

Testtheoretische und empirische Studien zur Frage der Dimensionalität des Konstruktes „Dispositioneller Optimismus“

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt dem
Fachbereich Psychologie und Sportwissenschaften
der Johann Wolfgang Goethe-Universität
Frankfurt am Main

von
Wolfgang Achim Rauch
aus Frankfurt am Main

Frankfurt am Main 2007

D30

vom Fachbereich Psychologie und Sportwissenschaften der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Helfried Moosbrugger

Gutachter: Prof. Dr. Helfried Moosbrugger, Prof. Dr. Volker Hodapp

Datum der Disputation: 5. Dezember 2007

Inhaltsverzeichnis

1	Einleitung	1
1.1	Interindividuelle Unterschiede im Optimismus	1
1.2	Die Optimismus-Pessimismus-Unabhängigkeits-Hypothese	1
1.3	Übersicht	4
2	Optimismus	5
2.1	Der Begriff „Optimismus“	5
2.1.1	Zur Problematik der Begriffsverwendung in der OP-Hypothese	5
2.1.2	Historische und philosophische Perspektiven	5
2.1.3	Schlussfolgerungen	7
2.2	Optimismus in der Psychologie	7
2.2.1	Die OP-Hypothese und ihre Basis	8
2.2.2	Dispositioneller Optimismus	9
2.2.3	Weitere psychologische Modelle zum Optimismus	12
2.3	Schlussfolgerungen der theoretischen Analyse	14
3	Das Problem des reflexiven Messmodells bei Selbstberichtsfragebogen	15
3.1	Die unreflektierte „Reflexivitätsannahme“	16
3.2	Die Prozedur der Fragebogenmethode	17
3.2.1	Probleme der theoretischen Explikation und der operationalen Definition	18
3.2.2	Probleme der Itemformulierung	19
3.2.3	Probleme des Antwortprozesses	19
3.3	Fazit: Probleme der Fragebogenmethode	22
4	Zur Frage der „Konfirmation“ bei „konfirmatorischen“ Faktorenanalysen	24
4.1	Konfirmation vs. Falsifikation	24
4.1.1	Das Induktionsproblem	24
4.1.2	Logisches Schließen nach Modus tollens	25
4.1.3	Falsifikation bei probabilistischen Hypothesen	26
4.2	Protektive Auxiliarannahmen	27
4.3	„Konfirmatorische“ Faktorenanalyse der OP-Hypothese	28
4.4	Zur Unterdeterminiertheit von faktorenanalytischen Ergebnissen	31
4.5	Schlussfolgerungen	33
5	Alternativen zur OP-Hypothese: Methodeneffekte im Zusammenhang mit Sozialer Erwünschtheit	34
5.1	Response sets als Erklärung für Richtungsfaktoren	35
5.2	Das Prinzip der Theoretischen Sparsamkeit	36
6	Invertierte Items in der Skala „Personaler Optimismus“	38
7	Diskussion	40
7.1	Zusammenführung	40
7.2	Methodologische Bewertung und Ausblick	40
8	Literatur	42
9	Schriftenanhang	50
9.1	Schrift A	51
9.2	Schrift B	52
9.3	Schrift C	53
10	Zusammenfassung	54

1 Einleitung

1.1 Interindividuelle Unterschiede im Optimismus

Es ist vorteilhaft, optimistisch zu sein, mindestens aus gesundheitspsychologischer Sicht: Optimisten sind nicht nur besser gelaunt, sondern auch physisch gesünder; vor allem aber zeigen sie ein höheres Beharrungs- und Durchhaltevermögen (Peterson, 2000). Optimismus wird zwar von manchen Autoren (nur) als eine Facette von seelischer Gesundheit aufgefasst und nicht als ein davon zu unterscheidendes Konstrukt (Hoyer, 2000); aber selbst wenn es so wäre, ließen sich die Unterscheidbarkeit zwischen Optimismus und anderen Facetten seelischer Gesundheit genauso wenig leugnen wie der Zusammenhang von Optimismus auch mit physischer Gesundheit. Optimismus ist also ein psychologisches Konstrukt, das nicht nur grundlagenwissenschaftlich interessant ist.

Menschen sind unterschiedlich optimistisch. Mehrere Zugänge zur Erfassung von Optimismus als interindividuell unterschiedlich ausgeprägter Disposition wurden vorgeschlagen: Neben der indirekten Messung von optimistischen Erwartungen mittels der Beurteilung des typischen Attributionsstils einer Person und der Messung des Konstruktes Hoffnung ist der wohl populärste Ansatz die mehr oder weniger direkte Erfassung von dispositionellem Optimismus mittels Selbstberichtsfragebogen. Sehr viele der relevanten gesundheitspsychologischen Untersuchungen verwendeten einen solchen Fragebogen, nämlich den „Life Orientation Test“ (LOT) von Scheier und Carver (1985) bzw. dessen revidierte Fassung, den LOT(-R) (Scheier, Carver & Bridges, 1994) zur Messung von Optimismus. Angesichts seiner großen Popularität (eine Internet-Recherche bei scholar.google.com etwa zeigt mehr als 450 Zitationen des Artikels von Scheier et al., 1994) verwundert es aber nicht, dass der LOT(-R) auch Ziel methodisch orientierter Kritik geworden ist. Von besonderer Bedeutung ist dabei die von verschiedenen Autoren aufgestellte Hypothese, dass der LOT(-R) nicht etwa ein einziges, sondern zwei Konstrukte erfasse, welche „Optimismus“ und „Pessimismus“ genannt werden, die aber konträr zum Alltagssprachgebrauch nicht als einander entgegengesetzt dargestellt werden. Diese „Optimismus-Pessimismus-Unabhängigkeits-Hypothese“ (im Folgenden der Kürze halber die OP-Hypothese genannt) stellt allerdings nicht nur die Validität des Instrumentes selbst, sondern auch die dahinter liegende Theorie des dispositionellen Optimismus in Frage, die von einem Kontinuum mit zwei Polen ausgeht (Scheier & Carver, 1985) – ein gleichzeitiges Vorliegen von Optimismus und Pessimismus ist in dieser Theorie unmöglich.

1.2 Die Optimismus-Pessimismus-Unabhängigkeits-Hypothese

Die OP-Hypothese ist offensichtlich zumindest aus theoretischer Perspektive bedeutsam: Seit sie (wahrscheinlich) zum ersten Mal von Marshall, Wortman, Kusulas, Herwig und Vickers (1992) vorgetragen wurde, ist sie von so vielen anderen Forschern aufgegriffen worden, dass eine vollständige Aufzählung in dieser Arbeit unangemessen erscheint. Die folgende Auswahl nennt aber wohl die einflussreichsten: Genau wie Marshall et al. (1992) veröffentlichten auch Robinson-Whelen, Kim, MacCallum und Kiecolt-Glaser (1997) im *Journal of Personality and Social Psychology*, in dem auch die Arbeit von Scheier et al. (1994) publiziert wurde. Eine weitere häufig zitierte Arbeit zur OP-Hypothese stammt aus der Zeitschrift *Health Psychology* (Mroczek, Spiro, Aldwin & Ozer,

1993). Darüber hinaus sind besonders die Arbeiten von E.C. Chang¹ und Kollegen (E.C. Chang, Maydeu-Olivares & D’Zurilla, 1997; E.C. Chang, D’Zurilla & Maydeu-Olivares, 1994), von Kubzansky, Kubzansky und Maselko (2004) sowie die von Herzberg, Glaesmer und Hoyer (2006) zu beachten, die die OP-Hypothese aus einer stärker methodisch orientierten Perspektive vertreten².

Die OP-Hypothese lässt sich wie folgt wiedergeben: Die Kovarianzen der Antworten auf die

Items des LOT(-R) lassen sich nicht durch einen einzigen Faktor repräsentieren, wie es der Fall sein sollte, wenn alle Itemantworten ein einziges latentes Konstrukt widerspiegeln (s. Kapitel 3). Stattdessen werden zwei Faktoren benötigt, um die Struktur der Kovarianzen abzubilden. Die Vertreter der OP-Hypothese setzen diese zwei Faktoren mit zwei Konstrukten gleich, die sie „Optimismus“ und „Pessimismus“ nennen. Allerdings erscheint die OP-Hypothese bei näherer Betrachtung schwach begründet: Zum einen stützt sie sich ausschließlich auf faktorenanalytische Befunde, d.h. auf eine Methode, deren Ergebnisse empirisch unterdeterminiert sind in dem Sinne, dass alternative Lösungen existieren, die eine genauso gute Erklärung der Daten ermöglichen. Zum anderen beruhen diese Befunde fast ausschließlich auf Erhebungen mit dem LOT(-R) (ausgenommen Chang et al., 1997, und Kubzansky et al., 2004, die Abwandlungen des LOT untersucht), d.h. das Vorliegen einer scheinbar für die OP-Hypothese sprechenden Faktorenlösung ist meist mit der Verwendung des LOT(-R) konfundiert. Zuletzt wird die OP-Hypothese auch dadurch geschwächt, dass sich die Indikatoren der „Optimismus“- und „Pessimismus“-Faktoren aus den jeweils positiv bzw. invertiert formulierten Items (d.h. Aussagesätze, deren *Bejahung* als indikativ für Optimismus gesehen wird, und Aussagen, deren *Verneinung* als indikativ für Optimismus gesehen wird) zusammensetzen. Das Phänomen solcher „Richtungsfaktoren“, d.h. einer Faktorenstruktur mit einem Faktor für positiv formulierte Items und einem zweiten Faktor für invertiert formulierte Items, ist allerdings keineswegs auf den LOT(-R) beschränkt, sondern tritt bei unterschiedlichsten Fragebogen auf (vgl. etwa Marsh, 1996).

Das zentrale Anliegen der vorliegenden Arbeit ist es, die OP-Hypothese kritisch zu prüfen und letztendlich zurückzuweisen. Dabei werden drei Argumentationslinien methodologischer und empirischer Art verfolgt, die sich in den folgenden drei Thesen zusammenfassen lassen:

1. Das Unterdeterminiertheitsargument: Die zur Unterstützung der OP-Hypothese herangezogenen korrelativen und faktorenanalytischen Methoden sind keineswegs geeignet, irgendeine Hypothese zu *bekräftigen* bzw. zu *bestätigen* – auch wenn solche Methoden „konfirmatorisch“ genannt werden. Zu kritisieren ist die Methodik nicht allein aufgrund der grundsätzlichen und in den empirischen Wissenschaften praktisch universell akzeptierten Überlegung, dass eine „Konfirmation“ von Hypothesen über induktive Schlüsse wegen der Begrenztheit der Beobachtungsmenge niemals möglich ist, sondern es wird auch gezeigt, dass faktorenanalytische

¹ Beachtenswert ist, dass E.C. Chang in einer neueren, inhaltlich orientierten Arbeit zwar den LOT-R verwendet, aber gar nicht auf die OP-Hypothese eingeht, sondern einzig den Gesamtwert „Optimismus“ in seine Analysen einbezieht.

² Andere methodisch orientierte Arbeiten zum Thema lassen sich eher der Gegenposition, also der Zurückweisung der OP-Hypothese, wie sie auch in der vorliegenden Arbeit vertreten wird, zuordnen, vor allem Røysamb und Strype (2002), McPherson und Mohr (2005) und Vautier, Raufaste und Cariou (2003); Stéphane Vautier scheint aber gewissermaßen die Seiten gewechselt zu haben, s. Vautier und Raufaste (2006).

lytische Methoden unter einer eigenen, speziellen Unterdeterminiertheit und Uneindeutigkeit leiden. Faktisch wird argumentiert, dass sog. „konfirmatorische“ Faktorenmodelle nicht nur *nicht* „konfirmatorisch“ sind, sondern zumindest im Rahmen gängiger Praxis sogar annähernd *unfalsifizierbar*. Das Attribut „konfirmatorisch“ in „konfirmatorischen“ Faktorenanalysen im Sinne der „Bestätigung“ eines Messmodells erscheint mehr als unangemessen schließlich deshalb, da für eine gegebene Datenlage grundsätzlich immer alternative Modelle existieren, die die gleiche oder unter Umständen auch eine bessere Lösung darstellen.

Die Hypothese, dass der LOT-R partiell unabhängige Konstrukte namens „Optimismus“ und „Pessimismus“ erfasse, wird mittels methodologischer und im engeren Sinne statistisch-methodischer Untersuchungen als zu gering begründet zurückgewiesen.

2. Das Methodeneffektmodell: Anhand einer eigenen empirischen Erhebung mit dem LOT-R wird eine alternative Faktorenlösung mit zwei Faktoren vorgestellt, die die theoretisch angenommene Struktur (eine Dimension mit „Optimismus“ und „Pessimismus“ als entgegengesetzten Polen) beibehält und dabei gleichzeitig sowohl eine psychologisch plausible Erklärung für die Zweidimensionalität der Korrelationsmatrizen liefert als eine bessere Anpassung im mathematisch-statistischen Sinne aufweist als das mit der OP-Hypothese verbundene Modell. Darüber hinaus lässt sich die alternative Faktorenlösung in die bestehende Literatur zum Thema „Antwortverzerrung in Persönlichkeitsfragebögen durch Soziale Erwünschtheit“ einordnen.

Die „OP-Hypothese“ wird deutlich entkräftet, da das „Methodeneffektmodell“ eine bessere faktorenanalytische Lösung für Daten liefert, die mit dem LOT-R erhoben wurden.

3. Die Alternativskala: In einer weiteren Untersuchung wird gezeigt, dass die Skala „Personaler Optimismus“ (PO-Skala) aus dem POSO-E-Fragebogen von Schweizer, Schneider und Beck-Seyffer (2001), die ebenfalls dispositionellen Optimismus erfasst, zumindest im Hinblick auf ihre psychometrischen Eigenschaften besser geeignet erscheint als der LOT(-R) und anders als dieser keine faktorenanalytische Bidimensionalität aufweist, obwohl auch die Skala „Personaler Optimismus“ invertierte Items enthält. Die OP-Hypothese ist also in Hinblick auf die PO-Skala nicht haltbar.

Die „OP-Hypothese“ wird zurückgewiesen, da die für diese Hypothese sprechenden Befunde ausschließlich auf dem LOT-R basieren, aber nicht durch alternative Instrumente für die Erfassung von Optimismus repliziert werden können.

Die OP-Hypothese wird also zunächst aus theoretischer Perspektive entkräftet; anschließend wird gezeigt, dass die von den Vertretern der OP-Hypothese verwendeten Methoden nicht geeignet sind, um eine empirische Entscheidung über die Hypothese herbeizuführen. Weiterhin wird mit dem Methodeneffektmodell eine empirisch besser fundierte Alternativerklärung vorgestellt. Zuletzt wird gezeigt, dass das Problem der Zweidimensionalität positiv und invertiert formulierter Optimismus-Items, auf dem die OP-Hypothese beruht, für die PO-Skala überhaupt nicht auftritt.

Für wesentliche theoretische und empirische Ergebnisse dieser Arbeit wird auf bereits veröffentlichte Schriften zurückgegriffen, nämlich

1. für die Darstellung wesentlicher mathematisch-statistischer Ergebnisse auf Schrift A:

Schrift A: Rauch, W. A. & Moosbrugger, H. (im Druck). Klassische Testtheorie. Grundlagen und Erweiterungen für heterogene Tests und Mehrfacettenmodelle. In M. Amelang & L.F. Hornke (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich B Methodologie und Methoden, Serie II Psychologische Diagnostik, Band 2 Methoden*. Göttingen: Hogrefe.

2. Die eigene empirische Erhebung mit dem LOT-R ist in Schrift B enthalten:

Schrift B: Rauch, W. A., Moosbrugger, H. & Schweizer, K. (2007). Method effects due to social desirability as a parsimonious explanation of the deviation from unidimensionality in LOT-R scores. *Personality and Individual Differences*, 42, 1597-1607.

3. Schließlich wurden die empirischen Befunde zur Skala „Personaler Optimismus“ in Schrift C vorgestellt:

Schrift C: Rauch, W. A., Moosbrugger, H. & Schweizer, K. (im Druck). An IRT Analysis of the Personal Optimism Scale. *European Journal of Psychological Assessment*.

1.3 Übersicht

In Kapitel 2 wird der inhaltlich-psychologische Hintergrund zum Thema Optimismus vorgestellt, dabei wird insbesondere auf die Theorie zum dispositionellen Optimismus von Scheier und Carver (1985; Scheier, Carver & Bridges, 1994; Carver & Scheier, 2002) und den Vergleich dieser Theorie mit anderen Konzeptionen von Optimismus eingegangen. In Kapitel 3 und 4 wird das Unterdeterminiertheitsargument dargelegt, wonach die faktorenanalytischen Ergebnisse zum LOT-R nicht rechtfertigen, die Wirkung eines partiell unabhängigen Konstruktes „Pessimismus“ anzunehmen. In Kapitel 3 wird die für eine faktorenanalytische Auswertung erforderliche Annahme untersucht, dass sich psychologische Konstrukte unmittelbar in den Antworten auf Selbstberichtsfragebogen widerspiegeln. Dieses „reflexive Messmodell“ wird mit gängigen sozialpsychologischen Ergebnissen über Einstellungsfragebogen kontrastiert. In Kapitel 4 wird aus testtheoretischer Perspektive dargelegt, dass faktorenanalytische Ergebnisse an einer fundamentalen Unterdeterminiertheit leiden und dass die Vertreter der OP-Hypothese wichtige Alternativerklärungen für die Bidimensionalität der LOT(-R)-Daten vernachlässigen. In beiden Kapiteln wird auf inhaltlich-theoretische Ergebnisse aus Kapitel 2 zurückgegriffen sowie auf testtheoretisch-statistische Befunde aus Schrift A. Kapitel 5 greift auf die empirische Studie zum LOT-R zurück, die in Schrift B beschrieben wurde (das Methodeneffektmodell) und Kapitel 6 auf die Studie zur Skala „Personaler Optimismus“ (die Alternativskala; Schrift C). In Kapitel 7 schließlich werden die Ergebnisse zusammenfassend diskutiert.

2 Optimismus

2.1 Der Begriff „Optimismus“

2.1.1 Zur Problematik der Begriffsverwendung in der OP-Hypothese

Die OP-Hypothese lässt sich als Aussage A wie folgt formulieren: „Optimismus und Pessimismus repräsentieren zwei ‚partiell unabhängige Dimensionen‘“. Eine pointierte Umformung der OP-Hypothese ist A': „Menschen können gleichzeitig optimistisch und pessimistisch sein“. Wie die Formulierung A' zeigt, ist die OP-Hypothese eine „bold conjecture“ im Popperschen Sinne, d.h. eine These mit einem gewissermaßen unerwarteten, neuartigen oder gar kontraintuitiven Charakter – der folgenden Darstellung der historischen Entwicklung der Begriffe „Optimismus“ und „Pessimismus“ ist zu entnehmen, dass „Optimismus“ und „Pessimismus“ in anderen Kontexten als der OP-Hypothese als entgegengesetzte, einander ausschließende Einstellungen aufgefasst werden. Die OP-Hypothese führt also eine Verwendung der Begriffe ein, die unüblich ist und daher auch zu Missverständnissen führen kann.

Eine sprachlich-historische Untersuchung erscheint auch daher nötig, dass die unübliche Begriffverwendung in der OP-Hypothese mit einer Uneindeutigkeit der Untersuchungsmethodik korrespondiert (s. vor allem Kap. 4). Schon überhaupt das, was benannt wird, ist ja Ergebnis einer Interpretation. Mit den Begriffen „Optimismus“ und „Pessimismus“ werden keineswegs direkte Beobachtungen belegt, sondern „Faktoren“ oder „latente Variablen“ als Ergebnisse der faktorenanalytischen Untersuchung der Korrelationen von Reaktionen auf Fragebogenitems, und die Fragebogenitems selbst sind auch wiederum nur indirekte Indikatoren (Kapitel 3). Um den Hintergrund für die im engeren Sinne methodischen und empirischen Argumente zu liefern, wird in der folgenden Darstellung zunächst auf die Verwendung der Begriffe „Optimismus“ und „Pessimismus“ in der historischen Entwicklung eingegangen; die Begriffe wurden zunächst verwendet, um theologische und philosophische Positionen zu kennzeichnen. Anschließend werden psychologische Theorien zum Optimismus vorgestellt.

2.1.2 Historische und philosophische Perspektiven

Die Verwendung der Begriffe „Optimismus“ und „Pessimismus“ ist für das 17. Jahrhundert zum ersten Mal dokumentiert. In ihrer damaligen Bedeutung bezeichneten sie unterschiedliche und gegensätzliche philosophische Positionen (Domino & Conway, 2002); erst später erweiterte sich ihre Bedeutung auf die Kennzeichnung interindividueller Unterschiede („X ist optimistischer als Y“). Als Begriffe zur Bezeichnung philosophischer und weltanschaulicher Positionen verwendet, ist die Gegensätzlichkeit von „Optimismus“ und „Pessimismus“ unzweifelhaft; eine Optimistin erwartet grundsätzlich das Gute, eine Pessimistin das Schlechte. Beachtenswert ist, dass religiöser, metaphysischer oder weltanschaulicher Optimismus keinesfalls „unrealistisch“ ist in dem Sinne, dass „das Schlechte“ geleugnet würde, wie die folgenden Ausführungen zeigen sollen.

2.1.2.1 Optimismus und Pessimismus in der Religion

„Das Schlechte“ ist wohl einer der wesentlichen Ausgangspunkte religiöser oder metaphysischer Auseinandersetzungen mit der Welt: Die erste „edle Wahrheit“ des Buddhismus etwa ist es, dass alles Leben Leiden sei, und in der christlichen Schöpfungsgeschichte

wird die Menschwerdung, nämlich das Verzehren des verbotenen Apfels, gleichgesetzt mit der Erkenntnis, dass überhaupt Gut und Böse sei: „Gott weiß vielmehr: Sobald ihr davon esst, gehen euch die Augen auf; ihr werdet wie Gott und erkennt Gut und Böse“ (Gen. 3,5). Im Christentum wird das Leiden, das durch Böses entsteht, also aktiv zugefügt wird, nicht grundsätzlich von einem Leiden durch Krankheit, Umwelt- oder Natureinflüsse unterschieden; alles Leiden und auch der Tod sind Folge der Trennung von Gott (von Brück, 2002). Religiöse Vorstellungen setzen also bei der Akzeptanz³ des Leidens an, versprechen aber, meist durch religiöse Praxis, gewissermaßen dialektisch den Ausweg aus dem Leiden in eine bessere, nämlich jenseitige Welt.

Sind Religionen aufgrund ihrer Dialektik sowohl pessimistisch als auch optimistisch? Die gängige Antwort ist Nein (Miller, Richards & Keller, 2002). Der Optimismus der Religionen äußert sich gerade in der Lebensbejahung oder zumindest im Beharren, in der Akzeptanz, *obwohl* alles Leben Leiden ist (vgl. von Brück, 2002). Zwar wird die Dialektik hinsichtlich der Akzeptanz des Leidens in der Welt von manchen als Verleugnung aufgefasst und von anderen erst recht als Pessimismus: Dass alles Leben Leiden sei, ist schließlich nicht unbedingt positiv. Die religiösen Gesamtsysteme aber sind immer optimistisch und nicht pessimistisch – und erst recht nicht beides gleichzeitig. Glauben impliziert die Erwartung eines „Besseren“, wenn es auch erst im Jenseits kommt. Aus einer philosophischen, abstrakten Perspektive heraus kann Glauben auch als eine dialektische Auflösung begriffen werden, als eine Synthese aus der Erwartung des Leidens im Leben angesichts der Erwartung des Besseren im Jenseits.

2.1.2.2 Optimismus und Pessimismus in der Philosophie

Die wohl bekannteste philosophisch-optimistische Position stammt von Gottfried Wilhelm Leibniz, dessen optimistische Haltung im Zusammenhang mit seinem „Gottesbeweis“ deutlich wird: Leibniz setzt voraus, dass Gott vollkommen sei; die Unvollkommenheit der Welt und die Existenz des „Übels“ in der Welt sind aber dennoch erklärbar, da schließlich Gott den vollkommenen Plan zur Gestaltung der Welt verfolgt habe, ja, da Gott vollkommen ist, ist es dem unvollkommenen Menschen gar nicht möglich, die Vollkommenheit der Schöpfung in Zweifel zu ziehen. In der Tat erfülle auch die Unvollkommenheit der Welt einen Zweck: In einer vollkommenen Welt ist das Empfinden von Glück nicht möglich, sondern nur durch den Wechsel zwischen mehr oder weniger guten Zuständen (Helferich, 1992).

Allerdings wird das Hinwenden ins Jenseitige von prominenten pessimistischen Philosophen verspottet: Leibniz' optimistische Position – oder vielmehr: eine (absichtlich) falsch verstandene Abwandlung seines Optimismus – wurde nicht zuletzt berühmt durch ihre satirische Verspottung bei Voltaire, der seine übermäßig ans Gute glaubende Romanfigur Candide eine lange Reihe von Unglücken und Missgeschicken erleben ließ, bevor dieser seine optimistische Einstellung aufgibt: „Gut gesagt! recht gut! sagte Candide, allein wir müssen unsern Garten bestellen.“ Falsch verstanden ist der Leibnizsche Optimismus hier aber, genau wie auch der allgemeinere religiöse Optimismus, insofern als sich weder der eine noch der andere auf jedes einzelne Individuum und die *Erfüllung* der Erwartungen bezieht – tatsächlich ist Optimismus ja per Definition zukunftsgerichtet. Eine Gegenposition zu Leibniz entwickelte Schopenhauer, der unter Bezug auf ersteren gar den Beweis

³ Man beachte, dass Akzeptanz nicht unbedingt Passivität impliziert; religiöse Praxis richtet sich schließlich sehr häufig auf eine Verringerung des Leidens Anderer, genau wie auch a-religiöse Praxis etwa gemeinnütziger Organisationen in fast allen Kulturen.

anstrebe, dass die Welt die schlechteste aller möglichen Welten sei; selbst Glücksgefühl sei nur eine Illusion aufgrund einer temporären Reduktion des konstanten Leidens.

Das Wechselspiel zwischen dem Streben nach dem „Besseren“ und der Akzeptanz des Leidens, also von Unvollkommenheit, unmoralischem Verhalten, Krankheit und Tod, wurde aber letztlich von den meisten Philosophen dialektisch aufgelöst. Immanuel Kant etwa hielt zwar sowohl den Beweis von Gottes Existenz als auch den Beweis seiner Nicht-Existenz für unmöglich, plädierte aber dennoch aus *rationaler* Sicht für den Glauben an Gott und auch für moralisches Verhalten selbst im Angesicht dessen, dass es Menschen, die unmoralisch handeln, offenbar häufig besser zu gehen scheint (Domino & Conway, 2002); auf diese Weise ersetzt Kant die metaphysische Begründung für eine optimistische Weltanschauung durch eine rationale Begründung. Bei Nietzsche schließlich spiegelt sich auf der einen Seite der Schopenhauersche Pessimismus wieder: Im dionysischen Element der Kunst und der Musik äußere sich das ziellose Streben des Lebens, dem das apollonische Element, das nach Individualität strebt, am Ende nicht widerstehen kann. Auf der anderen Seite finden sich bei Nietzsche aber auch immer wieder optimistische Tendenzen, etwa bei der Hoffnung auf den „Übermenschen“, der das Menschsein an sich überwindet: „Was ist der Affe für den Menschen? Ein Gelächter oder eine schmerzliche Scham. Und ebendas soll der Mensch für den Übermenschen sein: ein Gelächter oder eine schmerzliche Scham.“ (Nietzsche, 1887/1930).

2.1.3 Schlussfolgerungen

Im Bisherigen wurde gezeigt, dass die Aussage A': „Menschen können gleichzeitig optimistisch und pessimistisch sein“, also eine Umformulierung der OP-Hypothese, nach der historischen Entwicklung der Begriffe Optimismus und Pessimismus unüblich erscheint. Legt man die Verwendung der Begriffe in der Theologie oder Philosophie zugrunde, kann der Gehalt der OP-Hypothese nicht bewertet werden, oder sie muss sogar als falsch gewertet werden. Üblicherweise werden die Begriffe Optimismus und Pessimismus als *Gegensätze* gewertet, genauso wie etwa die Begriffe Hitze und Kälte. Die Aussage B': „Ein Gegenstand kann gleichzeitig heiß und kalt sein“ ist aber der üblichen Aussagenlogik gemäß falsch; wahr werden kann er höchstens durch Hinzunahme weiterer Vorbedingungen, indem etwa der Gegenstand als aus mehreren Teilen bestehend angenommen wird; eine weitere Möglichkeit, die Aussage B' logisch konsistent zu machen, wäre die Umdeutung der Begriffe „heiß“ und „kalt“, so dass sie *etwas anderes* meinen als im allgemeinen Sprachgebrauch.

2.2 Optimismus in der Psychologie

Im vorangegangenen Abschnitt wurde ausgeführt, dass die OP-Hypothese begrifflich-logisch fragwürdig ist, wenn die allgemein übliche Bedeutung der Begriffe „Optimismus“ und „Pessimismus“ herangezogen wird. Allerdings bezieht sich die vorliegende Arbeit nicht in erster Linie auf eine sprachliche Analyse. Sehr wahrscheinlich haben die Vertreter der OP-Hypothese, vor allem Marshall et al. (1992), Robinson-Whelen et al. (1997) sowie Chang (2002b; Chang, Maydeu-Olivares & D'Zurilla, 1997) etwas anderes im Sinn gehabt als die übliche Bedeutung der Begriffe Optimismus und Pessimismus, und selbstverständlich wäre dies nicht das erste Mal, dass ein Begriff in einer Spezialwissenschaft wie der Psychologie eine andere Bedeutung erhält als im Alltagsgebrauch. In der Psychologie ist es häufig sogar der Fall, dass ein Begriff für den einen Forscher etwas anderes bedeutet als für den anderen, oder dass ein und dasselbe Phänomen oder Konzept von unterschiedlichen Forschern mit unterschiedlichen Begriffen bezeichnet wird (vgl. etwa Block,

1995, zur Frage der Nicht-Übereinstimmung der „Big Five-Faktoren“ aus dem lexikalischen Ansatz der Persönlichkeitsforschung mit den Big Five nach dem NEO-Modell). Der Bedeutungsklärung wird in dieser Arbeit deshalb so viel Platz eingeräumt, weil die Vertreter der OP-Hypothese erstaunlich wenig darüber sagen, was „Optimismus“ und „Pessimismus“ innerhalb ihres Theoriegebäudes bezeichnen.

2.2.1 Die OP-Hypothese und ihre Basis

Die OP-Hypothese wurde explizit und an exponierter Stelle wohl zum ersten Mal von Marshall et al. (1992) vertreten.

„... [earlier studies using the LOT were based on] the commonly held assumption [...] that *optimism and pessimism are opposite poles of a single unipolar continuum* and – in particular – that the LOT itself reflects this unidimensional nature. In actuality, *factor analytic studies of the LOT* [...] *have consistently demonstrated the superiority of a two-dimensional model of optimism and pessimism* (i.e., Marshall & Lang, 1990; Scheier & Carver, 1985). To date, however, the distinction between *optimism and pessimism factors* has been tentatively discounted as due to individual differences in responding to item format rather than to item content. ... At best, however, analysis of internal relations among LOT items provides only a partial test of the independence of *the two dimensions*. A better empirical basis for the claim of unidimensionality would be provided by examining the relations of *optimism and pessimism* to external criteria.” (Marshall et al., 1992, S. 1067, Hervorhebungen durch den Autor)

Dieser Ausschnitt ist in seiner Kürze durchaus repräsentativ für die Vertreter der OP-Hypothese. Die OP-Hypothese entstand aufgrund von Faktorenanalysen von Daten, die mit dem LOT erhoben wurden. Der LOT wird fast als einziges Messinstrument im Rahmen der Theorie des dispositionellen Optimismus verwendet (s. unten). Ein tieferes Verständnis der Theorie des dispositionellen Optimismus wird aber von Marshall et al. (1992) genau wie von den anderen Vertretern der OP-Hypothese offenbar nicht vorausgesetzt. Deutlich werden an diesem Zitat aber vor allem zwei Aspekte:

1. Die OP-Hypothese basiert auf faktorenanalytischen Untersuchungen des LOT, und zwar so gut wie ausschließlich.

Dember (2002) beschreibt zwar einen eigenen „Optimismus-Pessimismus“-Fragebogen, aber dessen Validität wurde von Chang et al. (1994) – also von Vertretern der OP-Hypothese – aufgrund seiner Multidimensionalität in Frage gestellt. Chang et al. (1997) wiederum haben selbst einen Fragebogen, den ELOT, erstellt; dieser Fragebogen eignet sich aber *nicht* zur Überprüfung der OP-Hypothese: die Items des ELOT, die z.T. direkt aus dem LOT stammen, wurden absichtlich so ausgewählt, dass sie die postulierte Zweidimensionalität möglichst gut abbilden. Dass ein solchermaßen konstruierter Fragebogen zwei Dimensionen abbildet, ist eine Tautologie, und kein Mittel zur Überprüfung einer Hypothese.

2. Die Vertreter der OP-Hypothese vermischen theoretische Begriffe (*optimism/pessimism* in der ersten Hervorhebung) mit den Ergebnissen einer Faktorenanalyse (zweite Hervorhebung). In der Folge wird aber von Schlussfolgerungen auf der theoretischen Ebene (*optimism and pessimism* in der dritten Her-

vorhebungen) ausgegangen, d.h. ohne weitere Begründung wird die Faktorenlösung mit theoretischen Konstrukten gleichgesetzt.

Nicht ausgeführt wird allerdings, was „Pessimismus“ ist – es gibt so gut wie keine theoretische Darlegung dessen, was „Pessimismus“ denn sei, wenn *nicht* das Gegenteil von „Optimismus“. Dieser Mangel an theoretischer Begründung ist umso gravierender, als – wie oben gezeigt wurde – die Verwendung der Begriffe in der OP-Hypothese nicht mit der üblichen Verwendung der Begriffe übereinstimmt (am Beispiel: Wenn man den Begriff „heiß“ umdefinieren würde, so dass er nicht mehr das Gegenteil von „kalt“ ist, müsste man gut erklären, was das neue „heiß“ im Unterschied zum alten bedeuten soll; ansonsten würde man immer wieder auf Verständigungsprobleme stoßen). Stattdessen wird nur das Ergebnis, dass zwei Faktoren, also mathematisch-statistische Konstruktionen, für die Repräsentation der Zusammenhangsstruktur der LOT(-R)-Items gebraucht werden, unreflektiert als Beleg dafür akzeptiert, dass es zwei psychologisch bedeutsame Konstrukte gebe. Manche Vertreter der OP-Hypothese bemühen sich, die Lücke zu füllen; Chang et al. (1997) sowie Kubzansky et al. (2004) sprechen davon, dass Pessimismus „generalisierte negative Zukunftserwartungen“ darstelle. Allerdings bringt dies keine wirkliche Auflösung: „generalisiert“ bedeutet „verallgemeinert“, „abstrahiert“ oder ähnliches. Dass jemand gleichzeitig „verallgemeinerte positive“ und „verallgemeinerte negative“ Erwartungen haben kann, erscheint ebenso merkwürdig wie die Aussage, dass jemand gleichzeitig optimistisch und pessimistisch sei, wie deutlich wird, wenn die zuvor eingeführte Analogie erneut herangezogen wird: „Im Allgemeinen ist mir zu kalt und zu warm.“ Da die unübliche Verwendung der Begriffe „Optimismus“ und „Pessimismus“ von den Vertretern der OP-Hypothese nicht ausreichend erklärt wird, soll im Folgenden überprüft werden, ob andere psychologische Theorien im Zusammenhang mit Optimismus Anhaltspunkte dafür liefern können, was „Pessimismus“ denn sein könnte.

2.2.2 Dispositioneller Optimismus

Die OP-Hypothese basiert auf Untersuchungen mit dem LOT(-R), der als Messinstrument im Rahmen der Theorie des dispositionellen Optimismus entwickelt wurde. Diese Theorie wurde vor allem in Scheier und Carver (1985) sowie Carver und Scheier (2002) dargelegt. Zwar ist auch die Theorie des dispositionellen Optimismus kurz gefasst – der Artikel von Scheier und Carver (1985) etwa umfasst 29 Seiten, von denen knapp drei Seiten der eigentlichen Theorie gewidmet werden – im Vergleich zur „Theorie“ hinter der OP-Hypothese ist sie aber erfrischend gehaltvoll.

Im Wesentlichen besteht die Theorie aus Definitionen des dispositionellen Optimismus wie den folgenden, die anschließend vor dem Hintergrund einer allgemeineren Theorie der Selbstregulation erläutert werden.

„Optimisten sind Menschen, die dazu neigen, positive Erwartungen an die Zukunft zu haben“ (Scheier et al., 1994, S. 1063).

„Optimisten sind Menschen, die gute Erfahrungen in der Zukunft erwarten“ (Carver und Scheier, 2002, S. 31; Übersetzungen jeweils durch den Autor).

Zusammengenommen wird dispositioneller Optimismus auch als *generalisierte positive Zukunftserwartung* definiert (z.B. Scheier & Carver, 1985). Diese Definition wird auch für die vorliegende Arbeit übernommen (vgl. auch Schweizer & Koch, 2001; Schweizer et al., 2001).

Was der Begriff „generalisierte positive Zukunftserwartung“ bedeutet, wird klarer, wenn man das Konstrukt im allgemeinem Modell der Selbstregulation nach Carver und Scheier (1998) verankert (s. auch Carver & Scheier, 2002). In diesem Modell wird Verhalten als „Ausgabefunktion“ eines hierarchisch organisierten Feedback-Systems gesehen. Verhalten wird eingesetzt, um Ziele zu erreichen, und durch Kontrollfunktionen wird das Verhalten je nach Zielerreichungsgrad angepasst. Höher geordnete Ziele (etwa das Ziel der Selbsterhaltung) üben ihren Einfluss aus, indem die Kontrollfunktionen – und damit die Ziele – untergeordneter Feedbackschleifen angepasst werden. So könnte etwa das Ziel, Hunger zu verringern, erreicht werden, indem ein Bewegungssystem in Gang gesetzt wird, das wiederum ein Ziel hat, nämlich den Löffel voller Suppe zum Mund zu führen; dann, wenn der Hunger ausreichend verringert wurde, wird das Ziel des Löffel-Bewegungssystems wieder auf Null gesetzt.

Dieses Modell der Selbstregulation ist zwar keineswegs in Gänze empirisch überprüft (fraglich ist, ob es überhaupt geprüft werden kann), aber ähnliche Modelle sind in der Psychologie sehr weit verbreitet (vgl. etwa Frese & Zapf, 1994, oder Dörner, 2004), und für jeweils kleine Ausschnitte und spezifische Funktionen gibt es empirische Belege (Carver & Scheier, 1998). Emotionen haben im Selbstregulationsmodell von Carver und Scheier die Funktion der Überwachung der Qualität und/oder Geschwindigkeit dessen, mit der der Verhaltensregulationsprozess zum angestrebten Ziel führt (Carver & Scheier, 2002): Arbeitet der Verhaltensprozess wie beabsichtigt, dann resultieren positive Gefühle⁴, tut er es nicht, so resultieren negative Gefühle; positive Gefühle führen entsprechend zur Aufrechterhaltung des laufenden Prozesses, negative Gefühle zur Modifikation oder zur Zielaufgabe. Gefühle sind also in diesem Modell *nicht* Konsequenzen einer Zielerreichung auf unterer Ebene, sondern begleiten und modifizieren verhaltenssteuernde Feedbackschleifen (vgl. auch Dörner, 2004). Demnach beinhalten Gefühle also auch einen auf die Zukunft gerichteten Aspekt, eine Erwartung: positive Gefühle entstehen, wenn der Prozess *erfolgsversprechend* abläuft, nicht nur dann, wenn der Erfolg bereits eingetreten ist.

Innerhalb dieses von Gefühlen modifizierten Selbstregulationssystems haben vergangene Erfahrungen eine Auswirkung auf zukünftige Prozesse. Bisherige Erfolge (oder Misserfolge) dienen als Kontrollfunktion für die Emotionsregulation, insofern als die vorausschauende Überwachung der Zielerreichung durch Gefühle wiederum hinsichtlich eines bestimmten Ziels geprüft werden muss. Dispositioneller Optimismus entsteht nun aus der verallgemeinerten gesammelten Erfahrung hinsichtlich des Erfolges bisheriger Zielerreichungsversuche (Carver & Scheier, 2002): je mehr bisherige Erfolge, desto optimistischer die Erwartung für zukünftige Versuche. Die „Funktion“⁵ von Optimismus ist also diese: in einem gegebenen Selbstregulationsprozess den Ausgangspunkt und Kontrollwert für modifizierende Regulationsprozesse zu setzen. Je optimistischer die Zukunftserwartungen, desto eher wird eine Handlung in Gang gesetzt und desto eher wird eine Handlung aufrechterhalten. Innerhalb dieses Systems ist das gleichzeitige Vorliegen *generalisierter* positiver wie negativer Zukunftserwartungen unmöglich: Der Kontrollwert ist entweder positiv *oder* negativ ausgerichtet.

⁴ Carver und Scheier (2002) benutzen den Begriff „feelings“; es wird nicht deutlich, ob damit eine spezielle Eingrenzung des phänomenalen Bestandteils von Emotionen gemeint ist oder ob sich die Ausführungen auch auf die physiologischen und verhaltensmäßigen Bestandteile beziehen.

⁵ Selbstregulationstheorien wie der von Carver und Scheier (1998) liegt zumindest implizit ein aus der Biologie übernommenes funktionales Erklärungsmodell zugrunde: Physische und psychische Phänomene sind „funktional“, da sie dem „Zweck“ der Selbsterhaltung dienen.

Dass Optimismus zu einer höheren Handlungsbereitschaft und Persistenz führt, ist die zentrale medierende Hypothese über den Zusammenhang von Optimismus und Gesundheit. Und tatsächlich sprechen zahlreiche Befunde dafür, dass Optimisten beispielsweise angemessenere Coping-Strategien verwenden und entsprechend besser mit schweren Erkrankungen umgehen (Peterson & Bossio, 2002). Allerdings gibt es auch Forscher, die der grundlegenden Annahme einer positiven Auswirkung von Optimismus widersprechen. Auffällig sind etwa zwei Studien, die den Zusammenhang von dispositionellem Optimismus mit risikoreichem Sexualverhalten untersuchten und zu genau entgegengesetzten Ergebnissen gelangen: Während Perkins, Leserman, Murphy und Evans (1993) berichteten, dass optimistische homosexuelle Männer sich *höherem* Risiko einer HIV-Infektion aussetzen, kamen Carvajal, Garner und Evans (1998) theoriekonform zum umgekehrten Ergebnis.

An dieser Stelle soll keine lange Erörterung der positiven und negativen Auswirkungen von Optimismus erfolgen, sondern nur festgehalten werden, dass generalisierte positive Zukunftserwartungen nicht dysfunktional sind, sondern im Allgemeinen eine angemessene psychosoziale und psychophysische Anpassung des Individuums ermöglichen. Vor allem aber sollte eine Schlussfolgerung aus der Konzeption von Optimismus als *generalisierter* positiver Zukunftserwartung klar werden: Wenn sich das Ausmaß von Optimismus aus dem Ausmaß bisheriger Selbstregulationserfolge ergibt – je erfolgreicher in der Vergangenheit, desto optimistischer für die Zukunft – dann ist die Annahme einer „partiellen Unabhängigkeit von Optimismus und Pessimismus“ bedeutungslos bzw. logisch falsch. Generalisierte Erwartungen können nicht gleichzeitig positiv und negativ sein; eine Gleichzeitigkeit positiver und negativer Erwartungen ist höchstens dann möglich, wenn diese auf einer Ebene (etwa „generalisiert“) und jene auf einer anderen Ebene (etwa „spezialisiert“) stattfinden. Die OP-Hypothese ist also mit der Theorie des dispositionellen Optimismus unvereinbar.

Es liegt darüber hinaus eine Art der Unvereinbarkeit zwischen der Theorie des dispositionellen Optimismus und der OP-Hypothese vor, die allerdings von keiner Seite explizit kommentiert wird: Offenbar bezeichnet „Optimismus“ für Carver und Scheier bereits vor der operationalen Definition, d.h. der Umsetzung der theoretischen Begriffe in Fragebogenitems, mittels derer Daten erhoben werden, etwas anderes als „Optimismus“ im Sinne der Vertreter der OP-Hypothese:

- Bei *Carver und Scheier* ist „Optimismus“ ein theoretischer Begriff, d.h. seine Bedeutung lässt sich, wenn auch nicht besonders präzise, anderen Wissenschaftlern verständlich machen ohne direkten Bezug auf eine spezifische operationale Definition. Das bedeutet auch, dass Ableitungen aus der Theorie des dispositionellen Optimismus zumindest möglich sind, die sich nicht auf Untersuchungen mit dem LOT(-R) beziehen.
- Bei den *Vertretern der OP-Hypothese* bezieht sich „Optimismus“ auf eine spezifische operationale Definition des Begriffes, und zwar *nicht* auf den LOT(-R) insgesamt, sondern nur auf faktorenanalytisch gewonnene Faktoren.

Die OP-Hypothese stellt also keine Modifikation der Theorie des dispositionellen Optimismus dar – tatsächlich ist die OP-Hypothese mit der Theorie des dispositionellen Optimismus *inkommensurabel*, d.h. die Begriffe der einen lassen sich nicht in Begriffe der anderen überführen, es gibt „kein gemeinsames Maß“. Die Begriffe „Optimismus“ und „Pessimismus“ erfahren durch die Vertreter der OP-Hypothese eine Umdeutung, wie in Abschnitt 2.1.3 gezeigt wurde. Die OP-Hypothese lässt sich wegen der Inkommensurabi-

lität vor dem Hintergrund der Theorie des dispositionellen Optimismus nicht weiter erörtern; ganz ähnlich wie vor dem Hintergrund des alltagssprachlichen Gebrauches ist die OP-Hypothese auch vor dem Hintergrund der Theorie des dispositionellen Optimismus begrifflich-logisch fragwürdig.

2.2.3 Weitere psychologische Modelle zum Optimismus

Es wurde gezeigt, dass die OP-Hypothese mit der Theorie des dispositionellen Optimismus inkommensurabel ist. Bevor aber der Status der OP-Hypothese hinsichtlich ihres theoretischen Gehaltes grundsätzlich beurteilt werden kann, sollte überprüft werden, inwieweit sie in andere psychologische Auffassungen von Optimismus eingeordnet werden kann. Neben dem dispositionellen Optimismus wurden von Peterson (2000) vor allem die Erfassung optimistischer und pessimistischer Attributionen sowie das Konstrukt Hoffnung als wesentliche psychologische Zugänge zur konzeptuellen Einordnung der OP-Hypothese genannt; darüber hinaus soll noch die Konzeption des „defensiven Pessimismus“ behandelt werden.

2.2.3.1 Optimistischer Attributionsstil

Auf welche Weise Personen die Ursachen von Ereignissen beschreiben, d.h. auf welche Ursachen sie Ereignisse attribuieren, stellt seit mehreren Jahrzehnten eine wesentliche Forschungsfrage der allgemeinen Psychologie und der Sozialpsychologie dar. Besonders interessant sind dabei Attributionen von Ereignissen, die die Person selbst betreffen: Solche Attributionen wurden vor allem im Zusammenhang mit Leistungsmotivation und mit Depressionen untersucht. Aufbauend auf Tierversuchen zur „gelernten Hilflosigkeit“ wurde postuliert, dass eine Attribution von Misserfolgen auf *internale* (d.h. im Subjekt liegende), *stabile* (d.h. relativ unveränderbare) und *globale* (d.h. allgemeine, sich auf mehrere als nur auf die spezifische Instanz beziehende) Ursachen motivations- und leistungsabträglich ist und in der schlimmsten Form zu Depressionen führt. Ähnlich nachteilig ist die Attribution von Erfolgen auf *externale*, *instabile* und *spezifische* Ursachen. Wenn aber der Attributionsstil umgekehrt ist, also etwa Erfolge auf die eigene Person (internal) und auf allgemeine Eigenschaften (stabil und global) attribuiert werden, so spricht dies für eine gute psychische Gesundheit (Gillham, Shatté, Reivich & Seligman, 2002). Der Attributionsstil wurde auf ganz ähnliche Weise gesundheitspsychologisch untersucht wie der dispositionelle Optimismus und es zeigten sich auch sehr ähnliche Ergebnisse: Ein *optimistischer Attributionsstil* (Gillham et al., 2002; Peterson & Bossio, 2002), nämlich die internale, stabile und globale Attribution für positive Ereignisse zusammen mit einer externen, instabilen und spezifischen Attribution von negativen Ereignissen führt genau wie ein hoher dispositioneller Optimismus zu einer besseren psychosozialen und gesundheitlichen Anpassung auch im Zusammenhang mit schweren Erkrankungen oder kritischen Lebensereignissen.

Zur Untersuchung des Attributionsstils mittels Fragebogenverfahren oder mittels der Analyse gesprochener oder geschriebener spezieller Ereigniserklärungen werden sowohl die Attributionen für positive Ereignisse als auch die Attributionen für negative Ereignisse erfasst. Die so gewonnenen Zahlenwerte werden im Allgemeinen zu einem Gesamtwert zusammengefasst (etwa: Je größer die Anzahl internaler Attributionen von positiven Ereignissen und externaler Attributionen von negativen Ereignissen, desto höher der Gesamtwert). Allerdings sprechen empirische Befunde dafür, dass der Attributionsstil für positive Ereignisse nicht immer komplementär zum Attributionsstil für negative Ereignisse ist (Peterson, 2000; vgl. Gillham et al., 2002). Könnte die OP-Hypothese also in den

Rahmen der Untersuchung des Attributionsstils eingeordnet werden? Gillham et al. (2002) bemerken, dass es erstaunlich wenige Studien zum Zusammenhang zwischen dispositionellem Optimismus und Attributionsstilen gibt; die direkten Korrelationen seien auch nur schwach bis moderat. Tatsächlich gibt es zumindest von der Theorie her auch einen deutlichen Unterschied zwischen beiden Konstrukten: Der Attributionsstil bezieht sich viel mehr auch auf aktives Handeln der Person und hat demzufolge eine engere Verwandtschaft zu Konstrukten wie Kontrollüberzeugungen und „agency“ (s. unten, „Hoffnung“). Darüber hinaus hat der Attributionsstil – selbst wenn dieser über Ereignisse hinweg relativ stabil ist – einen spezifischeren Charakter, und Attributionen für bestimmte Ereignisse können durchaus „optimistisch“ sein, auch wenn die grundsätzlichen Erwartungen „pessimistisch“ sind (Gillham et al., 2002). Die Theorie des Attributionsstils gibt immerhin Hinweise auf eine Unabhängigkeit der Attributionsstile für positive und negative Ereignisse; was das allerdings im Zusammenhang mit der OP-Hypothese bedeutet, ist noch eine offene Frage – schließlich ist der Attributionsstil ein vom dispositionellen Optimismus unterscheidbares Konstrukt.

2.2.3.2 Hoffnung

Nach Peterson (2000) verbindet das Konstrukt „Hoffnung“ zwei Aspekte, die in den Theorien zum dispositionellen Optimismus sowie zum optimistischen Attributionsstil in unterschiedlichem Maße eine Rolle spielen: den Erwartungsaspekt (dispositioneller Optimismus) und den Handlungsaspekt (Attributionsstil). Ähnlich wie beim dispositionellen Optimismus geht es um die Frage der Zielerreichung. Die Theorie zum Konstrukt Hoffnung (Snyder, Sympson, Michael & Cheavens, 2002) unterscheidet dabei zwischen zwei Aspekten: „agency“, eine Erwartung bezüglich des erfolgreichen Einsatzes eigener Anstrengungen, und „pathways thinking“, eine Erwartung bezüglich des erfolgreichen Auffindens von „Wegen zum Ziel“. Das Konstrukt Hoffnung hat eine deutliche Überschneidung mit dem Konstrukt der Selbstwirksamkeit (Scholz, Gutierrez Doña, Sud & Schwarzer, 2002), das sich ja ebenfalls auf Erfolgserwartungen bezüglich eigener Anstrengung bei der Zielerreichung bezieht. Hinsichtlich des Gehaltes der OP-Hypothese ergibt sich aus der Theorie zum Konstrukt Hoffnung nichts Wesentliches: Die Theorie beschäftigt sich nicht mit negativen Erwartungen und kann dementsprechend nicht zur Klärung beitragen, was „Pessimismus“ sein könnte.

2.2.3.3 Defensiver Pessimismus

Eine letzte psychologische Konzeption soll noch besprochen werden, da sie sich ausdrücklich auf Pessimismus bezieht: das Konstrukt des *defensiven Pessimismus* von Norem (2002). Dieses Konstrukt unterscheidet sich in zweierlei Hinsicht von den übrigen: 1. Es wird davon ausgegangen, dass Erwartungen strategisch und in gewisser Hinsicht aktiv eingesetzt werden, selbst wenn dieser Einsatz nicht notwendigerweise bewusst ist. 2. Darüber hinaus können nach der Konzeption des defensiven Pessimismus auch pessimistische Erwartungen förderlich sein, zumindest für bestimmte Individuen: Die Theorie basiert nämlich in gewisser Weise auf einer Interaktionshypothese, nach der es vom jeweiligen Angstbewältigungsstil abhängt, welche Art von Erwartungen förderlich sind. Demnach gibt es Personen, die pessimistische Erwartungen „defensiv“ einsetzen. Indem sie sich die schrecklichen Konsequenzen ausmalen, wenn ihre negativen Erwartungen in Erfüllung gehen würden, erhalten sie aus der resultierenden Angst genügend Energien, um die Erfüllung der negativen Erwartungen abzuwenden. Umgekehrt wehren sich *strategische Optimisten* gegen die Angst, die aus negativen Erwartungen resultieren würde, indem sie sich ausschließlich auf optimistische Erwartungen beschränken; unbehindert von

Angst und beflügelt durch positive Gefühle steigern sie auf diese Weise ihre Leistungsfähigkeit. Die Annahmen, die bezüglich der Förderlichkeit von strategischem Optimismus getroffen werden, sind im Übrigen ganz ähnlich denen beim dispositionellen Optimismus und bei den übrigen Konzeptionen: Strategische Optimisten verleugnen in gewisser Weise negative Informationen und erreichen auf diese Weise angemessene Leistungen und Durchhaltevermögen. Die Verwandtschaft zu Modellen der Angstbewältigung ist offensichtlich, da solche Modelle ganz analog von zwei unterschiedlichen und einander entgegenstehenden Bewältigungsstilen ausgehen.

E.C. Chang (2002b) betont, dass seine Version der OP-Hypothese unter anderem durch das Modell des defensiven Pessimismus inspiriert sei; nicht zuletzt vertritt er auch die Ansicht, dass Pessimismus nicht immer nur negative Konsequenzen habe. Dennoch ist festzuhalten, dass Norem (2002) explizit von einander *entgegengesetzten* Arten der strategischen Erwartung ausgeht: Für manche Individuen ist defensiver Pessimismus förderlich, für andere strategischer Optimismus. Ein gleichzeitiges Vorliegen von defensivem Pessimismus und strategischem Optimismus ist also auch nach dieser Konzeption nicht möglich.

2.3 Schlussfolgerungen der theoretischen Analyse

Insgesamt hat die Untersuchung der theoretischen Grundlagen zum Begriff „Optimismus“ gezeigt, dass der Status der OP-Hypothese aus begrifflich-logischer Sicht fragwürdig ist: Die Aussage, dass „Optimismus“ und „Pessimismus“ partiell unabhängige Konstrukte seien, ist weder mit dem historischen Hintergrund der beiden Begriffe noch mit verbreiteten und ausgearbeiteten psychologischen Theorien zum Thema vereinbar. Die Definition von „Optimismus“ als *generalisierter* positiver Zukunftserwartung und „Pessimismus“ als *generalisierter* negativer Zukunftserwartung (Chang et al., 1997) ist kein Ausweg, sondern macht die begriffliche Problematik nur noch deutlicher. Das Problem der unklaren Begrifflichkeit wäre für sich genommen allerdings nicht so gravierend, wenn es denn eine Theorie des „Pessimismus“ gäbe – doch gerade diese liefern die Vertreter der OP-Hypothese nicht.

3 Das Problem des reflexiven Messmodells bei Selbstberichtsfragebogen

Die theoretische Analyse in Kapitel 2 hat deutlich gemacht, dass die OP-Hypothese begrifflich-logisch fragwürdig ist und dabei auch zu gering theoretisch untermauert. In diesem Kapitel soll nun eine erste an der üblichen psychologischen Methodik im engeren Sinne orientierte Untersuchung der OP-Hypothese erfolgen. Dabei wird auf wesentliche Ergebnisse aus Schrift A zurückgegriffen und es werden sozial- und allgemeinpsychologische Ergebnisse zu Antwortprozessen bei Selbstberichtsfragebogen einbezogen.

Die OP-Hypothese basiert ausschließlich auf faktorenanalytischen Befunden zu Daten, die mittels des LOT oder LOT-R erhoben wurden. Da die OP-Hypothese also untrennbar sowohl mit der Erhebungsmethode (dem LOT(-R)-Fragebogen) als auch mit der Auswertungsmethode (Faktorenanalyse) verbunden ist, erscheint eine Untersuchung dieser methodischen Grundlagen unverzichtbar. Zunächst soll eine kurze allgemeine Analyse der Erhebungsmethode „Selbstberichtsfragebogen“ erfolgen, die für die anschließende im engeren Sinne testtheoretische Untersuchung grundlegend ist. Auf welche Weise die Zahlen gewonnen werden, die faktorenanalytisch ausgewertet werden, ist zentral für den Status der OP-Hypothese: Dass diese Zahlen tatsächlich die Bedeutung haben, die ihnen im Rahmen faktorenanalytischer Untersuchungen zugesprochen werden („der Optimismus-Faktor“), muss auf eine Weise begründet werden, die außerhalb des Geltungsbereiches der im engeren Sinne testtheoretischen Modelle liegt (Maraun, 1996). Die Frage der „Gültigkeit“ dieser Gleichsetzung von Faktoren mit psychologischen Konstrukten ist die Frage der *Konstruktvalidität* (vgl. Schrift A, Abschnitt 1.3.2) oder die Frage „Inwieweit lassen sich Belege und Begründungen dafür finden, dass die Testwerte auch so interpretiert werden können, wie es bei der Testkonstruktion und -anwendung intendiert war?“ (vgl. Messick, 1989). Zur Beantwortung dieser Frage sollte sicherlich auch geklärt werden, auf welche Art und Weise sich Konstruktausprägungen in Itembeantwortungen übertragen. In der Persönlichkeitspsychologie wird dieses Problem allerdings so gut wie nie behandelt, wie Borsboom, Mellenbergh und van Heerden (2003) pointiert feststellen:

„It is obvious that this problem [by which processes are construct scores translated into test scores] is even more serious in personality testing, in which one generally does not even have the faintest idea of what happens between item administration and item response. For this reason, it would be difficult to conceive of a meaningful interpretation of such an elliptical causal statement [‘construct scores cause test scores’] without rendering it completely vacuous, in the sense that the position on the latent variable is shorthand for whatever process leads to person’s response.” (S. 214)

Die Zusammenfassung einiger Ergebnisse sozialpsychologischer Forschung bezüglich der Antwortprozesse bei Selbstberichtsfragebogen soll aufzeigen, dass fraglich ist, ob sich Konstrukte mehr oder weniger direkt⁶ in Antworten auf Fragebogenitems widerspiegeln – zumindest dürfte aus der folgenden Darstellung klar werden, dass sich wahrscheinlich nicht nur die Ausprägung des eigentlich zu erfassenden Konstruktes systematisch in den Antworten auf Selbstberichtsfragebogen widerspiegelt.

⁶ „Mehr oder weniger direkt“: nur durch zufällige Messfehler beeinträchtigt.

3.1 Die unreflektierte „Reflexivitätsannahme“

Der LOT(-R) wurde konstruiert, um dispositionellen Optimismus als generalisierte positive Zukunftserwartung zu erfassen. Wie die überwiegende Mehrheit von Selbstberichtsfragebogen, die in der Persönlichkeitspsychologie eingesetzt werden, besteht auch der LOT(-R) aus mehreren Aussagen, und die Probanden sollen auf einer mehrfach gestuften Antwortskala angeben, inwieweit diese Aussagen auf sie zutreffen. Angesichts der großen Popularität von solchen Selbstberichtsfragebogen – nicht zuletzt basieren die wichtigsten Theorien zumindest der „mainstream“-Persönlichkeitspsychologie auf Faktorenanalysen von Fragebogendaten – ist es recht erstaunlich, wie wenig *methodologische* Studien sich mit Selbstberichtsfragebogen befassen. Eine der wenigen Ausnahmen stellt das Herausgeberwerk von Angleitner und Wiggins (1986) dar, in dem sich Forscher aus unterschiedlichen Perspektiven mit der Fragebogenmethode befassen. Fiske (1986) zeichnet in diesem Buch ein gewissermaßen sehr pessimistisches Bild der Nützlichkeit der Methode. Er spricht Fragebogen zwar eine Nützlichkeit für die praktische Anwendung etwa im Bereich der Individualdiagnostik zu, aber negiert die Möglichkeit, wissenschaftliche Erkenntnisse aus Fragebogendaten *für sich genommen* zu gewinnen, zumindest so lange keine befriedigende Theorie für die Antwortprozesse und Messoperationen vorliegt, aus denen Fragebogendaten entstehen. Gab es entsprechende Fortschritte in den letzten zwanzig Jahren, die zur Revision von Fiskes These hätten beitragen können? Im Folgenden wird die Fragebogenmethode grundsätzlich und am Beispiel der OP-Hypothese diskutiert.

Warum eine solche Diskussion sich auch unter das Stichwort „Testtheoretische Untersuchung“ einfügen lässt, wird deutlicher, wenn man sich vergegenwärtigt, dass fast alle psychometrischen Modelle, seien es Faktorenanalysen oder Modelle der Item-Response-Theorie, auf einem sog. *reflexiven Messmodell* beruhen (Edwards & Bagozzi, 2000). Dabei wird mehr oder weniger implizit davon ausgegangen, dass die Einschätzungen der eigenen Person anhand der Fragebogenitems praktisch unmittelbar und nur mit nicht weiter erklärbar Abweichungen – den zufälligen Messfehlern – die jeweilige Ausprägung auf dem latenten Konstrukt „reflektieren“, also widerspiegeln. Diese Modellvorstellung stammt wahrscheinlich noch daher, dass die Klassische Testtheorie und auch deren moderne Nachfolger, die latent trait-Modelle (vgl. Schrift A, Abschnitt 5), auf einer Analogie zu einfachen Messungen in der Physik basieren (Traub, 1997; s. auch das „Uhrenbeispiel“ aus Abschnitt 2.2.1 in Schrift A). Bei solchen Messungen gibt es eine klar definierte Größe, etwa den Abstand zweier Objekte im Teleskop eines Astronomen, und die Messung dieses Abstandes durch den Astronomen; Abstand und Messergebnis weichen aber selbst bei größtmöglicher Konstanzhaltung anderer Umstände (also gewissermaßen bei einer möglichst großen Objektivität) voneinander ab, und es wird angenommen, dass die Abweichungen als „zufällig“, nämlich durch geringe Einflüsse der *nicht* kontrollierbaren Umstände zustande kommen.

Leider sind Fragebogenmessungen aber keinesfalls ein so einfaches Beispiel einer „Reflexion“ einer klar definierten Größe auf möglichst präzise Messgeräte. Stattdessen umfasst der Prozess der Beantwortung eines Fragebogenitems mehrere Teilschritte, von denen jeder für sich mit Unklarheiten und selten geprüften Zusatzannahmen verbunden ist. So werden psychologische Konstrukte keinesfalls direkt in den Fragebogenitems abgebildet; stattdessen werden Konstrukte typischerweise indirekt erfasst, weil auf der einen Seite davon ausgegangen wird, dass Versuchspersonen wenn überhaupt, dann nur unzureichend verstehen, um welche Konstrukte es denn geht, auf der anderen Seite aber auf

Selbstberichte eben dieser Versuchspersonen nicht verzichtet⁷ werden kann. Die Konstrukte werden von psychologischen Forschern allerdings meist leider nur in recht allgemeiner und abstrakter Form theoretisch definiert (wie am Beispiel von Optimismus und erst recht der OP-Hypothese in Kapitel 2 gezeigt wurde) und dann weiterhin (wenn überhaupt) unpräzise und abstrakt operationalisiert, bis dann schließlich nach meist immer noch unklaren Richtlinien Items formuliert werden, von denen gehofft wird, dass sich die Konstruktausprägung indirekt in ihnen abbildet. Auch der eigentliche Antwortprozess läuft aber nicht so ab, dass die Messinstrumente angelegt und „abgelesen“ werden; die Probanden werden keinesfalls gebeten, direkte Einschätzungen abzugeben, sondern ihnen werden schwer verständliche, uneindeutige, abstrakte Aussagen (vgl. Angleitner, John & Löhr, 1986; Johnson, 2004) vorgelegt. Schließlich spitzen sich die Probleme der Fragebogenmethode noch häufig dadurch zu, dass den Probanden das eigentliche Ziel der Messung verheimlicht wird oder diese zumindest im Unklaren darüber gelassen werden⁸. Dass sich nach so einem vielschrittigen Prozess am Ende das Konstrukt doch mehr oder weniger deutlich in den Antworten abbildet, ist keinesfalls selbstverständlich: An jeder Stelle des Prozesses können systematische Verzerrungen auftreten. In der Tat erscheint gerade der Versuch, die Probanden über das Ziel der Messung im Unklaren zu lassen, als fundamentaler Fehler: die eigentliche Konstruktausprägung wird so mit den Vermutungen der Probanden darüber, was denn erfasst werden soll, konfundiert.

Dennoch wird das reflexive Messmodell in der Persönlichkeitspsychologie normalerweise unreflektiert übernommen – vielleicht auch deshalb, weil angesichts der *mathematisch-statistischen* Komplexität von konfirmatorischen Faktorenanalysen oder Modellen der Item-Response-Theorie oft übersehen wird, dass das *konzeptuelle* Modell nur eine sehr vereinfachte Annahme über den Antwortprozess darstellt. In gewisser Weise lässt sich sicherlich das Zitat von Freedman (1999) zum Problem der Kausalität in diesem Zusammenhang anwenden: „Investigators often try to use statistical models not only to control for confounding, but also to correct basic deficiencies in the design or the data.“ (S. 248). In der Sozial- und Allgemeinen Psychologie, insbesondere in der Einstellungsforschung, ist die Erforschung von Antwortprozessen allerdings ein breites und äußerst aktives Forschungsfeld mit deutlich komplexeren Annahmen als denen des reflexiven Messmodells in der Persönlichkeitspsychologie (vgl. etwa Schwarz, 1999; Tourangeau, Rips & Rasinski, 2000).

3.2 Die Prozedur der Fragebogenmethode

Die Fragebogenmethode lässt sich in die folgenden Teile zerlegen (vgl. Fiske, 1986; Angleitner et al., 1986):

1. Der Untersucher hat eine Theorie über ein Konstrukt, das er erfassen will (vgl. Schrift A, insbesondere Abschnitt 1.3.2, zum Begriff des Konstruktes).
2. Aus dieser Theorie über das Konstrukt wird abgeleitet⁹, welche Verhaltensweisen oder per Introspektion zugänglichen Affekte und Kognitionen indikativ für das Konstrukt sein sollten¹⁰.

⁷ Es soll an dieser Stelle nicht die Tatsache kritisiert werden, dass es sich um Selbstbeobachtungen anstatt um Fremdbeobachtungen handelt. Es ist klar, dass Selbstbeobachtungen für die Psychologie völlig unverzichtbar sind.

⁸ Der LOT(-R) etwa wird typischerweise mit vier „Füllitems“ vorgegeben „in order to disguise (somewhat) the underlying purpose of the test“ (Scheier & Carver, 1985, S. 224).

⁹ Tatsächlich gibt es noch andere Vorgehensweise, etwa die „externale Konstruktionsstrategie“, nach der aus einer beliebig gewonnenen Menge von Items diejenigen ausgewählt werden, die mög-

3. Möglichst einfach zu verstehende Aussagen, die Items, werden generiert, in denen Verhalten, Affekte oder Kognitionen beschrieben werden, und zwar so, dass das Zutreffen (oder Nicht-Zutreffen im Falle invertierter Items) der Aussagen impliziert, dass das Konstrukt beim Probanden vorliegt (nicht vorliegt) bzw. hoch (niedrig) ausgeprägt ist. Allerdings werden Fragebogenitems auch häufig von anderen Autoren übernommen oder modifiziert, anstatt dass sie neu geschrieben werden (Angleitner et al., 1986).
4. Die Items werden Probanden vorgelegt, die fast immer mittels einer abgestuften „Antwortskala“ mit mindestens zwei Stufen angeben sollen, inwieweit die Aussagen auf sie zutreffen.
5. Die Antworten werden zusammengefasst, um eine Schätzung der Ausprägung des Konstruktes beim jeweiligen Antwortenden zu gewinnen. Die Zusammenfassung erfolgt fast immer (außer bei einem Item-Response-Modell, vgl. Schrift C) nach der folgenden naiven Regel: „Jeder Antwortstufe wird eine Zahl zugeordnet. Addiere die Zahlen, die der jeweils gewählten Antwortstufe zugeordnet wurde.“
6. Wenn eine Zusammenfassung der Antworten aufgrund bestimmter statistischer Kriterien unangemessen erscheint, werden Items modifiziert oder entfernt und eventuell neue Items generiert.

3.2.1 Probleme der theoretischen Explikation und der operationalen Definition

Schritt 1 der Fragebogenmethode erfordert eine theoretische Explikation und Schritt 2 so etwas Ähnliches wie eine operationale Definition des zu erfassenden Konstruktes. In Schritt 2 werden die theoretischen Begriffe aus Schritt 1 in zumindest introspektiv Beobachtbares „übersetzt“, indem festgelegt wird, welche Beobachtungen mit bestimmten Ausprägungen des Konstruktes verbunden sein sollen (vgl. etwa Wilson, 2003). Allerdings wird dieser Schritt auch häufig verkürzt oder in gewisser Weise übersprungen, und von der Theorie (die zudem manchmal sehr knapp gefasst sein kann, wie in Kapitel 2 am Beispiel der OP-Hypothese deutlich wurde) wird direkt auf Schritt 3 übergeleitet. Pytlik Zillig et al. (2002) etwa zeigen für das „Big Five“-Modell der Persönlichkeit, dass die Items, die unterschiedliche Faktoren des Modells erfassen sollen, sich systematisch darin unterscheiden, ob Verhalten, Affekte oder Kognitionen beschrieben werden; so beziehen sich bspw. Neurotizismusitems fast ausschließlich auf Affekte, Gewissenhaftigkeitsitems dagegen fast ausschließlich auf Verhalten. Die „Theorie“ der Big Five, und zwar insbesondere in der Form, die eine genetische Basis der Big Five postuliert, kann diese Divergenz nicht erklären bzw. geht überhaupt nicht darauf ein: McCrae et al. (2000) etwa konzeptualisieren die Big Five als „Temperamente“ (so etwas ähnliches wie Dispositionen hinsichtlich affektiver Reaktionen), deren spezifische Ausformungen sich aufgrund kultureller Überformungen in Verhalten und Kognitionen äußern – dass Gewissenhaftigkeitsitems sich aber fast ausschließlich auf Verhalten beziehen, wirft angesichts der theoretischen Konzeption eines „Gewissenhaftigkeits*temperaments*“ Fragen auf.

lichst hoch mit einem „externen“ Kriterium korrelieren; in der Praxis der Persönlichkeitspsychologie ist aber wohl das hier skizzierte Vorgehen das am häufigsten vorkommende.

¹⁰ Angleitner, John und Löhr (1986) verwendeten ein stärker an der sprachlich-inhaltlichen Form orientiertes Kategoriensystem, um Items der damals populärsten Fragebogen zu untersuchen. Die Darstellung hier orientiert sich aber an der Arbeit von Pytlik Zillig, Hemenover und Dienstbier (2002), die die in der Psychologie weit verbreitete „ABC“-Kategorisierung in „Affects“, „Behavior“ und „Cognition“ verwendeten, um typische Fragebogen zu analysieren.

Überhaupt wird die Angemessenheit einer Theorie in der Persönlichkeitspsychologie und die Angemessenheit ihrer Umsetzung in Beobachtbares, also die Angemessenheit der Schritte 1 und 2 in der obigen Übersicht, nur sehr selten unabhängig von den Fragebogenitems geprüft (vgl. etwa Eysenck, 1997). Auch im Falle des dispositionellen Optimismus gibt es keine wirklich operationale Definition des Konstruktes, und die Items des LOT wurden nach kaum nachvollziehbaren Regeln erstellt („The first step in constructing the optimism scale was to generate a pool of items aimed at assessing generalized outcome expectancies.” Scheier & Carver, 1985, S. 224).

3.2.2 Probleme der Itemformulierung

Akzeptiert man in der Folge, wie es die meisten Autoren stillschweigend tun, dass die Theorieformulierung und Operationalisierung nur unbefriedigend gelöst werden, stellt sich dennoch die Frage, ob die sprachliche Umsetzung in Items angemessen ist (Schritt 3). Wiederum beispielhaft ist in diesem Zusammenhang die Studie von Angleitner et al. (1986), die zusammenfassend feststellten, dass es in keinem der von ihnen untersuchten Fragebogenverfahren gelungen sei, vier Kriterien für „gute“ Fragebogenitems als erfüllt zu beurteilen: Verständlichkeit, Eindeutigkeit, Konkretheit, Bezug auf den Beantwortenden. Dass sich die sprachliche Qualität von Fragebogenitems zwanzig Jahre später grundlegend verbessert haben sollte, ist nicht abzusehen. In der Tat erscheint etwa das LOT-R-Item „Auch in ungewissen Zeiten erwarte ich normalerweise das Beste“ keinesfalls einfach verständlich, eindeutig und konkret. Die Bewertung des Zutreffens dieser Aussage erfordert, dass der Beantwortende sich vorstellen kann, was „ungewisse Zeiten“ sind, es erfordert, dass alle potentiellen Beantwortenden den gleichen Referenzrahmen für „normalerweise“ haben usw. usf.

3.2.3 Probleme des Antwortprozesses

Während also bereits die ersten drei Schritte einer Fragebogenmethode im Allgemeinen und auch im besonderen Fall des LOT(-R) mit ungeprüften und z. T. auch nur schwer überprüfbareren Annahmen verbunden sind, ist die wohl problematischste Annahme mit dem Antwortprozess in Schritt 4 verbunden. Zur Verdeutlichung dieses Problems wird die formale Definition eines reflexiven Messmodells herangezogen. Die Reflexivitätsannahme lässt sich mathematisch-statistisch mittels der Annahme der *konditionalen Unabhängigkeit* formalisieren (vgl. Schrift A, Abschnitt 5). Konditionale Unabhängigkeit bedeutet, dass die Antwortverteilung bei Konstanzhaltung der latenten Variablen zufällig ist, d.h. gewissermaßen, dass nach Kontrolle der latenten Variablen keine Information mehr über die Individuen aus den Items entnommen werden kann. Im engeren Sinne besteht die konditionale Unabhängigkeitsannahme aus zwei Teilen: *experimentelle Unabhängigkeit* (vgl. Schrift A, Abschnitt 2.5) und *lokale Homogenität* (Schrift A, Abschnitt 5.2). Experimentelle Unabhängigkeit erfordert, dass die Antwortverteilungen zweier (oder mehrerer) Items für eine gegebene Person unabhängig sind, während lokale Homogenität erfordert, dass die Antwortverteilung eines Items für alle Personen mit der gleichen Ausprägung der latenten Variablen gleich ist.

Experimentelle Unabhängigkeit und lokale Homogenität sind vielleicht für die zu Beginn des Abschnitts zitierten Beispiele einfacher physikalischer Messungen sinnvolle Annahmen. Differenziertere Untersuchungen über die Antwortprozesse bei Selbstberichtsfragebogen insbesondere aus der sozialpsychologischen Forschung lassen das reflexive Messmodell allerdings als extreme Vereinfachung erscheinen. Mehrere Arbeiten stimmen darüber ein, dass sich der Antwortprozess bei Selbstbeurteilungsfragebogen in drei bis vier

Phasen einteilen lässt (Angleitner et al., 1986; Schwarz, 2007; Tourangeau et al., 2000): 1. Verstehen und Interpretieren des Items; 2. Vergleich der Aussage mit Informationen aus dem Gedächtnis; 3. Einpassen des Ergebnisses des Item-Selbst-Vergleiches auf die vorgegebenen Antwortoptionen; 4. Anpassung der Antwort in Bezug auf Selbstdarstellungstendenzen. Das heißt also, dass sich die Konstruktausprägung nicht unmittelbar in den Antworten widerspiegelt, sondern dass zuvor ein mehr oder weniger bewusster und aktiver Verarbeitungsprozess auf Seiten der Probanden abläuft. Die Annahme eines reflexiven Messmodells muss daher kritisch analysiert werden: Wenn etwa die Antworten auf ein Item dazu führen, dass ein folgendes Item anders verstanden wird (1. Schritt des Antwortprozesses), als wenn die Reihenfolge umgekehrt wäre, ist das eine Verletzung der experimentellen Unabhängigkeit. Ähnliches gilt etwa dann, wenn der Gedächtnisabruf, der im zweiten Schritt erforderlich ist, durch Antworten auf vorangegangene Items erleichtert wird. Auf der anderen Seite wäre die Annahme der lokalen Homogenität nicht mehr haltbar, wenn zwei Personen mit gleicher Ausprägung der latenten Variablen in unterschiedlicher Weise zu einer Antwortanpassung in Bezug auf Selbstdarstellungstendenzen neigen. Umgekehrt ist das reflexive Messmodell nur dann haltbar, wenn alle Personen das Item gleich verstehen und interpretieren, der Vergleich des Items mit Gedächtnisinhalten für alle Personen mehr oder weniger gleichmäßig funktioniert, alle Personen die Antwortskala gleichmäßig nutzen und Selbstdarstellungstendenzen entweder keine Rolle spielen oder aber bei allen Personen mehr oder weniger in gleicher Weise wirksam sind – dann und nur dann ist das Ergebnis des Item-Gedächtnisvergleiches die einzige systematische Varianzquelle.

Allerdings sprechen viele empirische Befunde dagegen, dass der Antwortprozess auf diese Weise vernachlässigt werden kann. Knowles und Kollegen (Knowles, 1988; Knowles, Coker, Scott, Cook, & Neville, 1996) zeigten etwa, dass sich Itemmittelwerte und -trennschärfen von Ängstlichkeitsitems in Abhängigkeit von der Reihenfolge ihrer Vorgabe änderten (vgl. auch Hartig, Hölzel und Moosbrugger, 2007, sowie Keith, Hodapp, Schermelleh-Engel und Moosbrugger, 2003) und damit eine deutliche Verletzung der Annahme der experimentellen Unabhängigkeit vorliegt. Kontexteffekte können damit erklärt werden, dass das Verstehen und Interpretieren von Fragebogenitems ein viel stärker an sozialer Interaktion angelegter Prozess ist als etwa das Lösen einer Mathematikaufgabe. Knowles und Condon (2000) führen dazu die Unterscheidung zwischen dem „say system“ von Wortschatz und Grammatik auf der einen Seite und dem „implicature system“ auf der anderen Seite ein; letzteres dient dazu, Hintergrundinformationen zur Verfügung zu stellen, um den vom „say system“ verarbeiteten Sätzen Bedeutung zu geben. Beispielsweise erfordert das LOT-R-Item „Ich zähle selten darauf, dass mir etwas Gutes widerfährt“ zunächst eine oberflächliche Verarbeitung („Worauf zähle ich? Dass mir etwas Gutes widerfährt. Darauf zähle ich *selten*.“). Eine solche Aussage ist zunächst hinsichtlich des „say systems“ schwierig, wie durch die Hervorhebung in der Klammer deutlich gemacht werden sollte – das Attribut „selten“ stellt eine *psychologische Negation* dar, die die Verarbeitung des Items erschwert (Helfrich, 1986). Das Item stellt aber auch eine Herausforderung an das Funktionieren des „implicature systems“: Untersucher und Versuchsperson müssen – ohne direkt in Kontakt zu treten – in etwa das Gleiche meinen, wenn von „etwas Gutes“ die Rede ist und in etwa die gleiche Vorstellung haben, wie oft „selten“ ist.

Das „implicature system“ ist wahrscheinlich ungewollt beteiligt, wenn vorhergehende Items einen Einfluss auf folgende Items haben: So zeigte Knowles (1988), dass die Korrelation eines Items mit dem Gesamttest, also die sog. Trennschärfe, ansteigt, je mehr andere Items zum selben Konstrukt zuvor beantwortet wurden. Wahrscheinlich erschlie-

ßen die Probanden aus den zuvor beantworteten Items, welches Konstrukt der Untersucher erfassen will, und können sich daher besser hinsichtlich dieses Konstruktes einschätzen. Auch Schwarz und Kollegen haben in vielen Untersuchungen festgestellt, dass Versuchspersonen die Untersuchungssituation nutzen, um Hinweise zum Verständnis eines Items zu suchen, oder auch mehr oder weniger unbewusst ihre Antworten an den Untersuchungskontext anpassen (Schwarz, 1999); Norenzayan und Schwarz (1999) etwa stellen fest, dass Probanden unterschiedliche Erklärungen für das Phänomen von Serienmördern abliefern, je nachdem, ob sich der Forscher als „Persönlichkeitsforscher“ oder als „Sozialpsychologe“ ausgab. Kontexteffekte innerhalb eines Fragebogens führen auf jeden Fall zur Verletzung der experimentellen Unabhängigkeit. Da die Items eines Fragebogens aber fast immer (auch bei Untersuchungen mit dem LOT(-R)) in einer festen Reihenfolge vorgegeben werden, ist ihr Effekt mit den absoluten Itemeffekten (Schwierigkeit, Trennschärfe) konfundiert und nicht feststellbar.

Unter gewissen Umständen können Itemkontexteffekte und die daraus resultierende Verletzung der experimentellen Unabhängigkeit aber doch beobachtet werden. Knäuper (1999; Knäuper, Schwarz, Park & Fritsch, im Druck) zeigte, dass Effekte der Itemreihenfolge mit dem Alter und der Kapazität des Arbeitsgedächtnisses abnehmen. Gerade weil sich Personen mit weniger leistungsfähigem Arbeitsgedächtnis weniger an vorherige Fragen erinnern, *verringert* sich die Konsistenz ihrer Antworten auf zusammengehörige Fragen. Dieses Ergebnis ist sehr interessant besonders vor dem Hintergrund, dass Herzberg et al. (2006) behaupten, dass die Korrelation von „Optimismus“ und „Pessimismus“ mit dem Alter abnehme – die mögliche Alternativerklärung von alters- oder gedächtnisabhängigen Antworteffekten wird von den Autoren nicht in Betracht gezogen. Im übrigen ist diese Alternativerklärung sensu Knäuper et al. (im Druck), dass die im Alter verringerte Gedächtniskapazität zu weniger konsistenten Antworten und damit auch zu einer Verringerung der Korrelationen der positiv und invertiert formulierten Items führt, auch vereinbar mit dem Modell der *Akquieszenz* oder „Ja-Sage-Tendenz“ von Knowles und Condon (1999). Diese beiden Forscher zeigten, dass bestätigende Antworten („Ja“-Antworten) typischerweise schneller erfolgen als verneinende (vgl. auch Chesca & Hollemann, 2007) und insbesondere, dass die Anzahl von bestätigenden Antworten ansteigt, wenn die Probanden gleichzeitig eine Aufgabe bearbeiten müssen, die die „cognitive load“ ansteigen lässt und also die Arbeitsgedächtniskapazität verringert.

Es soll hier nochmals betont werden, dass die Verletzung der experimentellen Unabhängigkeitsannahme unter den üblichen Vorgabebedingungen für Selbstberichtsfragebogen mit fester Itemreihenfolge nicht erkannt werden kann, außer dann, wenn weitere Kovariaten einbezogen werden wie etwa bei Knäuper et al. (im Druck) das Alter oder die Arbeitsgedächtniskapazität. Feststellbar ist nur die Kovarianz von Itembeantwortungen; diese Kovarianz wird im reflexiven Messmodell darauf zurückgeführt, dass den Items eine gemeinsame „verursachende“ latente Variable zugrunde liegt. Die Kovarianz könnte aber auch einfach nur dadurch entstehen, dass die Versuchspersonen glauben, dass alle Items in gleicher Richtung beantwortet werden sollen, ohne das tatsächlich ein „echtes“ Konstrukt hinter den Antworten steht. Ein solcher Fall würde der üblichen Terminologie entsprechend als ein Modell korrelierter Fehler dargestellt werden müssen (vgl. Schrift A, Abschnitt 6.1), ohne dass entschieden werden kann, ob das eine oder das andere das „wahre“ Modell ist (s. Kapitel 4 und Schrift A, Abschnitt 4.3).

Während Kontexteffekte vor allem einen Einfluss auf die Phasen des Verstehens und Interpretierens sowie die Phase des Item-Selbst-Vergleichs haben, gibt es auch Befunde dafür, dass die mehrfach gestuften Antwortskalen keineswegs von allen Personen gleich-

mäßig genutzt werden. Ein einfach zu verstehendes Beispiel für solche Effekte liefert die Untersuchung von Rost, Carstensen und von Davier (1999). Diese zeigten, dass die fünf-fach gestufte Antwortskala eines Big Five-Fragebogens von zwei latenten Personenklassen unterschiedlich genutzt wurde, insofern als die eine Klasse eher zu extremen Antworten an den äußeren Enden der Skala tendierte, die andere Klasse aber eine „Tendenz zur Mitte“ aufwies, also die äußeren Enden der Skala vermied. Sofern solche differentiellen Effekte der Antwortskalennutzung nicht in ein Modell miteinbezogen werden, führen sie zu einer eindeutigen Verletzung der lokalen Homogenitätsannahme.

3.3 Fazit: Probleme der Fragebogenmethode

Die OP-Methode basiert auf der Beobachtung, dass sich die Kovarianzen der LOT(-R)-Itembeantwortungen nicht durch einen einzigen latenten Faktor repräsentieren lassen. Als Erklärung nehmen die Vertreter der OP-Hypothese an, dass sich zwei Konstrukte, nämlich „Optimismus“ und „Pessimismus“ in den Antworten widerspiegeln. Die Analyse der Fragebogenmethode hat allerdings aufgezeigt, dass die Gültigkeit eines solchen einfachen reflexiven Messmodells für Selbstberichtsfragebogen zumindest fraglich ist: Itemkontexteffekte und Annahmen der Probanden über die Untersuchungssituationen können einen deutlichen Einfluss auf die Itemkovarianzen haben. Vautier, Mullet und Bourdet-Loubère (2003) konnten so auch zeigen, dass die Korrelationen positiv und invertiert formulierter Ängstlichkeitsitems anstiegen, wenn die Probanden per Instruktion darauf verwiesen wurden, dass die Items ein eindimensionales Konstrukt erfassen sollen.

Die in diesem Kapitel vorgestellte Kritik an der Fragebogenmethode lenkt die Aufmerksamkeit auf die Schwäche eines in der persönlichkeitspsychologischen Forschung zentralen methodischen Ansatzes. Diese Methode ist der fast ausschließliche Weg zum Erkenntnisgewinn in der Mainstream-Persönlichkeitspsychologie. Von namhaften Forschern des Feldes werden weitgehende Hypothesen wie diejenige, dass Persönlichkeit im Wesentlichen genetisch determiniert und über die Lebensspanne hinweg äußerst stabil sei, mit Ergebnissen aus Fragebogenuntersuchungen begründet (McCrae et al., 2000). Wie die Geschichte nicht nur anderer Wissenschaften zeigt, kann es allerdings zuweilen sehr fruchtbar sein, eine alterprobtete Untersuchungsmethode zugunsten einer neuen aufzugeben. Ein prominentes Beispiel aus der Fachgeschichte der Psychologie ist der Perspektivenwechsel vom Behaviorismus zur kognitiven Psychologie: Die meisten heutigen Forscher würden wohl darin übereinstimmen, dass die Methoden der kognitiven Psychologie Erkenntnis brachten, die mit den Methoden der Behavioristen nicht möglich gewesen waren. In diesem Sinne könnte auch die Persönlichkeitspsychologie profitieren, wenn sie ihr wesentliches Instrument, den Selbstberichtsfragebogen, in Zweifel zöge, genauso wie es die experimentelle Psychologie mit dem behavioristischen Dogma des Verzichtes auf die Untersuchung von Kognitionen und Emotionen getan hat. Immerhin mehren sich auch in der Persönlichkeitspsychologie Befunde dafür, dass Selbstberichtsfragebogen keinesfalls unter allen Umständen den einzig möglichen Zugang zu interindividuellen Unterschieden darstellen. Die implizite Persönlichkeitsmessung etwa rückt immer mehr in den Vordergrund des Interesses und eignet sich zudem für experimentelle Studien; so konnte gezeigt werden, dass implizite Maße insbesondere bei eingeschränkter Motivation oder unter kognitiver Belastung eine bessere Verhaltensvorhersage ermöglichen als Selbstberichtsfragebogen (Asendorpf, Banse & Mücke, 2002; Dovidio, Kawakami & Gaertner, 2002; Hofmann, Rauch & Gawronski, 2007).

Im spezifischen Zusammenhang der vorliegenden Arbeit ist jedenfalls kaum zu bezweifeln, dass die OP-Hypothese einen besonders gravierenden Fall von holzschnittartigen

Vereinfachungen in der Persönlichkeitspsychologie darstellt: Korrelative, querschnittliche Fragebogenuntersuchungen werden herangezogen, um Aussagen über Konstrukte aufzustellen – es wird angenommen, dass sich ein Konstrukt namens „Pessimismus“ in den Kovarianzen der drei (!) invertierten LOT(-R)-Items widerspiegelt, ohne dass es theoretische Begründungen für „Pessimismus“ gibt (s. Kapitel 2) und ohne Effekte der Untersuchungsmethode Selbstberichtsfragebogen überhaupt in Betracht zu ziehen. Solche Nachlässigkeit leistet sich die moderne faktorenanalytische Persönlichkeitspsychologie im Allgemeinen nicht. Schließlich sind das Ausschalten oder die Kontrolle von alternativen Erklärungen definierende Voraussetzungen für wissenschaftliches Arbeiten: Wer Alternativerklärungen oder Zweifel an seinen Ergebnissen nicht berücksichtigt, wird zuweilen gar als „Pseudowissenschaftler“ beschimpft (vgl. Still & Dryden, 2004).

Im folgenden Kapitel soll allerdings im Rahmen einer Analyse faktorenanalytischer Untersuchungen unter anderem deutlich gemacht werden, dass ein „confirmation bias“ jedoch nicht nur äußerst menschlich ist, sondern im spezifischen Zusammenhang durch manche Methodiker ungewollt gefördert wird, wenn nämlich von „konfirmatorischen“ Faktorenanalysen die Rede ist.

4 Zur Frage der „Konfirmation“ bei „konfirmatorischen“ Faktorenanalysen

Zur Einleitung dieses Kapitels sei wiederholt, dass die OP-Hypothese auf dem Befund beruht, dass zur Repräsentation der Kovarianzen von Antworten auf LOT(-R)-Items mindestens zwei Faktoren erforderlich sind. Zunächst ist dieser Befund unabhängig davon, ob „explorative“ oder „konfirmatorische“ Faktorenanalysen (Moosbrugger & Schermelleh-Engel, 2006) verwendet werden. Scheier und Carver (1985) etwa berichten bereits für den ursprünglichen LOT die Ergebnisse sowohl einer „explorativen“ Hauptachsenanalyse, bei der nach dem „Eigenwerte-größer-1-Kriterium“ zwei Faktoren extrahiert wurden, als auch die Ergebnisse *mehrerer* „konfirmatorischer“ Faktorenanalysen, nämlich einer einfaktorischen Lösung mit korrelierten Messfehlervariablen und einer zweifaktorischen Lösung. Auch im Artikel über die Revision des LOT (Scheier et al., 1994) führten die Autoren sowohl „explorative“ als auch „konfirmatorische“ Faktorenanalysen durch und kamen mehr oder weniger zu den selben Ergebnissen wie in Scheier und Carver (1985). Marshall et al. (1992), deren Artikel meist als Ausgangspunkt der OP-Hypothese zitiert wird, nutzten zwar ebenfalls zunächst „explorative“ und anschließend „konfirmatorische“ Faktorenanalysen, sie *verzichteten* aber genau wie Herzberg et al. (2006) auf eine Betrachtung des einfaktorischen Modells mit korrelierten Fehlern, das Scheier et al. (1994) vorgeschlagen hatten. Die OP-Hypothese basiert zum Teil auch auf dem folgenden Argument: Das zweifaktorische Modell mit einem „Optimismus“-Faktor für die positiven Items und einem „Pessimismus“-Faktor für die invertierten Items müsse bei einer konfirmatorischen Faktorenanalyse nicht zurückgewiesen werden, das einfaktorische Modell aber schon. Der Gehalt dieses Arguments soll in diesem Kapitel aus methodologischer Sicht analysiert werden.

4.1 Konfirmation vs. Falsifikation

4.1.1 Das Induktionsproblem

„Die Falsifikation von Hypothesen ist der Königsweg zur Erkenntnis in den empirischen Wissenschaften“. Eine solche oder ähnliche Aussage würden wohl die meisten Psychologen akzeptieren. Die wissenschaftstheoretischen Grundlagen der gängigen empirischen und statistischen Methodologie werden in typischen Lehrbüchern (z.B. Bortz & Döring, 1995 oder Sarris & Reiß, 2005) nur recht kurz und unter fast ausschließlichem Bezug auf die wissenschaftstheoretischen Arbeiten von Sir Karl Popper behandelt. Für die vorliegende Arbeit ist es aber hilfreich, kurz darzulegen, warum die Falsifikation von Hypothesen erstrebt wird, aber nicht deren Bestätigung:

- Wissenschaftliche Aussagen, zumindest in den naturwissenschaftlichen Disziplinen, sind fast immer unbegrenzte *Allsätze*, d.h. eine wissenschaftliche Aussage soll sich nicht nur auf eine spezifische Beobachtungsmenge beziehen, sondern unbegrenzt auch für vergangene, gegenwärtige und zukünftige (potentielle) Beobachtungen der selben Art gültig sein.
- *Beobachtungsmengen* sind aber immer (notwendigerweise) begrenzt.
- Daher gibt es keine Möglichkeit zur Bestätigung (*Konfirmation*) einer wissenschaftlichen Aussage: Eine potentielle, bisher nicht beobachtete Untersuchungseinheit könnte der Aussage widersprechen.

Ein kurzes Beispiel sei angeführt: Das Hauskaninchen möge aus seinen bisherigen Beobachtungen das „Gesetz“ erschlossen haben, dass es jeden Tag gefüttert und auch sonst durch den Menschen gut versorgt wird – bis es eines Tages zum Schlachter kommt (ein ähnliches Beispiel wird von Ladyman, 2002, Bertrand Russell zugeschrieben). Dieses *Problem der Induktion*, das sich hinter der Unmöglichkeit der „Konfirmation“ von Hypothesen verbirgt, lässt sich aber auch in einem stärker naturwissenschaftlichen Kontext illustrieren. In der Physik etwa galt die Newtonsche Gravitationslehre lange Zeit als „wahr“, bis sie durch die Einsteinsche Theorie ersetzt wurde – tatsächlich *konnte* die Newtonsche Lehre aber lange Zeit nicht empirisch widerlegt werden, weil die Klasse von Beobachtungseinheiten, für die sie nicht gilt, nämlich für sehr hohe relative Geschwindigkeiten sowie für extrem kleine und extrem große Objekte, früher grundsätzlich nicht beobachtbar war (vgl. Ladyman, 2002). Die Bestätigung von Hypothesen aufgrund induktiver Schlüsse ist also aufgrund der Begrenztheit der jeweiligen Beobachtungsmenge unmöglich.

Das Induktionsproblem ist so umfassend, dass man zunächst versucht sein mag, es aufgrund seiner Allgemeinheit als irrelevant für die wissenschaftliche Praxis anzusehen; so kann man sich etwa das folgende Argument vorstellen: „Trotz des Induktionsproblems hat die Wissenschaft doch schon so viel Nützliches und Wichtiges herausgefunden“. Ein Aspekt des Induktionsproblems ist aber zentral für die wissenschaftliche Arbeit: das Problem der möglichen Alternativerklärungen. Daraus nämlich, dass eine Theorie in der Lage war, gewisse Beobachtungen vorherzusagen, folgt noch lange nicht, dass die Theorie „richtig“ ist, da es – wie etwa im Beispiel der Gravitationslehren – andere Theorien geben mag, die sehr viel vorhersagemächtiger sind. Die Folge aus dem Induktionsproblem ist also: Theorien oder Hypothesen (als kleine Bausteine von Theorien) können unter keinen Umständen *bestätigt* werden, sofern sie auch nur irgendeine Generalisierbarkeit über bereits erfolgte Beobachtungen hinaus beanspruchen. Stattdessen können Hypothesen höchstens widerlegt oder, um den in der Literatur üblicheren Begriff zu verwenden, *falsifiziert* werden, wenn sich zeigt, dass Vorhersagen aus den Hypothesen sich empirisch als falsch erweisen.

Aufgrund des Induktionsproblems in der skizzierten Struktur ist es die gängige Norm in den empirischen Wissenschaften, Hypothesen dem Risiko der Falsifikation auszusetzen (Meehl, 1978). Es wird geradezu als notwendiges Kriterium der Wissenschaftlichkeit angesehen, dass eine Theorie oder zumindest aus einer Theorie abgeleitete Hypothesen falsifizierbar sein müssen, und umgekehrt: Eine Theorie, die *nicht* falsifizierbar ist, ist aus dem Kanon der empirischen Wissenschaften auszuschließen (vgl. Still & Dryden, 2004). Ein berühmtes Beispiel: Popper verneinte der Psychoanalyse den Status als Wissenschaft, da sie in der Lage sei, jegliche Beobachtung menschlichen Verhaltens zu erklären, aber eben nicht zu Vorhersagen komme, die potentiell falsifizierbar seien (vgl. Ladyman, 2002).

4.1.2 Logisches Schließen nach Modus tollens

Da Hypothesen also aufgrund des Induktionsproblems niemals bestätigt werden können, sondern ausschließlich falsifiziert, stellt sich die Frage, wie solche Falsifikationsentscheidungen ablaufen. Das Prinzip solcher Entscheidungen ist der *Modus tollens*: Zu überprüfen ist der Wahrheitswert (wahr/falsch) eines Konditionalsatzes der Form: „Wenn Theorie T zutrifft, dann folgt Beobachtung O “, oder formalisiert: $T \rightarrow O$. Das Ergebnis einer empirischen Untersuchung ist dann entweder O oder „Nicht- O “ (formal: $\neg O$). Wie durch Einsetzen eines einfachen Beispiels deutlich wird, ist nur die Beobachtung $\neg O$

aussagekräftig hinsichtlich des Status von T : Der Konditionalsatz sei: „Wenn jemand schläft, hat er die Augen geschlossen“. Die Beobachtung sei: „Hans hat die Augen geschlossen.“ Aus Hansens geschlossenen Augen, also aus O , resultiert keine Information darüber, ob er schläft oder nicht. Nur dann, wenn Hansens Augen geöffnet wären, könnte man folgern, dass er *nicht* schläft, also dass die Theorie T nicht zutrifft, also dass $\neg T$.

Allerdings fällt den meisten Menschen die Form des logischen Schließens nach Modus tollens aber normalerweise sehr schwer, und zwar sowohl im Alltag als auch in der Wissenschaft. Es ist ein häufig repliziertes Ergebnis kognitionspsychologischer Untersuchungen, dass Menschen einem „confirmation bias“ unterliegen (Nickerson, 1998), nämlich der Tendenz, eher nach bestätigenden als nach negierenden Informationen bezüglich einer geäußerten Hypothese zu suchen. Der confirmation bias wird häufig mit der „2-4-6“-Aufgabe untersucht; diese Zahlenfolge wird vorgegeben und die Aufgabe der Versuchsperson ist es, Hypothesen aufzustellen und zu überprüfen, nach welcher Regel die Folge konstruiert wurde. Dabei gelingt es nur wenigen Personen, tatsächlich die Regel herauszufinden, die von den Forschern benutzt wurde; anstatt Beispiele zu überprüfen, die bestimmte Regeln ausschließen (also falsifizieren) würden, generieren die meisten Menschen nur solche Beispiele, die nicht zur Falsifikation geeignet sind. Wie aber das Beispiel Hansens geschlossener Augen demonstriert hat, ist eine bestätigende Beobachtung keineswegs informativ bezüglich des Status der Hypothese. In der Tat kann aber der Fehlschluss, aus der Bestätigung der Konsequenz die Bestätigung des Antezedens abzulesen, auch als eine Instanz des Induktionsproblems betrachtet werden.

4.1.3 Falsifikation bei probabilistischen Hypothesen

Psychologische Hypothesen sind im Allgemeinen probabilistisch, d.h. sie folgen der Form „Wenn A, dann *wahrscheinlich* B“. Die probabilistische anstatt deterministische Natur schützt aber selbstverständlich nicht vor dem Induktionsproblem (das wäre ja auch gar nicht im Sinne der allermeisten psychologischen Forscher): Eine probabilistische Hypothese ist nicht mehr haltbar, wenn zu viele Gegenbeispiele auftreten. Um eine probabilistische Hypothese falsifizierbar zu machen, bedarf es entsprechend einer Quantifizierung, ab wann sie nicht mehr haltbar ist.

Nach dem gängigen Vorgehen (dem „Null-Ritual“, Gigerenzer, 2004) in der Inferenzstatistik wird eine statistische „Nullhypothese“ (H_0) formuliert und die Wahrscheinlichkeit einer zusammenfassenden Aussage über die Daten O_D unter Gültigkeit der H_0 geschätzt, also die *konditionale Wahrscheinlichkeit* $P(O_D | H_0)$ der Daten gegeben die H_0 . Diese Wahrscheinlichkeit wird nun im Rahmen einer Entscheidungsregel verwendet: Wenn $P(O_D | H_0)$ einen per Konvention festgelegten Wert, nämlich meist 5%, unterschreitet, wird die H_0 zurückgewiesen, wenn nicht, wird sie beibehalten. Am Beispiel der Hypothese „Alle Schwäne sind weiß“ würde die Beobachtung von fünf schwarzen Exemplaren in einer Stichprobe von hundert Schwänen nach dem gängigen Wert für ein akzeptables „Irrtumsrisiko“ nicht zu einer Zurückweisung der H_0 führen, dass Schwäne weiß sind – ein schwarzer Schwan mehr aber würde zur Zurückweisung führen.

4.2 Protektive Auxiliarannahmen

Während die gerade noch akzeptierte Irrtumswahrscheinlichkeit $\max P(O_D | H_0)$ im Beispiel der Schwanhypothese eine einfache Quantifizierung dessen ist, wie stark inhaltliche Hypothesen durch einzelne Beobachtungen („schwarze Schwäne“) verletzt werden dürfen, um dennoch beibehalten zu werden, ist der typische Fall der Anwendung der entscheidenden Statistik ein anderer. Fast alle Verfahren der Entscheidungsstatistik, die in der Psychologie zur Anwendung kommen, basieren auf mehr oder weniger komplexen mathematisch-statistischen Modellen, z.B. dem Allgemeinen Linearen Modell (Moosbrugger & Rauch, im Druck) oder einem faktorenanalytischen Modell (vgl. Schrift A). Wie bereits in dem Begriff „eine zusammenfassende Aussage über die Daten O_D “ angedeutet, basieren statistische Entscheidungen im Allgemeinen nämlich nicht auf der Menge aller Daten, sondern auf *Statistiken* im engeren Sinne, d.h. auf mathematischen Funktionen, mit denen den erhobenen Daten eine zusammenfassende Kennziffer zugewiesen wird wie z.B. der Mittelwert oder aber der Anteil der schwarzen Schwäne (Ostmann & Wutke, 1994). O_D wird im Folgenden als allgemeiner Ausdruck für eine zur Hypothesenprüfung verwendete Statistik verwendet. Mathematisch-statistische Modelle dienen zur Übersetzung inhaltlicher Aussagen in Aussagen über Statistiken. Diese Übersetzung bedingt allerdings die Erweiterung des oben skizzierten Entscheidungsprinzips zur Falsifikation von Hypothesen: Es wird nämlich nun nicht mehr direkt die Hypothese T getestet, sondern es ist die Zusatz- oder Auxiliarannahme A_M erforderlich, dass T angemessen in das mathematische Modell übersetzt wurde, mittels dessen die Beobachtungen O_D geprüft werden sollen. Geprüft wird also nicht alleine T , sondern die Konjunktion von T und A_M .

Gleichzeitig bedingt die mathematische Modellierung aber auch eine bestimmte Art von zusammenfassenden Aussagen: Beschäftigt man sich etwa mit „explorativen“ Faktorenanalysen, so steht meist die Frage der Anzahl der zu extrahierenden Faktoren und der Rotation im Vordergrund, analysiert man aber „konfirmatorische“ Faktorenmodelle, so ist der „Modellfit“ (s. unten) das zentrale Kriterium (z.B. Moosbrugger & Schermelleh-Engel, 2006). Je nach Modell ist also eine andere Art von zusammenfassenden Aussagen O_D im Vordergrund; daher wird wie folgt notiert: $(T \wedge A_M) \rightarrow M \supset O_D$, d.h. die Konjunktion von T und A_M impliziert die mit Modell M verbundenen Beobachtungen O_D .

Wie etwa Meehl (1990) darlegt, schützt nun aber gerade die Konjunktivität von T und A_M die Theorie T in gewisser Hinsicht vor der Falsifikation: Die Beobachtung $\neg O_D$ führt nämlich zu $\neg T \vee \neg A_M$, d.h. es kann nur geschlossen werden, dass *entweder T oder A_M oder beide falsch sind*, aber nicht, welche der drei Möglichkeiten falsch ist¹¹. Wissenschaftliche Theorien sind fast immer mit einem solchen „protektiven Gürtel“ von Zusatzannahmen geschützt und daher im engeren Sinne so gut wie nie falsifizierbar, wie in der Wissenschaftstheorie unter dem Stichwort „Quine-Duhem-These“ bekannt ist (vgl. z.B. Michell, 2004).

¹¹ Meehl (1990) führt aus, dass typischerweise eine ganze Reihe von Auxiliarannahmen zur Prüfung einer Theorie erforderlich sind – das hier vorgestellte Schema ist eine deutliche Vereinfachung.

Für den Modelltest wird das Minimum der sog. Diskrepanzfunktion

$$F_{ML}(\boldsymbol{\vartheta}) = \ln |\boldsymbol{\Sigma}(\boldsymbol{\vartheta})| - \ln |\mathbf{S}| + \text{tr} [\mathbf{S}\boldsymbol{\Sigma}(\boldsymbol{\vartheta})^{-1}] - p$$

genutzt; das Minimum dieser Funktion, also $\hat{F}_{ML} = \min F_{ML}(\mathbf{J})$, wobei p = Anzahl der Zeilen/Spalten von $\boldsymbol{\Sigma}$, ist die zusammenfassende Aussage O_D . Die Entscheidungsregel lautet (zunächst) wie folgt:

Multipliziert man \hat{F}_{ML} mit $(N-1)$, wobei N = Anzahl der Beobachtungen, so resultiert eine χ^2 -verteilte Größe mit $p-s$ Freiheitsgraden mit s als Anzahl der Parameter $\boldsymbol{\vartheta}$. Die Wahrscheinlichkeit $P(O_D | H_0)$ entspricht 1 – derjenigen Fläche der χ^2 -Dichteverteilung mit $p-s$ Freiheitsgraden, die zwischen 0 und $(N-1)\hat{F}_{ML}$ liegt. Wenn $P(O_D | H_0)$ kleiner als ein bestimmter Betrag ist, wird das Modell zurückgewiesen.

Dass ein „konfirmatorisches“ Faktormodell so bezeichnet wird, basiert genau auf dieser Entscheidungsregel (Jöreskog, 1969). Der Begriff „konfirmatorisch“ ist übrigens eher untypisch für die Anwendung von statistischen Entscheidungen, etwa bei der Prüfung von Unterschiedshypothesen in Varianzanalysen oder bei der Prüfung von Regressionsmodellen. Ganz im Gegenteil legen viele der entsprechenden Lehrbücher darauf Wert, dass solche statistischen Entscheidungen höchstens zu einer Falsifikation führen können, nicht aber zu einer „Konfirmation“. Die Verwendung des sonst so unüblichen Begriffs der „Konfirmation“ im Zusammenhang mit Faktorenanalysen kann natürlich damit begründet werden, dass damit ein Kontrast zu sog. „explorativen“ Faktorenanalysen hergestellt wird.

Eine Übersicht über relevante Lehrbücher und Fachveröffentlichungen ergibt allerdings sehr schnell, dass typischerweise überhaupt kein Modelltest im engeren Sinne durchgeführt wird. Simulationsstudien (z.B. Curran, West & Finch, 1996) zeigen, dass der χ^2 -Test sehr sensitiv gegenüber Verletzungen der Voraussetzungen und zwar insbesondere derjenigen der multivariaten Normalverteilung ist: Bei nicht-normalverteilten Daten führt der χ^2 -Test meist auch dann zu einer Zurückweisung, wenn das Modell „eigentlich“ korrekt ist. Auch die Sensitivität gegenüber der Stichprobengröße wird als Manko des χ^2 -Test angesehen; dieses Argument erscheint allerdings nicht ganz stichhaltig, da fast alle Signifikanztests sensitiver („teststärker“) werden, wenn die Stichprobe erhöht wird.

Aufgrund der zu hohen Sensitivität des χ^2 -Tests empfehlen fast alle relevanten Veröffentlichungen die Evaluation eines „konfirmatorischen“ Faktormodells auf der Grundlage sog. *deskriptiver Fit-Indizes* (Schermelleh-Engel, Moosbrugger & Müller, 2003). Eine ganze Reihe solcher Fit-Indizes wurde entwickelt (für eine Übersicht s. Schermelleh-Engel et al., 2003), und mehrere Studien geben Empfehlungen über die Verwendung von Fit-Indizes zur Bewertung eines Modells ab (z.B. Hu & Bentler, 1999). Die Frage ist nun, wie es um den „konfirmatorischen“ Charakter steht, wenn eine solche „Evaluation“ auf der Grundlage von Fit-Indizes geschieht. In der Tat argumentieren Marsh, Hau und Wen (2004), dass die Verwendung von Fit-Indizes anstelle des χ^2 -Test zu einer logischen Inkon-

sistenz führt, da Fit-Indizes zuweilen im Sinne einer echten Entscheidungsregel verwendet werden („Wenn Index X kleiner als ein bestimmter Wert, weise das Modell zurück“), zum anderen aber nur im Sinne von „Daumenregeln“ verwendet werden („Wenn Index X größer als ein bestimmter Wert, ist das Modell schon in Ordnung“). Darüber hinaus zeigt Yuan (2005), dass auch die am häufigsten empfohlenen Fit-Indizes gegenüber solchen Eigenschaften der Untersuchung sensitiv sind, gegenüber denen sie es eigentlich nicht sein sollten, wenn die Probleme des χ^2 -Tests vermieden werden sollen. In der Praxis dominiert daher wohl fast immer ein nicht mehr wirklich „konfirmatorisch“ zu nennendes Vorgehen, in dessen Rahmen relativ unspezifische und uneindeutige Entscheidungsregeln angewendet werden, die von unterschiedlichen wissenschaftlichen Kreisen unterschiedlich gehandhabt und akzeptiert werden (vgl. z.B. Barrett, 2007). Von einem Test mit einer eindeutigen Entscheidungsregel kann also keine Rede sein.

Grundsätzlich wird durch die uneindeutige Handhabung von Fit-Indizes und damit verbundenen, meist idiosynkratischen Entscheidungsregeln das Attribut „konfirmatorisch“ *ad absurdum* geführt: Das einzige, was eine solche Bezeichnung halbwegs rechtfertigen würde, nämlich die Möglichkeit der *Falsifikation*, d.h. der Zurückweisung eines *a priori* festgelegten Modells, kommt in der Praxis so gut wie nie vor. Wenn ein Forscher sein Modell verteidigen will, kann er entweder versuchen, aus der großen Menge von Fit-Indizes einen zu finden, der sein Modell nicht zurückweist, oder aber er folgt der bereits von Jöreskog (1969) vorgeschlagenen Strategie, sein Modell so lange zu modifizieren, bis es nach irgendwelchen Kriterien angemessen erscheint. Was unterscheidet nun eine derart praktizierte „konfirmatorische“ noch von einer „explorativen“ Faktorenanalyse?

In welcher Weise wird die „konfirmatorische“ Faktorenanalyse im Zusammenhang mit der OP-Hypothese eingesetzt? Bereits Scheier und Carver (1985) stellten sowohl in „explorativen“ als auch „konfirmatorischen“ Faktorenanalysen fest, dass ein zweifaktorielles Modell „besser“ ist als ein einfaktorielles. In „explorativen“ Faktorenanalysen erschien die Extraktion eines zweiten Faktors als nach gängigen Kriterien erforderlich, und „konfirmatorische“ Faktorenanalysen beruhten darauf, dass „preliminary examination revealed ... a higher degree of shared disturbance among positively phrased items than among negatively phrased items“ (S. 226). Die Autoren vermuteten eine Konfundierung mit Akquieszenz, die diese höheren Kovarianzen unter den positiven Items verursachen könnte (vgl. auch die Diskussion von „Testlet-Faktoren“ in Schrift A, Abschnitte 4.2 und 6.3). Sie modellierten anschließend sowohl ein einfaktorielles Modell mit korrelierten Fehlern als auch ein Modell mit zwei Faktoren für die positiv bzw. invertiert formulierten Items. In der weiter oben eingeführten Notation könnte diese Annahme eines „Akquieszenzeffektes“ als eine zusätzliche, „schützende“ Auxiliarannahme A_A für die „Theorie“ T_o betrachtet werden, dass die LOT(-R)-Items Optimismus als eindimensionales Konstrukt erfassen: Die Konjunktion $(T_o \wedge A_M \wedge A_A)$ muss nicht mehr als falsch angenommen werden, während sie ohne A_A nicht beibehalten werden könnte.

Die Vertreter der OP-Hypothese dagegen berücksichtigen nur die Tatsache, dass ein einfaktorielles Modell nicht angemessen zu den Daten passt, die Auxiliarannahme A_A eines Akquieszenzeffektes dagegen berücksichtigen sie nicht. Sie schlagen stattdessen die alternative OP-Hypothese T_{op} vor und argumentieren, $(T_{op} \wedge A_M)$ führe zu einer besseren Anpassung als $(T_o \wedge A_M)$. Die Vertreter der OP-Hypothese streben also nicht eine Falsifikation der ursprünglichen „Optimismus-Hypothese“ mit Akquieszenzeffekt an, sondern sie stützen ihre Hypothese auf die Falsifikation einer irrelevanten Hypothese.

Der wiederholte Vergleich der OP-Hypothese mit dem einfaktoriellen Modell testet nichts, da das einfaktorielle Modell ja schon von Scheier und Carver (1985) selbst verworfen worden war. Erforderlich wäre stattdessen der Versuch einer Falsifikation von $(T_o \wedge A_M \wedge A_A)$.

Instruktiv ist schließlich noch die Frage, wie die OP-Hypothese eigentlich entstand. Im Rahmen von „explorativen“ Faktorenanalysen wird typischerweise in einem ersten Schritt die Anzahl der zu extrahierenden Faktoren festgelegt; anschließend wird eine Rotations- transformation durchgeführt, die die Interpretation der Faktoren erleichtern soll (vgl. Moosbrugger & Schermelleh-Engel, 2006). Praktisch alle gängigen Rotationskriterien basieren dabei auf dem Prinzip der *Einfachstruktur*. Eine Einfachstruktur liegt dann vor, wenn eine Teilmenge der Variablen auf einem Faktor hohe Ladungen aufweist und auf den übrigen Faktoren Ladungen nahe Null, und jeder Faktor durch solche Teilmengen hoch ladender Variablen gekennzeichnet ist. Einfachstrukturen sind besonders leicht interpretierbar, weil nur herausgefunden werden muss, was der jeweiligen Teilmenge von Variablen „gemeinsam“ ist. Wahrscheinlich ist es dieses Kriterium der Einfachstruktur, das die OP-Hypothese überhaupt erst populär gemacht hat: Konfrontiert mit der Tatsache, dass ein einfaktorielles Modell die Kovarianzen der LOT(-R)-Items nicht angemessen repräsentiert, erschien den Vertretern der OP-Hypothese wahrscheinlich ein Modell mit zwei zur Einfachstruktur rotierten Faktoren am angemessensten, weil Einfachstrukturen in der Psychologie so häufig angenommen werden – es gibt aber keinesfalls ein Naturgesetz, dass alle Faktorenlösungen dem Prinzip der Einfachstruktur folgen müssen. Auf der Suche, was den jeweils auf den zwei Faktoren ladenden Items gemeinsam ist, wurde festgestellt, dass es die positiv formulierten Items auf der einen, die invertiert formulierten Items auf der anderen Seite waren; bereits Scheier und Carver (1985) berichteten von einer solchen Lösung. Dieses explorative Ergebnis wurde nun von den Anhängern der OP-Hypothese als *a priori* formulierte „Theorie“ ausgegeben. Selbstverständlich ist es nicht verwerflich, Hypothesen durch Exploration zu gewinnen, um sie anschließend einem Test zu unterwerfen; ein solches Vorgehen ist wohl fast die Regel dafür, wie neue wissenschaftliche Erkenntnisse gewonnen werden (vgl. Musgrave, 1989). In der Wiederholung der konfirmatorischen Faktorenanalyse mit „Optimismus“- und „Pessimismus“- Faktoren und dem vorhersehbaren Ergebnis, dass dieses Modell besser passt als ein einfaktorielles Modell, zeigt sich auf der anderen Seite aber genau das Problem, das mit der Frage verbunden ist: „what experiment could *disprove* your hypothesis?“ (Platt, 1964, S. 352): Die Vertreter der OP-Hypothese setzen ihre Hypothese *keinem* Test aus, der diese falsifizieren könnte. Die Überprüfung eines „konfirmatorischen“ Faktorenmodells, das aufgrund einer „explorativen“ Faktorenanalyse erstellt wurde, ist ganz sicher keine „bold conjecture“ im Popperschen Sinne – das Risiko, dass ein solches Modell falsifiziert wird, ist keinesfalls besonders hoch. Leider ist damit auch der Erkenntnisgewinn, den die „Konfirmation“ mit sich bringt, sehr niedrig (Meehl, 1978). Darüber hinaus führt die Beschränkung auf das willkürliche Modell einer Einfachstruktur zur Vernachlässigung von Alternativmodellen.

4.4 Zur Unterdeterminiertheit von faktorenanalytischen Ergebnissen

Im vorangegangenen Abschnitt wurde das mit der OP-Hypothese verbundene konfirmatorische Faktorenmodell als Matrixfunktion $\Sigma(\boldsymbol{\theta})$ dargestellt mit der Annahme, dass $\Sigma = \Sigma(\boldsymbol{\theta})$, also dass die Kovarianzmatrix der LOT(-R)-Items durch eine Funktion der Modellparameter $\boldsymbol{\theta}$ dargestellt werden kann. Neben dem bisher festgestellten Problem,

dass die Falsifikation oder „Konfirmation“ einer solchen Hypothese im Rahmen „konfirmatorischer“ Faktorenanalysen typischerweise gar nicht so verläuft wie es der Falsifikationsansatz normalerweise anstrebt, soll nun gezeigt werden, dass der Status von $\Sigma = \Sigma(\Theta)$ selbst dann uneindeutig ist, wenn das Modell $\Sigma(\Theta)$ auch nach konventionellen Kriterien beibehalten werden könnte. Bei der Prüfung von faktorenanalytischen Modellen liegt nämlich eine spezifische Instanz des Induktionsproblems vor¹²: Zu einer gegebenen Kovarianzmatrix Σ gibt es prinzipiell unendlich viele Modelle, die dieselbe Anpassung im Sinne einer bestimmten Diskrepanzfunktion wie z.B. der oben eingeführten Funktion $F_{ML}(\Theta)$ erreichen (Raykov & Penev, 1999; Tomarken & Waller, 2003). Für „konfirmatorische“ Faktorenanalysen kann die Existenz solcher *äquivalenter Modelle* einfach durch die Einführung einer beliebigen Transformationsmatrix \mathbf{R} gezeigt werden (vgl. Schrift A, Abschnitt 4.3.1.1): Wenn angenommen wird, dass $\Sigma(\vartheta) = \Lambda\Phi\Lambda'$, wobei Λ die Ladungsmatrix, Φ die Korrelationsmatrix der Faktoren und Θ die Fehlerkovarianzmatrix ist, entsteht eine identische Lösung $\Sigma(\vartheta) = \Sigma^*(\vartheta^*) = \Lambda^*\Phi^*\Lambda'^* + \Theta$ mit $\Phi^* = \mathbf{R}\Phi\mathbf{R}'$ und $\Lambda^* = \Lambda\mathbf{R}^{-1}$ $\Lambda^* = \Lambda\mathbf{R}^{-1}$.

Am Beispiel der OP-Hypothese könnte eine solche Transformation in ein äquivalentes Modell etwa wie folgt aussehen: Es sei

$$\mathbf{R} = \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}$$

so dass

$$\Lambda^* = \Lambda\mathbf{R}^{-1} = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ \lambda_3 & 0 \\ 0 & \lambda_4 \\ 0 & \lambda_5 \\ 0 & \lambda_6 \end{bmatrix} \begin{bmatrix} .5 & 0 \\ .5 & 1 \end{bmatrix} = \begin{bmatrix} .5\lambda_1 & 0 \\ .5\lambda_2 & 0 \\ .5\lambda_3 & 0 \\ .5\lambda_4 & \lambda_4 \\ .5\lambda_5 & \lambda_5 \\ .5\lambda_6 & \lambda_6 \end{bmatrix} = \begin{bmatrix} \lambda_1^* & \\ \lambda_2^* & \\ \lambda_3^* & \\ .5\lambda_4^* & \lambda_4^* \\ .5\lambda_5^* & \lambda_5^* \\ .5\lambda_6^* & \lambda_6^* \end{bmatrix}$$

und

$$\Phi^* = \mathbf{R}\Phi\mathbf{R}' = \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \phi_{11} & \\ & \phi_{22} \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4\phi_{11} & \\ -2\phi_{11} + 2\phi_{12} & \phi_{11} + \phi_{22} \end{bmatrix} = \begin{bmatrix} \phi_{11}^* & \\ \phi_{12}^* & \phi_{22}^* \end{bmatrix}$$

D.h. die willkürlich gewählte Transformationsmatrix \mathbf{R} bewirkt, dass das ursprüngliche Modell mit zwei korrelierten Faktoren und Einfachstruktur in ein Modell transformiert wird, in dem ein Teil der Items auf beiden (weiterhin korrelierten) Faktoren lädt. Selbstverständlich könnten auch beliebige andere Transformationsmatrizen gewählt werden oder aber eine andere Modellstruktur; Raykov und Penev (1999) und Schrift A, Abschnitt 4.3.1.1 demonstrieren, wie Modelle mit mehreren Faktoren in Modelle mit korrelierten Fehlern transformiert werden können.

¹² Tatsächlich besteht nicht nur die hier diskutierte Problematik äquivalenter Modelle, sondern für ein gegebenes Modell noch die zusätzliche Unterdeterminiertheit von Faktorwerten (vgl. Maraun, 1996, und die darauf folgende Diskussion in der Zeitschrift *Multivariate Behavioral Research*).

Ein Beobachter wird vielleicht zunächst anmerken, dass eine Transformation wie die hier dargestellte doch höchst künstlich sei. Tatsächlich aber stellt sich die Frage, ob nicht vielmehr die Annahme einer Einfachstruktur wie im ursprünglichen Modell künstlich und unrealistisch ist. In Kapitel 3 wurden bereits einige Argumente dafür aufgeführt, dass es eher unwahrscheinlich erscheint, dass Antworten auf Fragebogenitems ausschließlich durch ein einziges dahinterliegendes Konstrukt erklärt werden. Die hier beispielhaft dargestellte Transformation ähnelt von ihrer Struktur her einem Methodeneffektmodell, wie es in Schrift B untersucht wurde – das transformierte Modell enthält einen Faktor, auf dem alle Items laden, und einen Faktor, auf dem nur ein Teil der Items lädt. Solche Strukturen sind spätestens seit Holzinger und Swineford (1937) ein akzeptierter Bestandteil des Arsenal an Vorlagen für die Struktur von Faktormodellen. Es sei nochmals betont, dass das Modell $\Sigma^*(\vartheta^*)$ zu genau derselben Modellanpassung führt wie das Modell $\Sigma(\vartheta)$, d.h. falls ein Forscher *a priori* die Hypothese aufgestellt hätte, dass $\Sigma^*(\vartheta^*) = \Sigma$, wäre er zu genau demselben Ergebnis gekommen wie die Vertreter der OP-Hypothese ohne eine Möglichkeit der Unterscheidung beider Lösungen.

4.5 Schlussfolgerungen

Zusammengenommen zeigen die Betrachtungen in diesem Kapitel, dass der methodologische Status der „konfirmatorischen“ Faktorenanalysen im Zusammenhang mit der OP-Hypothese bestenfalls schwach ist. Zunächst wurde gezeigt, dass eine „Konfirmation“ von Hypothesen grundsätzlich nicht möglich ist. Der Begriff „konfirmatorische“ Faktorenanalyse dient hauptsächlich zur Abgrenzung gegenüber „explorativen“ Faktorenanalysen. Hinsichtlich des Status der mit den beiden unterschiedlichen Zugängen gewonnenen Erkenntnisse lässt sich allerdings kein besonders großer Unterschied feststellen. Tatsächlich leiden „konfirmatorische“ Faktorenanalysen in der Praxis häufig sogar unter einer dialektischen Verkehrung des Methodeninventars eines falsifikationistischen Ansatzes:

Da der χ^2 -Test zu streng ist, wird im Allgemeinen auf präzise Entscheidungsregeln verzichtet. Die Entscheidungen auf der Basis von Fit-Indizes, die in der Praxis vorkommen, und dabei häufig noch mit *post hoc* durchgeführten Modellmodifikationen gepaart werden, haben eine fatale Ähnlichkeit mit dem Explorieren unterschiedlicher Lösungen in der „explorativen“ Faktorenanalyse.

In Abschnitt 4.4 wurde schließlich gezeigt, dass Faktorenlösungen auch dann uneindeutig bleiben, wenn man das Problem der Uneindeutigkeit der Entscheidungen über die Falsifikation eines Faktormodells ignoriert – einfache Transformationsregeln führen zu Modellen mit identischen Fit, aber jeweils inhaltlich unterschiedlichen, dabei potentiell sinnvollen Interpretationen. Aus der Sicht der Vertreter der OP-Hypothese gibt es fatalerweise auch keine außerhalb des faktorenanalytischen Ansatzes liegende Begründung dafür, warum ihr Modell $\Sigma(\vartheta)$ besser wäre als ein Modell $\Sigma^*(\vartheta^*)$. Grundsätzlich können Hypothesen nicht „konfirmiert“ werden, so dass methodologisch meist gefordert wird, Alternativerklärungen für eine gegebene Theorie zu prüfen. Genau darauf verzichten aber die Vertreter der OP-Hypothese. Im Speziellen gibt es unendlich viele alternative Modelle, die genauso gut passen wie das Modell der OP-Hypothese; auch diese werden von den Vertretern der OP-Hypothese nicht geprüft. Im folgenden Kapitel wird nun ein „konfirmatorisches“ Faktormodell vorgestellt, das nicht direkt äquivalent, sondern „approximativ äquivalent“ (Schrift A, Exkurs 2) zum Modell der OP-Hypothese ist; das „Methodeneffektmodell“ zeigt sogar eine bessere Anpassung als das Modell der OP-Hypothese.

5 Alternativen zur OP-Hypothese: Methodeneffekte im Zusammenhang mit Sozialer Erwünschtheit

In Kapitel 2 wurden unterschiedliche Konzeptionen von Optimismus vorgestellt, und es wurde gefolgert, dass die OP-Hypothese im Zusammenhang mit dispositionalem Optimismus weder theoretisch begründet ist noch sich vor dem Hintergrund der anderen Konzeptionen hinreichend begründen lässt. Angesichts dieser Tatsache und angesichts der statistischen und inhaltlichen Unterdeterminiertheit von Faktorenmodellen, die in den Kapiteln 3 und 4 dargestellt wurden, ist es erstaunlich, dass nur sehr selten alternative Faktorenmodelle bzw. theoretische Erklärungen für die Bidimensionalität von LOT(-R)-Daten entwickelt wurden. Besonders erstaunlich ist es deshalb, weil das Phänomen von *Richtungsfaktoren*, d.h. einer Bidimensionalität, die sich durch eine Einfachstruktur von zwei Faktoren für positiv formulierte Items auf der einen und invertierte Items auf der anderen Seite abbilden lässt, keinesfalls auf den LOT-R beschränkt ist (s. Schrift B). Instruktiv ist das Beispiel der „Rosenberg Self-Esteem Scale“ (RSES; Rosenberg, 1965), die hinsichtlich ihrer Dimensionalität wohl noch öfter untersucht wurde als der LOT(-R) (für eine Übersicht s. Marsh, 1996). Die Richtungsfaktoren im Falle der RSES, die also genauso offensichtlich mit der Itemformulierung zusammenhängen wie beim LOT(-R), werden nämlich nur äußerst selten auf so etwas wie „positive Selbstbewertung“ oder „negative Selbstbewertung“ zurückgeführt, sondern die meisten Forscher gehen von „Methodeneffekten“ aus (Marsh, 1996; vgl. Schrift A zum allgemeinen Begriff des Methodeneffektes).

Im Speziellen bedeutet das in Schrift B vorgeschlagene Methodeneffektmodell, dass die Items des LOT-R, wie von den Autoren Scheier und Carver (1985; Scheier et al., 1994) intendiert, alle einen gemeinsamen Anteil haben, d.h. alle auf einem allgemeinen Faktor laden, der mit „Optimismus“ bezeichnet wird. Angesichts der in den vorangegangenen Kapiteln vorgebrachten grundsätzlichen Kritik an faktorenanalytischen Ergebnissen kann die Interpretation dieses Faktors als „Optimismus“ natürlich nur vorläufig sein, aber sie ist sicherlich die beste angesichts der Messintention. In Anlehnung an die Hypothesen von Scheier und Carver (1985) sowie Vautier, Raufaste und Cariou (2003) wird zusätzlich zum allgemeinen Faktor ein zweiter Faktor eingeführt, nämlich der Methodeneffektfaktor, der orthogonal zum ersten ist und auf dem nur die positiven Items laden¹³. Dieser Faktor bündelt die Varianz, die den positiv formulierten Items noch gemeinsam ist, nachdem der allgemeine Faktor herauspartialisiert wurde; es wird angenommen, dass diese gemeinsame Varianz aufgrund der sprachlichen Unterschiedlichkeit der beiden Itemgruppen entsteht und *nicht*, weil ein weiteres Persönlichkeitsmerkmal im engeren Sinne erfasst wird. Im Speziellen wird vermutet, dass sich dieser Methodeneffektfaktor mit dem *response set* der sozialen Erwünschtheit in Verbindung bringen lässt.

In Schrift B wird eine eigene empirische Untersuchung vorgestellt, in deren Rahmen überprüft wurde, inwieweit sich das vorgestellte Methodeneffektmodell, d.h. ein Modell mit einem allgemeinen Faktor für alle Items und einem dazu orthogonalen „Methodeneffektfaktor“ mit Ladungen ausschließlich der positiven Items an LOT-R-Daten anpassen lässt. Das wesentlichste Ergebnis ist das folgende: Das Methodeneffektmodell in Schrift B erreichte eine bessere Anpassung als das „Optimismus-Pessimismus“-Modell der OP-Hypothese. Informationen zur Durchführung und Auswertung der Studie können Schrift B entnommen werden. An dieser Stelle sollen vor allem zwei Aspekte der Studie hervor-

¹³ Zumindest grundsätzlich wäre auch ein zweiter Methodeneffektfaktor für die invertierten Items denkbar. Solche Modelle sind aber fast immer empirisch unteridentifiziert (vgl. Eid, 2000; Eid, Lischetzke, Nussbeck & Trierweiler, 2003).

gehoben werden: 1. die Erklärung des Phänomens von Richtungsfaktoren mit response sets wie *sozialer Erwünschtheit* und 2. das Prinzip der *Theoretischen Sparsamkeit*.

5.1 Response sets als Erklärung für Richtungsfaktoren

In Schrift B wird angenommen, dass die sprachliche Unterschiedlichkeit der Itemgruppen (positiv vs. invertiert formulierte Items) unterschiedlich mit *response sets* der Probanden interagiert. Response sets sind systematische Effekte bei der Itembeantwortung, die unabhängig vom jeweils zu erfassenden Konstrukt auftreten und damit einen unerwünschten Einfluss auf die Itembeantwortung ausüben (Paulhus, 2002). Solche Effekte sind in der Persönlichkeitspsychologie schon lange bekannt, insbesondere das Problem der Antwortverzerrung durch *soziale Erwünschtheit* wird seit Jahrzehnten diskutiert. In den Rahmen des in Abschnitt 3.2.3 vorgestellten Modells des Antwortprozesses bei Selbstberichtsfragebogen lässt sich eine solche Antwortverzerrung folgendermaßen einordnen: Nachdem eine Person ein Item interpretiert und ihre eigene Position hinsichtlich des Items verortet hat, folgt noch eine letzte Überprüfung, ob die Antwort auch mit sozialen Konventionen hinsichtlich dessen, was „angemessenes“ oder „erwünschtes“ Verhalten ist, übereinstimmt; wenn nicht, wird die manifeste Antwort gegenüber der eigentlich richtigen Antwort in Richtung des erwünschten Verhaltens abgeändert. Wenn eine Person also glaubt, dass eine Aussage wie „Meine Zukunft sehe ich immer optimistisch“ sozial erwünscht ist, wird ihre manifeste Antwort unter Umständen positiver ausfallen als die Antwort, die aus dem Item-Selbst-Vergleich resultiert. Eine ähnliche Annahme wurde bereits von Vautier, Raufaste und Cariou (2003) untersucht, die herausfanden, dass ein Methodenfaktor für LOT-R Items positiv mit den sozial erwünschten „Füllitems“ (z.B. „Es fällt mir leicht, zu entspannen“) korreliert.

In Schrift B wurde diese Annahme überprüft, indem ein etabliertes Instrument zur Erfassung von sozialer Erwünschtheit herangezogen wurde. Das „Balanced Inventory of Desirable Responding“ (BIDR; Musch, Brockhaus & Bröder, 2002) erfasst zwei Komponenten sozial erwünschter Antwortverzerrung, nämlich „selbsttäuschende Übertreibung“ (self-deceptive enhancement) oder kurz *Selbsttäuschung* und die „Tendenz zur Eindruckssteuerung“ (impression management) oder kurz *Fremdtäuschung*. Im engeren Sinne stellt dabei eigentlich nur die Fremdtäuschung ein „echtes“ response set dar. Die Theorie zum BIDR (vgl. Paulhus, 2002) geht davon aus, dass Selbsttäuschung eine unbewusste Abwehrtendenz darstellt. Personen mit starker Neigung zur Selbsttäuschung vermeiden es, „schlechte“ Eigenschaften vor sich selbst zuzugeben, so dass sie manche Eigenschaften bei sich selbst nicht erkennen, und streben eine positive Selbstbewertung an, so dass sie sich selbst erwünschte Eigenschaften in einem höheren Maße zuschreiben, als dies ein neutraler Beobachter tun würde. Beide Mechanismen sind *unabhängig* von der Beurteilung durch Dritte – entsprechend finden sich auch keine wesentlichen Unterschiede hinsichtlich der Werte in der Selbsttäuschungsskala, wenn die Motivation zu einer positiven Selbstpräsentation in sog. „faking“ Studien experimentell erhöht wird (Musch et al., 2002). Im Unterschied dazu zeigen sich deutliche Effekte einer solchen erhöhten Motivation hinsichtlich der Skala zur Fremdtäuschung: Es ist diese Tendenz zu einer bewussten Manipulation des Eindrucks, der auf andere vermittelt wird, die im eigentlichen Sinne dem letzten Schritt des in Abschnitt 3.2.3 beschriebenen Antwortprozesses bei Fragebogenitems entspricht.

Die Ergebnisse aus Schrift B zeigen, dass der Methodeneffektfaktor, der den positiv formulierten Items des LOT-R gemeinsam ist, tatsächlich substantiell ($r = .35$) mit Fremdtäuschung korreliert, aber vernachlässigbar mit Selbsttäuschung ($r = -.09$). Dieses Er-

gebnis stärkt die Annahme, dass die positiv formulierten Items in höherem Maße zu einer sozial erwünschten Antwortverzerrung einladen als die invertiert formulierten Items, wie es bereits Vautier et al. (2003) vermutet hatten; die als unbewusst konzipierte Selbsttäuschung dagegen trägt nicht zum Methodeneffekt bei. Eine mögliche Erklärung dafür ist die leichtere Verständlichkeit der positiven Items (Scheier & Carver, 1985): Bei diesen Items fällt es leichter, eine sozial erwünschte Selbstpräsentation herzustellen als bei den invertierten Items, hinsichtlich derer der sozial erwünschte Status schwieriger herauszufinden ist. Dass überhaupt eine Motivation zu sozial erwünschter Selbstpräsentation im Zusammenhang mit Optimismus besteht, legen unterschiedliche Ergebnisse nahe: Taylor, Lerner, Sherman, Sage und McDowell (2003) zeigen, dass Optimismus und andere Komponenten eines Maßes zur „seelischen Gesundheit“ positiv mit sozialen Aktivitäten und sozialer Aktivität korrelieren, und in ganz ähnlicher Hinsicht berichten pessimistische Krebspatientinnen stärkere Einschränkungen ihrer sozialen Aktivitäten nach einer Operation als optimistische (Carver, Lehmann & Antoni, 2003).

5.2 Das Prinzip der Theoretischen Sparsamkeit

In den vorangegangenen Kapiteln, insbesondere in Kapitel 2, wurde herausgestellt, dass die OP-Hypothese unter einer mangelnden theoretischen Begründung leidet: Was genau „Pessimismus“ ist, wird so gut wie gar nicht erklärt. Dennoch behandeln die Vertreter der OP-Hypothese „Pessimismus“ genauso wie etabliertere Konstrukte, indem sie etwa den Zusammenhang von „Pessimismus“ mit negativer Affektivität/Neurotizismus untersuchen (Marshall et al., 1992; Robinson-Whelen et al., 1997; Chang et al., 1997; Kubzansky et al., 2004). Das Methodeneffektmodell aus Schrift B dagegen basiert auf einer ausgearbeiteten Theorie, der Theorie des dispositionellen Optimismus von Scheier und Carver (1985). Das Methodeneffektmodell liefert damit eine aus theoretischer Perspektive *spar-samere* Erklärung: Anstelle eines neuen Persönlichkeitsmerkmals im engeren Sinne wird angenommen, dass die lokale Abhängigkeit der positiv formulierten Items nach Herauspartialisieren des gemeinsamen Faktors durch Antwortverzerrungen, sog. *response sets* erklärt werden kann.

Die Sparsamkeit des Methodeneffektmodells lässt sich wie folgt illustrieren: Wenn die Vertreter der OP-Hypothese recht hätten, müsste das ohnehin bereits von Hunderten von Konstrukten bevölkerte „nomologische Netz“ (vgl. Campbell & Fiske, 1959) der Persönlichkeitspsychologie einen weiteren Bewohner aufnehmen. Welcher theoretische Status sollte einem Konstrukt „Pessimismus“ zukommen? Viele Persönlichkeitsforscher, sehr prominent etwa die Vertreter der Big Five-Persönlichkeitstheorie (McCrae et al., 2000), vertreten eine explizit realistische Position hinsichtlich des ontologischen Status von Persönlichkeitseigenschaften (vgl. Schrift A, Abschnitt 1.3.2), d.h. sie gehen davon aus, dass Persönlichkeitseigenschaften unabhängig von einem Beobachter und dessen Instrumenten existieren, wahrscheinlich eine biologische Entsprechung haben und in wesentlichen Teilen genetisch veranlagt sind. Der Einwand, dass trotz eines solchen biologischen *Erklärungsmusters* kaum echte biologische *Befunde* für Persönlichkeitseigenschaften vorliegen, kann aktuell noch mit der relativ geringen Kenntnis von neurophysiologischen Vorgängen abgewehrt werden. Die Vertreter der OP-Hypothese äußern sich nicht hinsichtlich des Status ihres neuen Konstruktes. Es erscheint eher unwahrscheinlich, dass die Vertreter der OP-Hypothese eine sozialkonstruktivistische Position einnehmen und „Pessimismus“ nur als hinsichtlich der geteilten sozialen Übereinkunft zwischen Forscher und Untersuchungsobjekten existierend betrachten – sonst würden sie kaum Korrelationen zwischen „Pessimismus“ und Neurotizismus/negativer Affektivität berichten (Marshall et al., 1992; Robinson-Whelen et al., 1997). Angesichts der Konsequenzen, die die

Annahme der objektiven Existenz von „Pessimismus“ mit sich bringen würde, ist die breit akzeptierte Erklärung für Abweichungen von der Eindimensionalität bei Fragebogen und anderen Instrumenten (Campbell & Fiske, 1959), die das Methodeneffektmodell liefert, eine sehr viel sparsamere Annahme als die OP-Hypothese.

Sparsamkeit ist ein wichtiges Kriterium der Theorieevaluation und auch der Bewertung komplexer mathematischer Modelle (vgl. etwa Forster, 2000; Hitchcock & Sober, 2004; Stine, 2004). Das Sparsamkeitsprinzip wird dabei vor allem hinsichtlich der Anzahl der Parameter eines Modells diskutiert (s. etwa Forster, 2000). In neueren Ansätzen wie dem „Prinzip der minimalen Beschreibungslänge“ (minimum description length; Pitt, Myung & Zhang, 2002; Stine, 2004) wird es verallgemeinert, um auch Modelle vergleichen zu können, die sich nicht nur hinsichtlich der Parameterzahl unterscheiden. Grundsätzlich besagt das Sparsamkeitsprinzip, dass nur so viele Ursachen für eine Beobachtung angenommen werden sollten wie unbedingt nötig. Im Falle eines einfachen Regressionsmodell bedeutet dies etwa, dass Prädiktorvariablen entfernt werden sollten, wenn sie nicht in hohem Maße zur Verbesserung der Anpassung beitragen. Auch eine falsifikationistisch geprägte Methodologie bevorzugt einfachere Modelle (vgl. Meehl, 1978): Einfache Modelle schränken den Raum erlaubter Ergebnisse im Allgemeinen stärker ein als komplexe Modelle (Roberts & Pashler, 2000) und sind als „bold conjectures“ zu bevorzugen, da sie zu einem größeren Erkenntnisgewinn führen.

Das Sparsamkeitsprinzip ist in der mathematisch-statistischen Modellierung breit akzeptiert und etabliert. In Schrift B wird darüber hinaus gefordert, dieses Prinzip auch in seiner ursprünglichen, allgemeineren Form als Instrument zur Theorieevaluation einzusetzen: Die Tatsache, dass mit „Pessimismus“ ein völlig neues (und, wie in Kapitel 2 gezeigt wurde, sehr schwach begründetes) Konstrukt in die Welt der Persönlichkeitspsychologie eingeführt wird, sollte zu besonderer Vorsicht mahnen, insbesondere da das Methodeneffektmodell eine sehr überzeugende und bewährte Alternativerklärung liefert.

6 Invertierte Items in der Skala „Personaler Optimismus“

In den vorangegangenen Kapiteln wurde die OP-Hypothese vor allem im Zusammenhang mit dem LOT(-R) behandelt, da die Vertreter der OP-Hypothese ihre Analysen fast ausschließlich auf Daten basieren, die mit diesem Instrument erhoben wurden. Die OP-Hypothese wurde aus theoretischer Perspektive entkräftet (Kapitel 2), es wurde gezeigt, dass die verwendeten Methoden nicht geeignet sind, eine eindeutige Entscheidung über die OP-Hypothese auf der Basis von LOT(-R)-Daten herbeizuführen (Kapitel 3 und 4) und schließlich wurde in Kapitel 5 eine theoretisch sparsame und besser passende Alternativerklärung, das Methodeneffektmodell, vorgestellt. In diesem Kapitel wird eine weitere Argumentationslinie gegen die OP-Hypothese vorgestellt: In Schrift C wurde gezeigt, dass die Skala „Personaler Optimismus“ (Schweizer & Koch, 2001; Schweizer, Schneider & Beck-Seyffer, 2001; im Folgenden PO-Skala) *keine* Zweidimensionalität mit „Optimismus“ und „Pessimismus“ aufweist, obwohl diese Skala aus dem gleichen theoretischen Hintergrund wie der LOT(-R) abgeleitet wurde und genauso wie der LOT(-R) aus positiven und invertiert formulierten Items zusammengesetzt ist.

Die PO-Skala basiert ebenfalls auf der Definition von Optimismus als generalisierter positiver Zukunftserwartung. Sie wurde von den Autoren explizit als Alternative zum LOT-R erstellt, und zwar, um Items zur Verfügung zu stellen, in denen ein stärkerer direkter Bezug zur Person des oder der Probanden/Probandin hergestellt wird. Die Anforderung, dass sich Fragebogenitems deutlich auf die Person der Befragten beziehen sollten, wurde nicht zuletzt auch von Angleitner et al. (1986) gefordert. Darüber hinaus stimmen aber beide Skalen in ihrer theoretischen Herleitung überein und korrelieren auch hoch miteinander (Schweizer et al., 2001). Die PO-Skala kann also für einen Test der OP-Hypothese herangezogen werden: Findet sich auch für die PO-Skala eine zweidimensionale Struktur mit „Optimismus“- und „Pessimismus“-Items?

In Schrift C wurden Daten, die mit der PO-Skala erhoben wurden, nach Samejimas (1969) „Graded Response Model“ (GRM) analysiert, einem eindimensionalen Item-Response-Theorie-Modell (IRT-Modell) für polytome, d.h. mehrfach gestufte Items. Detaillierte Informationen über Durchführung, Auswertung und Ergebnisse können Schrift C entnommen werden. Wie in Schrift C ausgeführt wird, stellt ein IRT-Modell tatsächlich noch höhere Anforderungen als ein konfirmatorisches Faktorenmodell, da im Rahmen einer IRT-Analyse direkt die ordinal abgestufte Antwortskala modelliert wird anstelle einer Approximation der Antwortverteilung durch die Normalverteilungsannahme in der konfirmatorischen Faktorenanalyse. Das Ergebnis ist deutlich: Für die PO-Skala kann das GRM beibehalten werden, obwohl diese Skala sowohl positiv formulierte als auch invertierte Items enthält. Wenn in den PO-Daten das Problem der Zweidimensionalität vorliegen würde, das LOT(-R)-Daten im Allgemeinen betrifft, wäre es nicht möglich gewesen, das GRM erfolgreich anzupassen. Zwar zeigt sich, dass auch in der PO-Skala gewisse Unterschiede zwischen den positiven und den invertiert formulierten Items bestehen; so haben die positiven Items im Mittel etwas niedrigere Diskriminationsparameter und niedrigere Antwortschwellen, d.h. den positiven Items wird im Mittel etwas eher zugestimmt. Außerdem zeigt Item 3 („Ich freue mich über jeden neuen Tag.“) eine gewisse Abweichung vom erwarteten Antwortmuster. Diese Abweichung könnte ebenfalls mit einer Verzerrung durch soziale Erwünschtheit erklärt werden, möglicherweise aber auch einfach damit, dass das Item auch etwas anderes erfasst als Optimismus: die

„Freude über einen neuen Tag“ ist sicherlich in gewisser Weise mit positiver Affektivität konfundiert. Insgesamt aber zeigt die Untersuchung der PO-Skala, dass invertiert formulierte Optimismus-Items keinesfalls immer „Pessimismus“ als ein von „Optimismus“ unabhängiges Konstrukt erfassen – stattdessen steht mit der PO-Skala eine von psychometrischer Perspektive her bessere Alternative zur Erfassung von dispositionellem Optimismus zur Verfügung. Natürlich sind die Ergebnisse aus Schrift C keine *Bestätigung* dafür, dass die Items der PO-Skala alle ein einziges Konstrukt Optimismus erfassen (man erinnere sich an das Induktionsproblem aus Kapitel 4), aber sie stellen eine deutliche Abwertung der OP-Hypothese her.

7 Diskussion

7.1 Zusammenführung

In dieser Arbeit wurden drei Argumente gegen die OP-Hypothese vorgestellt:

1. Die Hypothese, dass der LOT-R einigermaßen unabhängige Konstrukte namens „Optimismus“ und „Pessimismus“ erfasse, wurde mittels methodologischer und im engeren Sinne statistisch-methodischer Untersuchungen als zu gering begründet zurückgewiesen.
2. Die „OP-Hypothese“ wurde mittels einer „besseren“ faktorenanalytischen Lösung für Daten, die mit dem LOT(-R) erhoben wurden, zwar nicht im engeren Sinne falsifiziert, aber dennoch deutlich entkräftet. Das Methodeneffektmodell liefert eine theoretisch sparsame Erklärung für die Zweidimensionalität der LOT(-R)-Daten, da kein „neues“ Konstrukt zur Erklärung herangezogen wird, sondern ein möglicher Effekt von sozial erwünschter Antwortverzerrung einbezogen wird. Die Befunde zur Unterdeterminiertheit von konfirmatorischen Faktorenanalysen aus Kapitel 4 gelten natürlich auch für das Methodeneffektmodell; aber gerade angesichts der Unterdeterminiertheit ist die theoretische Sparsamkeit dieses Modells ein umso stärkeres Argument, da aktuell keine anderen Entscheidungskriterien vorliegen.
3. Die „OP-Hypothese“ basiert nur auf Befunden zum LOT(-R), aber nicht auf alternativen Instrumenten für die Erfassung von Optimismus. Es wird gezeigt, dass etwa mit der PO-Skala ein Instrument vorliegt, das eine eindimensionale Erfassung von dispositionellem Optimismus ermöglicht.

Das erste Argument basiert vor allem auf methodologischen Überlegungen, die in den Kapiteln 3 und 4 unter Rückgriff auf Schrift A vorgestellt wurden, sowie auf theoretischen Argumenten, die in Kapitel 2 aufgeführt wurden. Das zweite Argument beruht auf der empirischen Untersuchung in Schrift B und das dritte Argument auf der Untersuchung in Schrift C. Die Ergebnisse der Arbeit können einfach zusammengefasst werden: Die OP-Hypothese lässt sich in ihrem momentanen Status kaum halten, da es theoretisch sparsame Alternativerklärungen für das Problem der invertierten Items im LOT(-R) gibt und da sich das Problem der Zweidimensionalität in der PO-Skala überhaupt nicht zeigt.

7.2 Methodologische Bewertung und Ausblick

Die vorliegende Arbeit folgt nicht unbedingt dem typischen Schema empirischer psychologischer Untersuchungen, da das Hauptanliegen die Zurückweisung einer Hypothese ist. Allerdings zeigt sich ja gerade in den Veröffentlichungen zur OP-Hypothese, dass das übliche Vorgehen bei der Suche nach neuen Ergebnissen nicht immer fruchtbar ist: Aus unterschiedlichen Perspektiven wurde gezeigt, dass ein „confirmation bias“ hinsichtlich der OP-Hypothese zur Missachtung einfacher Alternativerklärungen führte. Möglicherweise erschien der in Kapitel 2 dargestellte mangelnde theoretische Gehalt der OP-Hypothese als vernachlässigbar angesichts der positiven Ergebnisse der „konfirmatorischen“ Faktorenanalysen. Es ist allerdings zu vermuten, dass die dargestellten Probleme der OP-Hypothese zum Teil auf der Herausgeberpraxis wissenschaftlicher Fachzeitschriften beruhen, die im Allgemeinen zur Veröffentlichung *positiver* Befunde zwingt. Einer

Studie mit dem Ergebnis „XY wurde nicht bestätigt“ bleibt normalerweise nur die Ablage im „file drawer“ der Meta-Analyse-Methodiker (vgl. Rosenthal, 1979).

Auch aus einer anderen Perspektive ist festzuhalten, dass der „confirmation bias“, der die OP-Hypothese durch eine ganze Reihe von Veröffentlichungen getragen hat, bedauerlicherweise gewissermaßen institutionalisiert ist. In Kapitel 4 wurde ausführlich dargestellt, dass die Bestätigung von Hypothesen nach Übereinkunft fast aller Methodologen aus den unterschiedlichsten Disziplinen *nicht möglich ist*. Unglücklicherweise wurde aber die Maximum-Likelihood-Faktorenanalyse mit *a priori* restringierten Parametern als „konfirmatorisch“ bezeichnet. Sicherlich war in den sechziger Jahren der Enthusiasmus über eine deutlich verbesserte faktorenanalytische Methode so groß, dass diese unglückliche Bezeichnung nicht großartig diskutiert, sondern im Gegenteil gerne akzeptiert wurde. Dabei leiden Faktorenanalysen, egal ob „konfirmatorisch“ oder „explorativ“, doch noch unter einer sehr viel größeren Unsicherheit und Underdeterminiertheit als etwa statistische Analysen von experimentellen Untersuchungen. Faktorenanalysen sind nun einmal korrelative Verfahren, die zudem in der Persönlichkeitspsychologie mit Daten durchgeführt werden, deren Zustandekommen gleichfalls unklar ist. Eine „konfirmatorische“ Faktorenanalyse erstellt also ein Modell, zu dem es unendlich viele genauso gute (im Sinne des statistischen Kriteriums der Diskrepanzfunktion) Alternativen gibt, und zwar für Daten, die auf unklare Weise gewonnen wurden. Ausgerechnet eine solche Prozedur „konfirmatorisch“ zu nennen, ist allenfalls optimistisch, in den ansonsten aber meist durch das skeptische Suchen nach Alternativerklärungen geprägten empirischen Wissenschaften eher ungewöhnlich.

Was also ist das positive Ergebnis der vorliegenden Arbeit? Vor allem ist es die Erkenntnis, dass die Integration unterschiedlicher Perspektiven auf eine Fragestellung sehr viel fruchtbarer ist als das Verharren in einer einzigen Position. Die Vertreter der OP-Hypothese haben, durch „confirmation bias“ geprägt, versucht, „Pessimismus“ mit anderen Konstrukten zu korrelieren – heraus kam eine „crud“-Faktor geprägte Korrelation von Allem mit Jedem (Meehl, 1990). Risikoreiche Vorhersagen („bold conjectures“) und Erkenntnisgewinn dagegen sind nur möglich, wenn man die Grenzen eines einzelnen Untersuchungsansatzes erkennt, eigene Positionen und Methoden immer wieder in Frage stellt und mögliche Alternativerklärungen berücksichtigt. Verbleiben persönlichkeitspsychologische Theorien im Schema der immer gleich ablaufenden Fragebogenentwicklung und „konfirmatorischen“ Analyse mit vorhersehbarem Ergebnis – es ist schwer, Konstrukte zu finden, die *nicht* mit Neurotizismus/negativer Affektivität oder Extraversion/positiver Affektivität korrelieren – so bleibt ihnen wohl das Meehlsche Diktum nicht erspart: „theories in the ‚soft areas‘ of psychology have a fate like Douglas MacArthur said of what happens to old generals, ‚They never die, they just slowly fade away.‘“ (1990, S. 196).

8 Literatur

- Angleiter, A., John, O. P. & Löhr, F.-J. (1986). It's What You Ask and How You Ask It: An Itemmetric Analysis of Personality Questionnaires. In A. Angleitner & J. S. Wiggins (Eds.), *Personality Assessment Via Questionnaires: Current Issues in Theory and Measurement* (p. 61-108). New York: Springer Verlag.
- Angleitner, A. & Wiggins, J. S. (Eds.) (1986). *Personality Assessment Via Questionnaires: Current Issues in Theory and Measurement*. New York: Springer Verlag.
- Asendorpf, J. B., Banse, R. & Mücke, D. (2002). Double Dissociation Between Implicit and Explicit Personality Self-Concept: The Case of Shy Behavior. *Journal of Personality and Social Psychology*, *83*, 380-393.
- Barrett, P. (2007). Structural Equation Modeling: Adjudging Model Fit. *Personality and Individual Differences*, *42*, 815-824.
- Block, J. (1995). A Contrarian View of the Five-Factor Approach to Personality Description. *Psychological Bulletin*, *117*, 187-215.
- Borsboom, D., Mellenbergh, G. J. & van Heerden, J. (2003). The Theoretical Status of Latent Variables. *Psychological Review*, *110*, 203-219.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Berlin: Springer.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, *56*, 81-105.
- Carvajal, S. C., Garner, R. L. & Evans, R. I. (1998). Dispositional Optimism as a Protective Factor in Resisting HIV Exposure in Sexually Active Inner-City Minority Adolescents. *Journal of Applied Social Psychology*, *28*, 2196-2211.
- Carver, C. S., Lehman, J. M. & Antoni, M. H. (2003). Dispositional Pessimism Predicts Illness-Related Disruption of Social and Recreational Activities Among Breast Cancer Patients. *Journal of Personality and Social Psychology*, *84*, 813-821.
- Carver, C. S. & Scheier, M. F. (1998). *On the Self-Regulation of Behavior*. New York: Cambridge University Press.
- Carver, C. S. & Scheier, M. F. (2002). Optimism, Pessimism, and Self-Regulation. In E.C. Chang (Ed.), *Optimism & Pessimism. Implications for Theory, Research, and Practice* (pp. 31-51). Washington: American Psychological Association.
- Chang, E. C. (Ed.) (2002a). *Optimism & Pessimism. Implications for Theory, Research, and Practice*. Washington: American Psychological Association.
- Chang, E. C. (2002b). Cultural Influences on Optimism and Pessimism: Differences in Western and Eastern Construals of the Self. In E.C. Chang (Ed.), *Optimism & Pessimism. Implications for Theory, Research, and Practice* (pp. 257-280). Washington: American Psychological Association.

- Chang, E. C. (2002c). Optimism–Pessimism and Stress Appraisal: Testing a Cognitive Interactive Model of Psychological Adjustment in Adults. *Cognitive Therapy and Research*, 26, 675-690.
- Chang, E. C., Maydeu-Olivares, A., & D’Zurilla, T. J. (1997). Optimism and Pessimism as Partially Independent Constructs: Relations to Positive and Negative Affectivity and Psychological Well-Being. *Personality and Individual Differences*, 23, 433-440.
- Curran, P. S., West, S. G. & Finch, J. F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16–29.
- Dörner, D. (2004). Der Mensch als Maschine. In G. Jüttemann (Hrsg.), *Psychologie als Humanwissenschaft. Ein Handbuch* (S. 32-45). Göttingen: Vandenhoeck & Ruprecht.
- Domino, B. & Conway, D. W. (2002). Optimism and Pessimism from a Historical Perspective. In E.C. Chang (Ed.), *Optimism & Pessimism. Implications for Theory, Research, and Practice* (pp. 13-30). Washington: American Psychological Association.
- Dovidio, J. F., Kawakami, K. & Gaertner, S. L. (2002). Implicit and Explicit Prejudice and Interracial Interaction. *Journal of Personality and Social Psychology*, 82, 62-68.
- Edwards, J. R. & Bagozzi, R. P. (2000). On the Nature and Direction of Relationships Between Constructs and Their Measures. *Psychological Methods*, 5, 155-174.
- Eid, M. (2000). A Multitrait-Multimethod Model With Minimal Assumptions. *Psychometrika*, 65, 241-261.
- Eid, M., Lischetzke, T., Nussbeck, F.W. & Trierweiler, L.I. (2003). Separating Trait Effects From Trait-Specific Method Effects in Multitrait–Multimethod Models: A Multiple-Indicator CT-C(M-1) Model. *Psychological Methods*, 8, 38-60.
- Eysenck, H.J. (1997). Personality and Experimental Psychology: The Unification of Psychology and the Possibility of a Paradigm. *Journal of Personality and Social Psychology*, 73, 1224-1237.
- Fiske, D. W. (1986). The Trait Concept and the Personality Questionnaire. In A. Angleitner & J. S. Wiggins (Eds.), *Personality Assessment Via Questionnaires: Current Issues in Theory and Measurement* (pp. 35-46). New York: Springer Verlag.
- Forster, M. R. (2000). Key Concepts in Model Selection: Performance and Generalizability. *Journal of Mathematical Psychology*, 44, 205-231.
- Freedman, D. (1999). From Association to Causation: Some Remarks on the History of Statistics. *Statistical Science*, 14, 243-258.
- Frese, M., & Zapf, D. (1994). Action as the Core of Work Psychology: A German approach. In H. C. Triandis, M. D. Dunnette, & J. M. Hough (Eds.), *Handbook of industrial and organizational psychology*, Vol. 4 (2nd ed.) (pp. 271-340). Palo Alto, CA: Consulting Psychologists Press.

- Gigerenzer, G. (2004). Mindless Statistics. *The Journal of Socio-Economics*, 33, 587-606.
- Gillham, J. E., Shatté, A. J., Reivich, K. J., & Seligman, M. E. P. (2002). Optimism, Pessimism, and Explanatory Style. In E.C. Chang (Ed.), *Optimism & Pessimism. Implications for Theory, Research, and Practice* (pp. 53-75). Washington: American Psychological Association.
- Hartig, J., Hölzel, B. & Moosbrugger, H. (2007). A Confirmatory Analysis of Item Reliability Trends (CAIRT): Differentiating True Score and Error Variance in the Analysis of Item Context Effects. *Multivariate Behavioral Research*, 42, 157-183.
- Helferich, C. (1992). *Geschichte der Philosophie*. Stuttgart: Metzler.
- Herzberg, P.-Y., Glaesmer, H. & Hoyer, J. (2006). Separating Optimism and Pessimism: A Robust Psychometric Analysis of the Revised Life Orientation Test (LOT-R). *Psychological Assessment*, 18, 433-438.
- Hitchcock, C. & Sober, E. (2004). Prediction Versus Accommodation and the Risk of Overfitting. *British Journal of the Philosophy of Science*, 55, 1-34.
- Hofmann, W., Rauch, W. A. & Gawronski, B. (2007). And Deplete Us Not Into Temptation: Automatic Attitudes, Dietary Restraint, and Self-Regulatory Resources as Determinants of Eating Behavior. *Journal of Experimental Social Psychology*, 43, 497-504.
- Holzinger, K. J. & Swineford, F. (1937). The Bi-Factor Method. *Psychometrika*, 2, 41-54.
- Holzkamp, K. (1970). Wissenschaftstheoretische Voraussetzungen kritisch-emanzipatorischer Psychologie (Teil I). *Zeitschrift für Sozialpsychologie*, 1, 5-21.
- Hoyer, J. (2000). Optimismus und Gesundheit: Überblick, Kritik und Forschungsperspektiven. *Zeitschrift für Gesundheitspsychologie*, 8, 111-122.
- Hu, L. & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling*, 6, 1-55.
- Johnson, J. A. (2004). The Impact of Item Characteristics on Item and Scale Validity. *Multivariate Behavioral Research*, 39, 273-302.
- Jöreskog, K. G. (1969). A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika*, 34, 183-202.
- Keith, N., Hodapp, V., Schermelleh-Engel, K., & Moosbrugger, H. (2003). Cross-sectional and longitudinal confirmatory factor models for the German Test Anxiety Inventory: A construct validation. *Anxiety, Stress, and Coping*, 16, 251-270.
- Knäuper, B. (1999). The Impact of Age and Education on Response Order Effects in Attitude Measurement. *Public Opinion Quarterly*, 63, 347-370.

- Knäuper, B., Schwarz, N., Park, D., & Fritsch, A. (im Druck). The Perils of Interpreting Cohort Differences in Attitude Research: Question Order Effects Decrease With Age. *Public Opinion Quarterly*.
- Knowles, E. S., Coker, M. C., Scott, R. A., Cook, D. A., & Neville, J. W. (1996). Measurement-induced Improvement in Anxiety: Mean Shifts With Repeated Assessment. *Journal of Personality and Social Psychology, 71*, 352–363.
- Knowles, E. S. & Condon, C. A. (1999). Why People Say “Yes”: A Dual-Process Theory of Acquiescence. *Journal of Personality and Social Psychology, 77*, 379-386.
- Knowles, E. S. & Condon, C. A. (2000). Does the Rose Still Smell As Sweet? Item Variability Across Test Forms and Revisions. *Psychological Assessment, 12*, 245-252.
- Kubzansky, L. D., Kubzansky, P. E., & Maseko, J. (2004). Optimism and pessimism in the context of health: Bipolar opposites or separate constructs? *Personality and Social Psychology Bulletin, 30*, 943–956.
- Ladyman, J. (2002). *Understanding Philosophy of Science*. London: Routledge.
- Maraun, M. D. (1996). Metaphor Taken As Math: Indeterminacy in the Factor Analysis Model. *Multivariate Behavioral Research, 31*, 517-538.
- Marsh, H. W. (1996). Positive and Negative Self-Esteem: A Substantively Meaningful Distinction or Artifacts? *Journal of Personality and Social Psychology, 70*, 810-819.
- Marshall, G. N. & Lang, E. L. (1990). Optimism, Self-Mastery, and Symptoms of Depression in Women Professionals. *Journal of Personality and Social Psychology, 59*, 132-139.
- Marshall, G. N., Wortman, C. B., Kusulas, J. W., Hervig, L. K., & Vickers, R. R. Jr, (1992). Distinguishing optimism from pessimism: Relations to fundamental dimensions of mood and personality. *Journal of Personality and Social Psychology, 62*, 1067–1074.
- McCrae, R. R., Costa, P.T., Jr., Ostendorf, F., Angleitner, A., Hřebíčková, M. et al. (2000). Nature Over Nurture: Temperament, Personality, and Life Span Development. *Journal of Personality and Social Psychology, 78*, 173-186.
- McPherson, J. & Mohr, P. (2005). The Role of Item Extremity in the Emergence of Keying-Related Factors: An Exploration With the Life Orientation Test. *Psychological Methods, 10*, 120-131.
- Meehl, P. E. (1978). Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology. *Journal of Consulting and Clinical Psychology, 46*, 806-834.
- Meehl, P. E. (1990). Why Summaries of Research on Psychological Theories are Often Uninterpretable. *Psychological Reports, 66*, 195–244.

- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). Washington, DC: American Council on Education and National Council on Measurement in Education.
- Michell, J. (2004). Item Response Models, Pathological Science, and the Shape of Errors. Reply to Borsboom and Mellenbergh. *Theory & Psychology, 14*, 121-129.
- Miller, L., Richards, P. S. & Keller, R R. (2002). Foreword. In E.C. Chang (Ed.), *Optimism & Pessimism. Implications for Theory, Research, and Practice* (pp. xiii-xvii). Washington: American Psychological Association.
- Moosbrugger, H. & Rauch, W. (im Druck). Das Allgemeine Lineare Modell in der Evaluationsforschung. In H. Holling & R. Schwarzer (Hrsg.), *Grundlagen und Methoden der Evaluationsforschung* (=Enzyklopädie der Psychologie, Serie IV, Band I). Göttingen: Hogrefe.
- Moosbrugger, H. & Schermelleh-Engel, K. (2006). Faktorenanalyse. In F. Petermann & M. Eid (Hrsg.), *Handbuch der Psychologischen Diagnostik* (S. 304-317). Göttingen: Hogrefe.
- Mroczek, D. K., Spiro, A., Aldwin, C. M., & Ozer, D. J. (1993). Construct Validation of Optimism and Pessimism in Older Men: Findings From the Normative Aging Study. *Health Psychology, 12*, 406-409.
- Musch, J., Brockhaus, R., & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit. *Diagnostica, 48*, 121–129.
- Musgrave, A. (1989). Deductive Heuristics. In K. Gavroglu, Y. Gouderoulis & P. Nicolacopoulos (Eds.), *Imre Lakatos and Theories of Scientific Change* (pp. 15-32). Dordrecht: Kluwer Academic Publishers.
- Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology, 2*, 175-220.
- Nietzsche, F. (1887/1930). *Also sprach Zarathustra* (Erstausgabe 1887). Leipzig: Kröner.
- Norem, J. K. (2002). Defensive Pessimism, Optimism, and Pessimism. In E.C. Chang (Ed.), *Optimism & Pessimism. Implications for Theory, Research, and Practice* (pp. 77-100). Washington: American Psychological Association.
- Norenzayan, A. & Schwarz, N. (1999). Telling What They Want to Know: Participants Tailor Causal Attributions to Researchers' Interests. *European Journal of Social Psychology, 29*, 1011-1020.
- Paulhus, D. L. (2002). Socially Desirable Responding: The Evolution of a Construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The Role of Constructs in Psychological and Educational Measurement* (pp. 49-69). Mahwah, NJ: Lawrence Erlbaum Associates.

- Perkins, D. O., Leserman, J., Murphy, C. & Evans, D.L. (1993). Psychosocial Predictors of High-Risk Sexual Behavior Among HIV-Negative Homosexual Men. *AIDS Education and Prevention*, 5, 141-152.
- Peterson, C. (2000). The Future of Optimism. *American Psychologist*, 55, 44-55.
- Peterson, C. & Bossio, L. M. (2002). Optimism and Physical Well-Being. In E.C. Chang (Ed.), *Optimism & Pessimism. Implications for Theory, Research, and Practice* (pp. 127-145). Washington: American Psychological Association.
- Platt, J. R. (1964). Strong Inference. *Science*, 146, 347-353.
- Pytlík Zillig, L. M., Hemenover, S. H. & Dienstbier, R. A. (2002). What Do We Assess When We Assess a Big 5 Trait? A Content Analysis of the Affective, Behavioral, and Cognitive Processes Represented in Big 5 Personality Inventories. *Personality and Social Psychology Bulletin*, 28, 847-858.
- Ostmann, A. & Wutke, J. (1994). Statistische Entscheidung. In T. Herrmann & W.H. Tack (Hrsg.), *Methodologische Grundlagen der Psychologie* (=Enzyklopädie der Psychologie, Serie I: Forschungsmethoden der Psychologie, Band 1, S. 694-737). Göttingen: Hogrefe.
- Pitt, M. A., Myung, I. J. & Zhang, S. (2002). Toward a Method of Selecting Among Computational Models of Cognition. *Psychological Review*, 109, 472-491.
- Raykov, T. & Penev, S. (1999). On Structural Model Equivalence. *Multivariate Behavioral Research*, 34, 194-244.
- Roberts, S. & Pashler, H. (2000). How Persuasive is a Good Fit? A Comment on Theory Testing. *Psychological Review*, 107, 358-367.
- Robinson-Whelen, S., Kim, C., MacCallum, R. C., & Kiecolt-Glaser, J. K. (1997). Distinguishing Optimism from Pessimism in Older Adults: Is It More Important to be Optimistic or not to be Pessimistic? *Journal of Personality and Social Psychology*, 73, 1345-1353.
- Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.
- Rosenthal, R. (1979). The "File Drawer Problem" and Tolerance for Null Results. *Psychological Bulletin*, 86, 638-641.
- Rost, J., Carstensen, C.H. & von Davier, M. (1999). Sind die Big Five Rasch-Skalierbar? Eine Reanalyse der NEO-FFI-Normierungsdaten. *Diagnostica*, 45, 119-127.
- Røysamb, E. & Strype, J. (2002). Optimism and Pessimism: Underlying Structure and Dimensionality. *Journal of Social and Clinical Psychology*, 21, 1-19.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph No. 17*.

- Sarris, V. & Reiß, S. (2005). *Kurzer Leitfaden der Experimentalpsychologie*. München: Pearson Studium.
- Scheier, M. F. & Carver, C. S. (1985). Optimism, Coping, and Health: Assessment and Implications of Generalized Outcome Expectancies. *Health Psychology, 4*, 219-247.
- Scheier, M. F., Carver, C. S. & Bridges, M. W. (1994). Distinguishing Optimism From Neuroticism (And Trait Anxiety, Self-Mastery, and Self-Esteem): A Reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67*, 1063-1078.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research – Online, 8*, 23-74.
- Scholz, U., Gutierrez Doña, B., Sud, S. & Schwarzer, R. (2002). Is General Self-Efficacy a Universal Construct? Psychometric Findings From 25 Countries. *European Journal of Psychological Assessment, 18*, 242-251.
- Schwarz, N. (1999). Self-Reports: How the Questions Shape the Answers. *American Psychologist, 54*, 93-105.
- Schwarz, N. (2007). Cognitive Aspects of Survey Methodology. *Applied Cognitive Psychology, 21*, 277-287.
- Schweizer, K. & Koch, W. (2001). The Assessment of Components of Optimism by POSO-E. *Personality and Individual Differences, 31*, 563-574.
- Schweizer, K., Schneider, R., & Beck-Seyffer, A. (2001). Personaler und Sozialer Optimismus. *Zeitschrift für Differentielle und Diagnostische Psychologie, 22*, 13-24.
- Snyder, C. R., Sympson, S. C., Michael, S. T. & Cheavens, J. (2002). Optimism and Hope Constructs: Variants on a Positive Expectancy Theme. In E.C. Chang (Ed.), *Optimism & Pessimism. Implications for Theory, Research, and Practice* (pp. 101-126). Washington: American Psychological Association.
- Still, A. & Dryden, W. (2004). The Social Psychology of “Pseudoscience”: A Brief History. *Journal for the Theory of Social Behavior, 34*, 265-290.
- Stine, R. A. (2004). Model Selection Using Information Theory and the MDL Principle. *Sociological Methods and Research, 33*, 230-260.
- Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M. & McDowell, N. K. (2003). Portrait of the Self-Enhancer: Well Adjusted and Well Liked or Maladjusted and Friendless? *Journal of Personality and Social Psychology, 84*, 165-176.
- Tomarken, A. J. & Waller, N. G. (2003). Potential Problems With „Well Fitting“ Models. *Journal of Abnormal Psychology, 112*, 578-598.
- Tourangeau, R., Rips, L. C., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

- Traub, R. E. (1997). Classical Test Theory in Historical Perspective. *Educational Measurement: Issues and Practice*, 16, 8-14.
- Urbach, P. (1989). The Bayesian Alternative to the Methodology of Scientific Research Programmes. In K. Gavroglu, Y. Goudaroulis & P. Nicolacopoulos (Eds.), *Imre Lakatos and Theories of Scientific Change* (pp. 399-412). Dordrecht: Kluwer Academic Publishers.
- Vautier, S., Mullet, E., & Bourdet-Loubère, S. (2003). The Instruction Set of Questionnaires Can Affect the Structure of the Data: Application to Self-Rated State Anxiety. *Theory and Decision*, 54, 249-259.
- Vautier, S. & Raufaste, E. (2006). Configural Instability of Data From the Life Orientation Test – Revised. *Personality and Individual Differences*, 40, 1511-1518.
- Vautier, S., Raufaste, E. & Cariou, M. (2003). Dimensionality of the Revised Life Orientation Test and the Status of Filler Items. *International Journal of Psychology*, 38, 390-400.
- von Brück, M. (2002). *Wie können wir leben? Religion und Spiritualität in einer Welt ohne Maß*. München: C.H. Beck.
- Watkins, J. (1989). The Methodology of Scientific Research Programmes: A Retrospect. In K. Gavroglu, Y. Goudaroulis & P. Nicolacopoulos (Eds.), *Imre Lakatos and Theories of Scientific Change* (pp. 3-13). Dordrecht: Kluwer Academic Publishers.
- Wilson, M. (2003). On Choosing a Model for Measuring. *Methods of Psychological Research – Online*, 8, 1-22.
- Yuan, K.-H. (2005). Fit Indices Versus Test Statistics. *Multivariate Behavioral Research*, 40, 115-148.

9 Schriftenanhang

Schrift A

Rauch, W. A. & Moosbrugger, H. (im Druck). Klassische Testtheorie. Grundlagen und Erweiterungen für heterogene Tests und Mehrfacettenmodelle. In M. Amelang & L.F. Hornke (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich B Methodologie und Methoden, Serie II Psychologische Diagnostik, Band 2 Methoden*. Göttingen: Hogrefe.

Schrift B

Rauch, W. A., Moosbrugger, H. & Schweizer, K. (2007). Method effects due to social desirability as a parsimonious explanation of the deviation from unidimensionality iLOT-R scores. *Personality and Individual Differences*, 42, 1597-1607.

Schrift C

Rauch, W. A., Moosbrugger, H. & Schweizer, K. (im Druck). An IRT Analysis of the Personal Optimism Scale. *European Journal of Psychological Assessment*.

9.1 Schrift A

Rauch, W. A. & Moosbrugger, H. (im Druck). Klassische Testtheorie. Grundlagen und Erweiterungen für heterogene Tests und Mehrfacettenmodelle. In M. Amelang & L.F. Hornke (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich B Methodologie und Methoden, Serie II Psychologische Diagnostik, Band 2 Methoden*. Göttingen: Hogrefe.

9.2 Schrift B

Rauch, W. A., Moosbrugger, H. & Schweizer, K. (2007). Method effects due to social desirability as a parsimonious explanation of the deviation from unidimensionality in LOT-R scores. *Personality and Individual Differences*, 42, 1597-1607.

9.3 Schrift C

Rauch, W. A., Moosbrugger, H. & Schweizer, K. (im Druck). An IRT Analysis of the Personal Optimism Scale. *European Journal of Psychological Assessment*.

10 Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Hypothese, dass der „Life Orientation Test (-Revised)“ (LOT-R), der als Selbstberichtsfragebogen zur Erfassung von dispositionellem Optimismus konstruiert wurde, nicht nur „Optimismus“ erfasse, sondern auch „Pessimismus“; Pessimismus wird im Rahmen dieser „OP-Hypothese“ nicht als der Gegenpol von Optimismus aufgefasst, sondern als ein partiell unabhängiges Konstrukt. Die OP-Hypothese wird typischerweise mit sog. „konfirmatorischen“ Faktorenanalysen untersucht.

Die OP-Hypothese wird zunächst aus theoretischer Perspektive analysiert, insbesondere auch deshalb, weil die Begriffe „Optimismus“ und „Pessimismus“ im Rahmen der OP-Hypothese nicht auf dieselbe Art verwendet werden wie sonst in der Psychologie oder in anderen Wissenschaften üblich. Dabei zeigt sich, dass keine angemessene theoretische Herleitung der OP-Hypothese möglich ist: Was „Pessimismus“ ist, wenn es nicht der Gegenpol von Optimismus ist, bleibt unklar. Ausgehend von dem Befund, dass die theoretische Herleitung der OP-Hypothese unangemessen ist, wird diese Hypothese aus methodologisch-testtheoretischer Perspektive und mittels zweier empirischer Arbeiten kritisch überprüft.

Aus methodologischer und mathematisch-statistischer Perspektive wird dargestellt, dass die empirischen Befunde, die für die OP-Hypothese sprechen, nicht eindeutig sind; die OP-Hypothese ist also unterdeterminiert. Die Gültigkeit eines reflexiven Messmodells, das für die Untersuchung der OP-Hypothese mittels konfirmatorischer Faktorenanalysen erforderlich wäre, kann nicht ohne Weiteres angenommen werden. Die mathematisch-statistischen Annahmen der „experimentellen Unabhängigkeit“ und der „lokalen Homogenität“, die ein reflexives Messmodell konstituieren, sind für Selbstberichtsfragebogen in vielen Fällen verletzt.

Noch gravierender ist das Ergebnis, dass sich aus „konfirmatorischen“ Faktorenmodellen, die *für* die OP-Hypothese sprechen, keine ausreichenden Informationen über den empirischen Gehalt der Hypothese ergeben. In grundsätzlicher Form und anhand einfacher Beispiele wird gezeigt, dass für eine gegebene Kovarianzmatrix, die die Datenbasis einer „konfirmatorischen“ Faktorenanalyse darstellt, grundsätzlich unendlich viele Alternativmodelle konstruiert werden können, von denen zumindest eine Teilmenge auch mit sinnvollen Interpretationen verbunden ist. Angesichts der Existenz solcher Alternativmodelle müssen andere Kriterien zur Überprüfung der OP-Hypothese herangezogen werden als „konfirmatorische“ Faktorenmodelle.

Anhand einer empirischen Untersuchung mit dem LOT-R wird gezeigt, dass ein „Methodeneffektmodell“ eine bessere Anpassung an die Daten ermöglicht als das mit der OP-Hypothese verbundene Modell. Das Methodeneffektmodell beinhaltet die Annahme, dass die Antworten auf LOT-R-Items nicht ausschließlich durch die individuelle Ausprägung in „Optimismus“ bedingt werden, sondern zum Teil auch unerwünschte Effekte der Itemformulierungen abbilden. Entsprechend lässt sich eine Korrelation des „Methodenfaktors“ mit sozial erwünschtem Antwortverhalten beobachten. Schließlich wird in einer weiteren empirischen Untersuchung gezeigt, dass mit der Skala „Personaler Optimismus“ ein Instrument zur Erfassung von dispositionellem Optimismus vorliegt, welches nicht unter der faktorenanalytischen Mehrdimensionalität leidet, die den LOT-R kennzeichnet.

In der Zusammenschau der methodologischen und empirischen Ergebnisse muss die OP-Hypothese als zu gering begründet zurückgewiesen werden. Über den Rahmen des speziellen Inhaltsgebietes „Optimismus“ hinaus weist die vorliegende Arbeit aber auch auf die Notwendigkeit multipler Betrachtungsebenen bei der Untersuchung persönlichkeitspsychologischer Fragestellungen hin. Nachteilig für die persönlichkeitspsychologische Forschung könnten sich insbesondere die unreflektierte Verwendung von Selbstberichtsfragebogen auf der Seite der Datenerhebung und die unreflektierte Verwendung von „konfirmatorischen“ Faktorenanalysen auf der Seite der Datenauswertung auswirken.

Lebenslauf

Wolfgang Rauch

geboren am 21. April 1975 in Frankfurt am Main

Schulausbildung

1981 – 1994 Schulbesuch in Langen/Hessen.

1994 Schulabschluss mit dem Abitur (Allgemeine Hochschulreife), Note „sehr gut“ (1,3)

Hochschulausbildung

1995 – 2002 Studium der Psychologie an der Johann Wolfgang Goethe-Universität Frankfurt

Hilfskraft in der Abteilung für Psychologische Methodenlehre, Forschungsmethoden und Evaluation (Prof. Dr. Moosbrugger)

2002 Abschluss des Studiums mit dem Diplom in Psychologie, Note „sehr gut“ (1,1)

seit 2003 Wissenschaftlicher Mitarbeiter am Institut für Psychologie der J.W. Goethe-Universität Frankfurt am Main in der Abteilung für Psychologische Methodenlehre, Forschungsmethoden und Evaluation (Prof. Dr. Moosbrugger)

Berufspraktische Erfahrungen

1998 Praktikum: ISM, Kronberg, Institut für Sozial- und Marktforschung, internationale Marktforschungsprojekte

1999 Praktikum: Flughafen Frankfurt Main AG (heute Fraport AG). Personal- und Organisationsentwicklung

2002 Berater: S & F Personalpsychologie GmbH, Stuttgart. Human Resource-Beratung

Zusatzqualifikationen

1999 „Train-the-Trainer“ Weiterbildung bei Frau Dr. Strittmatter-Haubold von der Akademie für wissenschaftliche Weiterbildung, Pädagogische Hochschule Heidelberg e.V.

2000 Seminar „Projektorientiertes Arbeiten“ bei Herrn Ralf Krüger von der Flughafen Frankfurt/Main AG (heute Fraport AG)

2005 Lizenz für berufsbezogene Eignungsbeurteilungen nach DIN 33430 von der Föderation Deutscher Psychologinnenvereinigungen

2007 Kurs „Introduction to Bayesian Data Analysis Using WinBUGS“ bei Prof. Dr. Best und Dr. Spiegelhalter, Cambridge (England)

Frankfurt am Main, 15.10.07