabileren Zustandsdepressivität als Vergleichsgrundlage von Persönlichkeitseinschätzungen gut geeignet ist.

Der Begriff "Depression" hat seine Wurzeln in dem lateinischen Verb "deprimere", das "niederdrücken" bedeutet. Mit depressiven Stimmungen unterschiedlichen Ausmaßes assoziierte Symptome sind u.a. Traurigkeit, Hoffnungs- und Mutlosigkeit, starker Pessimismus bezüglich der Zukunft, geringes und/oder vermindertes Selbstbewußtsein und Selbstvertrauen, Minderwertigkeits- und Wertlosigkeitsgefühle, fehlende Lebensfreude, Schuldgefühle, Selbstanklagen, gedankliches Kreisen um den Tod, suizidale Ideen oder Akte, Weinerlichkeit, innere Unruhe, Gereiztheit, Verärgertheit.

¹ Durch einen unverzeihlichen Flüchtigkeitsfehler im Layout des Untersuchungsmaterials wurde die Symptomstärke mit einer fünfstufigen Skala (nie/1-2-3-4-5/fast immer) erfaßt. Dadurch wird ein Vergleich der Stichprobe mit den BDI-V Normwerten unmöglich. Die Validierung des Inventars durch den Vergleich der Selbst- und Fremdeinschätzungen ist unbetroffen, da alle Beurteiler die gleiche Skala erhielten. Eine fundamentale Änderung der Meßeigenschaften des BDI-V ist durch die Veränderung der Skala nicht zu erwarten (mündl. Mitteilung von Herrn Prof. Dr. Schmitt).

Weitere Kennzeichen depressiver Zustände sind Trägheit, fehlende Energie, Antriebs- und Interessenlosigkeit bis hin zur Apathie, Konzentrations- und Aufmerksamkeitsstörungen, Entscheidungserschwerern, Unentschlossenheit, sozialer Rückzug, Erschöpfung, Müdigkeit, geringerer oder erhöhter Schlafbedarf, Schlafstörungen, Appetitlosigkeit, verringerte sexueller Energie, Sorgen um das äußere Erscheinungsbild und den Gesundheitszustand, Hypochondrie (RUPPERT, 2003).

Jeder Mensch dürfte einige der aufgezählten Symptome aus seinem eigenen Erleben kennen, zumindest als episodischen Zustand, und nahezu jeder Mensch dürfte in seinem Umfeld jemanden haben, der aktuell an Depressionen leidet und entsprechendes Verhalten zeigt. Die Einschätzung solcher Stimmungen und Handlungsweisen entspricht insofern einer realistischen Anforderung der sozialen Interaktion des Alltags und erfüllt FUNDERS (1995, 1999) Forderung nach Erforschung umfassender, lebensnaher und relevanter laienpsychologischer Persönlichkeitsbeurteilungen. Die Items des BDI-V sind weit genug gefächert, um Urteile über depressive Persönlichkeitseigenschaften vielschichtig und differenziert abzubilden und damit vergleichbar zu machen.

Tabelle 3: Wortlaut der Items des BDI-V

Item	Selbstauskunft (Standard-Formulierung)	Fremdauskunft (Umformulierung)
BDI 01	Ich bin traurig.	Er/Sie ist traurig.
BDI 02	Ich sehe mutlos in die Zukunft.	Er/Sie sieht mutlos in die Zukunft.
BDI 03	Ich fühle mich als Versager.	Er/Sie fühlt sich als Versager(in).
BDI 04	Es fällt mir schwer, etwas zu genießen.	Es fällt ihm/ihr schwer, etwas zu genießen.
BDI 05	Ich habe Schuldgefühle.	Er/Sie hat Schuldgefühle.
BDI 06	Ich fühle mich bestraft.	Er/Sie fühlt sich bestraft.
BDI 07	Ich bin von mir enttäuscht.	Er/Sie ist von sich enttäuscht.
BDI 08	Ich werfe mir Fehler und Schwächen vor.	Er/Sie wirft sich Fehler und Schwächen vor.
BDI 09	Ich denke daran, mir etwas anzutun.	Er/Sie denkt daran, sich etwas anzutun.
BDI 10	Ich weine.	Er/Sie weint.
BDI 11	Ich fühle mich gereizt und verärgert.	Er/Sie fühlt sich gereizt und verärgert.
BDI 12	Mir fehlt das Interesse an Menschen.	Ihm/Ihr fehlt das Interesse an Menschen.
BDI 13	Ich schiebe Entscheidungen vor mir her.	Er/Sie schiebt Entscheidungen vor sich her.
BDI 14	Ich bin besorgt um mein Aussehen.	Er/Sie ist besorgt um sein/ihr Aussehen.
BDI 15	Ich muß mich zu Tätigkeiten zwingen.	Er/Sie muß sich zu jeder Tätigkeit zwingen.
BDI 16	Ich habe Schlafstörungen.	Er/Sie hat Schlafstörungen.
BDI 17	Ich bin müde und lustlos.	Er/Sie ist müde und lustlos.
BDI 18	Ich habe keinen Appetit.	Er/Sie hat keinen Appetit.
BDI 19	Sex ist mir gleichgültig.	Sex ist ihm/ihr gleichgültig.
BDI 20	Ich mache mir Sorgen um meine Gesundheit.	Er/Sie macht sich Sorgen um seine/ihre Gesundheit.

Anmerkung: Die vollständigen Fragebögen befinden sich im Anhang B (Fremdauskunft) und C (Selbstauskunft).

Meßgüte des Instruments: SCHMITT UND MAES (2000) untersuchten die Meßeigenschaften des BDI-V an einer sehr großen, heterogenen längsschnittlichen Stichprobe von 2500 Personen und fanden Belege für die einfaktorielle Struktur der Skala, eine hohe interne Konsistenz von $\alpha=.90$, eine hohe Reliabilität des Summenwertes in Latent-State-Trait-Analysen von .95 und eine Traitkonsistenz von .64 über den Zeitraum von zwei Jahren sowie eine niedrige Methodenspezifität (Testhälfstenspezifität) von nur .05. Die Konstruktvalidität des modifizierten Instruments zeigte sich in hohen negativen Korrelationen mit Indikatoren seelischer Gesundheit wie Selbstwertgefühl, Lebenszufriedenheit und retrospektiver Glücklichkeit.

SCHMITT, BECKMANN, DUSI, MAES, SCHILLER UND SCHONAUER (2003) konnten an einer demographisch heterogenen Gelegenheitsstichprobe und drei klinischen Stichproben erneut belegen, daß Homogenität und Reliabilität des BDI-V hoch sind (Cronbachs Alpha von .93), sogar etwas höher als beim BDI-O und anderen gängigen Depressionsmaßen. Die konvergente Validität des BDI-V zeigte sich in ausreichend hohen Korrelationen mit vier anderen Depressionsskalen sowie Fremdurteilen ausgebildeter und erfahrener Diagnostiker. Das BDI-V trennte zwischen der Normalstichprobe und den klinischen Stichproben ähnlich gut wie das BDI-O, beide Skalen konvergierten auf der Ebene der einzelnen Symptome.

In einer ganz aktuellen Studie wurden Normwerte für das BDI-V an einer bevölkerungsrepräsentativen Stichprobe von insgesamt über 4400 Personen ermittelt (SCHMITT, ALTSTÖTTER-GLEICH, HINZ, MAES & BRÄHLER, 2006). Die Ergebnisse sind konform mit allgemeinen Erkenntnissen zur Prävalenz der Depression: Frauen haben im Schnitt eine höhere Symptomausprägung als Männer, dieser Geschlechtseffekt interagiert zusätzlich mit dem Alter, die Diskrepanzen zwischen den Geschlechtern sind bei Jugendlichen unter 20 und Älteren über 60 am stärksten.

Trotz der vielfältigen Evidenzen für die Güte der Skala regten SCHMITT ET AL. (2003) eine weitere Validierung der neuen Kurzform des Beck-Depressions-Inventars durch den empirischen Vergleich von BDI-V Selbsteinschätzungen mit BDI-V Fremdeinschätzungen an: "Angehörigen und Freunden bleiben depressive Symptome in der Regel nicht verborgen. Selbstbericht-Fremdbericht-Korrelationen sollten deshalb eine Höhe erreichen, wie sie für Persönlichkeitsskalen typischerweise gefunden werden" (S. 155). Dieses Anliegen wird von der vorliegenden Untersuchung aufgegriffen.

4.1.2 Empathie / Perspektivübernahme: Skala von Maes, Schmitt und Schmal

Die Übernahme der Perspektive eines anderen ist ein kognitiver Prozeß. Der emotionale Vorgang des empathischen Mitfühlens, der Änderung des eigenen Erlebens in Richtung auf die vom anderen empfundenen Emotionen, beruht auf Perspektivübernahme. Könnte man das Ausmaß aktueller Perspektivübernahme und aktueller Empathie in situ messen, wären die Empathiewerte also immer abhängig. Für die Frage der Beurteilungskompetenz eines Judges interessiert jedoch nicht das Ausmaß der aktuellen, sondern der habituellen Empathie. Diese beinhaltet eine motivationale Komponente: Die für eine Person typische Bereitschaft, empathisch zu empfinden und zu agieren.

Zur Erfassung von Perspektivübernahme und Empathie des Fremdbeurteilers wurde eine im Rahmen des Projektes "Gerechtigkeit als innerdeutsches Problem" von MAES, SCHMITT UND SCHMAL (1995) angefertigte und eingesetzte deutsche Übersetzung und Modifizierung von Teilen des "Interpersonal Reactivity Index" (IRI) von DAVIS (1983) genutzt. Adaptiert wurden die beiden IRI-Subskalen "Perspective Taking" und "Empathic concern". Sie enthalten in der deutschen Version 9 bzw. 8 Items, die Verhaltens- und Empfindungsweisen innerhalb der sozialen Interaktion thematisieren, wie z.B. "Ich kann die Überlegungen anderer gut nachvollziehen" (Perspektivübernahme) oder "Ich bin mitfühlend gegenüber Menschen, die Probleme haben" (Empathie). Laut Instruktion sollen die Aussagen von den Untersuchungsteilnehmern auf einer fünfstufigen Likertskala mit den verbalen Ankern 1 "nie" und 5 "immer" danach eingestuft werden, wie selten oder oft die Person entsprechend handelt oder empfindet. Der vollständige Fragebogen ist im Anhang B dokumentiert. Die genaue Definition der beiden Konstrukte nach MAES ET AL. (1995, S. 35) lautet:

Perspective Taking repräsentiert die Neigung einer Person, eine Angelegenheit (auch) mit den Augen des Gegenüber zu betrachten.

Empathic Concern steht für die Bereitschaft, sich um die Gefühle und Bedürfnisse anderer zu sorgen.

Meßgüte des Instruments: Leider machen MAES ET AL. (1995) keine Angaben zur Reliabilität oder Validität der neuen deutschen Version des IRI. Für das amerikanische Original berichtet DAVIS (1983) interne Konsistenzen der Subskalen zwischen $\alpha=.71$ und $\alpha=.77$ und etwas niedrigere Retest-Reliabilitäten. In der vorliegenden Untersuchung erzielte die Empathieskala eine interne Konsistenz von $\alpha=.83$ und die Perspektivübernahmeskala von $\alpha=.84$, was als gute Werte angesehen werden können.

4.1.3 Offenheit des Wertesystems: Skala O6 des NEO-PI-R

Zur Erfassung der Offenheit des Wertesystems des Fremdbeurteilers wurde die Skala O6 der deutschen Version des NEO-PI-R (OSTENDORF & ANGLEITNER, 2004) eingesetzt. Dieser mehrdimensionale Persönlichkeitsfragebogen basiert auf dem "Revised NEO Personality Inventory" von COSTA UND MCCRAE (1992) und dient der differenzierten Erfassung der Merkmalsbereiche des Fünf-Faktoren-Modells (Neurotizismus, Extraversion, Offenheit, Verträglichkeit und Gewissenhaftigkeit). Jede der fünf Hauptskalen des hierarchischen Persönlichkeitssystems wird durch sechs Facettenskalen zu je acht Items definiert, die Verhaltensorientierungen im zwischenmenschlichen Bereich und interindividuelle Unterschiede im Einstellungs-, Erlebens- und Motivationsbereich beschreiben.

Die Skala O6 mißt eine Facette des Faktors "Offenheit für Erfahrung", und zwar die "Offenheit des Normen- und Wertesystems", definiert als die Bereitschaft, soziale, politische und religiöse Werte kritisch zu hinterfragen. Offenheit der Werte kann laut OSTENDORF UND ANGLEITNER (2004) als Gegensatz von Dogmatismus, Traditionalismus und Konservatismus angesehen werden.

Die Items bestehen aus Aussagen wie "Ich glaube, daß die Treue zu den eigenen Idealen wichtiger ist als Aufgeschlossenheit" oder "Ich meine, daß sich die Gesetze und die Sozialpolitik ändern sollten, um den Bedürfnissen der sich ändernden Welt Rechnung zu tragen" und sollen auf einer fünfstufigen Skala mit den Bezeichnungen "völlige Ablehnung / Ablehnung / neutral / Zustimmung / völlige Zustimmung" eingestuft werden. Der vollständige Fragebogen ist im Anhang B dokumentiert.

Meßgüte des Instruments: Die interne Konsistenz der Facette O6 des NEO-PI-R beträgt $\alpha = .85$ und ist damit an und für sich gut. Eine Einschränkung der Testgüte könnte im vorliegenden Fall allerdings dadurch verursacht werden, daß die acht Items der Skala in der Originalversion des Fragebogens verstreut über insgesamt 241 Items auf 11 Seiten dargeboten werden, in dieser Untersuchung jedoch massiert auf nur einem Blatt. Die durch die große inhaltliche Nähe der Items drohende Redundanz könnte stereotype Antwortmuster, Testübungseffekte u.Ä. begünstigen. Tatsächlich erzielte die massierte Offenheitsskala in der vorliegenden Untersuchung mit lediglich $\alpha = .51$ einen niedrigen Konsistenz-Wert.

4.1.4 Soziale Erwünschtheit: Balanced Inventory of Desirable Responding

Zur differenzierten Erfassung sozialer Erwünschtheitstendenzen des Selbstbeurteilers wurde die deutsche Kurzfassung des "Balanced Inventory of Desirable Responding" (MUSCH, BROCKHAUS & BRÖDER, 2002) eingesetzt. Der deutschen Fassung liegt die sechste englische Version des BIDR von PAULHUS (1994) zugrunde, die zwei unabhängige Faktoren sozialer Erwünschtheit erfaßt: "Self-Deceptive Enhancement" (Selbsttäuschung) und "Impression Management" (Fremdtäuschung). Mit Selbsttäuschung ist ein selbstwertdienliches Antwortverhalten gemeint, das die Realität und die eigene Person in optimistischer Weise verzerrt wiedergibt. Selbsttäuschung ist als unbewußter Prozeß anzusehen, im Gegensatz zur Fremdtäuschung, die ein bewußt verfälschendes Antwortverhalten mit dem Ziel der günstigeren Außendarstellung ist.

PAULHUS (1994) legt den Begriff Impression Management damit sehr restriktiv aus: Im gängigen sozialpsychologischen Ansatz wird Impression Management als Strategiebündel verschiedener Taktiken der Eindruckssteuerung (*engl.* image control) in der Interaktion mit der sozialen Umwelt beschrieben. Einige dieser Strategien sind automatisiert und laufen unbewußt ohne ausdrückliche Täuschungsabsicht ab (FREY & IRLE, 2002).

Das BIDR umfaßt im amerikanischen Original zwei Skalen mit je 20 Items. Die Fremdtäuschungsskala besteht aus positiv gepolten Items, in denen die Befragten als wünschenswert konnotierte, tatsächlich aber nur wenig verbreitete Eigenschaften für sich in Anspruch nehmen, und aus negativ gepolten Items, in denen die Befragten unerwünschte, aber weit verbreitete Einstellungen und Handlungen zugeben. Die Selbsttäuschungsskala ist inhaltlich etwas heterogener. Sie umfaßt positiv gepolte Items, die übertriebene Kontroll-Wahrnehmung und dogmatische Überzuversicht anzeigen und negativ gepolte Items, die eigene Unzulänglichkeiten und Zweifel thematisieren.

Die relativ geringe Korrelation zwischen den beiden Skalen ($r=.20$ bis $r=.32$ in Validierungsstudien des englischen Originals) ist ein Beleg für die Unabhängigkeit der beiden Faktoren Selbst- und Fremdtäuschung (PAULHUS, 1994). Die diskriminante Validität der Konstrukte wird durch unterschiedliche, spezifische Korrelationsmuster zu anderen Konstrukten belegt, z.B. korreliert Fremdtäuschung deutlich höher mit traditionellen Lügenskalen, während Selbsttäuschung einen stärker negativen Zusammenhang mit Neurotizismus und einen stärker positiven mit Selbstwertgefühl und psychologischer Anpassung hat (PAULHUS, BRUCE & TRAPNELL, 1995).

Der große Vorteil des BIDR ist, daß er die beiden Komponenten sozialer Erwünschtheit empirisch abgrenzbar macht und damit deren konzeptuelle Trennung rechtfertigt. Das klassische Konstrukt der sozialen Erwünschtheit entspricht bei PAULHUS (1994) inhaltlich der Fremdtäuschung, die das Ausmaß markiert, in dem sich Menschen bewußt in einem positiveren Licht darstellen. Fremdtäuschung variiert situativ. Selbsttäuschung ist dagegen verhältnismäßig stabil. Sie ist ein unbewußter Prozeß der Schönfärbung. Personen mit hohen Selbsttäuschungswerten sind tendenziell emotional stabiler und ausgeglichener: Die Selbsttäuschung ist Teil der psychologischen Anpassung, auch wenn oder gerade weil sie zu verschiedenen egozentrischen und selbstwertdienlichen Verzerrungen der Informationsverarbeitung führt.

Für die deutsche Fassung übersetzten MUSCH ET AL. (2002) alle 40 Items und unterwarfen sie einer empirischen Itemselektion anhand der Itemschwierigkeit, der korrigierten Trennschärfe und der faktorenanalytischen Ladungsstruktur nach Varimax-Rotation. Übrig blieben nur die trennschärfsten, am reinsten auf den beiden Hauptfaktoren ladenden Items mittlerer Schwierigkeit, insgesamt 10 je Skala. Die siebenstufige Likertskala mit den Ankern 1 "völlige Ablehnung" und 7 "völlige Zustimmung" wurde vom Original übernommen. Der vollständige Fragebogen ist im Anhang C dokumentiert.

Die Fremdtäuschungskomponente des BIDR scheint gut zur Kontrolle von Verfälschungs- und Verstellungstendenzen in der Selbstauskunft geeignet. Die Selbsttäuschungskomponente mißt Verzerrungstendenzen. Sie entspricht dem in der Sozialpsychologie breit diskutierten self-serving Bias (vgl. Abschnitt 2.1.4).

Meßgüte des Instruments: In drei größeren Studien konnten MUSCH ET AL. (2002) auch für die deutsche Kurzfassung des BIDR eine gute konvergente Validität der beiden Skalen mit verwandten Konstrukten (Fremdtäuschung mit Lügenskalen, Selbsttäuschung mit Selbstwertgefühlsskalen) und eine gute diskriminante Validität der BIDR Skalen gegeneinander ($r = .13$) belegen. Die Struktur der Itemladungen nach Faktorenanalyse belegt die Unabhängigkeit der beiden Faktoren Fremdtäuschung und Selbsttäuschung. Die internen Konsistenzen der beiden Subskalen ($\alpha = .64$ bzw. $\alpha = .66$) sind mäßig, erscheinen laut MUSCH ET AL. aber "noch vertretbar" (2002, S. 127). In der vorliegenden Untersuchung erzielte die Selbsttäuschungsskala eine niedrigere interne Konsistenz von nur $\alpha = .49$, die Fremdtäuschungsskala erreichte immerhin $\alpha = .68$.

4.1.5 Privatheit: Skala N der deutschen Ausgabe des 16 PF-R

Zur Erfassung der Offenheit und Expressivität der Zielperson wurden einige Items der Skala N (Privatheit) der deutschen Ausgabe des 16 PF-R (SCHNEEWIND & GRAF, 1998) genutzt. Die Persönlichkeitsdimension Privatheit ist in diesem Inventar definiert als:

Die Neigung, anderen Menschen gegenüber geradeheraus und offen oder aber verschlossen und wenig selbstöffnungsbereit zu sein.

Das Konstrukt Privatheit steht in Verbindung zur Globalskala Extraversion des 16 PF-R, es bildet zusammen mit Reserviertheit, Schüchternheit und Selbstgenügsamkeit den Gegenpol der Skala, nämlich Introversion. Der besondere Fokus von Privatheit liegt dabei auf der persönlichen Öffnung anderen gegenüber, nicht auf den allgemeineren Aspekten wie Gesprächigkeit und Geselligkeit (SCHNEEWIND & GRAF, 1998).

Die Skala N umfasst im 16 PF-R insgesamt 11 Items, die über den gesamten Pool von 184 Items verteilt sind. Um in der vorliegenden Arbeit eine zu starke Redundanz bei der massierten Abfrage zu vermeiden, wurden lediglich 7 Items entnommen. Als Auswahlkriterien dienten dabei einerseits der Schwierigkeitsindex, um eine gute Streuung der Schwierigkeit über die verbleibenden Items zu gewährleisten, andererseits die Augenscheinvalidität, um sprachlich und inhaltlich nahezu identische Items zu eliminieren.

Die Fragen der Skala beziehen sich darauf, wie viel man bereit ist im Freundeskreis von sich preiszugeben, wie sehr man auf andere zugeht und wie offen man sich generell charakterisiert. Durch die Instruktion wird die Auswahl einer von zwei gegensätzlichen Antworten bzw. Aussagen forciert. Eine mittlere, neutrale Kategorie soll nur dann angekreuzt werden, wenn keine der beiden Aussagen besser passt. Der vollständige Fragebogen ist im Anhang C dokumentiert.

Meßgüte des Instruments: Die Originalskala hat im Rahmen des kompletten 16 PF-R mit $\alpha = .74$ eine befriedigende interne Konsistenz. Die Konstruktvalidität ist belegt durch korrelative Vergleiche mit dem NEO-Fünf-Faktoren-Inventar (NEO-FFI), dem Trierer Persönlichkeitsfragebogen (TRF-2) und anderen bekannten und bewährten Inventaren (SCHNEEWIND ET AL., 1998). Obwohl die Items in der vorliegenden Untersuchung zusammenhängend auf einer Seite dargeboten wurden und nicht über einen längeren Fragebogen verteilt, erzielte die massierte Privatheitsskala eine ebenfalls befriedigende interne Konsistenz von $\alpha = .70$.

4.1.6 Acquaintanceship: Aspekte des Bekanntschaftsgrades und Umgangs

Zur Erfassung qualitativer und quantitativer Aspekte der Beziehung zwischen Judge und Target dienten einige neu generierte Items, die beiden Beurteilern in identischer Form vorgelegt wurden und zusammen mit den demographischen Angaben die jeweils letzte Seite des Fragebogenheftes bildeten. Im Anhang B bzw. C sind die entsprechenden Blätter vollständig dokumentiert.

Quantitative Aspekte:

Sowohl der Fremdbeurteiler als auch der Selbstbeurteiler wurden gebeten, die Dauer der gemeinsamen Bekanntschaft in Jahren zu beziffern. Für die weitere Analyse wurde der Mittelwert der beiden Angaben genutzt. Da für die Bekanntschaftsdauer und die Urteilsübereinstimmung nur ein schwacher direkter linearer Zusammenhang erwartet wurde, wurde eine zusätzliche kategoriale Gruppierungsvariable erzeugt, die die Bekanntschaftsdauer in "weniger als ein Jahr / ein bis sechs Jahre / sechs bis 21 Jahre / über 21 Jahre" einteilte. Diese Kategorien entsprechen einem Quartilssplit der Stichprobe und erzeugen vier gleich große Gruppen, deren durchschnittliche Urteilsübereinstimmung dann verglichen werden sollte. Die Frequenz des Umgangs wurde mit fünf Kategorien erfaßt: "Mehrmals täglich / täglich / mehrmals wöchentlich / 1-2x wöchentlich / seltener". Bei dieser und allen weiteren kategorialen Variablen wurden abweichende Angaben der beiden Beurteiler sinnvoll verknüpft.

Qualitative Aspekte:

Der Status der Verbindung sollte einer der Kategorien, "(bluts)verwandt / liiert / befreundet / bekannt" zugeordnet werden. Zur globalen Messung der Bekanntschaftsenge diente die Aussage "Ich kenne die andere Person meiner Meinung nach..." mit den Multiple Choice Antwortoptionen "sehr gut / gut / kaum". Außerdem wurden die Untersuchungsteilnehmer gebeten, drei qualitative Fragen über die Bekanntschaft mit "Ja" oder "Nein" zu beantworten, nämlich ob es "sehr intensive gemeinsame Erlebnisse" gab, ob man sich in "privaten" und ob man sich in "alltäglichen, strukturierten, eher oberflächlichen" Situationen begegnet.

Im Abschnitt 5.5 befinden sich im Zusammenhang der Ergebnisdarstellung des Moderators Acquaintanceship detailliertere Angaben zur jeweiligen Verrechnung der Selbst- und Fremdanfragen und zur Verteilung der Werte in der Stichprobe.

4.1.7 Beobachtbarkeit: gemittelte Augenscheinvalidität

PAUNONEN (1989) weist ausdrücklich darauf hin, daß die Beobachtbarkeit von Persönlichkeitseigenschaften als eigenes Merkmal getrennt erfaßt werden muß, unabhängig von den Einschätzungen des Ausprägungsgrades dieser Eigenschaften bei einzelnen Merkmalsträgern einer Stichprobe. Der Einwand scheint überflüssig, in den achtziger Jahren wurden jedoch einige Studien durchgeführt, bei denen die konkrete Beobachtbarkeit spezifischer Persönlichkeitseigenschaften eines konkreten Targets als Personenmerkmal erhoben und anschließend als Moderator der Urteilsübereinstimmung getestet wurde (z.B. KENRICK & STRINGFIELD, 1980, CHEEK, 1982, KAMMANN ET AL., 1984). Auf solcher Basis postulierte Effekte oder Nicht-Effekte der Observability müssen mit Vorsicht interpretiert werden: Durch den u-förmigen Zusammenhang zwischen der Ausprägung einer Eigenschaft und ihrer Beobachtbarkeit (die extremen Ausprägungen sind salient, im Gegensatz zu den mittleren) kann es zu Konfundierungen kommen.

Deshalb wurde in der vorliegenden Studie ein gesonderter Online-Fragebogen eingesetzt, um die ausprägungsunabhängige, generelle Beobachtbarkeit der einzelnen Verhaltens- bzw. Erlebensweisen des BDI-V von einer unabhängigen Stichprobe einschätzen zu lassen. Dabei war auch zu beachten, daß die beiden Pole eines bipolaren Konstruktes in unterschiedlichem Maße beobachtbar sein können (PAUNONEN, 1989). Daher wurden in dem Fragebogen, der im Anhang D dokumentiert ist, alle 20 Items des BDI-V zunächst in der normalen Fremdform ('negativ' gepolt), dann noch einmal in inhaltlich entgegengesetzter Weise ('positiv' gepolt) dargeboten. Beispielsweise hieß es zuerst "Die Person sieht mutlos in die Zukunft" (Fremdform BDI-V) und später "Die Person sieht mutvoll in die Zukunft" oder zuerst "Die Person fühlt sich bestraft" (Fremdform BDI-V) und später "Die Person fühlt sich nicht bestraft".

Die Instruktion forderte die Untersuchungsteilnehmer auf, sich einen prototypischen guten Bekannten vorzustellen, keine Person aus dem engsten und intimsten Umfeld. Danach sollte jedes der insgesamt 40 Items auf einer neunstufigen Skala mit verbalen Ankern danach eingestuft werden, inwieweit die beschriebene Eigenschaft von außen sicht- und beobachtbar ist. Für die weitere Analyse wurden die mittleren Observability-Ratings der beiden Pole jedes BDI-V Items zum jeweiligen Observability-Gesamtwert des Items gemittelt. Diese Vorgehensweise wird von FUNDER UND DOBROTH (1987) vorgeschlagen. Angaben zur Meßgüte der Erfassung finden sich im Abschnitt 5.6 im Zusammenhang der Ergebnisdarstellung des Moderators Beobachtbarkeit.

4.2 Durchführung der Datenerhebung

Die Beobachtbarkeit der einzelnen Verhaltens- bzw. Erlebensweisen des BDI-V wurde gesondert mit einem Online-Fragebogen erhoben (vgl. Abschnitt 4.1.7 bzw. Anhang D), auf den in verschiedenen studentischen Foren und an anderen Orten im Internet aufmerksam gemacht wurde mit der Bitte um Unterstützung. Die Resonanz war sehr gut. Die Stichprobe von über 200 Personen kam innerhalb von nur einer Woche zusammen. Da sich bei der fortlaufenden Berechnung der Observability-Kennwerte ab der Menge von 100 Untersuchungsteilnehmern nur noch marginalste Verschiebungen zeigten, wurde die Erhebung nach 7 Tagen geschlossen.

Die Hauptstichprobe mit den BDI-V Selbst- und Fremdeinschätzungen wurde in Landau aus der Studierenden- und Mitarbeiterschaft der Universität rekrutiert, aus einer Abteilung der DPA in Hamburg und zuletzt aus dem weiteren Bekanntenkreis des Autors in Hamburg. Ein Materialsatz beinhaltete neben einer Hauptinstruktion (vgl. Anhang A) und den Fragebögen (vgl. Anhang B bzw. C) zwei Briefumschläge der Standardgröße DIN lang, in die jeweils die ausgefüllten Fragebögen von Judge bzw. Target verpackt werden sollten und einen weiteren Umschlag der Größe DIN C4, in dem dann das komplette Versuchsmaterial zurück gegeben wurde. Dieses Verfahren sollte Verzerrungen im Antwortverhalten vorbeugen, indem es die Uneinsehbarkeit der Antworten des Judges für das Target bzw. der Antworten des Targets für den Judge sicher stellte.

Die Abgabe der ausgefüllten Fragebögen im Umfeld der Universität Landau erfolgte persönlich an den Autor, die Rücklaufquote betrug hierbei ca. 75%. Das Versuchsmaterial in Hamburg wurde mit frankierten Rückumschlägen ausgeteilt und dem Autor später postalisch zugestellt, die Rücklaufquote betrug ca. 95%. Vermutlich wurde die Compliance der Untersuchungsteilnehmer durch die Vorfrankierung zusätzlich erhöht. Die Erhebung erstreckte sich insgesamt über einen Zeitraum von vier Monaten.

4.3 Stichprobenbeschreibung

Den Online-Fragebogen zur Beobachtbarkeit der Items füllten insgesamt N=219 Personen aus, darunter 93 Frauen (ca. 43%) und 122 Männer (ca. 57%). Von den Untersuchungsteilnehmern waren 173 ledig oder geschieden (ca. 79%), 39 waren verheiratet (ca. 18%) und 7 verwitwet (ca. 3%). Das Durchschnittsalter der Teilnehmer betrug 30 Jahre, der Altersrange erstreckte sich von 17 bis zu 64 Jahren.

Die Fragebögen mit den BDI-V Selbst- und Fremdeinschätzungen wurden von insgesamt 177 Urteilerpaaren bzw. N=354 Personen ausgefüllt, darunter waren 205 Frauen (ca. 58%) und 149 Männer (ca. 42%). Das Durchschnittsalter der Untersuchungsteilnehmer betrug 36 Jahre, der Altersrange erstreckte sich von 17 bis 90 Jahren. Von den Untersuchungsteilnehmern waren 238 ledig oder geschieden (ca. 67%), 115 waren verheiratet (ca. 33%) und einer verwitwet (< 1%).

Von den 177 Urteilerpaaren waren 104 liiert (ca. 59%), 11 blutsverwandt (ca. 6%) und 62 miteinander bekannt oder befreundet (ca. 35%). Die Urteilerpaare kannten sich im Schnitt seit etwa 13 Jahren, der Range der Bekanntschaftsdauer erstreckte sich dabei von 2 Monaten bis zu 60 Jahren.

4.4 Auswertungsstrategien

4.4.1 Zwei Ebenen der Analyse: Personen-Level und Item-Level

Die Urteilsübereinstimmung von Selbst- und Fremdeinschätzungen kann mit zwei grundsätzlich verschiedenen Analyseansätzen untersucht werden, deren Datenstrukturen orthogonal zueinander sind. Sie ermöglichen die Untersuchung verschiedener Aspekte der Übereinstimmung und damit unterschiedlich akzentuierter Fragestellungen.

Personen-Level:

Die Übereinstimmung wird für jedes Urteilerpaar über alle Items ermittelt. Dieser Wert wird als "within-dyad accuracy" oder "profile accuracy" bezeichnet. Profile Accuracy beziffert die dyadische Übereinstimmung eines Fremd- und eines Selbsturteils mit einem Gesamtwert. Aussagen über den relativen Beitrag der einzelnen Items zur Urteilspräzision können nicht getroffen werden.

Item-Level:

Die Übereinstimmung wird für jedes Item über alle Urteilerpaare ermittelt. Dieser Wert wird als "between-dyad accuracy" oder "trait accuracy" bezeichnet. Trait Accuracy beziffert die durchschnittliche Übereinstimmung der Urteilerpaare bei der Einschätzung eines spezifischen Items. Neben der Identifikation und Analyse genereller Unterschiede in der Beurteilbarkeit der Items (durch den Vergleich ihrer Trait Accuracy) kann auf Item-Level auch der Einfluß personenbezogener Moderatoren überprüft werden (vgl. Abschnitt 4.4.2). Aussagen über die Urteilspräzision einzelner Beurteilerpaare können nicht getroffen werden.

4.4.2 Moderatorvariablenansatz

Eine Moderatorvariable ist ganz allgemein ausgedrückt "eine Variable, von deren Ausprägung Form und/oder Enge von Prädiktor-Kriteriums-Relationen abhängen" (DALBERT & SCHMITT, 1986, S. 30). Die in dieser Studie untersuchte Relation ist der Zusammenhang von Selbst- und Fremdeinschätzungen, die Urteilsübereinstimmung. Je nachdem, ob sie auf Personen- oder Item-Level analysiert werden soll, ergeben sich verschiedene Möglichkeiten zur Prüfung von Moderatoreffekten. Der erste Ansatz ist nomothetisch, der zweite idiographisch (CHAPLIN, 1991, S.150f).

Personen-Level:

Für jedes Urteilerpaar werden Indizes berechnet, die das Ausmaß der dyadischen Übereinstimmung quantifizieren (vgl. Abschnitt 4.4.1). Handelt es sich bei dem zu untersuchenden Moderator um eine diskrete Variable, wird die Stichprobe anhand der Ausprägungen des Moderators in Subgruppen aufgeteilt. Anschließend werden die Übereinstimmungskoeffizienten der Urteilerpaare in allen Subgruppen getrennt gemittelt und die resultierenden durchschnittlichen Übereinstimmungen der Gruppen verglichen. Zeigen sich dabei signifikante Unterschiede zwischen den Subgruppen, ist das ein Hinweis auf die Wirkung von Moderatoreffekten.

Liegt der potentielle Moderator als kontinuierliche Variable vor, ist ebenfalls eine Subgruppenbildung möglich, z.B. per Median- oder Quartilsplit. Eine künstliche Dichotomisierung bzw. Diskretisierung bedeutet aber immer einen Informationsverlust. Sinnvoller ist daher die direkte Korrelation der Moderatorvariable mit den Indizes der dyadischen Übereinstimmung. Vorzeichen und Wert des resultierenden Bravais-Pearson-Korrelationskoeffizienten drücken Stärke und Richtung des Moderatoreffektes aus. Ist es theoretisch begründbar, können zusätzlich andere potentielle Moderatoren aus der Korrelation auspartialisiert werden, um den Zusammenhang deutlicher zu machen.

Item-Level:

Übereinstimmungswerte werden nicht für Urteilerpaare über Items, sondern für Items über Urteilerpaare berechnet (vgl. Abschnitt 4.4.1). Diskrete Moderatorvariablen können auf Item-Level ebenfalls über Subgruppenvergleiche geprüft werden, nur daß auf dieser Analyseebene nicht die durchschnittlichen dyadischen Übereinstimmungen in den Subgruppen verglichen werden, sondern die gruppenspezifischen Korrelationen der Selbst- und Fremdeinschätzungen jedes Items.

Bei dem Vergleich der Korrelationen innerhalb verschiedener Subgruppen über die Gruppen hinweg ist zu beachten, daß unterschiedliche Varianzen der Selbst- und Fremdeinschätzungen in den Gruppen dazu führen können, daß die Korrelationen verzerrt sind. Sie sollten daher vor dem Vergleich einer Minderungskorrektur unterzogen werden (DALBERT & SCHMITT, 1986). Zur Überprüfung diskreter Moderatorvariablen auf Item-Level läßt sich auch eine moderierte Regressionsanalyse durchführen, dabei können unterschiedliche Stichprobengrößen und Fehlervarianzen in den Subgruppen jedoch leicht zur Verschätzung der Stärke des Moderatoreffektes führen. Experimentell sind Gruppengrößen leicht kontrollierbar, in quasi-experimentellen Ansätzen sind sie aber typischerweise verschieden, daher ist hier der Vergleich der Korrelationen innerhalb der Gruppen vorzuziehen (FRAZIER, TIX & BARRON, 2004).

Angemessen ist die Methode der moderierten Regression zur Überprüfung der Wirkung kontinuierlicher Moderatorvariablen auf Item-Level. Geht es um die Analyse von Moderatoren seitens des Judges, werden in einer hierarchischen Regressionsanalyse zuerst die Selbsteinschätzungen auf einem Item (Kriterium) auf die Fremdeinschätzungen des gleichen Items (Prädiktor) regressiert, dann auf die Moderatorvariable und zuletzt auf das Kreuzprodukt von Prädiktor und Moderator. Wenn das Regressionsgewicht des Kreuzproduktes signifikant wird, es also zu einem signifikanten Zuwachs an Varianzaufklärung (i.e. einer signifikante Zunahme von R^2 nach Eingabe des Interaktionsterms) kommt, ist das ein Beleg für das Vorliegen eines Moderatoreffektes (CHAPLIN, 1991). Geht es um die Analyse von Moderatoren auf Seiten des Targets, werden die Selbsteinschätzungen zum Kriterium und die Fremdeinschätzungen zum Prädiktor.

Für Regressionsanalysen mit einem Interaktionsterm gilt, daß der resultierende unstandardisierte Regressionskoeffizient (B) höchster Ordnung invariant gegenüber additiv-linearen Transformationen der Ausgangsvariablen wie Zentrierung oder Standardisierung ist, wenn die Haupteffekte kontrolliert (i.e. konstant gehalten) werden (AIKEN & WEST, 1991). Die Semipartialkorrelation des Kreuzproduktes (bei Konstanthaltung von Prädiktor und Moderator) mit dem Kriterium ist invariant gegenüber linearen Transformationen von Prädiktor und Moderator, die einfache Korrelation des Produktes mit dem Kriterium ist es nicht. Eine Standardisierung der Rohdaten und Berechnung des Interaktionsterms auf Basis von z-Werten hat aber den Vorteil, daß das Problem der Multikollinearität von Prädiktor und Moderator entschärft wird (FRAZIER ET AL., 2004). Die B-Gewichte der Analyse entsprechen den standardisierten Regressionsgewichten.

4.4.3 Operationalisierung der Urteilsübereinstimmung

4.4.3.1 Zur Problematik einfacher Globalmaße

Die absolute Übereinstimmung intervallskalierter Antwortprofile kann sehr einfach über die mittlere Abweichung oder die mittlere quadrierte Abweichung zwischen den beiden Urteilern über alle Items operationalisiert werden. In diesem Wert sind jedoch vier Komponenten konfundiert, nämlich Elevation (konstanter Fehler, wörtlich: Höhenlage), Differential Elevation (Variabilitätsfehler, wörtlich: differentielle Höhenlage), Stereotype Accuracy (stereotype Genauigkeit) und Differential Accuracy (differentielle Genauigkeit). Unter der Voraussetzung, daß ein einzelner Judge mehrere Targets auf mehreren Items beurteilt hat und sich diese Targets auf den selben Items ebenfalls eingeschätzt haben, kann man die Komponenten rechnerisch trennen (CRONBACH, 1955).

Elevation reflektiert die zentrale Skalennutzungstendenz des Beurteilers, quasi seine 'Antwortstrenge' oder, anders ausgedrückt, seine individuelle Einschätzung dessen, was bei Persönlichkeitseigenschaften eine mittlere Ausprägung darstellt und wie eine gegebene Ausprägung auf der Antwortskala ausgedrückt werden soll. Rechnerisch ergibt sich die Elevation-Komponente aus der Differenz der Durchschnittswerte aller Fremdeinschätzungen des Judges und aller Selbsteinschätzungen der betreffenden Targets. Bei der Untersuchung von Urteilsübereinstimmungen soll aber im Regelfall nicht analysiert werden, ob Selbst- und Fremdeinschätzer die genutzte Skala gleich interpretieren, sondern ob sie die Persönlichkeit des Targets übereinstimmend bewerten. Der zu ermittelnde Übereinstimmungswert sollte also frei von Elevation-Effekten sein.

Differential Elevation ist die Differenz der Durchschnittswerte aller Einschätzungen des Beurteilers bzgl. eines spezifischen Targets und aller Selbsteinschätzungen dieses Targets. Die Komponente entspricht damit der über alle Items gemittelten absoluten Genauigkeit des Urteils in Bezug auf einzelne Beurteilte und man kann anhand des Wertes der Differential Elevation erkennen, ob einzelne Personen übermäßig streng oder zu milde beurteilt wurden. Hängt das Ausmaß der beurteilenspezifischen Antwortstrenge systematisch mit bestimmten Eigenschaften der Beurteilten wie Geschlecht sozialem Status oder ethnischer Herkunft zusammen, ist das ein Hinweis auf Vorurteile, die der Beurteiler gegenüber bestimmten Personengruppen hat. Bei einem Untersuchungsansatz, in dem ein Beurteiler genau eine Zielperson einschätzt, sind Elevation und Differential Elevation immer konfundiert und rechnerisch nicht trennbar.

Stereotype Accuracy ergibt sich aus der Differenz der Durchschnittswerte der Einschätzungen aller Targets bzgl. eines spezifischen Items durch den Judge und aller Selbsteinschätzungen aller Beurteilten auf diesem Item. Die Komponente ist ein Maß der Genauigkeit, mit der ein Beurteiler die Mittelwerte der Items getroffen hat, quasi die itemspezifische Antwortstrenge. TAFT (1955) bezeichnete Stereotype Accuracy als "mass empathy" (S. 16), BRONFENBRENNER, HARDING UND GALWAY als "sensitivity to the generalized other" (1958, zit. n. BERNIERI ET AL., 1994, S.368). Bei der Untersuchung von Urteilsübereinstimmungen soll im Regelfall nicht analysiert werden, welche Eigenschaften ein Fremdeinschätzer am populationskonformsten einschätzt, sondern ob die Persönlichkeitstruktur eines Targets insgesamt zutreffend bewertet wurde. Der zu ermittelnde Übereinstimmungswert sollte also frei von Stereotype Accuracy sein.

Differential Accuracy ist die Übereinstimmung der residualisierten Selbst- und Fremdeinschätzungen, aus deren Datenmatrizen das Grand Mean, die Item- und die Targetmittelwerte entfernt wurden. Damit ist sie ein Maß der Genauigkeit der Vorhersage des Judges, inwieweit die Antwort eines Targets auf ein spezifisches Item von der generellen Antworttendenz dieses Targets und der typischen Antworttendenz aller Targets auf diesem Item abweicht. Bei der Untersuchung von Urteilsübereinstimmungen hat Differential Accuracy im Regelfall die größte inhaltliche Relevanz, weil sie die strukturelle Genauigkeit des Urteils in Bezug auf einzelne Beurteilte beschreibt.

Fazit: Einfache Globalmaße konfundieren mehrere Varianzquellen, die unterschiedliche inhaltliche Bedeutung haben und teilweise Skalennutzungs- und Bewertungstendenzen (*engl.* response sets) widerspiegeln, die unabhängig von den Bewertungsobjekten sind und für die Analyse von Urteilsübereinstimmungen meistens irrelevant. Statt einfache Globalmaße zu nutzen, sollten die verschiedenen Übereinstimmungskomponenten in spezifischeren Maßen rechnerisch getrennt werden.

If we assume that all measures are scaled where more means something good, then in essence, elevation assesses the degree to which the perceiver is overly positive or negative, differential elevation assesses whether the perceiver knows which targets are good and bad, stereotype accuracy assesses whether the perceiver know people in general vary on the traits, and differential accuracy assesses whether the perceiver knows how the targets differ from each other on the traits. Most analysts treat differential elevation and differential accuracy as 'real' accuracy and the other two as artificial (KENNY, 1994, S. 122).

4.4.3.2 Auswahl spezifischer Maße

In der vorliegenden Studie soll die Urteilsgüte paarweiser Beurteilungen evaluiert werden: Ein Judge beurteilt immer genau ein Target und seine Einschätzung wird dann mit der Selbsteinschätzung dieses Targets verglichen. Bei einem solchen vereinfachten Untersuchungsansatz ist eine vollständig differenzierende Komponentenanalyse im Sinne Cronbachs nicht möglich (FUNDER, 1999). Spezifische Übereinstimmungsmaße können die Übereinstimmungskomponenten aber zumindest teilweise isolieren.

Die korrekte Einschätzung des Ausprägungsgrades der depressiven Symptome des BDI-V (i.e. der Profilhöhe) ist ein Aspekt der Urteilsgüte und sollte in einem Übereinstimmungsmaß Ausdruck finden (vgl. Abschnitt 2.1.3). Im Versuchparadigma sind die Komponenten Elevation und Differential Elevation aber zwangsläufig konfundiert, höhenlagenabhängige Übereinstimmungsmaße werden durch Response Sets und die urteilsobjektunabhängige, nicht meßbare Urteilsstrenge verfälscht. Daher müssen auch Maße ohne Elevation-Komponenten berechnet werden, die nur die Ähnlichkeit der Profilverläufe widerspiegeln, des relevanteren Ähnlichkeitsaspekts (vgl. Abschnitt 2.1.3). Durch den Vergleich der Werte kann der Einfluß der Höhenlage abgeschätzt werden.

Personen-Level:

Ein einfach interpretierbares Maß der dyadischen Übereinstimmung ist die Profilkorrelation der Selbst- und Fremdeinschätzungen eines Urteilerpaares über alle Items. Da korrelative Maße die Kovarianz von Variablen unabhängig von ihrem Mittelwert (hier: der mittleren Antworttendenz der Beurteiler) und der Varianz der beiden Profile quantifizieren, sind sie frei von Elevation-Komponenten. Sie messen jedoch neben Differential Accuracy auch Stereotype Accuracy bzw. Implicit Profile Accuracy: Es droht eine künstliche Erhöhung des Übereinstimmungswertes in dem Maße, wie das spezifische Target ein 'typisches' Persönlichkeitsprofil hat und der Judge sich dieses 'typischen' Persönlichkeitsprofils bewußt ist (BERNIERI ET AL., 1994, S. 369f). Zur statistischen Kontrolle kann zuerst das prototypische Profil über die mittleren Selbsteinschätzungen auf allen Items berechnet werden, anschließend die Profilkorrelation jeder Fremdeinschätzung mit diesem Prototyp zur Quantifizierung des spezifischen Varianzanteils der Implicit Profile Accuracy, um zuletzt diese spezifische Varianz aus der jeweiligen Profilkorrelationen der Selbst- und Fremdeinschätzungen aller Urteilerpaare zu entfernen. Diese residualisierte Profilkorrelation spiegelt wider, inwieweit Judge und Target in ihrer Abweichung von den prototypischen Beurteilungen übereinstimmen.

Der Nachteil der beschriebenen Methode ist, daß die Validität des berechneten prototypischen Profils von Umfang und Homogenität der Stichprobe abhängt. Da diese in der vorliegenden Untersuchung nur 177 Urteilerpaare umfaßt, wurde ein anderer Weg zur Bereinigung des Übereinstimmungsmaßes von Implicit Profile Accuracy gewählt, nämlich die Berechnung der Profilkorrelation, nachdem die Rohwerte der Selbst- und Fremdeinschätzungen über die Personen standardisiert wurden. Ein hoher Wert auf diesem Maß der ipsativen Profilkorrelation bedeutet eine hohe Übereinstimmung bezüglich der Ausprägungen der einzelnen BDI-V Erlebens- und Verhaltensweisen relativ zu den Durchschnittswerten der zugrundeliegenden Population und relativ zueinander in der Targetpersönlichkeit (BERNIERI ET AL., 1994).

Ein populäres Maß zur höhenlagenabhängigen Bestimmung der Ähnlichkeit von Antwortprofilen ist der Kongruenzkoeffizient c , der auf verschiedene Forscher zurück geht (BURT, 1948, TUCKER, 1951, WRIGLEY & NEUHAUS, 1955, zit. n. GUADAGNOLI & VELICER, 1991). Er berücksichtigt Profilverlauf und -höhe, relativiert aber die Meßwerte jeder Spalte an ihren Streuungen. Der Koeffizient hat einen Range von -1 bis +1, wobei ein Wert von $c=0$ absolute Nichtübereinstimmung signalisiert. Ein Signifikanztest für c existiert nicht, im Zusammenhang der Analyse von Faktorladungsstrukturen werden jedoch oft Werte von $c > .9$ als Deckungsgleichheit interpretiert (CURETON & D'AGOSTINO, 1983, zit. n. GUADAGNOLI & VELICER, 1991). Die exakte Berechnungsformel für c ist im Anhang E dokumentiert.

Schließlich sollte neben der Profilkorrelation und dem Kongruenzkoeffizienten ein Maß berechnet werden, das neben Profilverlauf und -höhe auch die Profilstreuung berücksichtigt, aber diese differenzierter integriert, als das bei einem einfachen Globalmaß der Fall ist. MCCRAE (1993) entwickelte einen Koeffizienten der Profilähnlichkeit I_{pa} , der auf Catells r_p basiert (der wiederum auf der Summe der Abweichungsquadrate beruht), aber Abweichungen in verschiedenen Bereichen der Skala unterschiedlich stark gewichtet. Die Idee dahinter ist, daß zufällige Übereinstimmungen eher im mittleren Skalenbereich zustande kommen (in dem die meisten Meßwerte liegen) als in den Extrembereichen. Der Koeffizient I_{pa} wird am höchsten, wenn die Selbst- und Fremdeinschätzungsprofile im Verlauf sehr ähnlich und gleichzeitig in extremeren Bereichen angesiedelt sind; er wird am niedrigsten, wenn die Profilverläufe unähnlich sind und in den mittleren Ausprägungsgraden liegen.

Durch eine einfache Transformation kann I_{pa} in den Koeffizienten r_{pa} überführt werden, der einen Range von -1 bis 1 hat. Ein gegebener Wert von r_{pa} hat eine ähnliche statistische Signifikanz wie die entsprechende Ausprägung des Bravais-Pearson-Korrelationskoeffizienten r , was die Interpretation erleichtert (McCRAE, 1993). Der Nachteil des Koeffizienten ist, daß selbst perfekte Übereinstimmung der Profile zu einem submaximalen Übereinstimmungswert führt, wenn die Itemwerte nahe an den jeweiligen Stichprobenmittelwerten liegen. Die exakten Berechnungsformeln für I_{pa} und r_{pa} sind im Anhang E dokumentiert.

Item-Level:

Ein einfaches höhenlagenabhängiges Maß für die durchschnittliche itemspezifische Urteilsübereinstimmung ist die Korrelation der Selbst- und Fremdeinschätzungen eines Items über alle Urteilerpaare. Auf Item-Level gibt es natürlich keine Itemvarianz und damit auch keine Stereotype Accuracy, Korrelationen konfundieren hier jedoch Elevation, Differential Elevation und Differential Accuracy. Im Riverside Accuracy Project wurde Anfang der achtziger Jahre vorgeschlagen, die beiden Elevation-Komponenten durch getrennte Standardisierung der Werte der Judges und Targets zu eliminieren (FUNDER 1980); dazu wurden in der Folge meistens Q-Sort-Daten erhoben, die durch Forced-Choice-Techniken automatisch eine Quasi-Normalverteilung erzeugen (z.B. COLVIN, 1993a, FUNDER, 1987, 1995, 1999, KOLAR, 1995, VOGT & COLVIN, 2003). Normale, intervallskalierte Antwortprofile können umgerechnet werden, indem von jedem Itemwert der individuelle Urteilermittelwert (über die gesamte Skala) abgezogen wird. Item-Level-Korrelationen auf Basis solcher personenzentrierten Abweichungswerte enthalten keine Elevation, Differential Elevation und Stereotype Accuracy mehr.

Although similar in its conceptualization (i.e. emphasizing the relative magnitude of items/traits), the component of variance involved here is not exactly the differential accuracy component described by Cronbach. True differential accuracy assesses the accuracy of a single judge over multiple targets. The self-other accuracy described above measures the accuracy of a sample of judges, each rating a single target. ... [The measurement nevertheless] reflects the fact that the absolute magnitude of a given trait is irrelevant. Rather, what is being assessed is the shape of one's personality profile, the relative strengths of the traits in relation to one another - in short, the constellation of traits (BERNIERI ET AL., 1994, S. 369).

4.4.4 Zusammenfassung Auswertungsansatz

Wichtig ist, daß es nicht eine beste Methode gibt, ... Übereinstimmung zu messen. Einzelne Koeffizienten bilden jeweils eine bestimmte Eigenschaft der erhobenen Datenstrukturen ab und spiegeln somit jeweils unterschiedliche Informationen wider (WIRTZ & CASPAR, 2002, S.23).

Auf die Berechnung einfacher Globalmaße der Urteilsübereinstimmung wird verzichtet. Die Inspektion einiger gängiger Indizes (Osgoodsches Distanzmaß D , City-Block-Distanz, Raw Agreement, Main Diagonal) an den vorliegenden Daten zeigte, daß diese aufgrund der starken Konfundierung der unterschiedlichen Übereinstimmungskomponenten nicht ausreichend zwischen den Urteilerpaaren differenzieren konnten und keine moderierenden Wirkungen widerspiegeln. Berechnet werden statt dessen verschiedene spezifischere Übereinstimmungsmaße mit und ohne Elevation-Komponenten.

Auf Personen-Level wird die höhenlagenunabhängige Urteilsübereinstimmung über die Profilkorrelation der Rohwerte der Selbst- und Fremdeinschätzung über alle Items berechnet. Zur Kontrolle von Stereotype Accuracy wird zusätzlich die ipsative Profilkorrelation der stichprobenzentrierten Abweichungswerte berechnet. Zur weiteren Analyse werden jeweils die Fisher-z-Werte dieser Korrelationen genutzt. Übereinstimmung unter Einbezug von Differential Elevation (und notgedrungen Elevation) wird über den Kongruenzkoeffizienten c und den Profilähnlichkeitskoeffizienten r_{pa} operationalisiert.

Auf Item-Level wird die höhenlagenunabhängige Urteilsübereinstimmung über die Korrelation der personenzentrierten Abweichungswerte der Selbst- und Fremdeinschätzungen jedes Items über alle Urteilerpaare berechnet. Urteilsübereinstimmung unter Einbezug von Differential Elevation (und notgedrungen Elevation) wird über die Korrelation der Rohwerte der Selbst- und Fremdeinschätzungen jedes Items über alle Urteilerpaare operationalisiert.

Die postulierten kategorialen Acquaintanceship-Moderatoren werden auf Personen-Level über Subgruppenvergleiche der durchschnittlichen dyadischen Übereinstimmung und über die Korrelationen (Spearman's Rho) der Moderatoren mit den verschiedenen Übereinstimmungsmaßen überprüft, auf Item-Level über Subgruppenvergleiche aller durchschnittlichen itemspezifischen Übereinstimmungen. Alle anderen postulierten, kontinuierlichen Moderatoren werden auf Personen-Level über die Korrelationen (Bravais-Pearson) mit den Maßen der dyadischen Urteilsübereinstimmung überprüft und auf Item-Level mittels moderierter Regression getestet.

5. Ergebnisteil

Die Berechnung aller Kennwerte und die Durchführung aller statistischen Verfahren zur Prüfung der einzelnen Hypothesen (vgl. Abschnitte 5.1 bis 5.6) erfolgte computergestützt mittels des Softwarepaketes SPSS für Windows (Version 12.0; SPSS Inc., 2003).

5.1 Meßgüte des BDI-V

Tabelle 4: Interne Konsistenzen der BDI-V Selbst- und Fremdeinschätzungsskalen

	Cronbachs Alpha	Inter-Item Korrelationen		
		Mean	Min	Max
BDI-V Selbsteinschätzung	.86	.24	-.13	.62
BDI-V Fremdeinschätzung	.84	.23	-.12	.63

Anmerkung: jeweils N=177.

Die internen Konsistenzen der BDI-V Selbsteinschätzungen und der BDI-V Fremdeinschätzungen fielen mit Werten von Cronbachs $\alpha = .86$ bzw. $\alpha = .84$ beide sehr gut aus (vgl. Tabelle 4).

Tabelle 5: Trennschärfen und Reliabilitäten der Items des BDI-V (Selbst- und Fremdeinschätzungen)

	Selbsteinschätzung		Fremdeinschätzung	
	Trennschärfe [#]	Reliabilität ^{##}	Trennschärfe [#]	Reliabilität ^{##}
BDI 01 'Ich bin traurig'	.53	.51	.61	.48
BDI 02 'Ich sehe mutlos in die Zukunft'	.62	.55	.52	.38
BDI 03 'Ich fühle mich als Versager'	.63	.58	.56	.51
BDI 04 'Es fällt mir schwer, etwas zu genießen'	.48	.40	.40	.28
BDI 05 'Ich habe Schuldgefühle'	.58	.58	.57	.55
BDI 06 'Ich fühle mich bestraft'	.52	.43	.54	.44
BDI 07 'Ich bin von mir enttäuscht'	.60	.61	.64	.57
BDI 08 'Ich werfe mir Fehler und Schwächen vor'	.64	.54	.60	.57
BDI 09 'Ich denke daran, mir etwas anzutun'	.42	.37	.38	.35
BDI 10 'Ich weine'	.24	.34	.36	.32
BDI 11 'Ich fühle mich gereizt und verärgert'	.41	.31	.44	.26
BDI 12 'Mir fehlt das Interesse an Menschen'	.24	.24	.39	.31
BDI 13 'Ich schiebe Entscheidungen vor mir her'	.42	.35	.37	.32
BDI 14 'Ich bin besorgt um mein Aussehen'	.46	.45	.07	.16
BDI 15 'Ich muß mich zu Tätigkeiten zwingen'	.33	.33	.42	.34
BDI 16 'Ich habe Schlafstörungen'	.66	.52	.42	.31
BDI 17 'Ich bin müde und lustlos'	.37	.29	.60	.48
BDI 18 'Ich habe keinen Appetit'	.32	.31	.33	.26
BDI 19 'Sex ist mir gleichgültig'	.35	.28	.30	.30
BDI 20 'Ich mache mir Sorgen um meine Gesundheit'	.31	.24	.28	.21

Anmerkungen: jeweils N=177, [#] korrigierte Item-Total Korrelation, ^{##} quadrierte multiple Korrelation.

Für alle Items der BDI-V Selbst- und Fremdeinschätzungsskalen gilt, daß ihre Elimination die Konsistenz (Cronbachs Alpha) der jeweiligen Gesamtskala um maximal .02 senken würde.

Insgesamt acht Items der Selbsteinschätzungsskala und zehn Items der Fremdeinschätzungsskala erzielten auf dem Trennschärfekoeffizienten mäßige Werte zwischen .30 und .50, jeweils zwei Items der Selbsteinschätzungsskala und der Fremdeinschätzungsskala erzielten sehr niedrige Trennschärfewerte unterhalb von .30, was oft als unterste zulässige Grenze für eindimensionale Testverfahren angesehen wird (BORTZ & DÖRING, 1995, S. 200). Dabei waren Trennschärfe und Reliabilität der Fremdeinschätzung des Items BDI14 ('Ich bin besorgt um mein Aussehen') extrem niedrig (.07 und .16).

5.2 Allgemeine Urteilsübereinstimmung (Hypothese 1)

5.2.1 Personen-Level Ergebnisse

Tabelle 6: Verteilungswerte der dyadischen Übereinstimmungsmaße

Koeffizient	Mean	SD	Min	Max	Kolmogoroff-Smirnoff
Profilkorrelation	.46 [#]	.27 ^{##}	.00	.87	.56
Ipsative Profilkorrelation ⁺	.33 [#]	.27 ^{##}	.00	.85	.09
Kongruenzkoeffizient c	.92	.03	.76	.98	.15
Ähnlichkeitskoeffizient r_{PA}	.29	.26	-.61	.97	.20

Anmerkungen: N=177 Urteilerpaare, ⁺ Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte, [#] Mittelwertbildung erfolgte über Fisher-z-Werte und anschließende Rücktransformation, ^{##} Berechnung der Standardabweichung auf Basis der Fisher-z-Werte und anschließende Rücktransformation, **fettgedruckt**: Annahme auf Normalverteilung nicht widerlegt, da nichtsignifikantes Testergebnis.

Insgesamt sechs der 177 Urteilerpaare erzielten eine negative Profilkorrelation zwischen -.01 und -.05, die aufgrund theoretischer Erwägungen auf null gesetzt wurde, da ein gegenläufiger Profilverlauf vermutlich auf Zufallseinschätzungen beruht und keine größere Unähnlichkeit darstellt als ein nullkorrelierter (Anmerkung: Vorgehen wurde mit Herrn Prof. Dr. Schmitt abgesprochen). Für alle weiteren Analysen, bei denen die Profilkorrelation oder die ipsative Profilkorrelation als Übereinstimmungskoeffizienten genutzt und mit potentiellen Moderatorvariablen oder untereinander korreliert wurden, wurde auf die Fisher-z-Werte zurückgegriffen, da nur intervallskalierte Maße korreliert werden dürfen. Eine inhaltliche Beschreibung der allgemeinen Urteilsübereinstimmung auf Personen-Level erfolgt in Abschnitt 5.2.3.

Tabelle 7: Korrelationen der dyadischen Übereinstimmungsmaße untereinander

Koeffizient	korrigierte Item-Total Korrelation	Profilkorrelation	Ipsative Profilkorrelation ⁺	Kongruenzkoeffizient c
Profilkorrelation	.88			
Ipsative Profilkorrelation ⁺	.88	.88 ^{**}		
Kongruenzkoeffizient c	.68	.72 ^{**}	.60 ^{**}	
Ähnlichkeitskoeffizient r_{PA}	.61	.54 ^{**}	.61 ^{**}	.48 ^{**}

Anmerkungen: N=177 Urteilerpaare, *p<.05, **p<.01 (2-seitig), ⁺ Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte. Profilkorrelation und ipsative Profilkorrelation gingen als Fisher-z-Werte in die Berechnungen ein.

Die hohe Korrelation von $r=.88^{**}$ zwischen der Profilkorrelation der BDI-V Rohwerte und der Profilkorrelation der stichprobenzentrierten BDI-V Abweichungswerte ist ein Hinweis darauf, daß der Anteil der Stereotype Accuracy an dem dyadischen Übereinstimmungsmaß der Profilkorrelation in der vorliegenden Stichprobe gering ist. Die niedrigeren Korrelationen des Profیلähnlichkeitskoeffizienten r_{PA} nach McCRAE (1993) mit den anderen Übereinstimmungsmaßen ($r=.48^{**}$ bis $r=.61^{**}$, vgl. Tabelle 7) belegen die Sonderstellung des Koeffizienten, der neben Höhenlage und Profilverlauf auf komplexe Weise auch die Profilstreuung berücksichtigt.

5.2.2 Item-Level Ergebnisse

Tabelle 8: Korrelationen der Selbst- und Fremdeinschätzungen aller BDI-V Items (Rohwerte, personenzentrierte Abweichungswerte)

	Rohwerte	Ipsative Werte
BDI 01 'Ich bin traurig'	.47**	.21**
BDI 02 'Ich sehe mutlos in die Zukunft'	.45**	.21**
BDI 03 'Ich fühle mich als Versager'	.30**	.17*
BDI 04 'Es fällt mir schwer, etwas zu genießen'	.38**	.27**
BDI 05 'Ich habe Schuldgefühle'	.26**	.11
BDI 06 'Ich fühle mich bestraft'	.26**	.16*
BDI 07 'Ich bin von mir enttäuscht'	.35**	.21**
BDI 08 'Ich werfe mir Fehler und Schwächen vor'	.30**	.15*
BDI 09 'Ich denke daran, mir etwas anzutun'	.32**	.34**
BDI 10 'Ich weine'	.50**	.45**
BDI 11 'Ich fühle mich gereizt und verärgert'	.38**	.26**
BDI 12 'Mir fehlt das Interesse an Menschen'	.26**	.22**
BDI 13 'Ich schiebe Entscheidungen vor mir her'	.41**	.46**
BDI 14 'Ich bin besorgt um mein Aussehen'	.33**	.39**
BDI 15 'Ich muß mich zu Tätigkeiten zwingen'	.27**	.23**
BDI 16 'Ich habe Schlafstörungen'	.53**	.50**
BDI 17 'Ich bin müde und lustlos'	.29**	.20**
BDI 18 'Ich habe keinen Appetit'	.29**	.28**
BDI 19 'Sex ist mir gleichgültig'	.44**	.41**
BDI 20 'Ich mache mir Sorgen um meine Gesundheit'	.33**	.33**
Gesamtscore BDI-V	.52**	-

Anmerkungen: N=177 Urteilerpaare, * $p<.05$, ** $p<.01$ (2-seitig).

Auf Item-Level zeigte sich eine passable Übereinstimmung zwischen den BDI-V Selbst- und Fremdeinschätzungen. Die aggregierten Gesamteinschätzungen des Depressivitätsgrades des Targets (i.e. die von Judge und Target erzielten BDI-V Gesamtscores) korrelierten mit $r=.52^{**}$. Die Korrelationen der Rohwerte der einzelnen Items lagen mit einer Ausnahme (BDI 16, 'Ich habe Schlafstörungen') darunter, erreichten aber dennoch befriedigende Werte: Die durchschnittliche Selbst-Fremd-Korrelation der BDI-V Items lag bei $r=.36$ (Mittelwertsbildung über Fisher-z-Werte und Rücktransformation), der Range erstreckte sich von $r=.26^{**}$ bis $r=.53^{**}$.

Die Selbst-Fremd-Korrelationen der um die Höhenlage bereinigten personenzentrierten BDI-V Abweichungswerte fielen bei 16 von 20 Items niedriger aus als die der BDI-V Rohwerte, teilweise deutlich. In einem Fall war die Korrelation identisch (BDI 20), bei drei Items (BDI 09, BDI 13, BDI 14) waren die Korrelationen der Abweichungswerte leicht höher. Die durchschnittliche Selbst-Fremd-Korrelation der Abweichungswerte lag bei $r=.28$ (Mittelwertsbildung erfolgte über Fisher-z Werte), der Range erstreckte sich von $r=.11$ bis $r=.50^{**}$. Tendenziell ist also die Übereinstimmung der Bewertungen der absoluten Ausprägungshöhe einzelner depressiver Symptome größer als die Übereinstimmung der Beurteilungen der relativen Ausprägungshöhe, i.e. der Stellung der einzelnen Items zueinander in der Persönlichkeitsstruktur des Individuums.

5.2.3 Fazit allgemeine Urteilsübereinstimmung

Methodisches: Das Maß der ipsativen Profilkorrelation zeigte wie erwartet eine etwas niedrigere durchschnittliche Übereinstimmung an als die Profilkorrelation der Rohwerte, da es frei von Stereotype Accuracy ist. Die Werte des Kongruenzkoeffizienten c , der die Ähnlichkeit des Profilverlaufs und der Profilhöhe konfundiert (und damit neben Differential Elevation auch Elevation), zeigten wie erwartet ein höheres Agreement an als die korrelativen Maße. Der auch noch die Profilstreuung berücksichtigende Ähnlichkeitskoeffizient r_{PA} erzielte den mit Abstand größten Range ($r_{PA} = -.61$ bis $r_{PA} = .97$, vgl. Tabelle 8). Für die weitere Analyse wird er mit Vorbehalt zu betrachten sein (Anmerkung: Zur Problematik dieses Koeffizienten vgl. auch Abschnitt 4.4.2).

Inhaltliche Ergebnisse: Persönlichkeitseinschätzungen können insgesamt unverzerrt und damit einigermaßen zutreffend sein, aber dennoch im Detail unpräzise (vgl. Abschnitt 2.1.3). Diese Tatsache spiegelt die Auswertung auf Item-Level wider, bei der die Korrelationen der Rohwerte größtenteils höher ausfielen als die der personenzentrierten Abweichungswerte (vgl. Tabelle 8). Die über alle Items gemittelte durchschnittliche Selbst-Fremd-Korrelation lag bei $r=.36$, was als passabler Wert angesehen werden kann. Die Auswertung auf Personen-Level zeigte, daß die dyadische Urteilsübereinstimmung der Beurteilerpaare sehr unterschiedlich ausfiel (Range der Profilkorrelation von $r=.00$ bis $r=.87$, Range der ipsativen Profilkorrelation der stichprobenzentrierten Abweichungswerte von $r=.00$ bis $r=.85$, vgl. Tabelle 6). Mit einem Durchschnittswert von $r=.46$ bzw $r=.33$ erreichte sie aber insgesamt ebenfalls ein befriedigendes Niveau.

Hypothese 1 wurde auf Item- und auf Personen-Level bestätigt.

Die großen Unterschiede zwischen den einzelnen Urteilerpaaren deuten darauf hin, daß es zwischen verschiedenen Menschen tatsächlich Unterschiede in der Beurteilbarkeit bzw. Beurteilungsfähigkeit gibt (vgl. Abschnitt 2.3.2 bzw. 2.3.3). Die Unterschiede der durchschnittlichen Selbst-Fremd-Übereinstimmung zwischen den einzelnen Items deuten darauf hin, daß es tatsächlich Unterschiede in der Fremdbeurteilbarkeit spezifischer Erlebens- und Verhaltensweisen gibt (vgl. Abschnitt 2.3.5). Inwieweit diese mit der Beobachtbarkeit der Items zusammenhängen, wird in Abschnitt 5.6.2 besprochen.

5.3 Judge: Moderatoren der Fremdbeurteiler (Hypothesen 2-5)

Tabelle 9: Verteilungswerte der potentiellen Moderatoren seitens des Judges

Eigenschaft	Mean	SD	Min	Max	Kolmogoroff-Smirnoff
Empathie	29,01 ⁺	4,61	11	39	.46
Perspektivübernahme	30,04 ⁺⁺	4,79	18	42	.13
Offenheit d. WS	28,85 ⁺	3,99	18	37	.17
Lebensalter [Jahre]	36,28	16,09	17	90	.00

Anmerkungen: N=177, ⁺ auf einer Skala von 8 bis 40, ⁺⁺ auf einer Skala von 9 bis 45, **fettgedruckt:** Annahme auf Normalverteilung nicht widerlegt, da nichtsignifikantes Testergebnis.

Die Korrelationen zwischen den Ausprägungen der Eigenschaften betragen für Empathie und Perspektivübernahme $r=.53^{**}$, für Empathie und Offenheit d. WS $r=.17^*$ und für Perspektivübernahme und Offenheit d. WS $r=.15^*$, das Lebensalter korreliert nahe Null und nichtsignifikant mit Empathie, mit den beiden anderen Moderatoren ergaben sich Werte von $r=-.16^*$ (Perspektivübernahme) und $r=-.21^{**}$ (Offenheit d. WS), mit $*p<.05$, $**p<.01$ (2-seitig).

Für die Empathie- und Perspektivübernahmeskalen von MAES ET AL. (1995) existieren leider keine Normwerte, beide Stichprobenmittelwerte erscheinen dennoch eher im oberen Bereich angesiedelt. Die Korrelation von $r=.53^{**}$ zwischen Empathie und Perspektivübernahmefähigkeit belegt die inhaltliche Verwandtschaft der beiden Konstrukte. Offenheit des Wertesystems korrelierte dagegen nur mäßig ($r=.17^*$ und $r=.15^*$) mit den anderen beiden Eigenschaften. Der Stichprobenmittelwert der Offenheitsskala von 28,85 liegt deutlich über dem bevölkerungsrepräsentativen deutschen Normwert des NEO-PI-R von 21,13 (OSTENDORF & ANGLEITNER, 2004).

Die Altersspanne der Fremdbeurteiler reichte von 17 bis 90 Lebensjahren. Die Alterswerte sind nicht normalverteilt, 37% der Judges waren zwischen 20 und 30 Jahren alt, der überwiegende Rest verteilte sich auf die höheren Altersstufen mit einem zweiten, kleineren Gipfel bei 65 Jahren. Das durchschnittliche Alter lag mit 36,3 Jahren deutlich über dem typischen Altersschnitt allgemein üblicher rein studentischer Stichproben. Das Lebensalter korrelierte mäßig negativ mit Perspektivübernahme und Offenheit des Wertesystems ($r=-.16^*$ und $r=-.21^{**}$), i.e. jüngere Beurteiler sind tendenziell etwas offener und neigen etwas stärker zur Perspektivübernahme.

5.3.1 Personen-Level Ergebnisse

Tabelle 10: Korrelationen der potentiellen Moderatoren seitens des Judges mit den Kennwerten der dyadischen Urteilsübereinstimmung

	Profilkorrelation	Ipsative Profilkorrelation [#]	Kongruenzkoeffizient c	Ähnlichkeitskoeffizient r_{PA}
Empathie	-.14	-.15 [*]	-.15 [*]	-.16 [*]
Perspektivübernahme	-.04	-.10	-.05	-.07
Offenheit d. WS	-.07	-.12	-.01	-.04
Lebensalter [Jahre]	.03	.05	.08	.08

Anmerkungen: N=177 Urteilerpaare, $*p<.05$, $**p<.01$ (2-seitig), [#] Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte. Profilkorrelation und ipsative Profilkorrelation gingen als Fisher-z-Werte in die Berechnungen ein.

Wie Tabelle 10 zeigt, korrelierte das Lebensalter der Fremdbeurteiler überhaupt nicht bzw. nahe null und nichtsignifikant mit den Maßen der dyadischen Urteilsübereinstimmung. Ähnliches gilt im wesentlichen für die Neigung zur Perspektivübernahme und die Offenheit des Wertesystems der Fremdbeurteiler. Diese beiden Persönlichkeitseigenschaften korrelierten immerhin schwach und nichtsignifikant mit der ipsativen Profilkorrelation der stichprobenzentrierten Abweichungswerte, die frei von Stereotype Accuracy ist. Dieser Trend fiel jedoch der Richtung nach nicht erwartungsgemäß aus, i.e. höhere Ausprägungen von Perspektivübernahme und Offenheit gingen wenn überhaupt eher mit niedrigerer Urteilsübereinstimmung einher.

Für Empathie ergaben sich der Richtung nach nicht erwartungsgemäße, schwache aber dennoch signifikante Korrelationen von $r=-.15^*$ bzw. $r=-.16^*$ mit drei von vier Übereinstimmungsmaßen, die vierte lag knapp über der Signifikanzgrenze. Durch die Berechnung von Semipartialkorrelationen, aus denen der Bekanntschaftsgrad von Judge und Target auspartialisiert wurde (vgl. Tabelle F1 im Anhang F), werden die Zusammenhänge zwischen Empathie und den vier Übereinstimmungsmaßen schwächer und allesamt nichtsignifikant. Bei dem in Tabelle 10 angezeigten negativen Zusammenhang von Empathie und Urteilsgüte könnte es sich insofern um einen Artefakt handeln, der aus einer unkontrollierten Konfundierung resultiert.

Hypothesen 2 bis 5 konnten auf Personen-Level nicht bestätigt werden.

5.3.2 Item-Level Ergebnisse

Zur Analyse der Moderatorwirkungen auf Item-Level wurden hierarchische Regressionsanalysen durchgeführt, bei denen zuerst die Selbsteinschätzungen auf einem BDI-V Item als Kriterium auf den Prädiktor, i.e. die Fremdeinschätzungen des gleichen BDI-V Items, regressiert wurden, dann auf die Moderatorvariable und zuletzt auf das Kreuzprodukt von Prädiktor und Moderator. Vorliegen und Stärke von Moderatorwirkungen wurden aus der eventuellen Zunahme von R^2 nach Eingabe des Interaktionsterms geschlossen (vgl. Abschnitt 4.4.2). Aus der Überprüfung von vier potentiellen Moderatoren seitens der Fremdbeurteiler auf jeweils allen 20 BDI-V Items sowie dem BDI-V Gesamtscore resultierten daher 84 separate Regressionsanalysen. Tabelle 11 gibt den jeweiligen Zuwachs an Varianzaufklärung wieder, der aus der Eingabe der jeweiligen Kreuzprodukte der Moderatoren Empathie bzw. Perspektivübernahme bzw. Offenheit des Wertesystems resultierte.

Die Eingabe des Lebensalters der Fremdbeurteiler führte bei keinem der 20 BDI-V Items zu einer signifikanten Zunahme von R^2 . Auf eine Darstellung der betreffenden Werte wurde aufgrund der Ergebnislosigkeit der Analysen und aus Platzgründen verzichtet.

Hypothese 5 konnte auf Item-Level nicht bestätigt werden.

Tabelle 11: Ergebnisse der Regressionsanalysen bzgl. der Moderatoren des Judges: Zusätzliche Varianzaufklärung, i.e. Veränderung von R^2 durch Eingabe des jeweiligen Interaktionsterms

	Empathie	Perspektiv- übernahme	Offenheit d. WS
BDI 01 'Ich bin traurig'	.00	.00	.00
BDI 02 'Ich sehe mutlos in die Zukunft'	.01	.00	.00
BDI 03 'Ich fühle mich als Versager'	.00	.00	.01
BDI 04 'Es fällt mir schwer, etwas zu genießen'	.00	.00	.02*
BDI 05 'Ich habe Schuldgefühle'	.01	.00	.00
BDI 06 'Ich fühle mich bestraft'	.01	.00	.00
BDI 07 'Ich bin von mir enttäuscht'	.01	.00	.00
BDI 08 'Ich werfe mir Fehler und Schwächen vor'	.01	.01	.00
BDI 09 'Ich denke daran, mir etwas anzutun'	.01	.00	.00
BDI 10 'Ich weine'	.00	.00	.00
BDI 11 'Ich fühle mich gereizt und verärgert'	.00	.00	.01
BDI 12 'Mir fehlt das Interesse an Menschen'	.02	.00	.00
BDI 13 'Ich schiebe Entscheidungen vor mir her'	.00	.00	.00
BDI 14 'Ich bin besorgt um mein Aussehen'	.01	.04**	.03*
BDI 15 'Ich muß mich zu Tätigkeiten zwingen'	.00	.00	.00
BDI 16 'Ich habe Schlafstörungen'	.02*	.00	.02*
BDI 17 'Ich bin müde und lustlos'	.00	.01	.00
BDI 18 'Ich habe keinen Appetit'	.00	.00	.01
BDI 19 'Sex ist mir gleichgültig'	.00	.02*	.00
BDI 20 'Ich mache mir Sorgen um meine Gesundheit'	.00	.00	.00
BDI-V Gesamtscore	.00	.00	.00

Anmerkungen: N=177 Urteilerpaare, *F-Test signifikant mit $p < .05$, **F-Test signifikant mit $p < .01$ (2-seitig).

Die Kriterien der Regressionsanalysen waren die BDI-V Selbsteinschätzungen, als Prädiktoren dienten die BDI-V Fremdeinschätzungen. Die Selbst- und Fremdeinschätzungen und die jeweiligen Moderatoren gingen als Standard-z-Werte in die Analyse ein, das jeweilige Kreuzprodukt wurde aus den Standard-z-Werten berechnet.

Wie Tabelle 11 zeigt, hing das individuelle Ausmaß der Empathie der Fremdbeurteiler lediglich bei einem Item (BDI 16 'Ich habe Schlafstörungen') signifikant mit dem Übereinstimmungsgrad der Selbst- und Fremdeinschätzungen zusammen. Perspektivübernahme beeinflusste die Selbst-Fremd-Übereinstimmung bei zwei Items (BDI 14 'Ich bin besorgt um mein Aussehen' und BDI 19 'Sex ist mir gleichgültig'), Offenheit des Wertesystems bei drei Items (BDI 04 'Es fällt mir schwer, etwas zu genießen', BDI 14 'Ich bin besorgt um mein Aussehen' und BDI 16 'Ich habe Schlafstörungen'). Die Regression der Gesamtscores zeigte keine moderierenden Wirkungen an.

Der in den erwähnten sechs Einzelfällen zusätzlich aufgeklärte Varianzanteil von 2% bzw. 3% bzw. 4% mag unbedeutend erscheinen, gilt jedoch bereits als relevante, wenn auch kleine Effektstärke (COHEN, 1992, MCCLELLAND & JUDD, 1993, vgl. auch DALBERT & SCHMITT, 1986, S.37). Ein schwacher Effekt eines potentiellen Moderators bei maximal drei von 20 Analysen und damit ein moderierender Einfluß auf maximal drei von 20 Items kann jedoch insgesamt nicht als Beleg für eine generelle Wirksamkeit der betreffenden Eigenschaft gewertet werden.

Hypothesen 2 bis 4 konnten auf Item-Level nicht bestätigt werden.

5.4 Target: Moderatoren der Selbstbeurteiler (Hypothesen 6-9)

Tabelle 12: Verteilungswerte der potentiellen Moderatoren seitens des Targets

Eigenschaft	Mean	SD	Min	Max	Kolmogoroff-Smirnoff
Fremdtäuschung	34,78 ⁺	10,15	12	62	.30
Privatheit	14,51 ⁺⁺	3,70	7	21	.17
Konsistenz (SD)	0,84	0,23	0,31	1,67	.72
Selbsttäuschung	41,60 ⁺	6,73	22	61	.23

Anmerkungen: N=177, ⁺ auf einer Skala von 10 bis 70, ⁺⁺ auf einer Skala von 7 bis 21, **fettgedruckt:** Annahme auf Normalverteilung nicht widerlegt, da nichtsignifikantes Testergebnis.

Die Korrelationen zwischen den Ausprägungen der Eigenschaften betragen für Selbst- und Fremdtäuschung $r=.21^{**}$, für Selbsttäuschung und Privatheit $r=-.07$ und für Fremdtäuschung und Privatheit $r=-.15^{*}$, die Konsistenz des Selbsturteils korrelierte nahe Null und nichtsignifikant mit Privatheit, mit den beiden anderen Moderatoren ergaben sich Werte von $r=-.17^{*}$ (Selbsttäuschung) bzw. $r=-.23^{**}$ (Fremdtäuschung), mit $*p<.05$, $**p<.01$ (2-seitig).

Für die Selbst- und Fremdtäuschungsskalen von MUSCH ET AL. (2002) existieren leider keine Normwerte, der Stichprobenmittelwert der Fremdtäuschung erscheint jedoch im mittleren Bereich angesiedelt, während der Stichprobenmittelwert der Selbsttäuschung etwas höher ausfiel. Die mäßige Korrelation von nur $r=.21^{**}$ zwischen Selbst- und Fremdtäuschungstendenz belegt die inhaltliche Abgrenzung der beiden Konstrukte voneinander. Der Stichprobenmittelwert der Privatheit von 14,51 entspricht in etwa dem bevölkerungsrepräsentativen Normwert der Skala N des 16 PF-R (SCHNEEWIND & GRAF, 1998).

Fremdtäuschung korrelierte schwach negativ mit Privatheit, i.e. Personen mit stärkerer Neigung zur Fremdtäuschung sind gleichzeitig tendenziell etwas offener bezüglich ihres Privatlebens oder geben dies zumindest vor. Zwischen den Ausprägungen von Selbsttäuschung und Privatheit ergab sich dagegen kein statistisch relevanter Zusammenhang und auch die Konsistenz des Selbsturteils hing nicht mit Privatheit zusammen. Sie korrelierte jedoch mäßig negativ mit Selbst- und Fremdtäuschung, i.e. höher selbst- und fremdtäuschende Personen zeigen eine tendenziell inkonsistentere Selbsteinschätzung depressiver Verhaltens- und Erlebensweisen.

5.4.1 Personen-Level Ergebnisse

Tabelle 13: Korrelationen der potentiellen Moderatoren seitens des Targets mit den Kennwerten der dyadischen Urteilsübereinstimmung

	Profilkorrelation	Ipsative Profilkorrelation [#]	Kongruenzkoeffizient c	Ähnlichkeitskoeffizient r_{PA}
Fremdtäuschung	-.16 [*]	-.05	-.05	-.03
Privatheit	.00	.00	.08	-.02
Konsistenz (SD)	.14 [*]	.20 ^{**}	-.12	.06
Selbsttäuschung	-.08	-.08	-.16 [*]	.04

Anmerkungen: N=177 Urteilerpaare, $*p<.05$, $**p<.01$ (2-seitig), [#] Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte. Profilkorrelation und ipsative Profilkorrelation gingen als Fisher-z-Werte in die Berechnungen ein.

Tabelle 13 zeigt, daß das Ausmaß der Fremdtäuschungstendenz der Selbstbeurteiler mäßig und signifikant mit immerhin einem der Übereinstimmungsmaße korrelierte, der Profilkorrelation der BDI-V Rohwerte. Durch die Berechnung von Semipartialkorrelationen, aus denen der Bekanntschaftsgrad von Judge und Target auspartialisiert wurde (vgl. Tabelle F2 im Anhang F), wird der Zusammenhang noch etwas stärker ($r=-.20^{**}$ statt $r=-.16^*$). Bezüglich der Profilverlaufsähnlichkeit besitzt Fremdtäuschung also eine schwache bis mäßige moderierende Wirkung: **Hypothese 6 wurde auf Personen-Level bestätigt.**

Die Privatheit des Targets korrelierte dagegen überhaupt nicht bzw. nahe null und nichtsignifikant mit den verschiedenen Maßen der dyadischen Urteilsübereinstimmung (vgl. Tabelle 13): **Hypothese 7 konnte auf Personen-Level nicht bestätigt werden.**

Die Konsistenz des Ausprägungsgrades depressiver Symptome beim Selbstbeurteiler, operationalisiert über die Standardabweichung der Selbsteinschätzung über alle 20 Items, korrelierte nicht erwartungsgemäß schwach bis mäßig positiv und signifikant mit der Profilkorrelation und der ipsativen Profilkorrelation ($r=.14^*$ und $r=.20^{**}$). Eine höhere Standardabweichung entspricht einer höheren Inkonsistenz des Selbsturteils (ergo einer niedrigeren Ausprägung des Metatraits), eine höhere Profilkorrelation bedeutet jedoch eine höhere Urteilsübereinstimmung: **Hypothese 8 wurde auf Personen-Level widerlegt.**

Der Ausprägungsgrad der Selbsttäuschung korrelierte lediglich mit dem höhenlagenabhängigen Kongruenzkoeffizienten c mäßig negativ und signifikant, mit den Maßen der Profilverlaufsähnlichkeit dagegen nur sehr schwach und nichtsignifikant (vgl. Tabelle 13). Durch die Berechnung von Semipartialkorrelationen ohne den Bekanntschaftsgrad von Judge und Target (vgl. Tabelle F2 im Anhang F) wird der Zusammenhang zwischen Selbsttäuschung und dem höhenlagenabhängigen Koeffizienten noch deutlicher ($r=-.20^{**}$ statt $r=-.16^*$), auch die Korrelationen mit den anderen Übereinstimmungsmaßen werden stärker, bleiben jedoch nichtsignifikant. Inhaltlich kann das als Hinweis auf die Richtigkeit der Hypothese 9 (self-serving Bias) auf Personen-Level gesehen werden werden, da eine verzerrte Selbstsicht zu einer positiv verschobenen Selbsteinschätzung führt, die aber dennoch einigermaßen präzise bleibt, weil die relativen Ausprägungen der einzelnen Attribute zueinander durchaus richtig erkannt werden. Daher hängt die Selbsttäuschungstendenz stärker mit der Höhenlagenähnlichkeit der Selbst- und Fremdeinschätzungen zusammen, als mit der Profilverlaufsähnlichkeit (vgl. Abschnitt 2.1.3, S.20).

Tabelle 14: Abweichung der BDI-V Selbst- und Fremdeinschätzungen (Gesamtscore)

	Mean Selbsturteil	Mean Fremdurteil	mittlere Abweichung Selbst- / Fremdurteil [#]	Min Abweichung Selbst- / Fremdurteil [#]	Max Abweichung Selbst- / Fremdurteil [#]
BDI-V Gesamtscore	42,12 ⁺	42,52 ⁺	0,4 ⁺	-25 ⁺	27 ⁺

Anmerkungen: N=177 Urteilerpaare, ⁺ auf einer Skala von 20 bis 100, [#] positive Werte entsprechen inhaltlich einer weniger depressiven Selbsteinschätzung des Targets im Vergleich zur Fremdeinschätzung.

Tabelle 15: Korrelation und Semipartialkorrelation der Selbsttäuschungstendenz mit der absoluten Abweichung der BDI-V Selbst- und Fremdeinschätzungen

	<u>Abweichung Selbst- / Fremdurteil (BDI-V Gesamtscore)</u>	
	Korrelation	Semipartialkorrelation [#]
Selbsttäuschung	.29**	.26**

Anmerkungen: N=177 Urteilerpaare, *p<.05, **p<.01 (2-seitig),
[#] Semipartialkorrelation mit statistischer Kontrolle der Fremdtäuschung.

Wie Tabelle 14 zeigt, gab es nur eine minimale durchschnittliche positive Abweichung der BDI-V Selbsteinschätzungen von den BDI-V Fremdeinschätzungen (lediglich 0,4 Punkte bei einer Skalenspannweite von 80 Punkten). In der vorliegenden Stichprobe lag demnach keine generelle Verzerrung der Selbsteinschätzungen vor. Der Wertebereich der dyadischen Abweichung der einzelnen Beurteilerpaare zeigt jedoch, daß es in beide Richtungen Diskrepanzen von bis zu 25 bzw. 27 Punkten gab. Die Einschätzungen des Depressivitätsgrades des Targets lagen also bei den Beurteilerpaaren teilweise erheblich auseinander. Die Korrelation der Selbsttäuschungstendenz mit dem Ausmaß der positiven Verzerrung des Selbsturteils lag bei $r=.29^{**}$, auspartialisieren der Fremdtäuschungstendenz senkte den Wert kaum (vgl. Tabelle 15).

Der erste Teil von Hypothese 9 konnte auf Personen-Level nicht belegt werden, da in der Stichprobe kein generell wirksamer self-serving Bias vorlag. Der zweite Teil der Hypothese wurde jedoch bestätigt, da die Selbstbeurteiler sich in unterschiedlichem Ausmaß positiv verzerrt sahen und das Ausmaß dieser Verzerrung mit dem Ausmaß der gemessenen Selbsttäuschungstendenz zusammen hing.

5.4.2 Item-Level Ergebnisse

Zur Analyse der Moderatorwirkungen auf Item-Level wurden analog der Erläuterungen zu Beginn von Abschnitt 5.3.2 auch für die vier potentiellen Moderatoren seitens der Selbstbeurteiler insgesamt 84 separate hierarchische Regressionsanalysen durchgeführt. Tabelle 16 gibt den jeweiligen Zuwachs an Varianzaufklärung wieder, der aus der Eingabe der jeweiligen Kreuzprodukte der jeweiligen Moderatoren resultierte.

Das individuelle Ausmaß der Fremdtäuschungstendenz der Selbstbeurteiler beeinflusste die Selbst-Fremd-Übereinstimmung dreier Items (BDI 01 'Ich bin traurig', BDI 11 'Ich fühle mich gereizt und verärgert' und BDI 18 'Ich habe keinen Appetit'). Die Privatheit des Targets hing dagegen bei lediglich zwei Items (BDI 04 'Es fällt mir schwer, etwas zu genießen' und BDI 09 'Ich denke daran, mir etwas anzutun') signifikant mit der Urteilsübereinstimmung zusammen.

Die Regression der Gesamtscores zeigte keine moderierenden Wirkungen an.

Tabelle 16: Ergebnisse der Regressionsanalysen bzgl. der Moderatoren des Targets: Zusätzliche Varianzaufklärung, i.e. Veränderung von R^2 durch Eingabe des jeweiligen Interaktionsterms

	Fremd- täuschung	Privatheit	Konsistenz	Selbst- täuschung
BDI 01 'Ich bin traurig'	.02*	.01	.00	.02*
BDI 02 'Ich sehe mutlos in die Zukunft'	.00	.00	.00	.00
BDI 03 'Ich fühle mich als Versager'	.00	.02	.00	.00
BDI 04 'Es fällt mir schwer, etwas zu genießen'	.00	.02*	.01	.02
BDI 05 'Ich habe Schuldgefühle'	.00	.02	.01	.00
BDI 06 'Ich fühle mich bestraft'	.00	.00	.02	.03*
BDI 07 'Ich bin von mir enttäuscht'	.00	.00	.01	.00
BDI 08 'Ich werfe mir Fehler und Schwächen vor'	.00	.01	.00	.01
BDI 09 'Ich denke daran, mir etwas anzutun'	.01	.04**	.02	.00
BDI 10 'Ich weine'	.01	.01	.01	.00
BDI 11 'Ich fühle mich gereizt und verärgert'	.03**	.00	.01	.00
BDI 12 'Mir fehlt das Interesse an Menschen'	.00	.00	.01	.01
BDI 13 'Ich schiebe Entscheidungen vor mir her'	.00	.01	.00	.00
BDI 14 'Ich bin besorgt um mein Aussehen'	.00	.00	.00	.02*
BDI 15 'Ich muß mich zu Tätigkeiten zwingen'	.01	.00	.01	.00
BDI 16 'Ich habe Schlafstörungen'	.00	.00	.01	.01
BDI 17 'Ich bin müde und lustlos'	.02	.01	.00	.02
BDI 18 'Ich habe keinen Appetit'	.02*	.01	.01	.00
BDI 19 'Sex ist mir gleichgültig'	.00	.00	.00	.02*
BDI 20 'Ich mache mir Sorgen um meine Gesundheit'	.00	.01	.01	.00
BDI-V Gesamtscore	.00	.00	.01	.00

Anmerkungen: N=177 Urteilerpaare, *F-Test signifikant mit $p < .05$, **F-Test signifikant mit $p < .01$ (2-seitig).

Die Kriterien der Regressionsanalysen waren die BDI-V Fremdeinschätzungen, als Prädiktoren dienten die BDI-V Selbsteinschätzungen. Die Selbst- und Fremdeinschätzungen und die jeweiligen Moderatoren gingen als Standard-z-Werte in die Analyse ein, das jeweilige Kreuzprodukt wurde aus den Standard-z-Werten berechnet.

Die Konsistenz des Selbsturteils hatte auf Item-Level überhaupt keine signifikanten Auswirkungen, i.e. hing bei keinem einzigen Item signifikant mit der Urteilsübereinstimmung zusammen. Das individuelle Ausmaß der Selbsttäuschungstendenz des Targets beeinflusste die Selbst-Fremd-Übereinstimmung von vier Items (BDI 01 'Ich bin traurig', BDI 06 'Ich fühle mich bestraft', BDI 14 'Ich bin besorgt um mein Aussehen' und BDI 19 'Sex ist mir gleichgültig').

Alles in allem zeigte sich bei den Regressionsanalysen bezüglich der Eigenschaften der Selbstbeurteiler ein ähnliches Bild wie zuvor bereits bei den potentiellen Moderatoren der Fremdbeurteiler: Die erhobenen Persönlichkeitseigenschaften der Targets haben mit Ausnahme der Konsistenz durchaus schwache aber signifikante Effekte auf den Übereinstimmungsgrad einzelner Items, diese Effekte sind jedoch insgesamt zu sporadisch, um auf eine generelle Wirksamkeit eines der potentiellen Moderatoren schließen zu können, da sie zu wenige Items betreffen.

Hypothesen 6 bis 9 konnten auf Item-Level nicht bestätigt werden.

5.5 Information: Moderator Acquaintanceship (Hypothesen 10/11)

5.5.1 Bekanntschaftsdauer der Urteilerpaare (quantitativer Aspekt)

Die Angaben der Selbst- und Fremdbeurteiler bezüglich der Dauer ihrer gemeinsamen Bekanntschaft wurden jeweils zu einem gemeinsamen Wert der dyadischen Bekanntschaftsdauer gemittelt. Dabei machten 76% der Paare vollkommen identische Angaben, bei weiteren 16% betrug die Differenz der Angaben höchstens ein Jahr. Lediglich 15 Urteilerpaare (8%) lagen mehr als ein Jahr auseinander. Die maximale Abweichung der beiden Angaben betrug sechs Jahre.

Die durchschnittliche Bekanntschaftsdauer der Urteilerpaare betrug 13,7 Jahre bei einem Range von 2 Monaten bis zu 60 Jahren (Standardabweichung 15,41 Jahre). Per Quartilssplit wurde eine vierstufige Gruppierungsvariable erzeugt, deren Klassengrenzen null bis einem Jahr, einem bis sechs Jahren, sechs bis 21 und mehr als 21 Jahren Bekanntschaftsdauer entsprechen.

Tabelle 17 zeigt, daß sich auf Personen-Level kein Zusammenhang zwischen Bekanntschaftsdauer und Urteilsübereinstimmung belegen ließ. Die Korrelationen der kontinuierlich gemessenen Bekanntschaftsdauer mit den Maßen der dyadischen Übereinstimmung fielen sehr niedrig und allesamt deutlich nichtsignifikant aus. Ein Vergleich der durchschnittlichen dyadischen Übereinstimmung zwischen den vier per Quartilssplit erzeugten Subgruppen ergab ebenfalls keine Ergebnisse: Lediglich die Übereinstimmungswerte der Gruppe null bis ein Jahr Bekanntschaftsdauer fielen etwas niedriger aus als die der anderen drei Gruppen, zwischen denen es keine relevanten Unterschiede gab (vgl. Tabelle F4 im Anhang F). Die Abweichung der ersten Gruppe kann wenn überhaupt nur als Trend gewertet werden, da sie nichtsignifikant blieb.

Auf Item-Level wurden die itemspezifischen Korrelationen der Rohwerte der BDI-V Selbst- und Fremdeinschätzungen innerhalb der jeweiligen Subgruppen miteinander verglichen. Dabei zeigte sich zwar ein Trend zur größeren Urteilsübereinstimmung in den Gruppen mit längerer Bekanntschaftsdauer, dieser lag jedoch nur bei 13 der 20 Items vor und verlief nicht durchgehend über alle vier Gruppen (vgl. Tabelle F5 im Anhang F). Ein Grund für die inkonsistenten Ergebnisse könnte in dem kleinen Umfang der Subgruppen liegen (N=44 bzw. N=45).

Tabelle 17: Korrelationen der kontinuierlichen Variable Bekanntschaftsdauer mit den Kennwerten der dyadischen Urteilsübereinstimmung

	Profil- korrelation	Ipsative Profil- korrelation [#]	Kongruenz- koeffizient c	Ähnlichkeits- koeffizient r _{PA}
Bekanntschaftsdauer in Jahren	.07	.07	.08	.13

Anmerkungen: N=177 Urteilerpaare, [#] Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte. Profilkorrelation und ipsative Profilkorrelation gingen als Fisher-z-Werte in die Berechnungen ein.

5.5.2 Treffensfrequenz der Urteilerpaare (quantitativer Aspekt)

Die Angaben der Selbst- und Fremdbeurteiler bezüglich der Frequenz ihres Umgangs wurden logisch zu einem Wert vereinigt. Leider ergab sich eine stark schiefe Verteilung. Von den 177 Urteilerpaaren sehen sich 112 (63,2%) "täglich" oder "mehrmals täglich", 52 (29,4%) noch "mehrmals wöchentlich" oder "1-2x wöchentlich" und nur 13 Paare (7,3%) "seltener als wöchentlich", was einen Subgruppenvergleich auf Personen- oder Item-Level unmöglich macht, da die Gruppengrößen zu stark differieren. Die korrelative Auswertung auf Personen-Level ergab keinen Zusammenhang zwischen der Treffensfrequenz der Urteilerpaare und dem Grad der Übereinstimmung ihrer Beurteilungen, die Korrelationen mit den Maßen der dyadischen Übereinstimmung lagen alle nahe null (vgl. Tabelle F6 im Anhang F).

5.5.3 Bekanntschaftsgrad der Urteilerpaare (qualitativer Aspekt)

Zur globalen Messung des Bekanntschaftsgrades im Sinne der Vertrautheit und Beziehungstiefe diente die Aussage "Ich kenne die andere Person meiner Meinung nach ..." mit den Multiple Choice Antwortoptionen "sehr gut / gut / kaum". Die Angaben der Selbst- und Fremdbeurteiler wurden logisch zu einer fünfstufigen kategorialen Variable vereinigt, die die Bekanntschaftsenge der Urteilerpaare differenziert abbildet (vgl. Tabelle 18).

Tabelle 18: Schema der Erzeugung eines Gesamtwertes des Bekanntschaftsgrades durch Kombination der beiden Einzelangaben von Judge und Target

Gesamtwert Bekanntschaftsgrad	Angabe 1 Bekanntschaftsgrad	Angabe 2 Bekanntschaftsgrad
"sehr gut"	"sehr gut"	"sehr gut"
"recht gut"	"sehr gut"	"gut"
"gut"	"gut"	"gut"
"etwas"	"gut"	"kaum"
"kaum"	"kaum"	"kaum"

Anmerkung: Die durch dieses Schema nicht abgedeckte, aber theoretisch mögliche Kombination der beiden Angaben "sehr gut" und "kaum" trat in der vorliegenden Stichprobe bei keinem der Urteilerpaare auf.

Wie Tabelle 19 zeigt, fiel auch die Verteilung des Bekanntschaftsgrades in der vorliegenden Stichprobe extrem schief aus. Die Gruppengrößen differieren leider so stark, daß auch hier ein Subgruppenvergleich auf Personen- oder Item-Level unmöglich ist. Eine Gegenüberstellung der Gruppe "sehr gut" und aller anderen Angaben in einer zweiten Gruppe wäre zwar vom Umfang her möglich, ist jedoch in theoretischer Hinsicht unsinnig, da dadurch die großen inhaltlichen Unterschiede des Bekanntschaftsgrades in der zweiten Gruppe statistisch eingeebnet würden. Die moderierende Wirkung des Bekanntschaftsgrades konnte aber in der korrelativen Auswertung auf Personen-Level eindeutig belegt werden.

Tabelle 19: Häufigkeitswerte der kategorialen Variable Bekanntschaftsgrad

	"sehr gut"	"recht gut"	"gut"	"etwas"	"kaum"
absolute Häufigkeit	90	29	44	8	6
relative Häufigkeit	50,85%	16,38%	24,86%	4,52%	3,39%

Anmerkung: N=177 Urteilerpaare.

Die Korrelation (Spearman's Rho) des Bekanntschaftsgrades mit dem höhenlagenunabhängigen Übereinstimmungsmaß der Profilkorrelation lag bei $r = .29^{**}$, die Korrelation mit dem um Stereotype Accuracy bereinigten Übereinstimmungsmaß der ipsativen Profilkorrelation bei immerhin noch $r = .22^{**}$ (vgl. Tabelle 20). Beide Ergebnisse sind ein klarer Beleg dafür, daß der Bekanntschaftsgrad zwischen Selbst- und Fremdbeurteiler einen mittleren Einfluß auf die Ähnlichkeit der Profilverläufe hat. Mit zunehmendem Bekanntschaftsgrad stimmen die Selbst- und Fremdeinschätzungen der relativen Ausprägungen der einzelnen Erlebens- und Verhaltensweisen des BDI-V zunehmend überein, es gibt einen monotonen Zusammenhang zwischen Acquaintance und Agreement. Die jeweiligen Werte von Spearman's Rho bewegen sich im Bereich der zur Kontrolle ebenfalls berechneten Bravais-Pearson-Korrelationskoeffizienten.

Tabelle 20: Korrelationen der kategorialen Variable Bekanntschaftsgrad mit den Kennwerten der dyadischen Urteilsübereinstimmung

	Profil- korrelation	Ipsative Profil- korrelation [#]	Kongruenz- koeffizient c	Ähnlichkeits- koeffizient r_{PA}
Bekanntschaftsgrad (kategorial, 5-stufig)	.29 ^{**}	.22 ^{**}	.21 ^{**}	.16 [*]

Anmerkungen: N=177 Urteilerpaare, Spearman's Rho mit $*p < .05$, $**p < .01$ (2-seitig),

[#] Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte.

Profilkorrelation und ipsative Profilkorrelation gingen als Fisher-z-Werte in die Berechnungen ein.

Die moderierende Wirkung des Bekanntschaftsgrades zeigte sich auch in signifikanten Korrelationen mit den beiden anderen, höhenlagenabhängigen Übereinstimmungsmaßen c und r_{PA} . Diese fielen jedoch etwas schwächer aus als der Zusammenhang mit der Profilkorrelation (vgl. Tabelle 20), was vermutlich an der Konfundierung der verschiedenen Übereinstimmungskomponenten in den Maßen und ihrer dadurch verursachten größeren Ungenauigkeit liegt. Daß auf dem Ähnlichkeitskoeffizienten r_{PA} überhaupt eine Wirkung angezeigt wird, spricht für den starken Einfluß des Bekanntschaftsgrades auf die Urteilsübereinstimmung, da sich der Koeffizient in den vorhergehenden Analysen als insgesamt zu grob erwiesen hatte, um feinere Unterschiede der Urteilsübereinstimmung abzubilden (vgl. Abschnitte 5.2.1, 5.3.1 und 5.4.1).

5.5.4 Lierte vs. nicht-lierte Urteilerpaare (qualitativer Aspekt)

Der Vergleich der lierten und nicht-lierten Urteilerpaare zeigte erwartungsgemäß, daß die durchschnittliche Übereinstimmung unter den Beziehungspartnern größer war. Auf Personen-Level erzielten die lierten Paare im Schnitt eine Profilkorrelation von $r=.50$ und eine ipsative Profilkorrelation von $r=.37$, die nicht-lierten Urteilerpaaren lagen signifikant ($p<.01$, T-Test) niedriger bei $r=.40$ bzw. $r=.27$ (vgl. Tabelle F3 im Anhang F). Auf Item-Level zeigte der Vergleich der itemspezifischen Selbst-Fremd-Korrelationen, daß die Gruppe der lierten Urteilerpaare bei 15 Items (BDI-V Rohwerte) bzw. 13 Items (personenzentrierte BDI-V Werte) um mindestens $r=.10$ höhere Werte erzielte als die Gruppe der nicht-lierten Paare. Bei drei bzw. sechs Items fielen die Korrelationen nahezu gleich aus und bei lediglich zwei Items bzw. einem Item waren die Werte in der Gruppe der nicht-lierten Paare größer (vgl. Tabelle 21). Bei den gegebenen kleinen Gruppengrößen wurden die Differenzen zwischen den Subgruppen allerdings nur bei neun bzw. fünf Items signifikant, daher sind sie teilweise nur als Trends zu werten.

Tabelle 21: Korrelationen der Selbst- und Fremdeinschätzungen aller BDI-V-Items (Rohwerte, personenzentrierte Abweichungswerte), Subgruppen liert ja/nein

	BDI-V Rohwerte		ipsative BDI-V Werte	
	liert [#]	nicht-liert ^{##}	liert [#]	nicht-liert ^{##}
BDI 01 'Ich bin traurig'	.56**	.35**	.26**	.13
BDI 02 'Ich sehe mutlos in die Zukunft'	.55**	.30**	.31**	.09
BDI 03 'Ich fühle mich als Versager'	.37**	.22	.24**	.05
BDI 04 'Es fällt mir schwer, etwas zu genießen'	.44**	.28*	.34**	.14
BDI 05 'Ich habe Schuldgefühle'	.40**	.06	.21*	-.03
BDI 06 'Ich fühle mich bestraft'	.33**	.19	.20*	.10
BDI 07 'Ich bin von mir enttäuscht'	.45**	.23	.23*	.15
BDI 08 'Ich werfe mir Fehler und Schwächen vor'	.42**	.15	.15	.18
BDI 09 'Ich denke daran, mir etwas anzutun'	.56**	.07	.48**	.11
BDI 10 'Ich weine'	.60**	.36**	.50**	.40**
BDI 11 'Ich fühle mich gereizt und verärgert'	.32**	.44**	.20*	.31**
BDI 12 'Mir fehlt das Interesse an Menschen'	.35**	.04	.34**	-.06
BDI 13 'Ich schiebe Entscheidungen vor mir her'	.44**	.38**	.52**	.37**
BDI 14 'Ich bin besorgt um mein Aussehen'	.40**	.26*	.42**	.36**
BDI 15 'Ich muß mich zu Tätigkeiten zwingen'	.30**	.25*	.33**	.07
BDI 16 'Ich habe Schlafstörungen'	.67**	.31**	.62**	.29**
BDI 17 'Ich bin müde und lustlos'	.38**	.17	.19	.22
BDI 18 'Ich habe keinen Appetit'	.37**	.18	.40**	.13
BDI 19 'Sex ist mir gleichgültig'	.43**	.43**	.44**	.35**
BDI 20 'Ich mache mir Sorgen um meine Gesundheit'	.26**	.42**	.31**	.35**
BDI-V Gesamtscore	.63**	.42**	-	-

Anmerkungen: [#] N=104 Urteilerpaare, ^{##} N=73 Urteilerpaare, * $p<.05$, ** $p<.01$ (2-seitig), **fettgedruckt:** mit $p<.05$ signifikante Korrelationsunterschiede der Subgruppen (Differenz der Fisher-z-Werte $> .27$).

5.5.5 Fazit Acquaintanceship

Die erhobenen quantitativen Variablen "Bekanntschafsdauer" (kontinuierlich) und "Treffensfrequenz der Urteilerpaare" (kategorial) erzielten in der vorliegenden Stichprobe auf Personen-Level keine korrelativen Zusammenhänge mit der Urteilsübereinstimmung (vgl. Tabelle 17 und Tabelle F6 im Anhang F). Nach einer Subgruppenbildung per Quartilssplit zeigte sich hinsichtlich der Bekanntschafsdauer zwar ein schwacher erwartungsgemäßer Trend auf Item-Level, dieser lag jedoch nur bei 13 von 20 BDI-V Items vor (vgl. Tabelle F5 im Anhang F). Die Verteilung der Treffensfrequenz war zu schief, um auf Personen- oder Item-Level eine sinnvolle Auswertung über Subgruppen vornehmen zu können.

Die erhobenen dichotomen, qualitativen Acquaintanceship-Variablen "Gemeinsame intensive Erlebnisse ja/nein" und "Gemeinsame private Situationen ja/nein" konnten ebenfalls nicht genutzt werden, da sie von nahezu allen Urteilerpaaren bejaht wurden (154 bzw. 168 von 177 Paaren).

Auch die erhobene kategoriale, qualitative Acquaintanceship-Variable "Bekanntschaftsgrad" ließ sich auf Item-Level nicht sinnvoll auswerten, ergab jedoch auf Personen-Level signifikante Ergebnisse. Der Zusammenhang der Globaleinschätzung des Bekanntschaftsgrades mit der Übereinstimmung der Selbst- und Fremdbeurteiler war eindeutig: Die Korrelationen von $r=.29^{**}$ mit dem Übereinstimmungsmaß der Profilkorrelation und $r=.22^{**}$ mit dem bereinigten Übereinstimmungsmaß der ipsativen Profilkorrelation (vgl. Tabelle 20) belegen den starken Einfluß der Bekanntschafstseuge auf die Güte laienpsychologischer Persönlichkeitsbeurteilungen. Eine Analyse der Wirkung einer zweiten erhobenen qualitativen Bekanntschafstseuge, des Bekanntschafstseutyps (Blutsverwandtschaft vs. Partnerschaft vs. Freundschaft vs. Bekantschaft), war leider nicht möglich, da das Merkmal in der vorliegenden Stichprobe zu stark schiefverteilt war.

Quantitative Acquaintanceship-Variablen wiesen also wie in einigen früheren Studien keinen klaren Zusammenhang mit der Urteilsübereinstimmung auf, dennoch konnte über die qualitative Einschätzung des Bekanntschafstseuges unter dem Strich ein starker Einfluß der Bekantschaftseuge auf das Ausmaß der Urteilsübereinstimmung belegt werden:

Hypothese 10 wurde auf Personen-Level bestätigt.

Eindeutig waren auch die Ergebnisse bezüglich der erhobenen dichotomen, qualitativen Acquaintanceship-Variable "Urteilerpaar liiert ja/nein": Sowohl auf Personen- als auch auf Item-Level zeigte sich, daß die Urteilsübereinstimmung bei den liierten Paaren größer war. Die Differenzen waren auf Personen-Level signifikant (vgl. Tabelle F3 im Anhang F), auf Item-Level teilweise signifikant und teilweise als Trend zu werten (vgl. Tabelle 21).

Hypothese 11 wurde auf Item- und auf Personen-Level bestätigt.

5.6 Trait: Moderator Beobachtbarkeit (Hypothese 12)

5.6.1 Erfassung der Observability

Reliabilität der Messung: Der Onlinefragebogen zur Beobachtbarkeit der einzelnen BDI-V Items wurde von insgesamt 219 Personen ausgefüllt. Alle Verhaltens- und Erlebensweisen wurden sowohl 'negativ' formuliert (Originalvariante der Items) als auch inhaltlich entgegengesetzt ('positiv' formuliert) dargeboten, so daß insgesamt 40 Beschreibungen bewertet werden mußten. Die Reliabilität der Messung wurde anschließend durch Korrektur der mittleren Inter-Item Korrelation mit der Spearman-Brown-Formel (Cronbachs α) berechnet und fiel sehr gut aus. Sie lag für die 20 Originalitems des BDI-V bei $\alpha=.86$, für die 20 umformulierten Aussagen bei $\alpha=.88$ und über alle 40 Items zusammen bei $\alpha=.90$.

Unterschiede zwischen den Items: Die eingeschätzte Beobachtbarkeit der verschiedenen Items des BDI-V differierte deutlich, der Range der Gesamtwerte reichte auf der neunstufigen Likert-Skala von 3,32 für das Item BDI 09 'Die Person denkt daran, sich etwas anzutun' bis 7,15 für das Item BDI 13 'Die Person schiebt Entscheidungen vor sich her'. Die Ergebnisse sind sehr stabil, gegenüber einer ersten Zwischenauswertung auf Basis von N=110 Untersuchungsteilnehmern hat sich in der endgültigen Auswertung mit N=219 kein einziger Wert um mehr als 0,2 auf der Skala von 1 ('gar nicht beobachtbar') bis 9 ('vollkommen beobachtbar') verschoben. In Tabelle 22 sind die Observability-Werte aller Items dargestellt.

Unterschiede zwischen den beiden Polungen der Items: Die Profilkorrelation der (mittleren) Observability-Ratings beider Formulierungen der Verhaltens- und Erlebensweisen über alle 20 Items lag bei $r=.73^{**}$ ($p<.01$, 2-seitig). Tendenziell wird die Beobachtbarkeit der "positiven" Pole bei den meisten Items (15 von 20) als höher angesehen, zumeist jedoch nur marginal. Lediglich bei fünf Items wich die mittlere Einschätzung der beiden Pole stärker (i.e. um mehr als einen Punkt auf der Observability-Skala von eins bis neun) voneinander ab: Nach Einschätzung der Untersuchungsteilnehmer ist es besser beobachtbar, wenn ein Mensch mit sich zufrieden und entscheidungsfreudig ist, Appetit hat sowie Interesse an Sex, als wenn er von sich enttäuscht ist, Entscheidungen vor sich her schiebt, keinen Appetit hat und ihm Sex gleichgültig ist. Bei einem Item fiel die Abweichung deutlich und inhaltlich entgegengesetzt aus: Gelassenheit soll schwieriger zu beobachten sein als Gereizt- und Verärgertheit (vgl. Tabelle 22).

Die verschiedenen Verhaltens- und Erlebensweisen, die das BDI-V thematisiert, werden als unterschiedlich gut beobachtbar angesehen, mit deutlichen Unterschieden zwischen den einzelnen Items. Die Einschätzungen haben eine geringe interindividuelle Varianz, i.e. Menschen bewerten die Beobachtbarkeit der verschiedenen Verhaltens- und Erlebensweisen jeweils verhältnismäßig konform. Bei manchen Items interagiert die Einschätzung der Beobachtbarkeit mit ihrer Ausprägung.

Tabelle 22: Absteigend angeordnete Observability-Werte aller BDI-V Items: Gesamtwert und Normalform, Abweichung der positiv umgepolten Form, Korrelation zwischen beiden Formen

	Observability-Werte [#]			
	Gesamtwert ⁺	Normalform (negativ)	Abweichung Positivform ⁺⁺	Korrelation der beiden Formen
BDI 10 'Person weint'	7,15	7,58	-0,91	.33**
BDI 01 'Person ist traurig'	6,45	6,66	-0,41	.17**
BDI 17 'Person ist müde und lustlos'	6,20	5,92	0,55	.45**
BDI 11 'Person fühlt sich gereizt und verärgert'	6,12	6,97	-1,69	.20**
BDI 12 'Person fehlt das Interesse an Menschen'	5,72	5,28	0,88	.37**
BDI 04 'Person fällt es schwer, etwas zu genießen'	5,70	5,23	0,95	.32**
BDI 14 'Person ist besorgt um ihr Aussehen'	5,64	5,76	-0,24	.18**
BDI 18 'Person hat keinen Appetit'	5,49	4,65	1,66	.36**
BDI 07 'Person ist von sich enttäuscht'	5,16	4,26	1,80	.18**
BDI 13 'Person schiebt Entscheidungen vor sich her'	5,05	4,19	1,71	.12*
BDI 02 'Person sieht mutlos in die Zukunft'	4,76	4,38	0,76	.50**
BDI 03 'Person fühlt sich als Versager'	4,53	4,13	0,82	.31**
BDI 15 'Person muß sich zu Tätigkeiten zwingen'	4,43	4,25	0,36	.38**
BDI 08 'Person wirft sich Fehler und Schwächen vor'	4,22	4,10	0,25	.23**
BDI 20 'Person macht sich Sorgen um ihre Gesundheit'	4,21	4,08	0,25	.43**
BDI 05 'Person hat Schuldgefühle'	4,09	4,07	0,04	.36**
BDI 06 'Person fühlt sich bestraft'	4,08	4,18	-0,19	.34**
BDI 16 'Person hat Schlafstörungen'	3,83	3,55	0,57	.48**
BDI 19 'Sex ist Person gleichgültig'	3,32	2,59	1,48	.37**
BDI 09 'Person denkt daran, sich etwas anzutun'	3,21	3,08	0,25	.11*

Anmerkungen: N=219, [#] Skala von 1 ('gar nicht beobachtbar') bis 9 ('vollkommen beobachtbar'), ⁺ Mittelwert der beiden Formen, ⁺⁺ positive Werte bedeuten, daß inhaltlich 'positive' (i.e. nicht-depressive) Ausprägungen der Eigenschaft höhere Observability-Werte haben, *p<.05, **p<.01 (2-seitig).

5.6.2 Zusammenhang mit Urteilsübereinstimmung

Die Unterschiede der durchschnittlichen Selbst-Fremd-Übereinstimmung zwischen den einzelnen Items deuten darauf hin, daß es tatsächlich Unterschiede in der Fremdbeurteilbarkeit der spezifischen Erlebens- und Verhaltensweisen des BDI-V gibt (vgl. Abschnitt 5.2.3 und Tabelle 8).

Zur Quantifizierung des Zusammenhangs von itemspezifischem Agreement und itemspezifischer Observability wurden zuerst die Korrelationen der Selbst- und Fremdeinschätzungen aller 20 Itempaare über die ganze Stichprobe berechnet und in Fisher-z-Werte umgewandelt. Diese normalisierten Übereinstimmungswerte wurden dann über alle 20 Items mit den jeweiligen Werten der Beobachtbarkeit korreliert (vgl. Tabelle F7 im Anhang F). Das Resultat war eine mäßige Bravais-Pearson-Korrelation von $r=.20$. Die Korrelation der beiden Rangreihen Observability und Agreement ergab ein Kendalls Tau von $\tau=.18$. Der Zusammenhang von Beurteilbarkeit und Beobachtbarkeit ist also vorhanden, aber nur moderat ausgeprägt.

Hypothese 12 konnte auf Item-Level bestätigt werden.

6. Diskussion

6.1 Zusammenfassung und Interpretation der Ergebnisse

Die beiden globalen Anliegen der Untersuchung wurden erreicht: Die inhaltliche Treffsicherheit des BDI-V wurde erneut bestätigt und damit gleichzeitig eine passable Güte der alltäglichen laienpsychologischen Persönlichkeitsbeurteilung belegt.

Auf Personen-Level erreichten die Selbst- und Fremdbeurteiler eine gute durchschnittliche Übereinstimmung (mittlere Selbst-Fremd-Profilkorrelation von $r=.46$, um Stereotype Accuracy bereinigte mittlere ipsative Profilkorrelation von $r=.33$, vgl. Tabelle 6). Auf Item-Level zeigte sich ein ähnliches Bild: Die aggregierten BDI-V Gesamtscores der Selbst- und Fremdeinschätzer korrelierten mit $r=.52^{**}$, die Werte der einzelnen Items lagen etwas niedriger, bei einer guten mittleren Korrelation von $r=.36$ (vgl. Tabelle 8).

Die Reliabilität des BDI-V fiel gut aus (vgl. Tabellen 4 und 5), die Ergebnisse der Studie sind aber vor allem als weitere Validierung des Instruments zu verstehen, da die Übereinstimmung der Selbst- und Fremdbeurteilungen ein für Persönlichkeitsinventare typisches Ausmaß erreichte (vgl. FUNDER, 1999, FUNDER & COLVIN, 1997, MCCRAE ET AL., 1998, WATSON ET AL., 2000) und damit die Tauglichkeit des BDI-V zur Erfassung depressiver Zustände durch Laienauskunft erneut bestätigt wurde. Ein weiterer Beleg für die inhaltliche Angemessenheit des Inventars ist die Tatsache, daß die ermittelte Übereinstimmung mit zunehmendem Bekanntschaftsgrad größer wird (vgl. Tabelle 20): Natürlich kriegen gute Freunde depressive Verstimmungen eher mit als Bekannte.

Interessant sind die Differenzen unter den BDI-V Items: Bei einigen fiel die Übereinstimmung der Selbst- und Fremdeinschätzungen deutlich höher aus als bei anderen, analog dazu wurde die Fremdbeobachtbarkeit der einzelnen Erlebens- und Verhaltensweisen von einer unabhängigen Stichprobe auch als unterschiedlich hoch eingeschätzt (vgl. Tabelle 22). Allerdings korrespondierten diese Einschätzungen nur mäßig mit der tatsächlichen itemspezifischen Übereinstimmung (vgl. Tabelle F7 im Anhang F).

In einer erweiterten Perspektive sind die Ergebnisse auch ein Zeugnis der Qualität alltäglicher menschlicher Persönlichkeitsbeurteilungen, die keineswegs solipsistisch sind, sondern erstaunlich kongruent ausfallen. Das bestätigt die Sichtweise der "positiven Psychologie" und des Präzisionsparadigmas (vgl. Abschnitt 2.1.4). Die pessimistische Annahme einer generellen Verzerrung von Selbsturteilen wurde dagegen von den Daten nicht gestützt.

Das nähere Anliegen der Studie, einige aus dem Realistic Accuracy Model (FUNDER, 1995, 1999) theoretisch abgeleitete Moderatoren der Urteilsgüte empirisch zu belegen, scheiterte weitestgehend. Auf Item-Level hing keine der erhobenen Eigenschaften relevant mit dem Ausmaß der Urteilsübereinstimmung zusammen, i.e. die Regressionsanalysen zeigten lediglich sporadische Wirkungen einiger potentieller Moderatoren auf vereinzelt Items an (vgl. Tabellen 11 und 16). Auf Personen-Level korrelierten nur wenige Eigenschaften mit der Urteilsübereinstimmung, von einer Ausnahme abgesehen nur schwach bis mäßig und teilweise in nicht erwartungsgemäßer Richtung. Tabelle 23 stellt noch einmal alle signifikanten Ergebnisse auf Personen-Level im Überblick dar.

Tabelle 23: Moderatorwirkungen auf Personen-Level: Signifikante Korrelationen der potentiellen Moderatoren mit den Maßen der dyadischen Urteilsübereinstimmung

	Profil- korrelation	Ipsative Profil- korrelation [#]	Kongruenz- koeffizient c	Ähnlichkeits- koeffizient r _{PA}
<u>Judge</u>				
Empathie	-	-.15*	-.15*	-.16*
Perspektivübernahme	-	-	-	-
Offenheit d. WS	-	-	-	-
Lebensalter	-	-	-	-
<u>Target</u>				
Selbsttäuschung	-	-	-.16*	-
Fremdtäuschung	-.16*	-	-	-
Privatheit	-	-	-	-
Konsistenz (SD)	.14*	.20**	-	-
<u>Acquaintanceship</u>				
Bekanntschafsdauer	-	-	-	-
Treffensfrequenz	-	-	-	-
Bekanntschaftsgrad	.29**	.22**	.21**	.16*

Anmerkungen: N=177, *p<.05, **p<.01 (2-seitig), [#] Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte. Profilkorrelation und ipsative Profilkorrelation gingen als Fisher-z-Werte in die Berechnungen ein.

Judge: Von den potentiellen Moderatoren der Fremdbeurteiler hingen Perspektivübernahme, Offenheit des Wertesystems und Lebensalter auch auf Personen-Level nicht mit der Urteilsübereinstimmung zusammen. Für Empathie ergaben sich mehrere der Richtung nach unerwartete Korrelationen (vgl. Tabelle 23), diese waren jedoch schwach und sanken durch Ausparialisierung des Bekantschaftsgrades unter die Signifikanzgrenze (vgl. Tabelle F1 im Anhang F). Vermutlich handelt es sich um einen Artefakt aufgrund einer unkontrollierten Konfundierung, die Validität der Empathiemessung scheint zweifelhaft (vgl. Abschnitt 6.2).

Target: Von den potentiellen Moderatoren der Selbstbeurteiler zeigte Privatheit auch auf Personen-Level keinerlei Wirkung. Dafür ergaben sich dort für die Fremd- und Selbsttäuschung erwartungsgemäße und signifikante, wenn auch schwache Zusammenhänge mit der dyadischen Übereinstimmung (vgl. Tabelle 23), die durch Auspartialisierung des Bekanntschaftsgrades noch etwas stärker wurden (jeweils $r = -.20^{**}$ statt $r = .16^*$, vgl. Tabelle F2 im Anhang F). Damit wurde das in der Testtheorie altbekannte Problem der Verfälschung von Selbstauskünften aufgrund sozialer Erwünschtheitstendenzen erneut aufgezeigt und die Existenz eines personenspezifischen self-serving Biases belegt. Interessanterweise hing Fremdtäuschung nur mit dem Ausmaß der unkorrigierten Profilverlaufsähnlichkeit zusammen, während die Korrelationen mit den anderen Übereinstimmungsmaßen nahe Null lagen. Die Selbsttäuschungstendenz hing dagegen nur mit dem höhenlagenabhängigen Kongruenzkoeffizienten c und mit der absoluten Profilhöhenunähnlichkeit zusammen (vgl. Tabellen 15 und 23).

Das Ausmaß der Selbsttäuschung hat also keinen Einfluß auf die Ähnlichkeit der Profilverläufe der Selbst- und Fremdeinschätzer, aber auf die Ähnlichkeit der Profilhöhen. Anders ausgedrückt: Bei einer stärker ausgeprägten Selbsttäuschungstendenz beurteilt sich eine Person zwar insgesamt positiver als sie von der Umwelt gesehen wird, kann aber dennoch ein differenziertes und stimmiges Selbstbild haben, das die relativen Ausprägungen einzelner Eigenschaften zueinander richtig einschätzt. Dagegen wirkt sich eine stärker ausgeprägte Fremdtäuschungstendenz nicht als gleichmäßige Verschiebung oder Verzerrung aus, sie verfälscht vielmehr das Gesamtbild der Persönlichkeit, eben das Antwortprofil, indem die Auskünfte bei einzelnen Aspekten und Merkmalen (i.e. Items) unterschiedlich stark an subjektiven Normen ausgerichtet werden.

Die Ergebnisse zum Targetmoderator Metatrait bzw. Konsistenz waren widersprüchlich: Auf Item-Level beeinflusste die Standardabweichung des Selbsturteils bei keinem einzigen Item signifikant das Ausmaß der Selbst-Fremd-Übereinstimmung, aber auf Personen-Level ergaben sich unerwartete positive Korrelationen mit zwei Übereinstimmungskoeffizienten (vgl. Tabelle 23). Damit scheint die Metatrait-Hypothese widerlegt, da eine größere Konsistenz des Selbsturteils eben nicht mit höherer, sondern mit niedrigerer dyadischer Urteilsübereinstimmung einherzugehen scheint. Allerdings dürfte es sich hierbei um einen weiteren Artefakt handeln, denn die betroffenen Maße sind die Profilkorrelation und die ipsative Profilkorrelation der Selbst- und Fremdeinschätzungen.

Eine geringe Varianz zu korrelierender Variablen wirkt sich begrenzend auf die Höhe der erzielbaren Korrelation aus (BORTZ, 1999). Insofern senkt eine als niedrige Standardabweichung des Selbsturteils definierte hohe Konsistenz automatisch den Wert von Übereinstimmungsmaßen, die auf korrelativer Basis aus dem gleichen Selbsturteil erzeugt werden. In Vorgriff auf Abschnitt 6.2 (kritische Bewertung und Grenzen der vorliegenden Arbeit) kann festgestellt werden, daß die Operationalisierung der Konsistenz im Rahmen des aktuellen Untersuchungsansatzes inadäquat war und eine Art statistische Zirkularität erzeugt hat. Mit dem Konstrukt Metatrait meinen BAUMEISTER UND TICE (1988) die Dimension, inwieweit eine Person überhaupt einem bestimmten Trait oder einer bestimmten Traitstruktur unterliegt und dadurch eine transsituationale Konsistenz bestimmter Verhaltensweisen zeigt. Seine Messung über die Standardabweichung des Antwortprofils auf einem Satz auf die gleiche Persönlichkeitseigenschaft bezogener Items ist zwar durchaus üblich, aber der ermittelte Wert sollte dann hinsichtlich seiner moderierenden Wirkung auf die prognostische Validität von Selbst- oder Fremdeinschätzungen überprüft werden, also den Zusammenhang von Verhaltensvorhersagen und unabhängig gemessenem Verhalten (BRITT, 1993).

Acquaintanceship: Die stärkste aufgefundene Moderatorwirkung ging wie erwartet vom Bekanntschaftsgrad der Selbst- und Fremdbeurteiler aus, der auf Personen-Level signifikante Korrelationen von bis zu $r=.29^{**}$ mit der dyadischen Urteilsübereinstimmung erzielte (vgl. Tabelle 23). Außerdem zeigte der Vergleich der durchschnittlichen Urteilsgüte von liierten und nicht-liierten Urteilerpaaren ebenso erwartungsgemäß, daß die liierten Paare klar besser abschnitten (vgl. Tabelle 21). Beide Ergebnisse untermauern die große Bedeutung des Faktors Acquaintanceship für die Güte der laienpsychologischen Persönlichkeitsbeurteilung. Kein Zusammenhang konnte jedoch zwischen den erhobenen quantitativen Bekanntschafts-Variablen (Bekannschaftsdauer, Treffensfrequenz) und der Urteilsübereinstimmung belegt werden.

Observability: Zwischen der unabhängig eingeschätzten Beobachtbarkeit der verschiedenen Items des BDI-V und ihrer itemspezifischen Übereinstimmung besteht ein moderater Zusammenhang. Die Berechnungen ergaben Korrelationen der beiden Werte bzw. Rangreihen der Observability und des Agreements von $r=.20$ bzw. $\tau=.18$ (vgl. Tabelle F7 im Anhang F). Die Ergebnisse des Traitmoderators Beobachtbarkeit fielen damit erwartungsgemäß aus, wenn auch nicht übermäßig stark ausgeprägt.

6.2 Kritische Bewertung, Grenzen vorliegender Arbeit

Die Übereinstimmung der Selbst- und Fremdeinschätzungen depressiver Erlebens- und Verhaltensweisen auf dem BDI-V erreichte insgesamt ein befriedigendes Niveau. Zwischen den verschiedenen Paaren gab es deutliche Unterschiede im Ausmaß der Urteilsübereinstimmung, deren Quellen jedoch kaum erhellt werden konnten: Die meisten Persönlichkeitseigenschaften und -merkmale, die als potentielle Moderatoren seitens der Selbst- und Fremdeinschätzer erhoben wurden, zeigten in der Analyse keinen Zusammenhang mit den Übereinstimmungsmaßen. Eine solche empirische Nichtbestätigung theoretisch abgeleiteter Einflußfaktoren bedeutet aber natürlich nicht zwangsläufig, daß die postulierten Zusammenhänge inexistent sind und damit die zugrundeliegenden Theorien falsch. Mangelnde Ergebnisse können auch durch allgemeine und spezifische Mängel des Untersuchungsdesigns verursacht werden.

Bereits in diversen Studien zuvor konnte ebenfalls kein Einfluß bestimmter theoriebasierter Moderatoren auf die Güte laienpsychologischer Persönlichkeitsbeurteilungen belegt werden: "Nonexperimentalists conducting field research have found moderator effects to be extremely difficult to detect" (MCCLELLAND & JUDD, 1993, S. 376). Meistens ergaben sich nur schwache bis mittlere Effekte für den Bekanntschaftsgrad der Beurteiler und die Beobachtbarkeit bzw. Sichtbarkeit der zu beurteilenden Eigenschaften (vgl. FUNDER, 1999, HASELTON & FUNDER, in press). Der Nachweis des Einflusses spezifischer Merkmale der Beurteiler und/oder Beurteilten auf die Qualität von Persönlichkeitseinschätzungen gelang zumeist nur experimentell, bei konstanter Niedrighaltung des Bekanntschaftsgrades (vgl. FUNDER, 1999, VOGT & COLVIN, 2003).

Hier zeigt sich die generelle Schwierigkeit, einzelne Facetten komplexer und multikausaler Alltagsphänomene statistisch nachzuweisen. Aus dem Realistic Accuracy Model wird ersichtlich, daß es eine potentiell sehr große Anzahl von Einflußfaktoren geben dürfte, die sich auf die Urteilsgüte auswirken können (vgl. Abschnitt 2.2.3). Daraus folgt, daß der quasiexperimentell meßbare Einfluß eines isolierten Faktors eher gering sein wird. Der im Allgemeinen bereits diffizile Nachweis eines solchen schwachen Effektes wird durch spezifische Mängel des Untersuchungsdesigns noch weiter erschwert. Im Zusammenhang der vorliegenden Arbeit gibt es vier kritische Aspekte, welche die Qualität der Ergebnisse negativ beeinflußt haben könnten: Die Erfassung der Moderatoren und deren jeweilige Verteilung in der Stichprobe, die prinzipielle Bestimmung der Urteilsgüte und die konkrete Berechnung der Urteilsübereinstimmung.

Erfassung der Moderatoren, Verteilung in der Stichprobe

Zweifel an einer adäquaten Erfassung scheinen insbesondere bei den Moderatoren Empathie, Perspektivübernahme, Offenheit des Wertesystems (alle Judge) und Privatheit (Target) angebracht, die alle keine erwartungsgemäßen Ergebnisse zeigten.

Die massierte Abfrage der Items zur Offenheit des Wertesystems des Judges und der Privatheit des Targets dürfte die Validität der an und für sich bewährten Subskalen des NEO-PI-R bzw. 16 PF-R beeinträchtigt haben. Für beide Merkmale gilt, daß die Items der Skalen von den meisten Untersuchungsteilnehmern intraindividuell eher stereotyp beantwortet wurden (Redundanz-Effekt). Außerdem weisen die aggregierten Gesamtwerte der beiden Skalen insgesamt nur eine mäßige interindividuelle Varianz auf. Vermutlich ist es nicht gelungen, die individuellen Ausprägungen der beiden Eigenschaften korrekt abzubilden. Vielleicht waren sich die Untersuchungsteilnehmer bezüglich dieser Merkmale aber auch zu ähnlich, so daß der moderierende Effekt der Offenheit bzw. der Privatheit nicht zum tragen kommen konnte. Hier wäre eine deutlich größere und heterogenere Stichprobe wünschenswert gewesen, aber leider war eine noch breiter angelegte Erhebung aus organisatorischen und logistischen Gründen nicht möglich.

Die Tauglichkeit von Selbstberichten zur Erfassung von Perspektivübernahme und Empathie ist zumindest fraglich. Vieles deutet darauf hin, daß es bei der Messung zu starken Verzerrungen aufgrund sozialer Erwünschtheitstendenzen kommt (STEINS, 1998). Das kontraintuitive Resultat des Moderators Empathie und die Nichtwirksamkeit des Moderators Perspektivübernahme in der vorliegenden Untersuchung könnten also damit zusammenhängen, daß die individuellen Ausprägungen dieser Eigenschaften der Fremdbeurteiler nicht adäquat abgebildet wurden, weil die Werte eine Konfundierung mit anderen Eigenschaften aufweisen. Leider wurden die sozialen Erwünschtheitstendenzen der Judges nicht miterhoben und konnten daher auch nicht statistisch kontrolliert werden. Eine validere Messung von Empathie ist über physiologische Parameter und standardisierte Beobachtung der Mimik bei instruierten Imaginationsaufgaben möglich (STEINS, 1998). Für die Erfassung von Perspektivübernahme gibt es eine Reihe elaborierter Verfahren, die nicht auf Selbstauskunft, sondern instruiertem Agieren der Untersuchungsteilnehmer beruhen (z.B. PIAGET & INHELDER, 1947, FLAVELL ET AL., 1968, HASS, 1984, zit. n. STEINS & WICKLUND, 1993). Die Anwendung der genannten Methoden an einer größeren Stichprobe wäre jedoch zu aufwendig gewesen und hätte den Rahmen einer Diplomarbeit gesprengt.

Die adäquate Erfassung der Fremdtäuschungs- und Selbsttäuschungstendenz der Targets scheint dagegen gelungen zu sein, beide zeigten erwartungsgemäße und signifikante Zusammenhänge mit der Urteilsübereinstimmung. Damit wurde die Verfälschung von Selbstauskünften aufgrund sozialer Erwünschtheit belegt und die Existenz eines selbstwertdienlichen, personenspezifischen Biases. Die Korrelationen der beiden Merkmale mit den Übereinstimmungsmaßen von jeweils $r=.16^*$ (vgl. Tabelle 23) erscheinen zunächst eher niedrig. Angesichts der Tatsache, daß von einzelnen, isolierten Einflußfaktoren nur eine schwache Wirkung zu erwarten ist (siehe oben), sind die erzielten Korrelationen und die darin liegende Bestätigung eines schwachen Zusammenhangs im Rahmen des vorliegenden Untersuchungsansatzes jedoch als befriedigend anzusehen: "Low power is a particular problem in nonexperimental studies, which have much less power for detecting interaction and moderator effects than do experiments" (FRAZIER ET AL., 2004, S. 117). Pessimistische Schätzungen gehen davon aus, daß sich Moderator-effekte im Bereich der Persönlichkeit immer nur im Bereich von höchstens $r=.10$ bis $r=.20$ bewegen (CHAPLIN, 1991, S. 143). Außerdem beziffern Korrelationen bekanntlich nur den linearen bzw. monotonen Zusammenhang zweier Merkmale. Es scheint jedoch theoretisch denkbar, daß höhere Werte bestimmter Moderatorvariablen zumindest in bestimmten Ausprägungsbereichen zu einer progressiven oder sprunghaften Erhöhung der Urteilsgüte führen könnten.

Nicht nur befriedigend sondern sogar erstaunlich gut fiel das Ergebnis bezüglich des Bekanntschaftsgrades von Selbst- und Fremdbeurteiler aus, für den der stärkste Zusammenhang mit der Urteilsübereinstimmung belegt werden konnte (bis zu $r=.29^{**}$, vgl. Tabelle 23). Das repliziert den Standardbefund zum Thema Personality Judgment: Acquaintanceship ist die einzige personenbezogene Moderatorvariable, die bisher wirklich stabil empirisch abgesichert werden konnte (vgl. FUNDER, 1999, HASELTON & FUNDER, in press). Wie in diversen Studien zuvor (z.B. McCrae, 1994, MCCRAE ET AL., 1998, PAULHUS & BRUCE, 1992, WATSON ET AL., 2000) gelang es auch in der vorliegenden Untersuchung nicht, einen linearen Zusammenhang zwischen quantitativen Acquaintanceship-Variablen und Agreement aufzuzeigen. Dafür wurde die moderierende Wirkung des Bekanntschaftsgrades über qualitative Variablen nachgewiesen, analog zu den Ergebnissen anderer Studien (z.B. COLVIN & FUNDER, 1991, PAUNONEN, 1989, WATSON & CLARK, 1991). Die Erfassung des Moderators Bekanntschaftsenge über eine fünf-stufige kategoriale Variable war insofern angemessen.

Der zuvor bereits dokumentierte Effekt der Beobachtbarkeit bzw. Sichtbarkeit von Persönlichkeitseigenschaften auf ihre Fremdbeurteilbarkeit zeigte sich in den Ergebnissen der vorliegenden Studie nur moderat. Die Erfassung der Observability scheint dennoch adäquat gelungen zu sein, wie die relative interindividuelle Stabilität der Werte zeigt (vgl. Abschnitt 5.6.1), auch wenn der Zusammenhang dieser Werte mit der itemspezifischen Urteilsübereinstimmung nur mäßig war (vgl. Abschnitt 5.6.2). Der Unterschied zu anderen Untersuchungen und damit der Grund für eine schwächere Auswirkung der Beobachtbarkeit dürfte eher in der verhältnismäßigen Homogenität der BDI-V Items liegen. Natürlich beziehen sich diese alle auf die gleiche Persönlichkeitsdimension, nämlich Eigenschaftsdepression, auch wenn sie Verhaltens- und Erlebensweisen beschreiben, deren 'Innerlichkeit' differiert. In Studien, die stärkere Zusammenhänge zwischen der Beobachtbarkeit von Traits und ihrer Beurteilbarkeit belegen konnten, wurden fast ausschließlich Unterschiede zwischen den Big Five Persönlichkeitsdimensionen untersucht (z.B. JOHN & ROBBINS, 1993, PAUNONEN, 1989, WATSON ET AL., 2000), die einen viel breiteren Bereich abdecken und sich dementsprechend stärker unterscheiden.

Bestimmung der Urteilsgüte

Unabhängig von der Frage, ob eine angemessene Erfassung der Moderatoren realisiert wurde oder nicht, ist bereits die zugrundeliegende Prämisse dieser Untersuchung heikel, die Quasi-Gleichsetzung dyadischer Urteilsübereinstimmung mit Urteilsgüte. Da es keine Quelle einer objektiven Beschreibung der 'wahren' Persönlichkeitseigenschaften einer Person gibt, muß die Exaktheit und Präzision subjektiver Urteile über eine Person notgedrungen durch den Vergleich mit anderen subjektiven Urteilen bestimmt werden, wenn keine unabhängigen Verhaltensdaten vorliegen. Solche Urteile können einerseits übereinstimmen und gleichzeitig 'unwahr' sein, bei Nichtübereinstimmung ist es andererseits schwierig zu ermitteln, welche Einzelurteile 'wahrer' sind (vgl. Abschnitt 2.1.2).

Diese Dilemmata können abgeschwächt werden, indem möglichst viele subjektive Urteile aggregiert werden, optimalerweise in einem Round Robin Design. Dazu müßten zu jedem Target Einschätzungen mehrerer Urteiler erfaßt werden und von jedem Urteiler Einschätzungen über mehrere Targets, um die personenspezifische Urteilsgüte differenzierter analysierbar zu machen (vgl. Abschnitt 2.1.3). In der vorliegenden Studie beurteilt jedoch ein Judge immer jeweils genau ein Target und dieses gleichzeitig sich selbst, es kommt zu einer Verschachtelung (*engl.* nesting) der Urteiler.

Durch das Nesting kann jede Variation der dyadischen Urteilsübereinstimmung gleichwahrscheinlich auf den Fremd- oder den Selbstbeurteiler zurückgehen, die Eigenschaften der beiden interagieren unkontrolliert miteinander und erzeugen einen konfundierten Accuracy-Score: "This fact creates, at best, a huge source of noise and, at worst, a serious interpretational confound for any attempt to detect properties of judges that are associated with their accuracy" (FUNDER, 1999, S. 143).

Natürlich war diese Problematik bei der Konzeption der Studie bekannt. Ein Round Robin Design oder die Hinzuziehung weiterer Fremdbeurteiler zur Erweiterung der Dyade wäre jedoch für eine Diplomarbeit zu aufwendig gewesen, ebenso die zusätzliche Erfassung von Verhaltensdaten. Für das Globalanliegen der Untersuchung, die Validierung des BDI-V durch den Nachweis einer guten durchschnittlichen Konvergenz von Selbst- und Fremdeinschätzungen, spielt die Verschachtelung der Urteiler keine Rolle. Der Nachweis von Moderatorwirkungen ist jedoch durch das Nesting stark erschwert: "The researcher must be very fortunate indeed to find anything strong or meaningful" (FUNDER, 1999, S. 143). Der stärkste Einfluß auf die Urteilsübereinstimmung konnte für den Bekanntschaftsgrad von Judge und Target belegt werden. Das hängt auch damit zusammen, daß Acquaintanceship der einzige der getesteten Moderatoren ist, der bei beiden Beurteilern jeweils gleich ausgeprägt ist: Das Nesting wird quasi aufgehoben.

Berechnung der Urteilsübereinstimmung

Auf Personen-Level waren die Profilkorrelationen, die nur die Verlaufsähnlichkeit quantifizieren und frei von Elevation sind, am besten geeignet, um Moderatorwirkungen widerzuspiegeln. Die ipsative Profilkorrelation der stichprobenzentrierten BDI-V Abweichungswerte, die frei von Stereotype Accuracy ist, zeigte theoriekonform niedrigere Übereinstimmung an als die Profilkorrelation der BDI-V Rohwerte. Analog dazu fielen auf Item-Level die Korrelationen der personenzentrierten Abweichungswerte niedriger aus als die der Rohwerte, da diese frei von Elevation sind.

Die Berechnung der Urteilsübereinstimmung in der vorliegenden Arbeit erscheint insgesamt angemessen, aus Gründen der Übersichtlichkeit hätten jedoch Maße, die mehrere Ähnlichkeitsaspekte konfundieren (Kongruenzkoeffizient c , Ähnlichkeitskoeffizient r_{PA}) weggelassen werden können. Auch in theoretischer Hinsicht ist die Profilverlaufsähnlichkeit, also die Übereinstimmung in der Bewertung der relativen Ausprägungen der Persönlichkeitsattribute zueinander, der wichtigere Aspekt.

6.3 Ausblick

An enduring problem in the study of accuracy in personality judgment is that consistent individual differences in judgmental ability have been surprisingly difficult to establish. Perhaps this because personality judgment is such an essential life skill that nearly everyone can do it well enough to get by (HASLTON & FUNDER, in press, S. 20).

Die vorliegende Arbeit zeigt erneut eine passable Güte der alltäglichen laienpsychologischen Persönlichkeitsbeurteilung auf. Damit hat sie ihr Hauptanliegen erreicht und neben einem kleinen Beitrag zur umfangreichen Validierung des BDI-V auch weitere Evidenz für die Sichtweise der "positiven Psychologie" und neuerer Strömungen der Sozialpsychologie geschaffen. Die Suche nach Moderatoren der Urteilsübereinstimmung gestaltete sich schwieriger. Auch hierin liegt die Studie im Einklang mit den aktuellen Erkenntnissen im Forschungsbereich Personality Judgment. Vielleicht unterscheiden sich Menschen nur graduell in ihrer allgemein recht gut ausgeprägten Beurteilungsfähigkeit, so daß sehr feine Untersuchungsmethoden angewandt werden müssen, um die Quellen dieser Unterschiede empirisch abbildbar zu machen.

Aus den Erfahrungen der vorliegenden Arbeit lassen sich zwei zentrale Anforderungen für zukünftige Studien zur Identifikation von Moderatoren der Urteilsgüte ableiten:

Erstens sollte ein konfundierter Accuracy-Score unbedingt vermieden werden, da sonst unkontrollierte Interaktionen die schwachen Moderatoreffekte überdecken können. Dazu können einerseits unabhängige Verhaltensdaten als Außenkriterien der Urteilsgüte hinzugezogen werden, andererseits können mehrere Fremdeinschätzungen der gleichen Zielpersonen aggregiert werden, um das Problem der Verschachtelung der Urteilerpaare zu mildern. Beide Methoden erzeugen einen erheblichen Aufwand, v.a. bei größeren Stichproben, dürften jedoch die Reliabilität der ermittelten Werte der Urteilsgüte stark erhöhen. Dadurch werden Moderatoreffekte potentiell leichter nachweisbar.

Zweitens sollte die Erfassung der zu überprüfenden Moderatorvariablen sehr sorgfältig angegangen werden. Um eine hohe Reliabilität und v.a. Validität der Messung zu erreichen, sollten sie optimalerweise nicht über Selbstauskünfte ermittelt werden, da sonst subjektive Verzerrungen drohen. Systematische Beobachtung, verschiedene Testverfahren und (aggregierte) Fremdeinschätzungen gewährleisten eine höhere Objektivität, erzeugen jedoch ebenfalls einen erheblichen Aufwand.

Um bei der Erforschung laienpsychologischer Persönlichkeitsbeurteilungen in Zukunft auch subtilere Einflußfaktoren zu belegen, werden insofern aufwendige und komplexe Studien nötig sein. Für die vorliegende Arbeit bleibt festzuhalten, daß trotz mancher Limitation des Untersuchungsansatzes immerhin schwache Einflüsse der Targetmoderatoren Fremd- und Selbsttäuschung auf die Urteilsübereinstimmung und ein stärkerer Einfluß des Moderators Bekanntschaftsgrad belegt werden konnten, außerdem ein moderater Zusammenhang zwischen der eingeschätzten Beobachtbarkeit der einzelnen BDI-V Items und ihrer tatsächlichen itemspezifischen Übereinstimmung.

Leider ergaben die auf Item-Level berechneten Regressionsanalysen kaum brauchbare Resultate, sicherlich auch aufgrund der teilweise unbefriedigenden Erfassung der Moderatorvariablen (vgl. Abschnitt 6.2). Dennoch deuten die Analysen an, daß die Fremdbeurteilbarkeit bestimmter BDI-V Items stärker von bestimmten Eigenschaften der Zielpersonen oder der Fremdbeurteiler beeinflußt wird als die Fremdbeurteilbarkeit anderer Items. Ein frappierendes Ergebnis war z.B. die Tatsache, daß der stärkste belegte Interaktionseffekt aller 168 einzelnen Regressionsanalysen ausgerechnet den Zusammenhang der Privatheit des Targets und der Selbst-Fremd-Übereinstimmung des Items "Ich denke daran, mir etwas anzutun" (BDI 09) betraf, was inhaltlich unmittelbar einleuchtend erscheint.

Die Binnenstruktur des BDI-V verdient also durchaus noch eine nähere Betrachtung. Weitere Untersuchungen mit größeren Stichproben wären denkbar, um die Unterschiede zwischen den einzelnen Items gezielter zu analysieren. Ein stärkeres Augenmerk auf die itemspezifischen Differenzen hätte den Umfang dieser Arbeit gesprengt, trotzdem ergeben die Daten verschiedene Ansatzpunkte. Beispielsweise sind die Ausprägungen der meisten Erlebens- und Verhaltensweisen des BDI-V besser fremdbeurteilbar, wenn Judge und Target besser miteinander bekannt sind. Bei einigen Items scheint es jedoch tendenziell umgekehrt zu sein. Ein anderes interessantes Phänomen zeigt sich in den Unterschieden der durchschnittlichen itemspezifischen Selbst-Fremd-Abweichung, manche der Erlebens- und Verhaltensweisen werden anscheinend in ihrem Ausmaß von Fremdbeurteilern systematisch über- bzw. unterschätzt.

Dazu zuallerletzt noch eine versöhnliche Nachricht: Am stärksten unterschätzen Menschen offenbar das Ausmaß, in dem sich andere Menschen Sorgen um ihr Aussehen machen (BDI 14), und das quer durch alle Altersklassen!

7. Literatur

- Aiken, L.S. & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Akert, R.M. & Panter, A.T. (1988). Extraversion and the ability to decode nonverbal communication. *Personality and Individual Differences*, 9, 965-972.
- Allport, G.W. (1937). *Pattern and growth in personality*. New York: Holt, Rinehart & Winston.
- Allport, G.W. & Odbert, H.S. (1936). Trait names: A psycholexical study. *Psychological Monographs*, 47, 1-171.
- Ambady, N., Bernieri, F.J. & Richeson, J.A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology*, 32, 201-271.
- Ambady, N., Hallahan, M. & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality & Social Psychology*, 77, 538-547.
- Ambady, N., Hallahan, M. & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, 69, 518-529.
- Ambady, N. & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality & Social Psychology*, 64, 431-441.
- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention* (3., erw. und aktualisierte Aufl.). Berlin: Springer.
- Anderson, J.R. (1996). *Kognitive Psychologie* (2. Aufl., Übers. der 4. amerikan. Aufl.). Heidelberg: Spektrum, Akad. Vlg.
- Asendorpf, J.B. (1999). *Psychologie der Persönlichkeit* (2. Aufl.). Berlin: Springer.
- Aspinwall, L.G. & Staudinger, U.M. (Eds.). (2003). *A psychology of human strengths: Fundamental questions and future directions for a positive psychology*. Washington, DC: American Psychological Association.
- Baumeister, R.F. & Tice, D.M. (1988). Metatraits. *Journal of Personality*, 56, 571-598.
- Beck, A.T. & Steer, R.A. (1987). *Beck Depression Inventory (BDI)*. San Antonio: The Psychological Corporation Inc.
- Bem, D.J. & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, 81, 506-520.
- Bernieri, F.J., Zuckerman, M., Koestner, R. & Rosenthal, R. (1994). Measuring person perception accuracy: Another look at self-other agreement. *Personality and Social Psychology Bulletin*, 20, 367-378.
- Blackman, M.C. & Funder, D.C. (1998). The effect of information on consensus and accuracy in personality judgment. *Journal of Experimental Social Psychology*, 34, 164-181.
- Borkenau, P. & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, 62, 645-657.
- Borkenau, P. & Liebler, A. (1993). Consensus and self-other agreement for trait inferences from minimal information. *Journal of Personality*, 61, 477-496.
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F.M. & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, 86, 599-614.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5., vollst. überarb. und aktualisierte Aufl.). Berlin: Springer.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (2., überarb. Aufl.). Berlin: Springer.

- Branje, S.J.T., Aken, M.A.G. van, Lieshout, C.F.M. van & Mathijssen, J.J.P. (2003). Personality judgments in adolescents' families: The perceiver, the target, their relationship, and the family. *Journal of Personality, 71*, 49-77.
- Brinkmann, R.D. (1998). *Vorgesetzten-Feedback*. Heidelberg: Sauer.
- Britt, T.W. (1993). Metatraits: Evidence relevant to the validity of the construct and its implications. *Journal of Personality and Social Psychology, 65*, 554-562.
- Brodtt, S. & Ross, L. (1998). The role of stereotyping in overconfident social prediction. *Social Cognition, 16*, 225-252.
- Brown, J.D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition, 4*, 353-376.
- Brunswik, E. (1956). *Perception and the representative design of experiments*. Berkeley: University of California Press.
- Cattell, R.B. (1964). *Personality and social psychology. Collected papers*. San Diego: Knapp.
- Chalmers, A.F. (2001). *Wege der Wissenschaft: Einführung in die Wissenschaftstheorie* (5., überarb. und erw. Aufl.). Berlin: Springer.
- Chaplin, W.F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality, 59*, 143-177.
- Cheek, J.M. (1982). Aggregation, moderator variables, and the validity of personality tests: A peer-rating study. *Journal of Personality and Social Psychology, 43*, 1254-1269.
- Cohen, R. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Colvin, C.R. (1993a). 'Judgable' people: Personality, behavior and competing explanations. *Journal of Personality and Social Psychology, 64*, 861-873.
- Colvin, C.R. (1993b). Childhood antecedents of young-adult judgability. *Journal of Personality, 61*, 611-635.
- Colvin, C.R., Block, J. & Funder, D.C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology, 68*, 1152-1162.
- Colvin, C.R. & Funder, D.C. (1991). Predicting personality and behavior: A boundary on the acquaintanceship effect. *Journal of Personality and Social Psychology, 60*, 884-894.
- Costa, P.T. & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory*. Odessa, FL: Psychological Assessment Resources.
- Cronbach, L.J. (1955). Processes affecting scores on 'understanding of others' and 'assumed similarity.' *Psychological Bulletin, 52*, 177-193.
- Dalbert, C. & Schmitt, M. (1986). Einige Anmerkungen und Beispiele zur Formulierung und Prüfung von Moderatorhypothesen. *Zeitschrift für Differentielle und Diagnostische Psychologie, 7*, 29-43.
- Davis, M. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113-126.
- Dawes, R.M. & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes, 65*, 201-211.
- Estes, S.G. (1938). Judging personality from expressive behavior. *Journal of Abnormal and Social Psychology, 33*, 217-236.
- Eye, A. von & Mun, E. Y. (2004). *Analyzing rater agreement: Manifest variable methods*. Mahwah, NJ: Lawrence Erlbaum.
- Farah, A. & Atoum, A. (2002). Personality traits as self-evaluated and as judged by others. *Social Behavior and Personality, 30*, 149-156.
- Festinger, L.A. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140.
- Festinger, L.A. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

- Flink, C. & Park, B. (1991). Increasing consensus in trait judgments through outcome dependency. *Journal of Experimental Social Psychology, 27*, 453-467.
- Frazier, P.A., Tix, A.P. & Barron, K.E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology, 51*, 115-134.
- Frey, D. & Irle, M. (Hrsg.). (1993). *Theorien der Sozialpsychologie, Bd. I: Kognitive Theorien*. Bern: Huber.
- Frey, D. & Irle, M. (Hrsg.). (2002). *Theorien der Sozialpsychologie, Bd. III: Motivations-, Selbst- und Informationsverarbeitungstheorien*. Bern: Huber.
- Funder, D.C. (1980). On seeing ourselves as others see us: Self-other agreement and discrepancy in personality ratings. *Journal of Personality, 48*, 473-493.
- Funder, D.C. (1983). The 'consistency' controversy and the accuracy of personality judgments. *Journal of Personality, 51*, 346-359.
- Funder, D.C. (1987). Errors and mistakes: Evaluating the accuracy of social judgement. *Psychological Bulletin, 101*, 75-90.
- Funder, D.C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science, 2*, 31-39.
- Funder, D.C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652-670.
- Funder, D.C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego: Academic Press.
- Funder, D.C. (2004). *The Personality Puzzle* (3rd Ed.). New York: W.W. Norton & Co.
- Funder, D.C. (2006). Towards a resolution of the personality triad: Persons, situations and behaviors. *Journal of Research in Personality, 40*, 21-34.
- Funder, D.C. & Colvin, C.R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgments. *Journal of Personality and Social Psychology, 55*, 149-158.
- Funder, D.C. & Colvin, C.R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology, 60*, 773-794.
- Funder, D.C. & Colvin, C.R. (1997). Congruence of others' and self-judgments of personality. In R. Hogan, J. Johnson & S. Briggs (Eds.), *Handbook of Personality Psychology* (pp. 617-647). San Diego: Academic Press.
- Funder, D.C. & Drobny, K.M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52*, 409-418.
- Funder, D.C. & Harris, M.J. (1986). On the several facets of personality assessment: The case of social acuity. *Journal of Personality, 54*, 528-550.
- Funder, D.C., Kolar, D.C. & Blackman, M.C. (1995). Agreement among judges of personality: Interpersonal relations, similarity and acquaintanceship. *Journal of Personality and Social Psychology, 69*, 656-672.
- Funder, D.C. & Sneed, C. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology, 64*, 479-490.
- Funder, D.C. & West, S.G. (1993). Consensus, self-other agreement and accuracy in personality judgment: An introduction. *Journal of Personality, 61*, 457-476.
- Gage, N.L. & Cronbach, L.J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review, 62*, 411-422.
- Gaugler, B.B. & Thornton, G.C. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology, 74*, 611-618.
- Gilovich, T., Griffin, D. W. & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.
- Glaserfeld, E. von. (1998). Konstruktion der Wirklichkeit und des Begriffs der Objektivität. In H. Gumin & H. Meier (Hrsg.), *Einführung in den Konstruktivismus* (S. 9-39). München: Piper Verlag.

- Gosling, S.D., John, O.P., Craik, K.H. & Robins, R.W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, 74, 1337-1349.
- Gosling, S.D., Ko, S.J., Mannarelli, T. & Morris, M.E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, 82, 379-398.
- Guadagnoli, E. & Velicer, W. (1991). A comparison of pattern matching indices. *Multivariate Behavioral Research*, 26, 323-343.
- Guilford, J. P. (1959). *Personality*. New York: McGraw-Hill.
- Hammond, K.R. (1996). *Human judgment and social policy*. New York: Oxford University Press.
- Haselton, M.G. & Buss, D.M. (2003). Biases in social judgment: Design flaws or design features? In J. Forgas, K. Williams & B. von Hippel (Eds.), *Responding to the social world: Implicit and explicit processes in social judgments and decisions* (pp. 23-43). New York: Cambridge University Press.
- Haselton, M.G. & Funder, D.C. (in press). The evolution of accuracy and bias in social judgment. In M. Schaller, D.T. Kenrick & J.A. Simpson (Eds.), *Evolution and social psychology*. New York: Psychology Press. Online im Internet verfügbar unter URL www.sscnet.ucla.edu/comm/haselton/webdocs/HaseltonFunder.pdf [20.Juni 2006].
- Haselton, M.G. & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, 10, 47-66.
- Hastie, R. & Rasinski, K.A. (1988). The concept of accuracy in social judgment. In D. Bar-Tal & A.W. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 193-208). Cambridge: Cambridge University Press.
- Hautzinger, M., Bailer, M., Worall, H. & Keller, F. (1994). *Beck-Depressions-Inventar (BDI)*. Bern: Huber.
- Havenstein, N.M. & Alexander, R.A. (1991). Rating ability in performance judgments: The joint influence of implicit theories and intelligence. *Organizational Behavior and Human Decision Processes*, 50, 300-323.
- Hayes, A.F. & Dunning, D. (1997). Construal processes and trait ambiguity: Implications for self-peer agreement in personality judgment. *Journal of Personality and Social Psychology*, 72, 664-677.
- Hofstee, W.K.B. (1994). Who should own the definition of personality? *European Journal of Personality*, 8, 149-162.
- Holzkamp, K. (1983). *Grundlegung der Psychologie*. Frankfurt: Campus-Vlg.
- Jaccard, J., Turrisi, R. & Wan, C.K. (1990). *Interaction effects in multiple regression*. Newbury Park, CA: Sage.
- Jahoda, M. (1958). *Current concepts of positive mental health*. New York: Basic Books.
- John, O.P. & Robins, R.W. (1993). Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61, 521-551.
- John, O.P. & Robins, R.W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, 66, 206-219.
- Jones, E.E. (1990). *Interpersonal perception*. New York: Freeman.
- Jones, E.E. & Nisbett, R. (1971). *The actor and the observer: Divergent perceptions of the causes of behavior*. Morristown, NJ: General Learning Press.
- Kahneman, D.T. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Kammann, R., Smith, R., Martin, C. & McQueen, M. (1984). Low accuracy in judgments of other's psychological well-being as seen from a phenomenological perspective. *Journal of Personality*, 52, 107-123.

- Kanning, U.P. (1999). *Die Psychologie der Personenbeurteilung*. Göttingen: Hogrefe.
- Kenrick, D.T. & Stringfield, D.O. (1980). Personality traits and the eye of the beholder: Crossing some traditional philosophical boundaries in the search for consistency in all of the people. *Psychological Review*, *87*, 88-104.
- Kenrick, D.T. & Funder, D.C. (1988). Profiting from controversy: Lessons from the person-situation-debate. *American Psychologist*, *43*, 23-34.
- Kenny, D.A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.
- Kenny, D.A. (2005). *Social relations model: References*. Online im Internet verfügbar unter URL <http://davidakenny.net/doc/srmbiblio.pdf> [20.Juni 2006].
- Kenny, D.A. & Acitelli, L.K. (2001). Accuracy and bias in the perception of the partner in a close relationship. *Journal of Personality and Social Psychology*, *80*, 439-448.
- Kleiter, E.F. (1987). *Generalität und Binnenstruktur von Persönlichkeitseigenschaften: Mehrebenen-Analysen als Ausweg aus der Personalismus-Situationismus-Debatte*. Weinheim: Deutscher Studien-Vlg.
- Kolar, D.W. (1995). *Individual differences in the ability to accurately judge the personality characteristics of others*. Unpublished doctoral dissertation, University of California, Riverside.
- Kolar, D.W., Funder, D.C. & Colvin, C.R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality*, *64*, 311-337.
- Krueger, J.I. & Funder, D.C. (2004). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, *27*, 313-327.
- Kruglanski, A.W. (1989). The psychology of being 'right': The problem of accuracy in social perception and cognition. *Psychological Bulletin*, *106*, 395-409.
- Lamnek, S. (2005). *Qualitative Sozialforschung* (4., vollst. überarb. Aufl.). Weinheim: Beltz PVU.
- Lester, D. (1991). Accuracy of recognition of genuine versus simulated suicide notes. *Personality and Individual Differences*, *12*, 765-766.
- Letzring, T.D., Wells, S.M. & Funder, D.C. (in press). Quantity and quality of available information affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*. Online im Internet verfügbar unter URL http://www.rap.ucr.edu/Info_Pers.pdf [20.Juni 2006].
- Levesque, M.J. & Kenny, D.A. (1993). Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. *Journal of Personality and Social Psychology*, *65*, 1178-1187.
- Lück, H.E. (2002). *Geschichte der Psychologie. Strömungen, Schulen, Entwicklungen* (3., überarb. u. erw. Aufl.). Stuttgart: Kohlhammer.
- Luft, J. & Ingham, H. (1955). *The Johari window: A graphic model of interpersonal awareness*. Proceeding of the Western Training Laboratory in Group Development, University of California Extension Office.
- Maes, J., Schmitt, M. & Schmal, A. (1995). *Gerechtigkeit als innerdeutsches Problem: Werthaltungen, Kontrollüberzeugungen, Freiheitsüberzeugungen, Drakonität, Soziale Einstellungen, Empathie und Protestantische Arbeitsethik als Kovariate* (Berichte aus der Arbeitsgruppe "Verantwortung, Gerechtigkeit, Moral" Nr. 85). Trier: Universität Trier, Fachbereich I - Psychologie.
- Marangoni, C., Garcia, S., Ickes, W. & Teng, G. (1995). Empathic accuracy in a clinically relevant setting. *Journal of Personality and Social Psychology*, *68*, 854-869.
- McArthur, L.Z. & Baron, R.M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*, 215-238.
- McClelland, G.H. & Judd, C.M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*, 376-390.
- McCrae, R.R. (1993). Agreement of personality profiles across observers. *Multivariate Behavioral Research*, *28*, 25-40.

- McCrae, R.R. (1994). The counterpoint of personality assessment: Self-reports and observer ratings. *Assessment, 1*, 159-172.
- McCrae, R.R. & Costa, P.T. (1995). Trait explanations in personality psychology. *European Journal of Personality, 9*, 231-252.
- McCrae, R.R., Stone, S., Fagan, P.J. & Costa, P.T. (1998). Identifying causes of disagreement between self-reports and spouse ratings of personality. *Journal of Personality, 66*, 285-313.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Moskowitz, D.S. & Schwarz, J.C. (1982). Validity comparison of behavior counts and ratings by knowledgeable informants. *Journal of Personality and Social Psychology, 42*, 518-528.
- Murphy, K. R. & Balzer, W.K. (1989). Rater errors and rating accuracy. *Journal of Applied Psychology, 74*, 619-624.
- Musch, J., Brockhaus, R. & Bröder, A. (2002). Ein Inventar zur Erfassung sozialer Erwünschtheit. *Diagnostica, 48*, 121-128.
- Myers, D.G. (2004). *Intuition: It's powers and perils*. New Haven: Yale University Press.
- Nisbett, R. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. New York: Prentice Hall.
- Nüse, R. (1995). *Über die Erfindung-en des radikalen Konstruktivismus: Kritische Gegenargumente aus psychologischer Sicht* (2., überarb. und erw. Aufl.). Weinheim: Deutscher Studien-Vlg.
- Oltmanns, T.F., Friedman, J.N.W., Fiedler, E.R. & Turkheimer, E. (2004). Perceptions of people with personality disorders based on thin slices of behavior. *Journal of Research in Personality, 38*, 216-229.
- Ostendorf, F. & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae, revidierte Fassung (NEO-PI-R)*. Göttingen: Hogrefe.
- Park, B., Kraus, S. & Ryan, C.S. (1997). Longitudinal changes in consensus as a function of acquaintance and agreement in liking. *Journal of Personality and Social Psychology, 72*, 604-616.
- Patterson, M.L. (1994). Interaction behavior and person perception: An integrative approach. *Small Group Research, 25*, 172-188.
- Paulhus, D.L. (1994). *Balanced Inventory of Desirable Responding: Reference Manual for BIDR Version 6*. Vancouver, BC, Canada: Department of Psychology, University of British Columbia.
- Paulhus, D.L. & Bruce, M.N. (1992). The effect of acquaintanceship on the validity of personality impressions: A longitudinal study. *Journal of Personality and Social Psychology, 63*, 816-824.
- Paulhus, D.L., Bruce, M.N. & Trapnell, P.D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21*, 100-108.
- Paulhus, D.L. & Reynolds, S. (1995). Enhancing target variance in personality impressions: Highlighting the person in person perception. *Journal of Personality and Social Psychology, 69*, 1233-1242.
- Paunonen, S.V. (1989). Consensus in personality judgment: Moderating effects of target-rater acquaintanceship and behavior observability. *Journal of Personality and Social Psychology, 56*, 823-833.
- Richter, P. (1991). *Zur Konstruktvalidität des Beck-Depressionsinventars bei der Erfassung depressiver Verläufe*. Regensburg: Roderer.
- Rogers, C.R. (1961). *On becoming a person*. Boston: Houghton Mifflin.
- Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 174-214). New York: Academic Press.

- Ross, L., Greene, D. & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13, 279-301.
- Ruppert, F. (2003). *Depressionen – Symptome, Ursachen und Verläufe aus Sicht einer systemischen Psychotraumatologie*. München: Katholische Stiftungsfachhochschule München.
- Schmitt, M. (1982). *Empathie: Konzepte, Entwicklung, Quantifizierung* (Berichte aus der Arbeitsgruppe "Verantwortung, Gerechtigkeit, Moral" Nr. 9). Trier: Universität Trier, Fachbereich I – Psychologie.
- Schmitt, M. (1989). Ipsative Konsistenz (Kohärenz) als Profilähnlichkeit. *Trierer Psychologische Berichte*, 16, Heft 2.
- Schmitt, M., Altstötter-Gleich, C., Hinz, A., Maes, J. & Brähler, E. (2006). Normwerte für das vereinfachte Beck-Depressions-Inventar (BDI-V) in der Allgemeinbevölkerung. *Diagnostica*, 52, 51-59.
- Schmitt, M., Beckmann, M., Dusi, D., Maes, J., Schiller, A. & Schonauer, K. (2003). Messgüte des vereinfachten Beck-Depressions-Inventars (BDI-V). *Diagnostica*, 49, 147-156.
- Schmitt, M. & Maes, J. (2000). Vorschlag zur Vereinfachung des Beck-Depressions-Inventars (BDI). *Diagnostica*, 46, 38-46.
- Schmitt, M. & Maes, J. (2001). *Gerechtigkeit als innerdeutsches Problem: Abschlussbericht an die DFG* (Berichte aus der Arbeitsgruppe "Verantwortung, Gerechtigkeit, Moral" Nr. 132). Trier: Universität Trier, Fachbereich I - Psychologie.
- Schmitt, M. & Montada, L. (Hrsg.). (1999). *Gerechtigkeitserleben im wiedervereinigten Deutschland*. Opladen: Leske + Budrich.
- Schneewind, K. A. & Graf, J. (1998). *Der 16-Persönlichkeits-Faktoren-Test, revidierte Fassung (16 PF-R) - Deutsche Ausgabe des 16 PF Fifth Edition*. Bern: Huber.
- Schönpflug, W. (2000). *Geschichte und Systematik der Psychologie*. Weinheim: Beltz, PVU.
- Sedikides, C. & Skowronski, J.J. (1993). The self in impression formation: Trait centrality and social perception. *Journal of Experimental Social Psychology*, 29, 347-357.
- Seligman, M.E.P. & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, 55, 5-14.
- Snyder, M. (1987). *Public appearances, private realities: The psychology of self-monitoring*. New York: Freeman.
- Snyder, M. & Ickes, W. (1985). Personality and social behavior. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd Ed., Vol. 2, pp. 883-948). Boston: McGraw-Hill.
- Snyder, C.R. & Lopez, S.J. (2002). *Handbook of positive psychology*. New York: Oxford University Press.
- Spain, J.S., Eaton, L.G. & Funder, D.C. (2000). Perspectives on personality: The relative accuracy of self vs. others for the prediction of behavior and emotion. *Journal of Personality*, 68, 837-867.
- Steele, C.M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 261-302). New York: Academic Press.
- Steins, G. (1998). Diagnostik von Empathie und Perspektivübernahme: Eine Überprüfung des Zusammenhangs beider Konstrukte und Implikationen für die Messung. *Diagnostica*, 44, 117-129.
- Steins, G. & Wicklund, R.A. (1993). Zum Konzept der Perspektivenübernahme: Ein kritischer Überblick. *Psychologische Rundschau*, 44, 226-239.
- Sullivan, M.P. & Chapdelaine, A. (2001). Agreement in personality trait ratings: At first and over time. *Social Behavior and Personality*, 29, 635-647.
- Swann, W.B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, 91, 457-477.

- Swann, W.B. & Gill, M.J. (1997). Confidence and accuracy in person perception: Do we know what we think we know about our relationship partners? *Journal of Personality and Social Psychology*, *73*, 747-757.
- Taft, R. (1955). The ability to judge people. *Psychological Bulletin*, *52*, 1-23.
- Vogt, D.S. & Colvin, C.R. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality*, *71*, 267-295.
- Vogt, D.S. & Colvin, C.R. (2005). Assessment of accurate self-knowledge. *Journal of Personality Assessment*, *84*, 239-151.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Watson, J.B. (1976). *Behaviorismus* (2. unv. Aufl.). Frankfurt a. M.: Fachbuchhandlung f. Psychologie.
- Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology*, *57*, 120-128.
- Watson, D. & Clark, L.A. (1991). Self- versus peer-ratings of specific emotional traits: Evidence of convergent and discriminant validity. *Journal of Personality and Social Psychology*, *60*, 927-940.
- Watson, D., Hubbard, B. & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, *78*, 546-558.
- Zuckerman, M., Bernieri, F., Koestner, R. & Rosenthal, R. (1989). To predict some of the people some of the time: In search of moderators. *Journal of Personality and Social Psychology*, *57*, 279-293.
- Zuckerman, M., Koestner, R., Deboy, T., Garcia, T., Maresca, B.C. & Sartorius, J.M. (1988). To predict some of the people some of the time: A reexamination of the moderator variable approach in personality theory. *Journal of Personality and Social Psychology*, *54*, 1006-1019.

8. Anhang

Anhang A:	Hauptinstruktion Fragebögen	113
Anhang B:	Fragebögen Judge	114
	Umschlag, Instruktion Judge	114
	BDI-V (Fremdeinschätzungsform)	115
	Empathie / Perspektivübernahme.....	116
	Offenheit des Wertesystems	117
	Bekannschaftsgrad, Demographie.....	118
Anhang C:	Fragebögen Target	119
	Umschlag, Instruktion Target	119
	BDI-V (Selbsteinschätzungsform).....	120
	Selbst- und Fremdtäuschung	121
	Privatheit.....	122
	Bekannschaftsgrad, Demographie.....	123
Anhang D:	Screenshots Onlinefragebogen Beobachtbarkeit	124
Anhang E:	Übereinstimmungskoeffizienten	130
Anhang F:	Weitere Ergebnistabellen	131

Liebe Untersuchungsteilnehmer!

Zunächst einmal möchten wir Ihnen danken, daß Sie uns bei dieser Studie unterstützen wollen. Dazu bitten wir Sie, einige kurze Fragebögen auszufüllen. Der Zeitaufwand sollte pro Person maximal 10 Minuten betragen.

Da wir bestimmte Effekte der Selbst- und Fremdeinschätzung untersuchen wollen, benötigen wir **PAARE VON UNTERSUCHUNGSPERSONEN**, die sich kennen sollten:

Die eine Person gibt einige **Selbsteinschätzungen** ab, dh: beantwortet Fragen über sich selbst.

Die andere Person hingegen gibt (neben Selbsteinschätzungen) zunächst einige **Fremdeinschätzungen** ab, dh: beantwortet Fragen über die erste Person.

Deshalb gibt es zwei verschiedene, jeweils einzeln geheftete Fragebogenversionen:

Wie auf den Deckblättern zu ersehen, ist der eine für den Selbsteinschätzer und der andere für den Fremdeinschätzer. Nach dem Ausfüllen der jeweiligen Fragebögen bitten wir Sie, diese zu falten und in den jeweils anhängenden, weißen Briefumschlag zu tun und zu verschließen.

Die beiden weißen Briefumschläge kommen dann in den größeren braunen Umschlag. Dieser ist bereits adressiert und frankiert und wir bitten Sie, ihn in einen Briefkasten der Deutschen Post einzuwerfen.

Es ist für die Aussagekraft der Ergebnisse wichtig, daß Sie die Fragen ehrlich und offen beantworten. Selbstverständlich ist völlige Anonymität gewährleistet!

Die verschlossenen Umschläge, die bei uns eingehen, werden zunächst gesammelt und dann zu einem späteren Zeitpunkt alle auf einmal von einer neutralen Person am PC erfaßt.

Es findet keinerlei Auswertung auf Individuums-Ebene statt, dh: einzelne Werte einzelner Personen finden keine Beachtung, untersucht werden lediglich bestimmte Regelmäßigkeiten innerhalb des gesamten Datensatzes über alle Personen.

Da die Daten völlig anonym sind, ist es leider nicht möglich, Ihnen eine Rückmeldung über persönliche Ergebnisse zu geben. Sollten Sie jedoch an den Ergebnissen der Studie insgesamt interessiert sein, schicken Sie bitte eine kurze Email an die Adresse PsychStudie@kukli.de und wir mailen Ihnen im Sommer einen ausführlichen Bericht.

Nun müssen Sie nur noch entscheiden, wer eine Selbstauskunft abgeben möchte und wer die Fremdauskunft, und Sie können beginnen. Nochmals herzlichsten Dank!!

Fragebögen

Fremdauskunft

Um ein vollständiges Bild zu erzielen, ist es wichtig, daß Sie möglichst alle Fragen beantworten. Denken Sie nicht zu lange über die einzelnen Fragen nach. Geben Sie bitte jeweils die Antwort, die Ihnen am ehesten in den Sinn kommt.

Es gibt natürlich keine richtigen oder falschen Antworten!

Wie ist das gegenwärtiges Lebensgefühl der anderen Person?

In diesem Fragebogen geht es um das gegenwärtige Lebensgefühl der anderen Person, die sie beurteilen sollen. Bitte geben Sie zu jeder Frage an, wie häufig diese andere Person die genannte Stimmung oder Sichtweise Ihrer Meinung nach erlebt.

	<div style="display: flex; justify-content: space-between; align-items: center;"> Nie Fast Immer </div> 				
Er/Sie ist traurig.	1	2	3	4	5
Er/Sie sieht mutlos in die Zukunft.	1	2	3	4	5
Er/Sie fühlt sich als Versager(in).	1	2	3	4	5
Es fällt ihm/ihr schwer, etwas zu genießen.	1	2	3	4	5
Er/Sie hat Schuldgefühle.	1	2	3	4	5
Er/Sie fühlt sich bestraft.	1	2	3	4	5
Er/Sie ist von sich enttäuscht.	1	2	3	4	5
Er/Sie wirft sich Fehler und Schwächen vor.	1	2	3	4	5
Er/Sie denkt daran, sich etwas anzutun.	1	2	3	4	5
Er/Sie weint.	1	2	3	4	5
Er/Sie fühlt sich gereizt und verärgert.	1	2	3	4	5
Ihm/Ihr fehlt das Interesse an Menschen.	1	2	3	4	5
Er/Sie schiebt Entscheidungen vor sich her.	1	2	3	4	5
Er/Sie ist besorgt um sein/ihr Aussehen.	1	2	3	4	5
Er/Sie muß sich zu jeder Tätigkeit zwingen.	1	2	3	4	5
Er/Sie hat Schlafstörungen.	1	2	3	4	5
Er/Sie ist müde und lustlos.	1	2	3	4	5
Er/Sie hat keinen Appetit.	1	2	3	4	5
Sex ist ihm/ihr gleichgültig.	1	2	3	4	5
Er/Sie macht sich Sorgen um seine/ihre Gesundheit.	1	2	3	4	5

Wie erleben Sie sich im Umgang mit anderen Menschen ?

Nun geht es um Sie selbst!

Die folgenden Aussagen beschreiben Handlungen und Empfindungen im Umgang mit anderen Menschen. Kreuzen Sie bitte für jede Aussage an, wie selten oder oft Sie so handeln oder empfinden.

	nie				immer
Ich kann die Gefühle anderer leicht nachempfinden.	1	2	3	4	5
Bevor ich andere kritisiere, bemühe ich mich um Verständnis für ihre Sicht der Dinge.	1	2	3	4	5
Wenn Leute ausgenutzt werden, habe ich das Bedürfnis, sie zu schützen.	1	2	3	4	5
Bei Meinungsverschiedenheiten versetze ich mich in die Lage meines Gegenüber.	1	2	3	4	5
Ich versuche, meine Mitmenschen besser zu verstehen, indem ich die Dinge aus ihrem Blickwinkel betrachte.	1	2	3	4	5
Es geht mir nahe, wenn andere ein Mißgeschick erleiden.	1	2	3	4	5
Auch wenn ich mir meiner Sache sicher bin, bedenke ich die Argumente der anderen.	1	2	3	4	5
Ich bin mitfühlend gegenüber Menschen, die Probleme haben.	1	2	3	4	5
Bei Auseinandersetzungen bemühe ich mich, die Ansichten aller Beteiligten zu verstehen.	1	2	3	4	5
Die Sorgen und Nöte anderer machen mir zu schaffen.	1	2	3	4	5
Bevor ich mich über jemanden aufrege, versuche ich, das Problem mit seinen Augen zu sehen.	1	2	3	4	5
Ich finde es leicht, Dinge vom Standpunkt anderer aus zu sehen.	1	2	3	4	5
Es tut mir weh, wenn andere ungerecht behandelt werden.	1	2	3	4	5
Ich kann die Überlegungen anderer gut nachvollziehen.	1	2	3	4	5
Ich denke mich in andere hinein.	1	2	3	4	5
Ich mache mir Sorgen um Menschen, denen es schlechter geht als mir.	1	2	3	4	5

**Nun noch einige Aussagen,
die sich zur Beschreibung Ihrer eigenen Person eignen könnten.**

Lesen Sie bitte jede dieser Aussagen aufmerksam durch und überlegen Sie, ob diese Aussage auf Sie persönlich zutrifft oder nicht.

Zur Bewertung steht Ihnen eine fünffach abgestufte Skala zur Verfügung, kreuzen Sie bitte an!

	völlige Ableh- nung	Ableh- nung	neutral	Zustim- mung	völlige Zustim- mung
Ich bin aufgeschlossen und tolerant für die Lebensgewohnheiten anderer Menschen.	-2	-1	0	+1	+2
Ich glaube, daß es Schüler oft nur verwirrt und irreführt, wenn man sie Redner hören läßt, die kontroverse Standpunkte vertreten.	-2	-1	0	+1	+2
Ich meine, daß sich die Gesetze und die Sozialpolitik ändern sollten, um den Bedürfnissen der sich ändernden Welt Rechnung zu tragen.	-2	-1	0	+1	+2
Ich glaube, daß wir bei ethischen Entscheidungen auf die Ansichten unserer religiösen Autoritäten achten sollten.	-2	-1	0	+1	+2
Ich glaube, daß die anderen Wertvorstellungen in fremden Gesellschaften für die Menschen dort richtig sein könnten.	-2	-1	0	+1	+2
Wenn Leute im Alter von 25 Jahren noch keine feste Meinung haben, dann stimmt mit ihnen etwas nicht.	-2	-1	0	+1	+2
Ich glaube, daß die Treue zu den eigenen Idealen und Prinzipien wichtiger ist als „Aufgeschlossenheit“.	-2	-1	0	+1	+2
Ich glaube, daß die „neue Moral“ der Freizügigkeit überhaupt keine Moral ist.	-2	-1	0	+1	+2

Nun möchten wir Sie noch um einige Angaben bitten über die **Beziehung zu der Person, die Sie am Anfang beurteilt haben**

Ich bin mit der anderen Person

- (bluts)verwandt
- liiert
- befreundet
- bekannt

Ich kenne die andere Person
meiner Meinung nach ...

- sehr gut
- gut
- kaum

Wir sehen uns normalerweise pro Woche
bzw. haben miteinander zu tun:

- mehrmals täglich
- täglich
- mehrmals wöchentlich
- 1-2x wöchentlich
- seltener

Wir kennen uns jetzt
seit ca. _____ Jahren.

Wir haben zusammen sehr intensive Erlebnisse gehabt.

- Ja
- Nein

Wir sehen uns in privaten Situationen.

- Ja
- Nein

Wir sehen uns in alltäglichen, strukturierten, eher oberflächlichen Situationen.

- Ja
- Nein

Zuallerletzt bitte noch drei **demographische Angaben:**

Alter: *Ich bin _____ Jahre alt.*

Geschlecht: *männlich / weiblich*

Familienstand: *ledig / liiert / verheiratet*

Das war's, Sie haben es geschafft. Bitte jetzt den Fragebogen in den weißen Umschlag stecken, diesen verschließen und in den großen braunen Umschlag tun.
Vielen Dank!!!!

Fragebögen

Selbstauskunft

Um ein vollständiges Bild zu erzielen, ist es wichtig, daß Sie möglichst alle Fragen beantworten. Denken Sie nicht zu lange über die einzelnen Fragen nach. Geben Sie bitte jeweils die Antwort, die Ihnen am ehesten in den Sinn kommt.

Es gibt natürlich keine richtigen oder falschen Antworten!

Wie ist Ihr gegenwärtiges Lebensgefühl?

In diesem Fragebogen geht es um ihr gegenwärtiges Lebensgefühl. Bitte geben Sie zu jeder Frage an, wie häufig Sie die genannte Stimmung oder Sichtweise erleben.

	Nie				Fast Immer
					
Ich bin traurig.	1	2	3	4	5
Ich sehe mutlos in die Zukunft.	1	2	3	4	5
Ich fühle mich als Versager(in).	1	2	3	4	5
Es fällt mir schwer, etwas zu genießen.	1	2	3	4	5
Ich habe Schuldgefühle.	1	2	3	4	5
Ich fühle mich bestraft.	1	2	3	4	5
Ich bin von mir enttäuscht.	1	2	3	4	5
Ich werfe mir Fehler und Schwächen vor.	1	2	3	4	5
Ich denke daran, mir etwas anzutun.	1	2	3	4	5
Ich weine.	1	2	3	4	5
Ich fühle mich gereizt und verärgert.	1	2	3	4	5
Mir fehlt das Interesse an Menschen.	1	2	3	4	5
Ich schiebe Entscheidungen vor mir her.	1	2	3	4	5
Ich bin besorgt um mein Aussehen.	1	2	3	4	5
Ich muß mich zu jeder Tätigkeit zwingen.	1	2	3	4	5
Ich habe Schlafstörungen.	1	2	3	4	5
Ich bin müde und lustlos.	1	2	3	4	5
Ich habe keinen Appetit.	1	2	3	4	5
Sex ist mir gleichgültig.	1	2	3	4	5
Ich mache mir Sorgen um meine Gesundheit.	1	2	3	4	5

Geben Sie bitte für jede der folgenden Aussagen an, wie sehr Sie ihr zustimmen

	<div style="display: flex; justify-content: space-between;"> völlige Ablehnung völlige Zustimmung </div> 						
Der erste Eindruck, den ich von anderen Menschen gewinne, bewahrheitet sich meistens.	1	2	3	4	5	6	7
Ich bin mir selber gegenüber nicht immer ganz ehrlich gewesen.	1	2	3	4	5	6	7
Ich weiß immer, warum ich etwas mag.	1	2	3	4	5	6	7
Es fällt mir schwer, einen beunruhigenden Gedanken beiseite zu drängen.	1	2	3	4	5	6	7
Manchmal verpasse ich etwas, weil ich mich einfach nicht schnell genug entscheiden kann.	1	2	3	4	5	6	7
Ich bin ein vollkommen rational denkender Mensch.	1	2	3	4	5	6	7
Ich kann Kritik selten vertragen.	1	2	3	4	5	6	7
Ich bin mir meiner Urteile sehr sicher.	1	2	3	4	5	6	7
An meinen Fähigkeiten als LiebhaberIn habe ich schon gelegentlich gezweifelt.	1	2	3	4	5	6	7
Ich weiß immer die Gründe für meine Handlungen.	1	2	3	4	5	6	7
Manchmal lüge ich, wenn ich muß.	1	2	3	4	5	6	7
Es ist schon einmal vorgekommen, daß ich jemanden ausgenutzt habe.	1	2	3	4	5	6	7
Ich fluche niemals.	1	2	3	4	5	6	7
Manchmal zahle ich es lieber anderen heim, als daß ich vergebe und vergesse.	1	2	3	4	5	6	7
Ich habe schon einmal zuviel Wechselgeld heraus bekommen, ohne es der Verkäuferin zu sagen.	1	2	3	4	5	6	7
Ich gebe grundsätzlich alles an, was ich zu verzollen habe.	1	2	3	4	5	6	7
Manchmal fahre ich schneller, als es erlaubt ist.	1	2	3	4	5	6	7
Ich habe Dinge getan, von denen ich anderen nichts erzähle.	1	2	3	4	5	6	7
Ich nehme niemals Dinge an mich, die mir nicht gehören.	1	2	3	4	5	6	7
Ich bin schon einmal wegen angeblicher Krankheit nicht zur Arbeit oder Schule gegangen.	1	2	3	4	5	6	7

**Lesen Sie bitte die folgenden Aussagen durch
und wählen die Antwort, die am ehesten auf Sie zutrifft**

Versuchen Sie bitte, möglichst eine "a"- oder "c"-Antwort zu wählen.

Kreuzen Sie das mittlere Fragezeichen "?" bitte nur an,
wenn weder "a" noch "c" besser zu ihnen passt.

Ich ziehe es vor, ...	[a] meine Probleme mit Freunden zu besprechen [b] ? [c] meine Probleme für mich zu behalten
Wenn jemand mir eine Frage stellt, die mir zu persönlich ist, versuche ich irgendwie, mich vor einer Antwort zu drücken.	[a] stimmt meistens [b] ? [c] stimmt meistens nicht
Ich gehöre zu den Menschen, die ...	[a] natürlich und direkt sind [b] ? [c] sich überlegt und diplomatisch verhalten
Im Gespräch mit Bekannten ziehe ich es vor, ...	[a] mich an Tatsachen und unpersönliche Dinge zu halten [b] ? [c] über Leute und deren Einstellungen zu reden
Im Umgang mit Menschen ist es besser, ...	[a] alle seine Karten auf den Tisch zu legen [b] ? [c] seine Karten verdeckt zu halten
Über meine Gefühle spreche ich ...	[a] bereitwillig, wenn es andere zu interessieren scheint [b] ? [c] nur, wenn es sich nicht vermeiden läßt
Ich bin sehr vorsichtig, wenn es darum geht, wem ich mich wirklich anvertrauen will.	[a] stimmt [b] ? [c] stimmt nicht

Nun möchten wir Sie noch um einige Angaben bitten über die **Beziehung zu der anderen Person, die Sie beurteilen soll**

Ich bin mit der anderen Person

- (bluts)verwandt
- liiert
- befreundet
- bekannt

Ich kenne die andere Person
meiner Meinung nach ...

- sehr gut
- gut
- kaum

Wir sehen uns normalerweise pro Woche
bzw. haben miteinander zu tun:

- mehrmals täglich
- täglich
- mehrmals wöchentlich
- 1-2x wöchentlich
- seltener

Wir kennen uns jetzt
seit ca. _____ Jahren.

Wir haben zusammen sehr intensive Erlebnisse gehabt.

- Ja
- Nein

Wir sehen uns in privaten Situationen.

- Ja
- Nein

Wir sehen uns in alltäglichen, strukturierten, eher oberflächlichen Situationen.

- Ja
- Nein

Zuallerletzt bitte noch drei **demographische Angaben:**

Alter: *Ich bin _____ Jahre alt.*

Geschlecht: *männlich / weiblich*

Familienstand: *ledig / liiert / verheiratet*

Das war's, Sie haben es geschafft. Bitte jetzt den Fragebogen in den weißen Umschlag stecken, diesen verschließen und in den großen braunen Umschlag tun.
Vielen Dank!!!!

Fragebogen zur Beobachtbarkeit spezifischer Verhaltens- und Erlebensweisen - Mozilla Firefox

Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

file:///F:/Fragebogen/Observability.htm

Sehr geehrter Untersuchungsteilnehmer,

dieser Online-Fragebogen ist Teil einer Studie zur Validierung eines Depressions-Inventars, die zur Zeit an der Uni Landau durchgeführt wird.

Die Untersuchung ist völlig anonym, dh: es werden keine persönlichen Daten gespeichert, außer drei freiwilligen, sehr allgemeinen demographischen Angaben.

Es geht hier nicht um die individuelle Testung einzelner Personen, sondern um die Ermittlung durchschnittlicher Einschätzungen einer größeren Anzahl von Personen.

Deshalb gibt es keine richtigen oder falschen Antworten, lediglich unterschiedliche Einschätzungen!

Das Ausfüllen des kompletten Fragebogens sollte nicht länger als drei bis fünf Minuten dauern.

Wir danken Ihnen für Ihre Mitarbeit und wünschen weiterhin schöne Feiertage & einen guten Rutsch!

Beobachtbarkeit spezifischer Verhaltens- und Erlebensweisen

Mit diesem Fragebogen soll ermittelt werden, für wie beobachtbar ein "normaler Mensch" spezifische Stimmungen, Zustände, Einstellungen und Handlungen bei anderen Menschen hält.

Im folgenden werden daher 40 Verhaltens- und Erlebensweisen genannt, die andere Personen auszeichnen können.

Bitte klicken Sie zu jeder dieser 40 Aussagen an, inwieweit der Zustand Ihrer Meinung nach von außen beobachtet und erkannt werden kann.

Bitte stellen Sie sich dazu als andere Person einen **prototypischen guten Bekannten** vor, aber keine Person, mit der Sie sehr engen und intimen Umgang haben, also z.B. keinen Beziehungspartner und kein Familienmitglied.

Bitte denken Sie nicht allzu lange über die Bewertung der einzelnen Aussagen nach, sondern klicken einfach auf der Skala von 1 ("gar nicht beobachtbar") bis 9 ("vollkommen beobachtbar") an, **was Sie spontan für zutreffend halten**.

Hier können Sie alle Eingaben noch einmal löschen

Und nun zu den 40 inhaltlichen Fragen:

Wie gut sind diese Verhaltens- bzw. Erlebensweisen von außen beobachtbar?

gar nicht beobachtbar vollkommen beobachtbar

Die Person ist traurig.

1 2 3 4 5 6 7 8 9

Fragebogen zur Beobachtbarkeit spezifischer Verhaltens- und Erlebensweisen - Mozilla Firefox

Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

file:///F:/Fragebogen/Observability.htm

Die Person sieht mutlos in die Zukunft.

1 2 3 4 5 6 7 8 9

Die Person fühlt sich als Versager.

1 2 3 4 5 6 7 8 9

Der Person fällt es schwer, etwas zu genießen.

1 2 3 4 5 6 7 8 9



gar nicht beobachtbar vollkommen beobachtbar

Die Person hat Schuldgefühle.

1 2 3 4 5 6 7 8 9

Die Person fühlt sich bestraft.

1 2 3 4 5 6 7 8 9

Die Person ist von sich enttäuscht.

1 2 3 4 5 6 7 8 9

Die Person wirft sich Fehler und Schwächen vor.

1 2 3 4 5 6 7 8 9



gar nicht beobachtbar vollkommen beobachtbar

Die Person denkt daran, sich etwas anzutun.

1 2 3 4 5 6 7 8 9

Fragebogen zur Beobachtbarkeit spezifischer Verhaltens- und Erlebensweisen - Mozilla Firefox

Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

file:///F:/Fragebogen/Observability.htm

Die Person weint.

1 2 3 4 5 6 7 8 9

Die Person fühlt sich gereizt und verärgert.

1 2 3 4 5 6 7 8 9

Der Person fehlt das Interesse an Menschen.

1 2 3 4 5 6 7 8 9



gar nicht beobachtbar vollkommen beobachtbar

Die Person schiebt Entscheidungen vor sich her.

1 2 3 4 5 6 7 8 9

Die Person ist besorgt um ihr Aussehen.

1 2 3 4 5 6 7 8 9

Die Person muss sich zu jeder Tätigkeit zwingen.

1 2 3 4 5 6 7 8 9

Die Person hat Schlafstörungen.

1 2 3 4 5 6 7 8 9



gar nicht beobachtbar vollkommen beobachtbar

Die Person ist müde und lustlos.

1 2 3 4 5 6 7 8 9

Fragebogen zur Beobachtbarkeit spezifischer Verhaltens- und Erlebensweisen - Mozilla Firefox

Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

file:///F:/Fragebogen/Observability.htm

Die Person hat keinen Appetit.

1 2 3 4 5 6 7 8 9

Sex ist der Person gleichgültig.

1 2 3 4 5 6 7 8 9

Die Person macht sich Sorgen um ihre Gesundheit.

1 2 3 4 5 6 7 8 9



gar nicht beobachtbar vollkommen beobachtbar

Die Person ist nicht traurig.

1 2 3 4 5 6 7 8 9

Die Person sieht mutvoll in die Zukunft.

1 2 3 4 5 6 7 8 9

Die Person fühlt sich nicht als Versager.

1 2 3 4 5 6 7 8 9

Die Person kann etwas genießen.

1 2 3 4 5 6 7 8 9



gar nicht beobachtbar vollkommen beobachtbar

Die Person hat keine Schuldgefühle.

1 2 3 4 5 6 7 8 9

Fragebogen zur Beobachtbarkeit spezifischer Verhaltens- und Erlebensweisen - Mozilla Firefox

Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

file:///F:/Fragebogen/Observability.htm

Die Person ist mit sich zufrieden.

1 2 3 4 5 6 7 8 9

Die Person wirft sich keine Fehler und Schwächen vor.

1 2 3 4 5 6 7 8 9



gar nicht beobachtbar vollkommen beobachtbar

Die Person denkt nie daran, sich etwas anzutun.

1 2 3 4 5 6 7 8 9

Die Person weint nicht.

1 2 3 4 5 6 7 8 9

Die Person ist immer gelassen.

1 2 3 4 5 6 7 8 9

Die Person hat Interesse an Menschen.

1 2 3 4 5 6 7 8 9



gar nicht beobachtbar vollkommen beobachtbar

Die Person ist entscheidungsfreudig.

1 2 3 4 5 6 7 8 9

Die Person ist zufrieden mit ihrem Aussehen.

1 2 3 4 5 6 7 8 9

Fragebogen zur Beobachtbarkeit spezifischer Verhaltens- und Erlebensweisen - Mozilla Firefox

Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

file:///F:/Fragebogen/Observability.htm

Die Person muss sich zu keiner Tätigkeit zwingen.

1 2 3 4 5 6 7 8 9

Die Person schläft gut.

1 2 3 4 5 6 7 8 9



gar nicht beobachtbar vollkommen beobachtbar

Die Person ist ausgeruht und voller Elan.

1 2 3 4 5 6 7 8 9

Die Person hat Appetit.

1 2 3 4 5 6 7 8 9

Sex ist der Person wichtig.

1 2 3 4 5 6 7 8 9

Die Person ist nicht besorgt um ihre Gesundheit.

1 2 3 4 5 6 7 8 9

Das waren die 40 inhaltlichen Fragen, vielen Dank für Ihre Mühe!!



Nun bitten wir Sie noch um drei demographische Angaben zu Ihrer Person:
(Dies ist freiwillig und nicht zwingend nötig, hilft uns jedoch weiter)

Alter Mann Frau Familienstand

Wenn Sie möchten, können Sie hier einen Kommentar hinterlassen:
(Ebenfalls rein freiwillig und nicht zwingend nötig)

Alles fertig, bitte absenden!

Coefficient of Congruence**[c]**

(BURT, 1948, TUCKER, 1951, WRIGLEY & NEUHAUS, 1955)

$$c = (\sum x_j y_j) \cdot (\sum x_j^2 \sum y_j^2)^{-1/2}$$

mit:

x = Selbsteinschätzung

y = Fremdeinschätzung

j = Anzahl der Items

→ Berechnung erfolgt auf Basis von Rohwerten

Indices of Profile Agreement**[I_{pa} / r_{pa}]**

(MCCRAE, 1993)

$$I_{pa} = (j + 2 \sum M_j^2 - \sum d_j^2) \cdot (10j)^{-1/2}$$

$$r_{pa} = I_{pa} \cdot ((j-2) + I_{pa}^2)^{-1/2}$$

mit:

M = Mittelwert der Fremd- & Selbsteinschätzung

d = Differenz der Fremd- & Selbsteinschätzung

j = Anzahl der Items

→ Berechnung erfolgt auf Basis von Standard-z-Werten

Tabelle F1: Semipartialkorrelationen (ohne Bekanntschaftsgrad) der potentiellen Moderatoren seitens des Judges mit den Kennwerten der dyadischen Urteilsübereinstimmung

Eigenschaft	Profil- korrelation	Ipsative Profil- korrelation ⁺	Kongruenz- koeffizient c	Ähnlichkeits- koeffizient r _{PA}
Empathie	-.09	-.12	-.12	-.13
Perspektivübernahme	-.01	-.08	-.03	-.05
Offenheit d. WS	-.03	-.09	.02	-.01
Lebensalter	-.09	-.04	.02	.02

Anmerkungen: N=177 Urteilerpaare, *p<.05, **p<.01 (2-seitig), + Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte. Profilkorrelation und ipsative Profilkorrelation gingen als Fisher-z-Werte in die Berechnungen ein.

Tabelle F2: Semipartialkorrelationen (ohne Bekanntschaftsgrad) der potentiellen Moderatoren seitens des Targets mit den Kennwerten der dyadischen Urteilsübereinstimmung

Eigenschaft	Profil- korrelation	Ipsative Profil- korrelation ⁺	Kongruenz- koeffizient c	Ähnlichkeits- koeffizient r _{PA}
Selbsttäuschung	-.14	-.12	-.20**	.01
Fremdtäuschung	-.20**	-.07	-.07	-.05
Privatheit	.05	-.03	.11	.01
Konsistenz (SD)	.16*	.21**	-.12	.07

Anmerkungen: N=177 Urteilerpaare, *p<.05, **p<.01 (2-seitig), + Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte. Profilkorrelation und ipsative Profilkorrelation gingen als Fisher-z-Werte in die Berechnungen ein.

Tabelle F3: Mittelwerte der dyadischen Urteilsübereinstimmung, Subgruppen liierte und nicht-liierte Urteilerpaare

	Profil- korrelation	Ipsative Profil- korrelation ⁺	Kongruenz- koeffizient c	Ähnlichkeits- koeffizient r _{PA}
liiert [#]	.50	.37	.93	.34
nicht-liiert ^{##}	.40	.27	.92	.22

Anmerkungen: [#] N=104 Urteilerpaare, ^{##} N=73 Urteilerpaare, + Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte. Die Mittelwertbildung erfolgte im Fall der Profilkorrelation und der ipsativen Profilkorrelation über Fisher-z-Werte und anschließende Rücktransformation in Bravais-Pearson-Korrelationen.

Fettgedruckt: mit p<.01 signifikante Mittelwertsunterschiede der beiden Subgruppen (T-Test für heterogene Varianzen, berechnet auf Basis der Fisher-z-Werte).

Tabelle F4: Mittelwerte der dyadischen Urteilsübereinstimmung, Subgruppen
Bekanntschafsdauer 0 bis 1, 1 bis 6, 6 bis 21 und mehr als 21 Jahre

Dauer der Bekanntschaft	Profilkorrelation	Ipsative Profilkorrelation ⁺	Kongruenzkoeffizient C	Ähnlichkeitskoeffizient r_{PA}
0 - 1 Jahr [#]	.40	.26	.91	.22
1 - 6 Jahre [#]	.48	.35	.92	.31
6 - 21 Jahre [#]	.48	.36	.92	.29
> 21 Jahre ^{##}	.47	.34	.93	.34

Anmerkungen: [#] N=44 Urteilerpaare, ^{##} N=45 Urteilerpaare, ⁺ Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte. Die Mittelwertbildung der Profilkorrelation und der ipsativen Profilkorrelation erfolgte über Fisher-z-Werte und anschließende Rücktransformation in Bravais-Pearson-Korrelationen.

Keiner der Mittelwertsunterschiede zwischen den Subgruppen ist signifikant (T-Test für heterogene Varianzen, berechnet auf Basis der Fisher-z-Werte).

Tabelle F5: Fisher-z-Werte der Selbst-Fremd-Korrelationen aller BDI-V-Items (Rohwerte), Subgruppen
Bekanntschafsdauer 0 bis 1 Jahr, 1 bis 6 Jahre, 6 bis 21 Jahre und mehr als 21 Jahre

	Bekanntschafsdauer			
	0-1 J [#]	1-6 J [#]	6-21 J [#]	> 21 J ^{##}
BDI 01 'Ich bin traurig'	.30	.47	.60	.69
BDI 02 'Ich sehe mutlos in die Zukunft'	.32	.33	.40	.81
BDI 03 'Ich fühle mich als Versager'	.05	.38	.42	.37
BDI 04 'Es fällt mir schwer, etwas zu geniessen'	.05	.14	.68	.71
BDI 05 'Ich habe Schuldgefühle'	.12	.24	.28	.41
BDI 06 'Ich fühle mich bestraft'	.41	.14	-.07	.66
BDI 07 'Ich bin von mir enttäuscht'	.18	.42	.18	.62
BDI 08 'Ich werfe mir Fehler und Schwächen vor'	.27	.27	.18	.54
BDI 09 'Ich denke daran, mir etwas anzutun'	.01	.45	.29	1.05
BDI 10 'Ich weine'	.40	.66	.46	.66
BDI 11 'Ich fühle mich gereizt und verärgert'	.31	.51	.30	.24
BDI 12 'Mir fehlt das Interesse an Menschen'	.10	.37	.35	.16
BDI 13 'Ich schiebe Entscheidungen vor mir her'	.46	.31	.51	.59
BDI 14 'Ich bin besorgt um mein Aussehen'	.31	.39	.40	.09
BDI 15 'Ich muß mich zu Tätigkeiten zwingen'	.22	.31	.10	.42
BDI 16 'Ich habe Schlafstörungen'	.12	.66	.66	.76
BDI 17 'Ich bin müde und lustlos'	.17	.27	.42	.38
BDI 18 'Ich habe keinen Appetit'	.23	.08	.52	.45
BDI 19 'Sex ist mir gleichgültig'	.33	.34	.65	.46
BDI 20 'Ich mache mir Sorgen um meine Gesundheit'	.29	.40	.41	.19

Anmerkungen: [#] N=44 Urteilerpaare, ^{##} N=45 Urteilerpaare, **fettgedruckt:** mit $p < .05$ signifikante Korrelationsunterschiede der Subgruppen (Differenz der Fisher-z-Werte $> .36$).

Tabelle F6: Korrelationen (Spearman's Rho) der kategorialen Variable Treffensfrequenz mit den Kennwerten der dyadischen Urteilsübereinstimmung

	Profil- korrelation	Ipsative Profil- korrelation [#]	Kongruenz- koeffizient c	Ähnlichkeits- koeffizient r_{PA}
Treffensfrequenz (ordinal, fünfstufig)	.04	.01	-.02	-.04

Anmerkungen: N=177 Urteilerpaare, [#] Profilkorrelation auf Basis stichprobenzentrierter Abweichungswerte. Profilkorrelation und ipsative Profilkorrelation gingen als Fisher-z-Werte in die Berechnungen ein.

Tabelle F7: Absteigend angeordnete Agreement-Werte aller BDI-V-Items und deren Observability: Fisher-z-Werte, Observability-Scores und jeweilige Rangplätze

	Agreement ⁺		Observability ⁺⁺	
	z'-Wert	Rangplatz	Score	Rangplatz
BDI 16 'Ich habe Schlafstörungen'	.59	01	3,83	18
BDI 10 'Ich weine'	.55	02	7,15	01
BDI 01 'Ich bin traurig'	.51	03	6,45	02
BDI 02 'Ich sehe mutlos in die Zukunft'	.48	04	4,76	11
BDI 19 'Sex ist mir gleichgültig'	.47	05	3,32	19
BDI 13 'Ich schiebe Entscheidungen vor mir her'	.44	06	5,05	10
BDI 11 'Ich fühle mich gereizt und verärgert'	.40	07	6,12	04
BDI 04 'Es fällt mir schwer, etwas zu genießen'	.40	08	5,70	06
BDI 07 'Ich bin von mir enttäuscht'	.37	09	5,16	09
BDI 14 'Ich bin besorgt um mein Aussehen'	.34	10	5,64	07
BDI 20 'Ich mache mir Sorgen um meine Gesundheit'	.34	11	4,21	15
BDI 09 'Ich denke daran, mir etwas anzutun'	.33	12	3,21	20
BDI 03 'Ich fühle mich als Versager'	.31	13	4,53	12
BDI 08 'Ich werfe mir Fehler und Schwächen vor'	.31	14	4,22	14
BDI 17 'Ich bin müde und lustlos'	.30	15	6,20	03
BDI 18 'Ich habe keinen Appetit'	.30	16	5,49	08
BDI 15 'Ich muß mich zu Tätigkeiten zwingen'	.28	17	4,43	13
BDI 12 'Mir fehlt das Interesse an Menschen'	.27	18	5,72	05
BDI 05 'Ich habe Schuldgefühle'	.27	19	4,09	16
BDI 06 'Ich fühle mich bestraft'	.27	20	4,08	17

Anmerkungen: ⁺Bravais-Pearson-Korrelation der Selbst- und Fremdeinschätzungen des jeweiligen Items über die gesamte Stichprobe (N=177 Urteilerpaare) als Fisher-z-Wert, ⁺⁺durchschnittliche Einschätzung der Beobachtbarkeit des jeweiligen Items über die gesamte Stichprobe (N=219) auf einer Skala von 1 bis 9.

Die Korrelation (Bravais-Pearson) der beiden Wertereihen über alle 20 Items beträgt $r=.20$, die Korrelation (Kendalls Tau) der beiden Rangreihen über alle 20 Items beträgt $\tau=.18$.

Danksagung

Ganz herzlich möchte ich mich für die Unterstützung auf den verschiedensten Ebenen, die mir bei der Durchführung der Erhebung und beim Verfassen der Diplomarbeit zuteil wurde, bei folgenden Personen bedanken:

Herr Prof. Dr. Schmitt, mein Betreuer, gab den Anstoß zu der Untersuchung und wertvolle Literaturhinweise zur Einarbeitung in den Forschungsbereich Personality Judgment. Soweit Probleme oder Fragen auftauchten, war er jederzeit ansprechbar und half bei der Lösung, so daß ich mich immer gut betreut fühlte. Hervorheben möchte ich auch seine motivationale Unterstützung durch manches Lob und manche Relativierung allzu selbstkritischer Tendenzen des Autors.

Frau Dr. Altstötter-Gleich erklärte sich freundlicherweise bereit, die Zweitbegutachtung dieser Arbeit zu übernehmen. Auch sie war jederzeit ansprechbar und half bei der Klärung einiger formaler Aspekte und Fragen. Außerdem danke ich ihr für die gute und anschauliche Einführung in die Welt der psychologischen Methodik in Form der Grundlagenvorlesungen im Rahmen meines Studiums.

Herr Prof. Dr. Ingmar Hosenfeld versorgte mich mit einigen guten Tipps zu verschiedenen Aspekten der Berechnung von Übereinstimmungskoeffizienten.

Herr Dipl.-Psych. Andreas Berg gab mir mit seinen wunderbaren Bergschen Axiomen eine hoffnungstiftende Leitlinie beim Verfassen der Arbeit.

Frau Dipl.-Psych. Heike Koch half mir dabei, einige Tücken und Bugs des Programms Microsoft Word zu überwinden und verhinderte so schwerere mentale Krisen.

Frau BTA Nadine Kirsch sorgte für eine wertvolle naturwissenschaftliche Erdung während der ganzen psychologisch-geisteswissenschaftlichen Lektüre, außerdem machte sie sich manches mal um mein seelisches und auch leibliches Wohl verdient.

Schließlich danke ich meinen lieben Eltern,

Frau Elke Hübner und Herrn Dipl.-Ing. Klaus Hübner, die mich wie schon so oft bei so vielen Dingen in meinem Leben auch bei diesem Vorhaben organisatorisch, logistisch, finanziell und motivational vorbildlich unterstützt haben.

Eidesstattliche Erklärung

Ich erkläre hiermit, daß ich die vorliegende Prüfungsarbeit selbständig verfaßt und keine anderen als die angegebenen Hilfsmittel und Quellen verwendet sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe.

Landau in der Pfalz, am 17. Juli 2006