

Soziale Erwünschtheit im Licht des Rational-Choice Ansatzes

Dissertation
zur Erlangung des Doktorgrades
der Sozialwissenschaftlichen Fakultät
der Georg-August-Universität Göttingen

vorgelegt von
Julia Lischewski

geboren in Nordhausen

Göttingen, 2014

Erstgutachter: Prof. Dr. Steffen M. Kühnel

Zweitgutachterin: Prof. Dr. Dagmar Krebs

Weitere Mitglieder der Prüfungskommission:

Drittprüferin: Prof. Dr. Stefanie Eifler

Tag der mündlichen Prüfung: 01.10.2014

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	5
1 Einleitung.....	6
2 Definition des sozial erwünschten Antwortverhaltens.....	11
2.1 SD-Bias im Rahmen der Klassische Testtheorie.....	11
2.2 Problemfelder bisheriger Definitionsansätze	14
2.2.1 Soziale Erwünschtheit von Merkmalen, Verhaltensweisen und Einstellungen.....	14
2.2.2 Unterschiede zwischen Frageformen.....	16
2.2.3 Konsequenzen für Wissens- und Beurteilungsfragen.....	19
2.2.4 Notwendigkeit präziser Definitionen.....	21
3 Theoretischer Forschungsstand zur Anwendung des Rational-Choice-Ansatzes auf das Befragtenverhalten.....	26
3.1 RC-Ansatz und SEU-Theorie.....	27
3.2 Anwendungsmodelle zur Erklärung der sozial erwünschten Antwortverzerrung.....	29
3.2.1 Modell nach Crowne-Marlow.....	29
3.2.2 Modell nach Esser.....	33
3.2.3 Modell nach Stocké.....	42
3.2.4 Modell nach Skarbek-Kozietulska et al.....	49
3.2.5 Modell nach Wolter.....	51
3.2.6 Erweitertes Frame-Selektion Modell nach Stocké.....	55
3.3 Zusammenfassung und Bewertung	62
4 Operationalisierungsvarianten der Erklärungsfaktoren im Rahmen des RC-Ansatzes.....	66
4.1 Erklärungsfaktoren der Handlungskonsequenz: soziale Anerkennung.....	66
4.1.1 Anerkennungsbedürfnis	67
4.1.1.1 Edwards Skala.....	68
4.1.1.2 Marlow Crowne Skala.....	69
4.1.1.3 SES-17 Skala.....	72
4.1.1.4 BIDR-Skala.....	73
4.1.1.5 Bewertung	76
4.1.2 Anonymität.....	79
4.1.2.1 Objektive Anonymität.....	79
4.1.2.1.1 Befragungstechniken.....	80
4.1.2.1.2 Erhebungsmodus.....	82
4.1.2.2 Subjektive Anonymität.....	86
4.1.3 Erwünschtheitswahrnehmung.....	88
4.1.3.1 Bezugsgruppen	88
4.1.3.2 Persönliche und extrinsische Bewertung sozialer Erwünschtheit.....	89
4.1.3.3 Bewertung	90
4.2 Erklärungsfaktoren für die Handlungskonsequenz: Authentizität.....	93
4.2.1 Authentizität.....	94
4.2.1.1 Selbstüberwachung.....	95
4.2.2 Einstellungsstärke.....	99
4.2.2.1 „objektive“ Einstellungsstärke.....	100
4.2.2.2 Subjektive Einstellungsstärke.....	102
4.2.3 Passgenauigkeit.....	102
4.3 Erklärungsfaktoren des Frame-Selektion Modells.....	105
4.3.1 Kooperative Befragtenrolle.....	105

4.3.1.1 Allgemeine Umfrageeinstellung.....	106
4.3.1.2 Umfrageerfahrung.....	107
4.4 Zusammenfassung und Bewertung	110
5 Empirische Analyse.....	113
5.1 Überblick über die Fragestellungen	113
5.2 Strukturelle Stabilität und Dimensionen der Täuschungsskalen.....	118
5.2.1 Hypothesen.....	118
5.2.2 Operationalisierung.....	119
5.2.3 Methode.....	121
5.2.4 Ergebnisse.....	123
5.2.5 Zusammenfassung und Bewertung.....	132
5.3 Erklärungsbeitrag des Marlow-Crowne Modells innerhalb von Bevölkerungsumfragen.....	135
5.3.1 Hypothesen.....	135
5.3.2 Operationalisierung	138
5.3.3 Methode.....	139
5.3.4 Ergebnisse.....	140
5.3.4.1 Deskriptive Ergebnisse	140
5.3.4.2 Multivariate Ergebnisse	141
5.3.5 Zusammenfassung und Bewertung.....	147
5.4 Anwendung der RC-Modelle.....	149
5.4.1 Eckpunkte der Modellvariation.....	150
5.4.2 Operationalisierung.....	152
5.4.3 Methode.....	154
5.4.4 Ergebnisse	156
5.4.4.1 Deskriptive Ergebnisse	156
5.4.4.2 Multivariate Ergebnisse.....	157
5.4.4.2.1 Ergebnisse der RC-Modelle mit der SES-17 Skala als Messung des Anerkennungsbedürfnisses	157
5.4.4.2.2 Ergebnisse der RC-Modelle mit der überarbeiteten SES-17 Skala als Messung des Anerkennungsbedürfnisses.....	164
5.4.4.2.3 Ergebnisse der RC-Modelle mit der FT-Skala als Messung des Anerkennungsbedürfnisses	166
5.4.4.2.4 Ergebnisse der Frame-Selektion-Modelle mit der überarbeiteten SES17-Skala als Messung des Anerkennungsbedürfnisses.....	167
5.4.5 Zusammenfassung und Bewertung.....	169
6 Zusammenfassung und Diskussion.....	172
Literatur.....	179
Anhang.....	189

Tabellenverzeichnis

Tabelle 1: Kommunikationskanäle.....	83
Tabelle 2: Grad der Anonymität nach Befragungsmodus.....	84
Tabelle 3: Dimensionen der Erwünschtheitswahrnehmung Beispiele.....	90
Tabelle 4: Übersicht über die verwendeten Skalen der einzelnen Surveys	114
Tabelle 5: Itemübersicht der Täuschungsskalen in dem SST Survey.....	120
Tabelle 6: Mittelwerte der SD-Skalen für das SST-Survey.....	123
Tabelle 7: Ergebnisse der konfirmatorischen Faktorenanalyse.....	124
Tabelle 8: Standardisierte Schätzer nach der konfirmatorischen Faktorenanalyse für die SES-17 Skala.....	125
Tabelle 9: Korrelationsmatrix für die Mittelwerte der SD-Skalen.....	126
Tabelle 10: Faktorenstruktur der explorativen Faktorenanalyse auf Basis des ML-Schätzers und der Promaxrotation.....	127
Tabelle 11: Standardisierte Schätzer nach der konfirmatorischen Faktorenanalyse für die SD-Skalen der SST.....	130
Tabelle 12: Mittelwerte der Täuschungs- und Vorurteilsskalen	140
Tabelle 13: Lineare Regression des offenen Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest).....	144
Tabelle 14: Lineare Regression des subtilen Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest.....	145
Tabelle 15: Lineare Regression der Israelkritik auf die Einflussfaktoren im Rahmen des RC- Modells von Marlow und Crowne (GMF Graduierten Pretest).....	145
Tabelle 16: Lineare Regression der Missing Values für den offenen Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne auf die Missing Values für den offenen Antisemitismus (GMF Graduierten Pretest).....	145
Tabelle 17: Mittelwerte der Erwünschtheitswahrnehmung für verschiedene Vorurteile (GMF Graduierten Pretest).....	147
Tabelle 18: Vergleich der RC-Modelle bezüglich der erklärenden Variablen.....	154
Tabelle 19: Operationalisierungsmöglichkeiten und Modellvarianten.....	155
Tabelle 20: Deskriptive Werte der abhängigen und unabhängigen Variablen	156
Tabelle 21: Lineare Regression des RC-Grundmodells auf die Fremdenfeindlichkeit – nach Grad der Anonymität getrennt	158
Tabelle 22: Vorhersagewerte der Regressionanalyse des RC-Modells nach Stocké auf den Antisemitismus.....	159
Tabelle 23: Prozentuale Verteilung der wahrgenommenen	162
Anonymität in der SST-Stichprobe.....	162
Tabelle 24: Vorhersagewerte der Regressionanalyse des RC-Modells nach Stocké auf negative Einstellungen gegenüber Langzeitarbeitslosen.....	165
Tabelle 25: Vorhersagewerte der Regressionanalyse des Frame-Slektion-Modells nach Stocké auf die Fremdenfeindlichkeit.....	168
Tabelle 26: Vorhersagewerte der Regressionanalyse des Frame-Slektion-Modells nach Stocké auf die Fremdenfeindlichkeit getrennt nach Umfrageerfahrung.....	168

Abbildungsverzeichnis

Abbildung 1: Erweiterung der klassischen Testtheorie.....	12
Abbildung 2: Vergleich des Erklärungsmodells von Stocké und Skarbek-Kozietulska et al.....	50
Abbildung 3: Einflussfaktoren des SD-Bias nach dem Frame-Selektion-Modell	58
Abbildung 4: Erklärungsfaktoren und Operationalisierungen der Handlungskonsequenz „soziale Anerkennung“	67
Abbildung 5: Profile der Erwünschtheitswahrnehmung.....	91
Abbildung 6: Erklärungsfaktoren und Operationalisierungen der Handlungskonsequenz „Authentizität“	93
Abbildung 7: Wechselbeziehungen zwischen Person und Publikum im Sinne der Impression Management Theorie.....	95
Abbildung 8: Erklärungsfaktoren und Operationalisierungen des Framing-Modells.....	105

Abkürzungsverzeichnis

Abkürzung	Bezeichnung
AS	Antisemitismus
BIDR	Balanced Inventory of Social Desirability Responses
FT	Fremdtäuschung
ST	Selbsttäuschung
FF	Fremdenfeindlichkeit
IP	Islamophobie
LZA	Negative Einstellungen gegenüber Langzeitarbeitslosen
SÜ	Selbstüberwachung
FZSÜ	Fähigkeit zur Selbstüberwachung
POSÜ	Personen-Orientierung
SKSÜ	Selbstkonsistenz
MC	Marlow-Crowne
RC	Rational Choice
RRT	Randomized Response Technic
SD	Social Desirability
SES-17	Soziale-Erwünschtheits-Skala-17
SES-17_k	Kurzskala Soziale-Erwünschtheits-Skala-17
SEU	Subjectivly Expected Utility-Theorie

1 Einleitung

Sozial erwünschtes Antwortverhalten ist ein Phänomen im Rahmen von Interviews, mit dem sich die Forschung seit mehr als 60 Jahren auseinandersetzt. Im Allgemeinen versteht man unter sozialer Erwünschtheit¹ die Tendenz von Befragten sich gegenüber dem Interviewer, Dritten oder einer fiktiven Öffentlichkeit in einem „guten Licht“ zu präsentieren, insbesondere wenn sensible Themenbereiche angesprochen werden. Im extremsten Fall wird dabei bewusst die eigentliche Meinung zurückgehalten und stattdessen eine Antwort gegeben, die dem Handlungsziel der positiven Selbstdarstellung dient. Was ist nun der Sinn und Zweck der Messung der sozialen Erwünschtheit?

Neben der rein methodischen Fragestellung soll die Kontrolle des sozial erwünschten Antwortverhaltens in der Regel bei der Erhebung von sensitiven Items, eine unverzerrte Schätzung von Einstellungen, Merkmalen oder Verhaltensweisen ermöglichen. Die Anwendung ist dabei besonders in der Surveyforschung, bspw. in Form von Bevölkerungsumfragen, relevant. Seit den 1950er Jahren haben sich deshalb verschiedene Formen der Kontrolle und Vermeidung von sozial erwünschtem Antwortverhalten entwickelt. Neben Skalen, die die allgemeine Täuschungstendenz von Befragten erfassen sollen, gibt es eine Reihe von Techniken, die durch erhöhte Anonymität der Interviewsituation die Antwortverzerrungen verringern. Auf vielen Forschungsgebieten werden diese Skalen und Techniken seit Jahren eingesetzt, um Effekte des SD-Bias zu kontrollieren oder experimentell zu überprüfen (Familiensoziologie: Mehlkop/Neumann 2012; Umweltsoziologie: Preisendörfer 2011; Wohnortwechsel: Kalter 1997; Gesundheitsvorsorge: Jungbauer-Gans/Kriwy 2004).

Oft steht bei Umfragen jedoch nicht das sozial erwünschte Antwortverhalten im Vordergrund, sondern vielmehr ein inhaltliches Thema mit sensitivem Hintergrund (Gesundheit, Vorurteile, Umweltverhalten, usw.). Aus Kosten- und Zeitgründen ist damit der Platz für Kontrollvariablen wie der Messung des SD-Bias² beschränkt. Dies ist sicherlich auch ein Grund dafür, warum der Ansatz von Marlow und Crowne (MC), der sich nur auf einen Erklärungsfaktor (Anerkennungsbedürfnis) beschränkt, das am häufigsten eingesetzte Modell ist³. Der Nutzen dieses Modells ist aber fraglich. Neuere Modelle von Stocké, Skarbek-Kozietulska et al. und Wolter

1 Im Folgenden auch Social Desirability (SD).

2 Der SD-Bias bezeichnet die Differenz zwischen der unverzerrten und sozial erwünschten Antwort.

3 So erfolgt die Kontrolle des SD-Bias bspw. im Allbus 2006 sowie den Cella und GMF-Surveys ausschließlich über sogenannte Lügenskalen.

wurden bisher noch nicht ausreichend getestet, um bewerten zu können, ob sie eine bessere Alternative darstellen.

Das Forschungsfeld zu diesem Phänomen ist breit, aber genau hier liegt die Problematik. Viele der Studien beschäftigen sich lediglich mit isolierten Effekten des SD-Bias, ohne dabei weitere wichtige Einflussfaktoren zu berücksichtigen. Im deutschen Kontext gibt es nur eine geringe Anzahl an Forschungsarbeiten, die sich dem Thema aus theoretischer Perspektive annähern. Zu ihnen zählen die Dissertationen von Hartmann (1990), Reinecke (1991a) sowie die Arbeiten von Esser (1986a, 1986b, 1990, 1991). Im Anschluss an Esser entwickelten Stocké (2004), Skarbek-Kozietulska et al. (2012) und Wolter (2012) den rc-orientierten Ansatz in den letzten Jahren weiter. Bisher fehlt es an einer systematischen Aufarbeitung der über die letzten Jahrzehnte entwickelten Ansätze. Ein weiteres Problem, besonders der neueren Arbeiten von Stocké, Skarbek-Kozietulska et al. und Wolter, sind die teilweise widersprüchlichen Ergebnisse, obwohl deren Analysen sich auf die selben theoretischen Annahmen, nämlich den Rational-Choice-Ansatz (RC) nach Hartmut Esser (1986a, 1986b, 1990, 1991), stützen. Wie gezeigt werden wird, ist es fraglich, ob bedingt durch zentrale Unterschiede bei der abhängigen Variable und der Operationalisierung überhaupt dieselbe Form des sozial erwünschten Antwortverhaltens in den Analysen betrachtet wird. Die fehlende Übereinstimmung der Ergebnisse wäre damit nicht mehr überraschend, sondern vielmehr die Folge der Betrachtung des SD-Bias aus unterschiedlichen Sichtweisen. An beiden Analysen wird das zentrale Problem der bisherigen SD-Forschung deutlich: Die Ergebnisse der bisherigen Forschung zur Erklärung sozial erwünschten Antwortverhaltens sind äußerst heterogen und unsystematisch (vgl. Wolter 2012). Dies liegt möglicherweise daran, dass nur wenige der Studien konsequent theoriegeleitet angelegt wurden. Stattdessen werden Erhebungen durchgeführt, bei denen erst am Ende versucht wird, die Ergebnisse mit ad hoc Erklärungen zu begründen. Diese empiristische Form der Forschung hat den Nachteil, dass die Ergebnisse der Studien kaum miteinander vergleichbar sind, da Operationalisierungen und Erhebungsdesigns stark variieren. Außerdem werden aktuelle Forschungsergebnisse bei der Umsetzung neuer Erhebungen nicht berücksichtigt. In jedem Survey werden andere Skalen verwendet. Zum anderen variiert die Fragestellung in Form der abhängigen Variable, sowie der jeweilige Erhebungskontext. Dies erschwert den Vergleich der Ergebnisse. Es wäre daher dringend notwendig die bisher existierenden Operationalisierungsmöglichkeiten und Modelle systematisch miteinander zu vergleichen, um die Effektivität der einzelnen Konstrukte besser beurteilen zu können. Gerade bei einem Phänomen, das nur geringe Erklärungsbeiträge liefert, können schon kleinste Unterschiede ausschlaggebend sein. Obwohl bereits seit über 50 Jahren

Forschung betrieben wird, gibt es deshalb kaum Konsens über Ursache und Wirkungsweise der sozialen Erwünschtheit.

Die vorliegende Arbeit beschäftigt sich vor allem mit der Frage, wo diese Lücken liegen und welche Folgen sie für das Forschungsgebiet insgesamt haben. Zu Beginn der Arbeit (Kapitel 2) steht zunächst die Auseinandersetzung mit dem Begriff „soziale Erwünschtheit“. Bereits Hartmann (1990) stellte in ihrer Dissertation fest, dass mit „sozialer Erwünschtheit“ unterschiedliche Definitionen und Begriffe verbunden werden. Ein Großteil der Forschungsarbeiten auf diesem Gebiet bezieht sich zwar auf die klassische Testtheorie (siehe dazu exemplarisch Esser 1986a, 1986b, Stocké 2004), jedoch wird nur in seltenen Fällen eine exakte Definition des zu untersuchenden Phänomens vorgenommen. Bei genauer Auseinandersetzung mit der Thematik wird deutlich, dass das Phänomen der „sozialen Erwünschtheit“ eine Reihe unterschiedlicher Dimensionen umfasst, die berücksichtigt werden müssen. Je nachdem, welcher Aspekt betrachtet wird, verändert sich der Blickwinkel, aus dem heraus soziale Erwünschtheit analysiert wird. Um die Vergleichbarkeit der Forschungsergebnisse zu gewährleisten, ist es jedoch notwendig, das zu untersuchende Phänomen „soziale Erwünschtheit“ und seine jeweiligen Dimensionen so exakt wie möglich zu beschreiben.

Unter welchen Bedingungen tritt sozial erwünschtes Antwortverhalten auf? Verschiedene Ansätze aus der Ökonomie (Rational-Choice), Sozialpsychologie (geplantes und spontanes Verhalten) und Persönlichkeitspsychologie (Selbstüberwachung, Selbstaufmerksamkeit) wurden auf das Befragtenverhalten angewendet um Determinanten für den SD-Bias abzuleiten. In der vorliegenden Arbeit liegt in Kapitel 3 der Schwerpunkt auf den Erklärungsmodellen, die auf Basis des RC-Ansatzes entwickelt wurden⁴. Zunächst werde ich auf die Genese und Weiterentwicklung dieses Ansatzes im Rahmen der SD-Forschung eingehen, im Anschluss daran die Interpretation der einzelnen Faktoren betrachten, um dann darauf aufbauend die Vergleichbarkeit zwischen den einzelnen Forschungsarbeiten zu beleuchten. Darüber hinaus ist es zunächst notwendig, den Stand der Forschung zu diesem theoretischen Ansatz darzustellen. Da die aktuellen empirischen Umsetzungen des RC-Ansatzes sowie die theoretischen Grundlagen in Deutschland durchgeführt und veröffentlicht wurden, beschränke ich mich hier größtenteils auf die deutsche Forschung. Im ersten Schritt erfolgt die Darstellung des allgemeinen RC-Modells in seiner ursprünglichen Form, um anschließend seine Anwendungsformen zu diskutieren. Be-

4 Der sozialpsychologische Ansatz von Fishbein/Ajzen der von Reinecke (1991a) auf das sozial erwünschte Antwortverhalten angewendet wurde, wird hierbei nicht weiter berücksichtigt. Da es sich, wie von Kühnel (1993) dargestellt, bei der Theorie von Fishbein/Ajzen auch um eine Form der Wert Erwartungstheorie handelt, umfasst die Vorstellung der RC-Modelle auch diesen Ansatz bis zu einem gewissen Grad.

reits das Modell von Marlow und Crowne lässt sich als RC-Modell interpretieren, wenn auch die Autoren in ihrem Artikel von 1964 nicht explizit auf den RC-Ansatz Bezug nehmen (vgl. hierzu auch Hartmann 1990). Deutlich näher an den RC-Ansatz bewegen sich die Arbeiten von Hartmut Esser, der in seinen Artikeln die theoretischen Grundlagen für die Forschung der folgenden Jahre auf diesem Gebiet legte. Da seine Begrifflichkeiten teilweise von dem klassischen Ansatz abweichen, wird zunächst Essers allgemeines Verständnis des RC-Modells dargestellt, um darauf aufbauend seine Modellierung des Befragtenverhaltens zu charakterisieren. Stocké, Skarbek-Kozietulska et al. und Wolter haben diese Modellierung für ihre Forschungsarbeiten aufgegriffen und weiterentwickelt. Ziel dieses Kapitels ist vor dem Hintergrund der vergleichenden Darstellung, die Modelle einer kritischen Analyse zu unterziehen, die insbesondere auf Lücken und Problemen der Ansätze abzielt. Basierend auf den Modellen werden in Kapitel 4 die Operationalisierungsmöglichkeiten der zentralen Determinanten des SD-Bias vorgestellt und kritisch diskutiert. Gerade vor dem Hintergrund der Vergleichbarkeit der Ergebnisse ist dieser Schritt notwendig, denn trotz unterschiedlicher Skalen werden die Analysen oft unreflektiert miteinander gleichgesetzt. Es ist jedoch fraglich, ob die unterschiedlichen Operationalisierungsvarianten tatsächlich austauschbar sind.

In Kapitel 5 der Arbeit werden die bei der Aufarbeitung des theoretischen Stands der Forschung herausgearbeiteten Aspekte, empirisch überprüft. Zunächst erfolgt die Überprüfung der Operationalisierungsvarianten des „Anerkennungsbedürfnisses“, denn die SD-Skalen zählen zu den in der Forschung am häufigsten eingesetzten Operationalisierungen für die Messung sozial erwünschten Antwortverhaltens. In den letzten 50 Jahren Forschung wurden eine Reihe unterschiedlicher Skalen entwickelt, die bisher noch keiner systematischen Überprüfung unterzogen wurden. Ziel der geplanten Analysen soll es daher sein im ersten Schritt die Faktorenstruktur der wichtigsten Skalen zu überprüfen, um eine Aussage über die Reliabilität der Skalen treffen zu können. Bisher gibt es aber keine Studie die alle zentralen Skalen vergleichend überprüft. In der folgenden Analyse sollen deshalb die gemeinsamen Dimensionen bzw. unterschiedlichen Strukturen der Skalen im Vergleich miteinander getestet werden. Im Anschluss werden häufig verwendete Kurzversionen der Täuschungsskalen überprüft.

Die zentrale Frage für den zweiten empirischen Teil der vorliegenden Arbeit lautet: Welchen Beitrag leisten die RC-Modelle bei der Erklärung von Vorurteilen in Surveys und wie stabil sind die Effekte der einzelnen Faktoren? Die Fragestellung untergliedert sich in zwei Teilschritte:

1. Das Modell von Marlow und Crowne ist der am häufigsten in der Forschung verwendete Ansatz. Bisherige Analysen zeigten allerdings keine homogenen Ergebnisse bei der Kon-

trolle des SD-Bias auf Grundlage des Marlow-Crowne-Modells. Während in einigen Studien theoriekonforme Zusammenhänge Befunde auftraten, konnten in anderen Arbeiten die theoretischen Erwartungen nicht bestätigt werden (siehe bspw.: Rattaazzi et al. 2003; Heerwig/McCabe 2009; Duck/Hunsberger 1999). Darüber hinaus konnten in einigen Fällen Effekte konträr zur erwarteten Kausalrichtung festgestellt werden (siehe bspw. Stocké 2004). Für Deutschland ergibt sich die Forschungslücke, dass es zwar eine Reihe von Bevölkerungsumfragen gibt, die nach dem Prinzip von Marlow und Crowne das Anerkennungsbedürfnis erheben, bisher jedoch kaum Veröffentlichungen zu diesen Daten zu finden sind. Aus diesen Gründen erscheint es notwendig, die bereits vorhandenen Bevölkerungsdaten zu analysieren und die Erklärungskraft, unter Berücksichtigung der Kritik an der unsystematischen Forschung, des Modells von Marlow und Crowne zu überprüfen.

2. Auch die Modelle von Stocké, Skarbek-Kozietulska et al. sowie Wolter sind bisher nur unzureichend getestet. Für alle drei Modelle liegt jeweils nur eine Untersuchung vor. Da zudem zentrale Unterschiede in den Definitionen der sozialen Erwünschtheit bestehen, sind weder die Ergebnisse miteinander vergleichbar, noch lassen sich auf Grundlage der einzelnen Studien sichere Erkenntnisse über die Einflussfaktoren ableiten. Um die Theorien ausreichend zu testen, bedarf es einer größeren Anzahl an Analysen um die Reproduzierbarkeit der Ergebnisse zu überprüfen. Das Ziel der vorliegenden Arbeit ist es deshalb, an einem Datensatz die Modelle⁵ von Stocké und Skarbek-Kozietulska et al. auf Einstellungsfragen am Beispiel der Vorurteile gegenüber Minderheiten anzuwenden⁶, um die Stabilität sowie den Erklärungsbeitrag der Ansätze zu überprüfen.

Im abschließenden Fazit (Kapitel 6) werden auf Grundlage der theoretischen Vorarbeiten und unter Berücksichtigung der empirischen Befunde die Probleme bei der Erklärung sozial erwünschten Antwortverhaltens auf der Basis des RC-Ansatzes aufgezeigt und diskutiert. Ziel ist es, den Sinn und Nutzen der SD-Forschung zu bewerten und Schlussfolgerungen sowie Hinweise für folgende Analysen zu formulieren.

5 Unter die beiden Modelle fällt auch ihre jeweilige Erweiterung über das Modell der Frame-Selektion, die hier ebenfalls überprüft werden.

6 Das Modell von Wolter kann im Rahmen dieser Arbeit nicht überprüft werden, da in der reinen Onlinestichprobe das Randomized-Response-Experiment nicht umsetzbar war.

2 Definition des sozial erwünschten Antwortverhaltens

Die meisten Arbeiten, die sich mit der theoretischen Erklärung des SD-Bias im Rahmen des RC-Ansatzes beschäftigen, berufen sich auf die klassische Testtheorie (siehe dazu exemplarisch Esser 1986a, 1986b; Stocké 2004). Der SD-Bias wird dabei zunächst als systematische Antwortverzerrung in Rahmen von Befragungen definiert. Bereits 1991 beschäftigte sich Hartmann in ihrer Dissertation mit der Definitionsvielfalt und den Abgrenzungsproblemen in Bezug auf den SD-Bias. Sie trennt dabei zwischen zwei Definitionsebenen: Wie werden Verhaltensweisen, Einstellungen oder Merkmale als sozial erwünscht definiert und wie wird die Verzerrung durch soziale Erwünschtheit als solche näher bestimmt. Hartmann geht äußerst detailliert auf die unterschiedlichen Definitionsvarianten der damaligen Forschung ein und diskutiert ihre jeweiligen Grenzen. Im folgenden Abschnitt soll auf der Arbeit von Hartmann aufgebaut werden. Dabei werden die von Hartmann herausgearbeiteten Kritikpunkte nicht nochmals dargestellt, sondern vielmehr durch aktuelle Probleme insbesondere in Bezug auf die Frageformen und Präzision der Definitionen ergänzt.

2.1 SD-Bias im Rahmen der Klassische Testtheorie

Die immer wieder in der Literatur zu findende Verbindung von klassischer Testtheorie und der Erklärung des SD-Bias, lässt es erforderlich erscheinen, zunächst die klassische Testtheorie vorzustellen, um diese Form der Definition anschließend zu bewerten.

Die Diskussion der klassischen Testtheorie geht auf Gulliksen (1950) zurück. In seinem Buch „Theory of Mental Tests“ formuliert er verschiedene Axiome für die Güterkriterien psychometrischer Verfahren. Nach dieser Auffassung besteht jeder gemessene Wert im Rahmen einer Befragung aus dem wahren Wert sowie einer Fehlerkomponente. Das zentrale Axiom der klassischen Testtheorie lautet, dass es sich bei diesen Fehlerwerten um unkorrelierte Zufallsfehler⁷ handelt. Auf Grundlage dieser Annahme kann davon ausgegangen werden, dass sich mit steigender Zahl statistisch unabhängiger Messungen der Durchschnittswert der gemessenen Werte dem wahren Wert annähert (Diekmann 2010: 262).

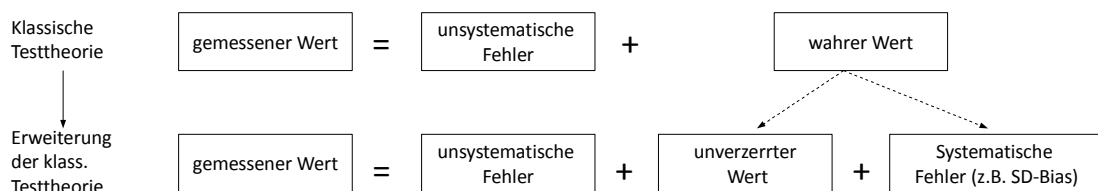
Bei systematischen Fehlern handelt es sich um keine zufälligen Verzerrungen, sondern um eine Korrelation zwischen wahren Wert und Messfehler, die auf einen Einfluss von wichtigen

7 Bei einem Zufallsfehler handelt es sich um einen Messfehler, dessen Erwartungswert gleich null ist und dessen Fehlerwerte voneinander und in Bezug auf das zumessende Konstrukt statistisch unabhängig sind (Diekmann 2010: 261ff.). Ob es sich bei der Unabhängigkeit voneinander um einen Grundbestandteil der Theorie handelt oder ein weiteres Axiom darstellt, ist durchaus umstritten.

Drittvariablen hinweist. Die klassische Testtheorie geht jedoch davon aus, dass diese Beziehung auf der Individualebene nicht besteht. Betrachtet man hingegen die wahren Werte von Populationen, ist es durchaus möglich, dass sich gegensätzliche systematische Verzerrungen gegenseitig aufheben und deshalb keine Beziehung zwischen Messfehler und wahren Wert zu finden ist. Würde man Verzerrungen wie den SD-Bias dem Fehlerwert der klassischen Testtheorie zuordnen, wäre das Axiom, dass es sich bei den Fehlerwerten um Zufallsfehler handelt, verletzt (Diekmann 2010: 264). Stattdessen werden systematische Fehler in der klassischen Testtheorie dem „wahren Wert“ zugeordnet. Sie haben damit zwar keinen Effekt auf die Reliabilität einer Messung, können sich aber negativ auf die Validität auswirken (Schumann 2010: 248). Gerade im Zusammenhang mit sensiblen Themen ist jedoch die Validität einer Messung entscheidend.

Bei sensiblen Themen, die in der Öffentlichkeit breit diskutiert werden und im Interesse konfligierender politischer Akteure liegen, ist die wahre Einstellung von Befragten von zentraler Bedeutung. Denn nur mit der wahren Einstellung lassen sich zuverlässige Aussagen über die Verteilung innerhalb der Bevölkerung treffen. Ein Beispiel ist hierfür das Thema Antisemitismus. Aufgrund der deutschen Geschichte gilt dieses Thema weiterhin als eine der wichtigsten Herausforderungen für die politische Bildung (Gebhardt et al. 2012: 7ff.). Unter anderem durch die Tabuisierung antisemitischer Positionierungen ist aber gerade die Äußerung von negativen Aussagen gegenüber Juden in der breiten Bevölkerung sozial unerwünscht. Aus diesem Grund werden antisemitischen Einstellungen im öffentlichen Raum und damit möglicherweise auch in Umfragen oft verheimlicht. Um einschätzen zu können, wie stark auch heute noch antisemitische Einstellungen, bspw. unter Jugendlichen, verbreitet sind, versucht man den wahren Wert der Einstellung (ohne systematische Verzerrungen wie der sozialen Erwünschtheit) möglichst genau zu messen. Gelingt dies nicht, ist die Validität der Daten gefährdet. Aufgrund dieser Problematik wurde die klassische Testtheorie um einen Messfehlerterm für systematische Effekte erweitert.

Abbildung 1: Erweiterung der klassischen Testtheorie



Quelle: Eigene Darstellung nach Diekmann (2008: 262)

Aus dieser Gleichung ergibt sich, dass der wahre Werte ein lineares Konzept darstellt. Außerdem wird implizit ein metrisches Skalenniveau der einzelnen Konstrukte unterstellt. Aus der Er-

weiterung der klassischen Testtheorie kann damit folgende Begriffsklärung abgeleitet werden: Bei der sozialen Erwünschtheit handelt es sich um einen systematischen Messfehler, der sich, im Gegensatz zu Zufallsfehlern, nicht durch unabhängige Messwiederholungen kontrollieren lässt. In der Methodenforschung versucht man, diesen Fehler schon bei der Datenerhebung zu vermeiden (bspw. durch einen hohen Grad an Anonymität) oder nachträglich durch zusätzliche Messungen (bspw. der Tendenz zu sozial erwünschten Verhalten) mathematisch zu kontrollieren. Wichtig ist an dieser Stelle festzuhalten, dass diese Interpretation des SD-Bias eben nicht, wie so oft behauptet (bspw. bei Reinecke 1991b: 293f.) der klassischen Testtheorie entspricht. Sie leitet sich nur aus einer modifizierten Version ab. Weiterhin setzt diese Definition eine wichtige Bedingung voraus: die Existenz eines wahren Wertes, der impliziert, dass es sich bei dem zu messenden Sachverhalt um ein stabiles und messbares Konstrukt handelt. Ob davon ausgegangen werden kann, dass diese Voraussetzungen immer erfüllt sind, darf zumindest als fraglich betrachtet werden. Das Problem des wahren Wertes in der praktischen Anwendung wird daher in Abschnitt 2.2.3 nochmals eingehen diskutiert.

Die Definition des SD-Bias als systematische Antwortverzerrung ist aus mehreren Gründen nicht hinreichend. Zum einen beschreibt sie nur die Stellung der Verzerrung im Kontext der klassischen Testtheorie, die jedoch, wie oben beschrieben, nur in ihrer modifizierten Form überhaupt auf den SD-Bias angewendet werden. Die klassische Testtheorie kann also allenfalls den Ausgangspunkt für eine Definition, aber nicht die Grundlage bilden. Zum anderen bleibt offen, welche Form der Verzerrung unter die Kategorie der sozialen Erwünschtheit fällt. Bisher fehlt es an einer allgemeingültigen Definition der sozialen Erwünschtheit, bestehende Begriffsklärungen sind, wie bereits DeMaio (1984: 257) feststellt, oftmals uninformativ und ungenau. In einigen Publikationen werden zudem sinnverwandte Begriffe verwendet (z.B. Lügertendenz und Anerkennungsbedürfnis), genauere Erläuterungen der Bezeichnungen unterbleiben oftmals (so z.B. bei: Batson et al. 1978, Duck/Hunsberger 1999, Couper et al. 2001, Coutts/Jann 2008, Hülshager et al. 2004)⁸. Im Folgenden werden verschiedene Problemfelder in Bezug auf die Definition des sozial erwünschten Antwortverhaltens aufgezeigt. Im Anschluss erfolgt daraus die Ableitung der zentralen Dimensionen einer präzisen Definition, die notwendig sind, um die verschiedenen Ebenen des Phänomens soziale Erwünschtheit differenziert zu betrachten.

8 Die gleiche Problematik beschrieb auch schon Hartmann (1991) vor 23 Jahren.

2.2 Problemfelder bisheriger Definitionsansätze

2.2.1 Soziale Erwünschtheit von Merkmalen, Verhaltensweisen und Einstellungen

Eine Gemeinsamkeit aller Definitionsversuche ist, dass die Antwortverzerrung in Richtung einer sozial erwünschten Norm erfolgt. Was genau ist sozial erwünscht? Ab wann entspricht ein Verhalten, ein Merkmal oder eine Einstellung der sozialen Norm? Dieser Teil der Definition müsste einen zentralen Bestandteil jeder Begriffsklärung des SD-Bias sein. Nur wenn aus der Definition deutlich wird, aus welcher normativen Grundlage sich die Verzerrung ergibt, kann sie präzise zu anderen Phänomenen abgegrenzt werden.

Eine Verzerrung vom unbeeinflussten true score in Form des SD-Bias setzt voraus, dass es Verhaltensweisen, Merkmale oder Einstellungen⁹ gibt, die als sozial erwünscht oder als sozial unerwünscht betrachtet werden. Soziale Erwünschtheit ist zunächst ein Bewertungskriterium, das auf Eigenschaften von Personen angewendet werden kann (Hartmann 1991: 38). Eigenschaften können dabei sowohl Verhaltensweisen als auch Einstellungen sein. Eine Voraussetzung dafür ist, dass jede Person individuelle Vorstellungen darüber hat, welche Eigenschaften erwünscht sind. Jedoch erst wenn diese Bewertung von einer Gruppe von Personen geteilt wird, ergeben diese Einschätzungen eine gesellschaftliche Zuschreibung von sozial erwünschten (z.B. Ehrlichkeit, Wahlteilnahme, Spenden) und sozial unerwünschten (z.B. Kriminalität, Fremdenfeindlichkeit, Alkohol- und Drogenmissbrauch) Verhaltensweisen bzw. Einstellungen. Groves definiert soziale Erwünschtheit als „a judgment about how highly valued a particular attribute is in the culture“ (Groves 2004: 437). Aber auch diese Definition ist nicht vollständig, denn es bleibt offen, was genau unter „culture“ zu verstehen ist. Zudem ist die Richtung der Bewertung (ob positiv oder negativ) nicht eindeutig. Gerade die Bestimmung dieser Erwünschtheitswahrnehmung ist jedoch für eine präzise Definition notwendig. Je nachdem, welches Kollektiv als „culture“ bzw. Bezugsrahmen für die soziale Erwünschtheit herangezogen wird, unterscheiden sich sowohl Effekt als auch Interpretation des jeweiligen SD-Bias. Zwei verschiedene Ansätze der Operationalisierung versuchen, sich der Messung anzunähern. Ausgangspunkt sind dabei jeweils individuelle Bewertungen (Hartmann 1991: 37f.).

1. Zum einen ist es möglich, zunächst die individuelle Erwünschtheitswahrnehmung einer Antwort für eine einzelne Person zu erheben und anschließend die verschiedenen individuellen Einschätzungen zu einem Kollektivwert zusammenzufassen (siehe Edwards 1967).

9 Im Folgenden wird aufgrund der Lesbarkeit auf eine vollständige Aufzählung aller drei Varianten verzichtet und stellvertretend nur noch die Einstellung genannt.

Ziel des Ansatzes ist es, mit dieser Technik einen Querschnitt über die Bewertung eines Sachverhaltes zu erhalten. Diese Form der Beurteilung wird auch als „normative social desirability“ bezeichnet (Taylor 1959 bei Hartmann 1991).

2. Die andere Variante besteht darin, die Perzeptionen der Individuen zu erfassen. Hierbei lässt man die Befragten einschätzen, wie die Mehrzahl der Bevölkerung über ein bestimmtes Verhalten oder eine Einstellung urteilen würde (Hartmann 1991: 40). Anstatt zu versuchen, die tatsächliche gesellschaftliche Meinung zu erfassen (d.h. über die individuelle Einstellung der Befragten), wird in diesem Fall die von den Befragten wahrgenommene Erwünschtheit einer Einstellung bei einer definierten Gruppe von Personen gemessen. Edwards (1967:35) bezeichnet diese Form der Bewertung als „personal desirability“. Hierbei kann zusätzlich unterschieden werden, ob die Befragten die Einstellung der Mehrheitsbevölkerung, ihrer Freunde oder Verwandten einschätzen sollen¹⁰.

Die beiden Konzepte haben verschiedene Konsequenzen: Bei der ersten Variante kann soziale Erwünschtheit nur entstehen, wenn es einen gesellschaftlichen Konsens über Normen gibt und Abweichung sanktioniert wird. Soziale Erwünschtheit im Sinne der zweiten Variante erfasst hingegen auch Phänomene, bei denen es keinen solchen Konsens gibt. Vielmehr ist hierbei entscheidend, welche Norm die Befragten selbst wahrnehmen. Die erste Variante bildet damit eine klassische soziologische Sichtweise ab. Auf Grundlage von Sanktionen werden soziale Normen durchgesetzt. Das Individuum wird durch den Druck der Gesellschaft an diese Normen gebunden, unabhängig davon, wie es die Norm wahrnimmt und bewertet. Die zweite Variante verfolgt hingegen einen eher konstruktivistischen Ansatz. Entsprechend dem „Thomas Theorem“¹¹ empfinden die Akteure hierbei einen sozialen Druck, wenn sie eine subjektive Norm wahrnehmen, unabhängig davon, ob diese Norm tatsächlich existiert. Was bedeutet diese Unterscheidung in Bezug auf die soziale Erwünschtheit?

Aus der Perspektive der klassischen Testtheorie stellt die sozial erwünschte Verzerrung einer Antwort einen systematischen Fehler im Rahmen von Befragung dar, die den eigentlich interessierenden Wert einer Aussage in Richtung einer tatsächlich existierenden bzw. einer wahrgenommenen Norm verzerrt. Für eine genaue Beschreibung der sozial erwünschten Antwortverzerrung muss zusätzlich eine Definition vorliegen, nach welchen Kriterien Einstellungen als sozial erwünscht bzw. sozial unerwünscht beurteilt werden. Handelt es sich um eine individuelle oder kollektive Bewertung und von welcher Bezugsgruppe geht die Normierung aus – dem In-

10 Eine umfangreiche Darstellung der möglichen Operationalisierungen findet sich in Kapitel 4.

11 Der zentrale Satz des Theorems lautet im Original „If men define situations as real, they are real in their consequences.“ (Thomas/Thomas 1928: 572)

interviewer, Freunden, Verwandte oder der Öffentlichkeit? Nicht über alle Eigenschaften gibt es einen allgemeinen Konsens. Es ist sinnvoll, im Rahmen einer Befragung, den Grad der Erwünschtheitswahrnehmung sowohl individuell als auch kollektiv zu erfassen, um anschließend eine differenzierte Aussage darüber treffen zu können, welche der abgefragten Einstellungen als sozial erwünscht oder unerwünscht gelten. Beide Definitionsansätze (Bewertungskriterium und Verzerrung) bilden eine Einheit und sind bei der Analyse des SD-Bias stets in Verbindung miteinander zu berücksichtigen.

2.2.2 Unterschiede zwischen Frageformen

Ein weiteres Problemfeld bezieht sich auf die Annahme des wahren Wertes und seine Übertragbarkeit auf unterschiedliche Frageinhalte. Fragebogen-Items lassen sich nach ihrem jeweiligen Inhalt in verschiedene Typen unterteilen. Es lassen sich dabei zwei grundsätzliche Formen unterscheiden: Handelt es sich um die Erfassung von Wissensbeständen oder soll der Befragte eine Bewertung abgeben? Bewertungsfragen beziehen sich damit zumeist auf die Erhebung von Einstellungen, Meinungen, Anschauungen und/oder Wertorientierungen. Bei den Wissensfragen steht hingegen die Erfassung von konkreten Merkmalen einer Person (z.B. Einkommen, Bildung, Alter, Geschlecht) oder retrospektiv berichteter Handlungen und Verhaltensweisen im Vordergrund. Zwischen diesen zwei Extremen finden sich Mischformen, wie bspw. die Fragen nach Handlungsintentionen. Hierbei wird sich zwar auf konkrete Situationen bezogen, da die Handlung jedoch noch nicht ausgeführt wurde, müssen sich die Befragten erst über ihre Intention klar werden. Dabei wird ein weiteres Unterscheidungsmerkmal deutlich: der zeitliche Bezug des Frageinhalts. Während Bewertungsfragen in der Regel auf weniger stabile Phänomene abstellen und die Antworten zumeist situationsabhängig generiert werden, erfassen Wissensfragen häufig Merkmale, die sich durch eine größere zeitliche Stabilität auszeichnen. Eine besondere Problematik von weniger stabilen und eher situativ generierten Positionierungen ergibt sich in Bezug auf soziale Erwünschtheit. Im Falle von sogenannten Non-Attitudes besitzen die Respondenten eigentlich keine feste Meinung zu einem bestimmten Thema, werden aber in der Befragungssituation dazu aufgefordert eine Meinung zu äußern. Es stellt sich die Frage, ob in diesem Fall ein SD-Bias entstehen kann oder ob eine Beeinflussung ausgeschlossen werden kann, da von vornherein kein true score existiert? Diese Frage ist bei der theoretischen Auseinandersetzung bisher unberücksichtigt geblieben. Für eine hinreichende Beantwortung wäre eine eigenständige theoretische sowie empirische Auseinandersetzung mit der Problematik der Non-Attitudes erforderlich.

Die klassische Testtheorie ist nicht der einzige Ansatz, der sich mit dem SD-Bias beschäftigt. In der Psychologie finden sich Arbeiten, die sich mit der Antwortverzerrung aus einem anderen Blickwinkel, unabhängig von einem wahren Wert, auseinandersetzen. Durch diese psychologischen Ansätze ist es möglich, den Entscheidungsprozess vor der Antwortäußerung differenzierter zu betrachten. Ein wichtiges Beispiel hierfür sind die Überlegungen zur Entstehung des SD-Bias im Rahmen des kognitiven Antwortprozesses.

Der kognitive Antwortprozess

Die Theorien zum kognitiven Antwortprozess fokussieren für die Bestimmung der Entstehung und Wirkung von Antworteffekten zunächst auf die Konkretisierung des kognitiven Prozesses der Befragung selbst. Die Grundlagen für die Entwicklung der einzelnen psychologischen Prozesse wurden bereits in den 20er Jahren mit den psychometrischen Theorien gelegt (Tourangeau et al. 2000: 3ff.). Darauf aufbauende Ansätze konkretisierten das Modell und seine einzelnen Komponenten¹².

Eine Zusammenfassung aktuellerer Ansätze findet sich bei Podsakoff et al. (2003), bei dem der kognitive Antwortprozess insgesamt fünf Phasen umfasst. In der ersten Phase ist die Aufmerksamkeit der Respondenten auf das Verstehen der Fragestellung gerichtet. Das Verständnis der Formulierung umfasst dabei die Erfassung der Logik der Frage sowie die Verbindung von Schlüsselwörtern mit relevanten Konzepten. In der zweiten Phase werden auf der Basis des verstandenen Inhalts mittels Stichwörtern und einer Abrufstrategie gezielt Informationen aus dem Langzeitgedächtnis erinnert. Da diese bloße Informationssammlung jedoch selten eine konkrete Antwort im Sinne der Frage möglich macht, ist die dritte Phase der Beurteilung notwendig. Hierbei wird die Vollständigkeit und Exaktheit der Informationen bewertet. Es werden, sofern erforderlich, Schlussfolgerungen, die eventuelle Lücken füllen, gezogen, eine allgemeine Beurteilung und eine Entscheidung hinsichtlich der Wissens-, Verhaltens- oder Einstellungsfragen getroffen. Die Abgleichung der eigenen Antwort mit den vorgegebenen Antwortmöglichkeiten findet dann anschließend in der vierten Phase der Antwortauswahl statt. In der letzten Phase wird die Antwort gegeben, wobei in dieser fünften Phase die Antwort nach bestimmten Kriterien editiert werden kann. Neben der Konsistenz können vor allem diese Fragen der Akzeptanz der Antwort im Sinne sozialer Erwünschtheit sein (Podsakoff et al. 2003: 885 f.). Während nach Podsakoff et al. (2003) der SD-Bias also ausschließlich in der fünften Phase zum Tragen kommt, gibt es nach Holtgraves (2004:162) drei Möglichkeiten, in denen es zu einem SD-Bias kommen kann:

12 Eine Übersicht über die verschiedenen Modelle findet sich bei Tourangeau et al. (2000: 3ff.).

1. Die erste Variante besteht darin, dass der Anreiz zur sozial erwünschten Antwort die Phase der Informationsgewinnung außer Kraft setzt. In diesem Fall zieht der Befragte nicht alle, ihm zu Verfügung stehenden Informationen, für die Beurteilung heran, sondern interpretiert den Sachverhalt ausschließlich vor dem Hintergrund sozial erwünschter Aspekte.
2. Die zweite Möglichkeit besteht darin, dass die soziale Erwünschtheit die Art und Weise der Informationssammlung beeinflusst. Die Befragten wählen ihre Informationen nach der Wirkung aus, die sie damit zu erzielen hoffen. Werden Personen z.B. zu ihren Persönlichkeitseigenschaften befragt, lügen sie also nicht per se, sondern wählen unbewusst oder bewusst Informationen über sich selbst aus, die sie in einem positiv Licht erscheinen lassen.
3. Nach Holtgraves kann es, wie im Modell von Podsakoff et al., auch erst in der Antworteditierung – der letzten Phase – zu einem SD-Bias kommen. In diesem Fall würde der Respondent zunächst Informationen zu der Frage sammeln, sortieren und bewerten. Im Rahmen der Evaluation müssten dann auch Aspekte einer als sozial erwünscht wahrgenommene Norm zur Beurteilung herangezogen werden.

Während es sich bei den ersten zwei Varianten um unbewusste Verzerrungen handelt, ist nur die Editierung durch den Befragten als „Lüge“ zu interpretieren. Denn nur in dieser Variante findet ein bewusstes Abwägen zwischen der generierten Meinung und der wahrgenommenen Norm statt. Um genauer zwischen beiden Varianten differenzieren zu können, werden an dieser Stelle die Bezeichnungen „soziale Beeinflussung“ (Eingriffe in Phase eins bis vier) und „sozial erwünschte Verzerrung“ (Eingriffe in Phase fünf) als Begrifflichkeiten vorgeschlagen. Für die Einschätzung des Geltungsbereichs und die Interpretation der Effekte von sozialer Erwünschtheit ist es notwendig, zwischen diesen beiden Formen klar zu unterscheiden. Denn während bei Ersterem eher unbewusste Verzerrungen berücksichtigt werden, basiert die zweite Form ausschließlich auf Täuschungen im Sinne einer bewussten Lüge.

Je nach Definition des SD-Bias sind sozial erwünschte Verzerrungen aber auch auf anderen Ebenen des Entscheidungsprozesses denkbar. Es ist denkbar, dass bereits das Frageverständnis und die Informationssammlung vom sozialen Kontext beeinflusst werden. Hängt die Verbindung von Schlüsselwörtern und Konzepten sowie die anschließende Informationssammlung nicht möglicherweise davon ab, von welcher Person die Frage gestellt wird und in welcher Situation das Interview stattfindet? Ist die Interpretation der Frage nach den persönlichen Stärken abhängig davon, ob sie ein potenzieller Arbeitgeber stellt oder ein Bekannter? Sicherlich

sind die Antworten in solchen Fällen nicht identisch, sondern orientieren sich an den Erwartungshaltungen, die dem Fragenden unterstellt werden. Wird die Verzerrung in Richtung auf sozial erwünschte Aspekte schon in Phase eins und zwei ausgelöst, weil die Befragten nach anderen Informationen suchen, oder wird erst in der dritten Phase die Menge von möglichen Antworten unter dem Aspekt der Angemessenheit beurteilt? Diese Frage ist zwar theoretisch interessant, jedoch für die Beurteilung des SD-Bias zunächst als eher zweitrangig zu bewerten. Welche Rolle spielt nun diese kognitionspsychologische Sicht für den theoretischen Kontext?

2.2.3 Konsequenzen für Wissens- und Beurteilungsfragen

Im Gegensatz zur klassischen Testtheorie geht der kognitionspsychologische Ansatz nicht per se von der Existenz eines eindeutig wahren Wertes aus. Schon der Abruf der Information sowie deren Bewertung kann durch soziale Effekte beeinflusst werden und damit in unterschiedlichen Situationen zu unterschiedlichen „wahren Antworten“ führen. Außerdem wird deutlich, dass die durch den Respondenten präferierte Antwort nicht zwingend zu den Skalen und Formulierungen der Umfrage passen muss. Im Folgenden soll die Konzeption des wahren Wertes vor dem Hintergrund des vorgestellten kognitionspsychologischen Ansatzes und mit Bezug auf Wissens- und Einstellungsfragen genauer herausgearbeitet werden.

Wissensfragen

Es wird üblicherweise davon ausgegangen, dass es sich bei den Antworten auf Wissensfragen um objektive Fakten handelt (bspw. Geschlecht, Alter, durchgeführte Handlungen). Das Konzept des wahren Wertes wird in Bezug auf Wissensfragen deshalb nicht kritisch diskutiert, da beim Vorliegen eines in der Regel beobachtbaren objektiven Wertes, keine wirkliche SD-Problematik entstehen kann. Gegen diese Position kann eingewendet werden, dass auch die Antworten auf Wissensfragen von den Befragten bis zu einem gewissen Grad generiert werden. Dieser kognitive Prozess ist nicht beobachtbar. Auch hier kann es bei der Informationssammlung und Bewertung zu Verzerrungen z.B. durch soziale Beeinflussungen kommen. Wie bereits erwähnt, wird dieser Bias jedoch nicht selbstverständlich durch das Phänomen der sozialen Erwünschtheit erfasst. Zwei Beispiele sollen dies verdeutlichen. Werden Respondenten nach ihrem Einkommen befragt, handelt es sich bei der Antwort um einen objektiven Fakt, der im Zweifelsfall nachgeprüft werden kann. Abweichende Antworten müssen jedoch nicht automatisch als Antwortverzerrungen im Sinne der sozialen Erwünschtheit angesehen werden. Vielmehr ist es möglich, dass eine befragte Person ihr Einkommen falsch erinnert. Fraglich ist nun, welches Einkommen in diesem Fall den unverzerrten Wert darstellt: Das tatsächliche Einkom-

men oder das Einkommen, das der Befragte zum Zeitpunkt der Erhebung glaubt zu erhalten? Dieselbe Problematik ergibt sich bei der Frage zu retrospektiven Verhaltensweisen. Bei retrospektiven Fragen können die Angaben der Befragten mit zuvor erhobenen amtlichen Daten verglichen werden. Solche Validierungsstudien werden bspw. in kriminologisch angelegten Untersuchungen eingesetzt (Heijden et al. 2000 sowie Köllisch & Oberwittler 2004, Skarbek-Kozietulska et al. 2012). In diesen Fällen ist der „wahre Wert“ also bereits vor der Erhebung bekannt. Die Antwortverzerrung im Sinne der sozialen Erwünschtheit wird dann als Differenz zwischen tatsächlichem Verhalten und berichtetem Verhalten definiert. Ob es sich hier allerdings wirklich immer um einen SD-Bias handelt, bleibt offen. Möglich sind bei retrospektiv angelegten Fragen Erinnerungsfehler, die durch soziale Beeinflussung oder anderes hervorgerufen sein können.

Bewertungsfragen

Anders ist es hingegen bei Bewertungsfragen. Bei dieser Frageform gibt es keinen objektiven Wert, mit dem die gegebene Antwort verglichen werden kann. Bei Bewertungsfragen muss das Konzept des wahren Wertes als noch unschärfer bewertet werden. In der Literatur lassen sich zwei grundlegende Definitionsansätze unterscheiden:

1. Ein Ansatz geht von der Existenz eines wahren Wertes auch bei Bewertungsfragen aus. Aus dieser Perspektive können im Idealfall die Bewertungen eines Einstellungsobjektes korrekt aus dem Gedächtnis abgerufen und entsprechend formuliert werden. Die Voraussetzungen hierfür ist die eindeutige Frageformulierung (Wilson et al. 1996: 105). Diese Interpretation liegt einer Reihe von Studien zugrunde, die sich mit dem Problem sozialer Erwünschtheit und Einstellungen befassen (z.B. Stocké 2004).
2. Der zweite Definitionsansatz geht davon aus, dass Bewertungsfragen keinen eindeutigen wahren Wert besitzen, da sie nicht extern beobachtbar und somit nicht nachprüfbar sind. Die Bewertung existiert demnach nur im Kopf der Befragten und kann damit zwar unklar oder inkonsistent, aber nie falsch sein (Bradburn et al. 2004: 117). Die Respondenten besitzen eine bestimmte Menge an Informationen zu einem Einstellungsobjekt. Die gegenwärtige Einstellung ist nur ein Ausschnitt aus diesen Informationen zu einem bestimmten Zeitpunkt. Bei dieser konstruktivistischen Sichtweise wird davon ausgegangen, dass es überhaupt keine wahre Bewertung gibt, sondern vielmehr viele Bewertungsbebenen, die durch den jeweiligen Kontext aktiviert werden. Bewertungen können sich damit aus affektiven, kognitiven oder verhaltensbezogenen Informationen ergeben (Wilson et al. 1996: 105f.).

Beide Definitionsansätze haben unterschiedliche Konsequenzen für die Interpretation des SD-Bias. Während man bei Ersterem nur von sozial erwünschter Antwortverzerrung spricht, wenn eine Abweichung von einem festen Wert beobachtet wird, kann eine wahrgenommene soziale Norm, aus der Perspektive des zweiten Ansatzes auch im Prozess der Antwortgenerierung eine entscheidende Rolle spielen. Im Rahmen des Rational-Choice-Ansatzes wird üblicherweise die erste Perspektive implizit vertreten. Auf die besondere Problematik bei Einstellungsfragen wird dabei nicht weiter eingegangen¹³. Vielmehr wird die Formulierung „wahrer Wert“ von den Autoren in Anführungsstriche gesetzt. Vermutlich mit dem Ziel die nicht ganz exakte Übertragbarkeit dieser Annahme für Einstellungen deutlich zu machen (siehe bspw. Stocké 2004).

Als Fazit lässt sich festhalten, dass die Existenz eines wahren Wertes in Bezug auf verschiedene Frageformen als durchaus problematisch bewertet werden muss. Für die Interpretation von Ergebnissen ist es notwendig, möglichst exakt zu definieren, was genau unter dem wahren Wert zu verstehen ist. Erst eine solche präzise Definition macht es möglich klar zu bestimmen, welche Phänomene als sozial erwünschte Antwortverzerrung zu betrachten sind.

2.2.4 Notwendigkeit präziser Definitionen

Eine Folge des unpräzisen Umgangs mit Definitionen ist, dass die verschiedenen Erklärungsansätze sich zwar auf dasselbe Wort „Soziale Erwünschtheit“ beziehen, dahinter jedoch unterschiedliche Konzepte stehen. Einige eher empiristische Autoren gehen noch einen Schritt weiter und sparen sich den Bezug zu einer theoretischen Grundlage völlig (siehe exemplarisch Banse/Gawronski 2003; Couper et al. 2003; Li/Bagger 2006). Ist das Ziel eine möglichst exakte empirische Auseinandersetzung mit einem Phänomen, so ist es unabdingbar, dass eine präzise Definition mit eindeutigen Begrifflichkeiten gewählt wird. Aus den vorgestellten Problematiken lassen sich vier Dimensionen ableiten, die für eine korrekte Begriffsklärung notwendig sind.

1. **Bezugsgruppen Dimension:** In den Ausführungen aus Abschnitt 2.2.1 ist deutlich geworden, dass Bestandteil der Definition sozialer Erwünschtheit ist, zu klären, was als sozial erwünscht verstanden wird. Dafür spielt die Bezugsgruppe für die soziale Norm eine entscheidende Rolle. Je nachdem welche Gruppe bzw. Ebene im Rahmen einer Analyse betrachtet wird, ergeben sich andere Interpretationen der Einflussfaktoren und Effekte. Ein zentraler Unterschied besteht bspw. darin, ob sich der Befragte nur an den Normen des Interviewers orientiert oder auch auf

13 Eine Ausnahme bildet die Diskussion der Implikationen des „wahren Wertes“ von Esser (1986a).

gesellschaftliche Bezüge zurückgreift. Während bei ersterer Variante der SD-Bias einen reinen Interviewereffekt darstellt, würde eine Analyse im Rahmen der zweiten Variante ein viel breiteres Feld der Antwortverzerrungen erfassen.

2. **Inhaltliche Dimension:** Zweitens ist es wichtig zu unterscheiden, auf welchen Sachverhalt sich der SD-Bias bezieht. Wie bereits dargestellt, ist die Differenzierung zwischen Bewertungs- und Wissensfragen hierbei von grundlegender Bedeutung. Handelt es sich bei dem abgefragten Sachverhalt um eine latente oder eine beobachtbare bzw. prüfbare Variable. Für beide Frageformen ergeben sich unterschiedliche Konsequenzen hinsichtlich sozial erwünschter Antwortverzerrung. Am problematischsten sind in diesem Zusammenhang Bewertungsfragen, zu denen auch die Abfrage von Einstellung zählt, einzuschätzen. Bewertungen lassen sich sowohl als konstante und exakt abrufbare Informationen betrachten als auch in Form von variablen Konzepten, die in Abhängigkeit von situationalen Einflüssen nach anderen Gesichtspunkten bewertet werden. Vor dem Hintergrund dieser Unterschiede zwischen den Fragetypen erscheint es sinnvoll, die theoretischen Erklärungsansätze nicht pauschal zu verallgemeinern. Eine Anwendung und Interpretation ohne Berücksichtigung der charakteristischen Unterschiede kann möglicherweise zu unbefriedigenden Ergebnissen führen, wenn je nach Definition andere Einflussfaktoren zum Tragen kommen.
3. **Kognitive Dimension:** Die dritte angesprochene Dimension betrifft die Form der Antwortverzerrung. Für eine präzise Definition des SD-Bias muss deutlich gemacht werden, ob nur bewusste oder auch unbewusste Verzerrungen in Richtung einer sozialen Norm berücksichtigt werden sollen. Eine differenzierte Betrachtung der kognitiven Abläufe, die im Rahmen des Antwortprozesses möglich sind, macht deutlich, dass zwischen der bewussten sozial erwünschten Verzerrung und der unbewussten sozialen Beeinflussung unterschieden werden muss. Für die Interpretation des Phänomens ist von entscheidender Bedeutung, ob auch unbewusste Verzerrungen als Response-Bias interpretiert werden oder ob diese als Bestandteil des wahren Wertes gelten sollen.
4. **Zeitliche Dimension:** Als letzte Dimension ist hier die Wirkung der Antwortverzerrung zu nennen. Bei einem Interview handelt es sich um eine soziale Situation. Im Rahmen einer solchen Situation kann es zu reaktivem Verhalten kommen, d.h. die Befragungssituation übt einen Einfluss auf die Antwort des Respondenten aus, in

dem er ihn bspw. zum Nachdenken über bestimmte Sachverhalte anregt. So kann auch der Einfluss der sozialen Norm erst durch die Interviewsituation zum Tragen kommen. Beeinflusst das Bewusstsein der Respondenten über die sozialen Normen ihre Einstellung langfristig, also über die Interviewsituation hinaus, oder zeigt die Norm nur einen kurzfristigen Effekt. Hier gilt es zu klären, ob nur kurzfristige Beeinflussungen oder auch langfristige Effekte als SD-Bias definiert werden sollen.

Eine Begriffsklärung, die einzig und allein auf die klassische Testtheorie Bezug nimmt und undifferenziert systematische Antwortverzerrungen behandelt, kann nicht als hinreichend bewertet werden. Denn wie die Differenzierung zwischen den verschiedenen Dimensionen gezeigt hat, gibt es nicht das eine sozial erwünschte Antwortverhalten. Der Begriff umfasst viele verschiedene Ebenen eines Phänomens. Aus einer präzisen Definition muss hervorgehen, welche Bezugsgruppen, inhaltlichen, kognitiven und zeitlichen Aspekte untersucht bzw. analysiert werden. Ansonsten sind die Vergleichbarkeit der Ergebnisse und deren sinnvolle Interpretation nur schwer möglich.

Forschungsstand zu den Definitionsansätzen

In den bisherigen Arbeiten zur sozialen Erwünschtheit wurde die Präzisierung in seltenen Fällen gar nicht und in der Regel eher unzureichend vorgenommen. Dazu einige exemplarische Beispiele.

In den Arbeiten von Batson et al. (1978) sowie Duck/Hunsberger (1999), die sich mit dem Einfluss der sozialen Erwünschtheit als Moderatorvariable bei dem Zusammenhang von Religiosität und Vorurteilen beschäftigen, werden nur die zu erwartenden Effekte der sozialen Erwünschtheit benannt. Eine Definition des Konstruktes fehlt völlig. Gleiches gilt für die Arbeiten von Couper et al. (2001) und Coutts/Jann (2008), die sensitive Fragen in Onlinesurveys untersuchen.

Neben diesen Untersuchungen, die gar keine Definition des Konstruktes vornehmen, gibt es auch Artikel, in denen die Erläuterung zur sozialen Erwünschtheit unpräzise bleiben, da sie nur begrenzt auf die oben genannten Dimensionen eingehen. Folgende Zitate können als exemplarische Beispiele für eine eher „unscharfe“ Begriffsklärung angeführt werden:

„[...] the tendency to provide answers which cause the respondent to look good.“
(Hancock/Flowers 2001: 6)

„Social desirability refers to a tendency to respond to self-report items in a manner that makes the respondent look good rather than to respond in an

accurate and truthful manner.“ (Holtgraves 2004: 161)

„The concept of social desirability rests on the notions that there are social norms governing some behaviors and attitudes and that people may misrepresent themselves to appear to comply with these norms.“ (Kreuter et al. 2008: 848)

„[...] Bias in the results of interviews or surveys when the respondents are trying to answer as ‘good’ people ‘should’ rather than in a way that reveals what they actually believe or feel.“ (Vogt 1999: 268 in Rossiter 2008: 3)

Die meisten dieser Definitionen führen an, dass es sich beim SD-Bias um eine Tendenz der Befragten handelt, sich in einem möglichst guten Licht darzustellen. Was unter „look good“ genau verstanden werden kann, bleibt entweder völlig offen (Hancock/Flowers 2001, Holtgraves 2004) oder wird auf soziale Normen (Kreuter et al. 2008) bzw., dass „was gute Leute tun sollten“ (Vogt 1999), zurückgeführt. Offen bleibt bei diesen Definitionen, welche Bezugsgruppe die sozialen Normen vorgibt, für welche Frageformen die Hypothesen abgeleitet wurden, ob es sich um bewusste oder unbewusste Verzerrungen handelt oder welche zeitliche Wirkung die Verzerrung hat.

Die umfangreichste deutschsprachige Arbeit zu diesem Thema findet sich bei Hartmann (1991). Sie geht ebenfalls auf die verschiedenen Ebenen des SD-Bias genauer ein. Hartmann arbeitet drei Dimensionen heraus: die soziale Erwünschtheit als Merkmal, als Antwortverzerrung und als individuell variierende Antworttendenz (Hartmann 1991: 235). Bei den Unterschieden zwischen den verschiedenen SD-Konzepten beschränkt sie sich jedoch hauptsächlich auf die Relevanz unterschiedlicher Bezugsgruppen, die die Norm vertreten. Aufbauend auf der Arbeit von Hartmann entwickelte Wolter (2012: 39 ff.) ebenfalls eine eigenständige Definition des SD-Bias, die auf drei Teildefinitionen beruht: Menschen besitzen ein Bedürfnis nach sozialer Anerkennung, dass zwischen den Individuen variiert. Zweitens nehmen Personen bestimmte Einstellungen oder Verhaltensweisen als sozial erwünscht oder sozial unerwünscht wahr. Sozial erwünschtes Antwortverhalten stellt darauf aufbauend die Abweichung des Antwortenden vom wahren Wert in Richtung der Erwünschtheitswahrnehmung dar. Ansonsten sind in der deutschsprachigen Forschung die Begriffsklärungen nur unzureichend:

Hülshager et al. (2004) umgehen eine Definition, in dem sie sich ausschließlich auf die Operationalisierung konzentrieren. Dabei werden die Begriffe soziale Erwünschtheit und Verfälschungstendenz synonym füreinander verwendet. Bei Stocké (2004) wird der SD-Bias zwar als systematische Antwortverzerrung des wahren Wertes beschrieben, auf die Problematik dieses Konzeptes bei Einstellungsfragen wird jedoch nicht näher eingegangen. Die Folge dieser unprä-

zisen Definitionen werden bei einem Vergleich der Forschungsarbeiten von Stocké (2004) und Skarbek-Kozietulska et al. (2012) deutlich. Obwohl beide denselben theoretischen Ansatz verwenden, kommen ihre Analysen zu unterschiedlichen Ergebnissen in Bezug auf die Erklärungsfaktoren für den SD-Bias. Während die Operationalisierung der Determinanten des SD-Bias bei den Studien weitestgehend gleich ist, beziehen sich die Erklärungsmodelle bei Stocké auf Einstellungselemente und bei Skarbek-Kozietulska et al. und Wolter auf Verhaltensfragen. Aufgrund der bereits diskutierten Unterschiede zwischen diesen beiden Frageformen, stellt sich aber die Frage, ob sich die theoretischen Vermutungen von Stocké überhaupt auf Verhaltensfragen übertragen lassen oder ob beide Arbeiten nicht vielmehr auf unterschiedliche Phänomene der sozialen Erwünschtheit Bezug nehmen und damit nicht direkt vergleichbar sind. Die jeweiligen Definitionen des Phänomens „soziale Erwünschtheit“ bleiben jedoch zu vage, um diese Fragen abschließend beantworten zu können.

Zusammenfassend lässt sich festhalten, dass eine klare Definition des Phänomens der sozialen Erwünschtheit und der damit verbundenen Begrifflichkeiten bisher in der Forschungsliteratur ignoriert wurde. Die genaue Eingrenzung des Untersuchungsgegenstandes gehört jedoch zu einem der ersten und wichtigsten Schritte im Rahmen einer empirischen Forschungsarbeit. Lücken und Ungenauigkeiten an dieser Stelle wirken sich auch auf Theorie, Untersuchung und die Ergebnisse aus. Im folgenden Kapitel wird der Stand der Forschung zur Anwendung des Rational-Choice-Ansatzes auf das Befragtenverhalten dargestellt. Im Hinblick auf die theoretischen Ansätze ist es ebenfalls von zentraler Bedeutung, welche Form der sozialen Erwünschtheit den Erklärungsmodellen zugrunde gelegt wird. Aus diesem Grund dienen die in diesem Abschnitt aufgestellten Kriterien auch als Bezugsrahmen für die Diskussion der folgenden Modelle.

3 Theoretischer Forschungsstand zur Anwendung des Rational-Choice-Ansatzes auf das Befragtenverhalten

Verschiedene Ansätze aus der Ökonomie (Rational-Choice), der Sozialpsychologie (geplantes und spontanes Verhalten) und der Persönlichkeitspsychologie (Selbstüberwachung, Selbstaufmerksamkeit) wurden auf das Befragtenverhalten angewendet um Determinanten für den SD-Bias abzuleiten. In der vorliegenden Arbeit liegt der Schwerpunkt auf den Erklärungsmodellen, die im Rahmen des RC-Ansatzes entstanden sind. Außerdem beschränkt sich die Darstellung, bis auf eine Ausnahme, auf den deutschen Forschungskontext.

Der Rational-Choice-Ansatz mit seiner Spezialisierung in Form der Subjectively Expected Utility (SEU) und der Frame-Selektion Theorie ist der in der deutschen quantitativen Forschung am häufigsten diskutierte und angewendete Erklärungsansatz für sozial erwünschtes Antwortverhalten (Esser 1986a, 1986b, 1990, 1991; Hartmann 1991; Stocké 2004; Skarbak-Kozietulska et al. 2012, Wolter 2012). Dabei basieren alle genannten Studien auf der von Hartmut Esser entwickelten Form des RC-Ansatzes, die Interpretation der einzelnen Elemente und deren Umsetzung ist jedoch eher als heterogen zu bewerten. Dies spiegelt sich insbesondere in den empirischen Ergebnissen zur Prüfung der jeweiligen Modellvariante wieder, die nicht nur widersprüchliche Befunde aufweisen, sondern durch die jeweilige Interpretation des Modells nur schwer vergleichbar sind (siehe bspw. Stocké 2004, Skarbak-Kozietulska et al. 2012 und Wolter 2012). Es ist deshalb notwendig, die wichtigsten Modelle getrennt voneinander zu betrachten. Zunächst wird das allgemeine Rational-Choice Modell sowie die Konkretisierung durch die SEU-Theorie vorgestellt. Im zweiten Schritt erfolgt die Darstellung der jeweiligen Modellinterpretation zur Erklärung des sozial erwünschten Antwortverhaltens in historischer Reihenfolge. Im Zentrum der Betrachtung stehen dabei das Modell von Marlow und Crowne (in der Interpretation von Hartmann) sowie das Grundmodell von Esser und seine Modifizierungen durch Stocké, Skarbak-Kozietulska et al. sowie Wolter. Bei Marlow und Crowne handelt es sich zwar nicht um deutsche Autoren, da ihr Erklärungsansatz jedoch zu den auch heute noch am häufigsten eingesetzten Modellen gehört, wird er in diesem Kapitel der Vollständigkeit halber ebenfalls behandelt. Alle fünf Modelle werden nach den Kriterien, die in Kapitel 2.2.4 aufgestellt wurden, bewertet. Insgesamt soll es die Vorgehensweise ermöglichen, Lücken und Problemfelder der einzelnen Ansätze herauszuarbeiten. Im Anschluss werden alle, aus den RC-Modellen abgeleiteten, Erklärungsfaktoren zusammenfassend dargestellt, um im folgenden Kapitel die Operationalisierungsvarianten vorzustellen.

3.1 RC-Ansatz und SEU-Theorie

Das Rational-Choice-Modell stellt einen allgemeinen Ansatz zur Erklärung sozialen Handelns dar. Ursprünglich nur in der Ökonomie verwendet, wird der RC-Ansatz auch immer häufiger für soziologische Fragestellungen genutzt (Familiensoziologie: Mehlkop/Neumann 2012; Umweltsoziologie: Preisendörfer 2011; Wohnortwechsel: Kalter 1997; Gesundheitsvorsorge: Jungbauer-Gans/Kriwy 2004). Grundkonzept des RC-Modells ist, dass Personen zum einen über Ressourcen verfügen und zum anderen Präferenzen in Bezug auf die zur Verfügung stehenden Handlungsoptionen¹⁴ haben (Diekmann/Voss 2004: 14f.). Darauf aufbauend besagt der RC-Ansatz, dass alle Handlungen das Ergebnis von Entscheidungen zwischen Alternativen darstellen. Die Akteure wählen die Alternative, die nach Abwägung der relevanten Vor- und Nachteile am günstigsten erscheint. Ein zentraler Bestandteil der Theorie ist damit die Entscheidungsregel, die determiniert, nach welchem Muster der Nutzen einer Handlung bestimmt wird. Die unterschiedlichen Theorien des RC-Ansatzes unterscheiden sich danach, ob bei dieser Entscheidungsregel nur bewusste oder auch unbewusste Prozesse berücksichtigt werden, ob der Entscheidungsprozess selbst Bestandteil der Abwägung ist und welche Bewertungskriterien zur Beurteilung herangezogen werden. Zumeist folgt die Entscheidungsregel dem Maximierungsprinzip, das Ausdruck in verschiedenen Theorien findet (Diekmann/Voss 2004: 15f.). Eine dieser Theorien ist die Maximierung des subjektiv erwarteten Nutzens (SEU-Theorie), bei der einzig die vermuteten Handlungskonsequenzen die Entscheidungen determinieren.

Die folgende Darstellung der SEU-Theorie soll es ermöglichen, Unterschiede und Gemeinsamkeiten der SD-Modelle von Marlow/Crowne, Esser, Stocké, Skarbek-Kozietulska et al. und Wolter herauszuarbeiten. Da alle Autoren auf unterschiedliche Begrifflichkeiten zurückgreifen, ist es notwendig allgemeine Definitionen der einzelnen Determinanten festzulegen, um die Vergleichbarkeit der Ansätze zu gewährleisten. In den anschließenden Kapiteln wird deshalb auf die folgende Darstellung der SEU-Theorie immer wieder Bezug genommen.

Nach der SEU-Theorie verfügt jeder Akteur über mindestens zwei **Handlungsalternativen (A_i)**. Sie stellen die in einer Situation wahrgenommenen Handlungsmöglichkeiten dar. Jede dieser Handlungsalternativen ist mit verschiedenen **Handlungskonsequenzen** (Ereignissen, die nach der Handlung eintreten können) verbunden. Diese Konsequenzen setzen sich multiplikativ aus zwei Aspekten zusammen: zum einen aus der **subjektiven Eintrittswahrscheinlichkeit (p_{ij})** der j-ten **Konsequenz** der i-ten Alternative und zum anderen aus der j-ten **Bewertung der Konsequenz (U_j)**. Diese Bewertungen können dabei einerseits positiv in Form von Nutzen, aber

14 Die Begriffe Handlungsoption und Handlungsalternative werden im Folgenden synonym verwendet.

auch negativ über die Kosten einer Handlung erfolgen. Der **Handlungsnutzen (SEU)** setzt sich aus der Summe der Produkte aus Wahrscheinlichkeiten und Bewertungen zusammen (vgl. Kühnel 1993b: 77f.).

$$SEU(i) = \sum_j p_{ij} * U_j \quad (3.1)$$

Die SEU-Theorie ist damit an drei Bedingungen geknüpft: Mathematische Voraussetzung ist, dass sowohl die Eintrittswahrscheinlichkeit als auch die Bewertung auf einer Ratio-Skala gemessen werden¹⁵. Inhaltliche Bedingung ist, dass der Akteur zwischen mindestens zwei Handlungsalternativen auswählen kann und jede dieser Alternativen mit mindestens einer Konsequenz verbunden ist, die sich in der Wahrscheinlichkeit und/oder der Bewertung von den anderen Alternativen unterscheidet.

Dieser allgemeine Ansatz der SEU-Theorie lässt sich auf das Befragtenverhalten anwenden. In einem Interview sieht sich der Respondent mit verschiedenen Handlungsalternativen konfrontiert. Bei geschlossenen Fragen besteht das sichtbare Set von Alternativen aus den möglichen Antwortoptionen und eventuellen Ausweichkategorien („weiß nicht“ und „keine Angabe“). Beim SD-Bias stehen die inhaltlichen Antwortalternativen im Mittelpunkt des Interesses¹⁶. Jede der Antwortoptionen ist mit einer oder mehreren Handlungskonsequenzen verknüpft. Berücksichtigt werden dabei nur solche Konsequenzen, bei denen die Eintrittswahrscheinlichkeit zwischen den Antwortoptionen variiert. Auch bei der Wahl einer Antwortoption entscheidet sich der Respondent für die Alternative mit dem größten Gesamtnutzen. Die SEU-Theorie setzt damit den idealen Akteur voraus, der in jeder Situation alle perzipierten Handlungskonsequenzen und Alternativen betrachtet. Im Rahmen einer Befragung gibt es verschiedene Handlungskonsequenzen, die von den Respondenten abgewogen werden können. Neben der Möglichkeit der Meinungsäußerung kann vor allem die Anerkennung durch den Interviewer oder Dritten eine wichtige Rolle bei Beantwortung der Fragen spielen. Wie diese allgemeinen Handlungskonsequenzen bewertet werden und welche Kriterien in die Berechnung der

15 Zunächst ist diese Festlegung auf die Ratio-Skala jedoch nur theoretisch gedacht. Die Problematik des Ratio-Niveaus ergibt sich erst, wenn die Determinanten tatsächlich gemessen werden sollen.

16 Es ist denkbar, dass Respondenten um eine sozial unerwünschte Antwort zu vermeiden auch auf die Non-Response Kategorien ausweichen. Gleichzeitig ist es möglich, dass Personen mit tatsächlicher Non-Attitude aus Gründen der sozialen Erwünschtheit eine inhaltliche Antwort geben, da sie es in der Befragungssituation als erwünscht empfinden. Im Zentrum der theoretischen Betrachtung steht jedoch die inhaltliche Verzerrung und nicht eventuelle Ausweichstrategien, denn es ist möglich, dass sich die Erklärungsfaktoren zwischen diesen beiden Varianten grundlegend unterscheiden. Aus diesem Grund konzentrieren sich die folgenden Ausführungen auf die inhaltlichen Verzerrungen und damit auf die Handlungsalternativen in Form der inhaltlichen Antwortoptionen.

subjektiven Eintrittswahrscheinlichkeiten eingehen, ist Bestandteil verschiedener Modelle zur Erklärung des SD-Bias.

Einer der ersten Versuche, die soziale Erwünschtheit im Rahmen eines Kosten-Nutzen Ansatzes zu interpretieren stammt von Crowne und Marlow (1964). Die Autoren benennen zwar nicht explizit die SEU-Theorie sondern im Gegenteil die Lerntheorie, ihre Definition des Konstruktes „soziale Erwünschtheit“ lässt sich jedoch, wie von Hartmann (1991: 124ff.) beschrieben, durchaus als RC-Modell interpretieren. Die wissenschaftliche Auseinandersetzung mit dem Befragtenverhalten im Rahmen der SEU-Theorie ist im deutschsprachigen Raum von Hartmut Esser geprägt. Im Folgenden sollen die Modellvarianten (Crowne/Marlow, Esser, Stocké, Skarbek-Kozietulska et al. sowie Wolter) zur Erklärung der Antwortverzerrungen im Rahmen des RC-Ansatzes vorgestellt und diskutiert werden.

3.2 Anwendungsmodelle zur Erklärung der sozial erwünschten Antwortverzerrung

3.2.1 Modell nach Crowne-Marlow

Eine der ersten theoretischen Arbeiten zur Erklärung des sozial erwünschten Antwortverhaltens stammt von Crowne und Marlow (1964). In „The Approval Motive“ leiten sie auf Basis einer Begriffsdefinition Erklärungsfaktoren für den SD-Bias ab. Im Zentrum ihrer Arbeit steht dabei die Entwicklung einer Skala zur Messung der Tendenz zur sozial erwünschten Antwortverzerrung. Als theoretische Grundlage wird nicht der RC-Ansatz, sondern die soziale Lerntheorie von Rotter (1960) herangezogen. In dieser Theorie ist das Verhalten von Befragten eine Funktion aus der Eintrittswahrscheinlichkeit des Ziels bei einer bestimmten Handlung und der Bedeutung der Situation für den Akteur. Die Nähe des lerntheoretischen Modells zum RC-Ansatz wird bei Crowne und Marlow dadurch deutlich, dass im Zentrum erlernte Erwartungen über den Zusammenhang von Handlung und Handlungsfolgen stehen (Crowne/Marlow 1964: 30f.).

Der zentrale Unterschied zwischen der Lerntheorie und dem Rational-Choice-Ansatz bezieht sich auf die zeitliche Komponente. Während der RC-Ansatz zukunftsorientiert ist, indem er den Akteuren unterstellt, Vor- und Nachteile bevorstehender Handlungen gegeneinander abzuwägen, ist die Lerntheorie an der Vergangenheit ausgerichtet, da bestimmte Handlungsalternativen basierend auf bisherigen Erfahrungen bewertet werden. Diese Unterscheidung zwischen Lern- und Handlungstheorie ist für die vorliegende Fragestellung zentral. Generell ist davon auszugehen, dass Befragte bei vielen Fragen keine bewusste Abwägung der Antwortalternati-

ven vornehmen, sondern lediglich erlernte Verhaltensweisen anwenden. Gerade wenn es um sensitive Themen geht, ist es möglich, dass Befragte aufgrund bisheriger Interviews die Erfahrung gemacht haben, dass der Interviewer auf sozial unerwünschte Antworten negativ reagiert. Ist dies dem Respondenten unangenehm, wird er in zukünftigen Situationen seine tatsächliche Meinung zurückhalten. Dabei wägt er nicht bei jeder Frage erneut Kosten und Nutzen ab, sondern orientiert sein Verhalten an seinen Erfahrungen. Auch hier kann das Verhalten durchaus auf rationalen Abwägungen basieren. Der Unterschied zur Handlungstheorie ist jedoch, dass der Abwägungsprozess auf Erfahrungen beruht, die in der Vergangenheit liegen¹⁷. Damit beinhaltet die Lerntheorie auch gleichzeitig einen wichtigen Mechanismus, der in RC-Modellen oft unterstellt wird (Lambert 2000: 188). Das Abwägen von Vor- und Nachteilen von Handlungsalternativen setzt voraus, dass die Akteure ein bestimmtes Wissen über die Situation und ihre Folgen besitzen. Dieses Wissen muss jeder Mensch basierend auf den Erfahrungen, die er in seinem Leben macht, erst erlernen. Die Lerntheorie kann deshalb als Wert-Erwartungsmodell, das situationsspezifische Erwartungen und Persönlichkeitszüge berücksichtigt (Krampen 2005), verstanden werden. Damit entspricht die soziale Lerntheorie von Rotter dem Grundgedanken der SEU-Theorie und kann somit als Variante des RC-Ansatzes aufgefasst werden (siehe Kunz 2004:14).

Wie genau stellt sich nun das Erklärungsmodell von Marlow und Crowne dar? Sie definieren soziale Erwünschtheit „as behavior motivated by a need for approval and the expectancy that approval can be attained by behaving in culturally acceptable ways [...]“ (Crowne/Marlow 1964: 40). Aus dieser Begriffsbestimmung lassen sich zwei Bedingungen für soziale Erwünschtheit ableiten: Zum einen muss der Akteur ein bestimmtes Bedürfnis nach sozialer Anerkennung haben und zum anderen muss eine Erwartung darüber bestehen, dass dieses Bedürfnis durch kulturell akzeptiertes Verhalten erlangt werden kann (Marlow/Crowne 1964: 40 Übersetzung nach Hartmann 1991: 125). Befragte unterscheiden sich in der Stärke ihrer Motivation, sich selbst in einem guten Licht darzustellen. Aus dieser Motivation lässt sich die Stärke des Bedürfnisses nach sozialer Anerkennung ableiten, dass eine grundsätzliche Tendenz beschreibt, sich sozialen Normen zu unterwerfen. Solche generalisierten Erwartungen, die auf Erfahrungen in ähnlichen Situationen beruhen, manifestieren sich in Persönlichkeitseigenschaften. Bezogen auf das Grundmodell der SEU-Theorie bedeutet dies, dass soziale Anerkennung eine Handlungskonsequenz darstellt, die bei allen Handlungsalternativen in einer Interviewsituation eine Rolle spielen kann. Die Relevanz des Nutzens der Handlungskonsequenz für die Handlungsalternative ergibt

17 Aus RC-theoretischer Sicht treten die erlernten positiven bzw. negativen Konsequenzen dann aber mit der Wahrscheinlichkeit 1 ein.

sich zum einen aus der Wahrscheinlichkeit, dass nach einer Handlung (in diesem Fall eine Antwortabgabe) soziale Anerkennung als Reaktion erwartet werden kann und zum anderen aus der Bewertung des Nutzens. Der Nutzen der Konsequenz ist Crowne und Marlow zufolge umso höher, je größer das Bedürfnis nach sozialer Anerkennung ist.

Diese theoretische Interpretation wird auch von Hartmann (1991: 124 ff.) geteilt. Das Modell von Crowne und Marlow beschreibt, obgleich es auf der Basis der Lerntheorie von Rotter konzipiert wurde, einen wichtigen Bestandteil der RC-Erklärung sozial erwünschten Antwortverhaltens: Die Befriedigung des Bedürfnisses nach sozialer Anerkennung. Unter der Annahme, dass das Anerkennungsbedürfnis die einzige perzipierte Konsequenz darstellt und die Wahrscheinlichkeit den Wert eins besitzt, würde ausschließlich die Tendenz das Antwortverhalten determinieren. Zur Kontrolle des SD-Bias entwickelten Crowne und Marlow eine Persönlichkeitsskala (MC-Skala) zur Messung des Anerkennungsbedürfnisses, die unabhängig von situationalen Einflüssen das sozial erwünschte Antwortverhalten erklären soll. Diese MC-Skala wurde in den letzten Jahrzehnten in verschiedensten Kontexten zur Kontrolle des sozial erwünschten Antwortverhaltens eingesetzt.

Stand der Forschung

Nur wenige Studien, die die MC-Skala verwenden, beziehen sich auf das konkrete theoretische Konstrukt, geschweige denn auf den RC-Ansatz. In vielen der Aufsätze wird einzig auf die Entwicklung der Täuschungsskalen beginnend bei Marlow und Crowne hingewiesen. Welche theoretischen Vermutungen mit diesen Skalen im Zusammenhang stehen, wird hingegen selten diskutiert. Im Folgenden sollen zusammenfassend wichtige Ergebnisse zum Einsatz der MC-Skala im Kontext der Vorurteilsforschung dargestellt werden. Die folgenden Studien beziehen sich zu meist nur implizit auf Vorarbeiten von Marlow und Crowne, da entweder die MC-Skala selbst oder eine ihrer Modifikationen in Form der BIDR (Balanced Inventory of Social Desirability Responses)- oder SES-17 Skala (Soziale-Erwünschtheits-Skala-17) eingesetzt wird.

Eine lange Forschungstradition findet sich in Bezug auf den Zusammenhang von Vorurteilen und Religiosität. Verschiedene Autoren vermuten, dass mit steigender intrinsischer Religiosität¹⁸ die negativen Stereotypen gegenüber out-groups rückläufig sind (Allport/Ross 1967). Duck und Hunsberger (1999) gehen davon aus, dass dieser Zusammenhang zu einem Großteil von der sozialen Erwünschtheit moderiert wird. Je größer die intrinsisch motivierte Religiosität ist, desto eher neigen Personen dazu Vorurteile, die gegen die Lehren ihrer Religion verstoßen, zu verschweigen. Um diese These zu überprüfen, erhoben sie bei einer Studierendenstichprobe

18 Im Gegensatz zur extrinsischen Religiosität stellt die intrinsische den verinnerlichten Glauben dar.

neben einer Skala zur religiösen Orientierung auch die Marlow-Crowne-Skala. Ihre Analyse konnte jedoch den erwarteten Zusammenhang der Täuschungsskala mit der Rassismus-Skala nicht bestätigen.

Rattaazzi et al. (2003) untersuchten den Zusammenhang zwischen subtilen und offenen Vorurteilmessungen mit der MC-Skala. Ihre Stichprobe setzte sich aus Schülern einer Highschool und Studierenden zusammen. Die Ergebnisse zeigten keinerlei Zusammenhang zwischen den geäußerten Vorurteilen und der Täuschungstendenz. Heerwig und McCabe (2009) beschäftigten sich mit dem Zusammenhang von Bildung und Toleranz. Sie vermuteten, wie schon Duck und Hunsberger in Bezug auf Religion, dass der positive Zusammenhang hauptsächlich durch die soziale Erwünschtheit moderiert wird. Dass das Phänomen hohe Bildung zu mehr Toleranz führt, wird dadurch vorhergesagt, dass hochgebildete Personen sich der sozial erwünschten Antwort bewusster sind. Ihre Analyse eines Listenexperiments konnte diese Annahmen jedoch nicht bestätigen. Einen geringen, aber signifikanten bivariaten Zusammenhang zwischen einer Kurzversion der MC-Skala und Rassismus bzw. religiöser Intoleranz konnten im Gegenteil Aosved et al. (2009) an einer Studierendenstichprobe feststellen.

Bewertung des Modells

Schon Hartmann bemängelt die intersubjektiv konstante Erwartung, von der Crowne und Marlow ausgehen. In ihrem Modell benennen Crowne und Marlow sowohl Bewertung als auch Erwartung, dem Anerkennungsbedürfnis wird jedoch letztlich der entscheidende Einfluss auf das sozial erwünschte Antwortverhalten zugeschrieben. Die Eintrittswahrscheinlichkeit spielt bei der Umsetzung und Konkretisierung des Konstruktes nur eine untergeordnete Rolle. Damit bleiben nicht nur alternative Handlungsoptionen unberücksichtigt, sondern auch die Variabilität der individuellen Erwartungen (Hartmann 1991: 126).

Crowne und Marlow haben damit eher ein rudimentäres Modell zur Erklärung des SD-Bias vorgelegt. Die theoretische Begründung der Autoren für die Einflussfaktoren spielt in den Ausführungen nur eine untergeordnete Rolle. In Kapitel 2.2.4 wurden Kriterien für eine präzisere Definition aufgestellt, nach denen sich alle unterschiedlichen SD-Modelle einordnen lassen müssten. Die Definition des Konstruktes bleibt bei Crowne und Marlow jedoch so vage, dass sich die Kriterien kaum anwenden lassen. In ihren Arbeiten finden sich keine Hinweise darauf, ob es sich nach ihrem Begriffsverständnis bei dem SD-Bias um eine Verzerrung oder auch eine Beeinflussung handelt, ob der Bias nur zeitlich instabile Veränderungen umfasst und ob ihr Modell auf alle Frageformen anwendbar ist. Ebenfalls bleibt offen, welche Bezugsgruppen für die Normwahrnehmung herangezogen werden können. In ihrer Definition sprechen sie zwar von

„culturally acceptable ways“ (Crowne/Marlow 1964: 40), was aber genau unter Kultur zu verstehen ist – auf wen sie sich bezieht und wie sie transportiert wird – darauf gehen die Autoren nicht weiter ein.

Obwohl die Unvollständigkeit des Modells von Crowne und Marlow bereits seit vielen Jahren kritisch diskutiert wird, ist es in der Forschung weiterhin der am häufigsten angewendete Ansatz zur Kontrolle sozial erwünschter Antwortverzerrungen. Die Studien zur Erklärung des SD-Bias bei Vorurteilsfragen über das Marlow und Crowne Modell zeigen aber kein konstantes Bild. Während sich in wenigen Studien theoriekonforme Zusammenhänge fanden, konnten in anderen Arbeiten die Erwartungen nicht bestätigt werden. Hinzu kommt, dass bei einer Reihe von Analysen die Skalen einen Effekt konträr zur erwarteten Kausalrichtung aufweisen. Auch die Umsetzung der Operationalisierung innerhalb der Studien ist zu kritisieren. Zwar wird die Problematik der sozial erwünschten Antwort mittlerweile in den meisten Forschungen zu sensiblen Bereichen thematisiert. Bei der Analyse werden jedoch häufig die jeweiligen Skalen als Kontrollvariable einfach kritiklos und ohne theoretischen Hintergrund berücksichtigt. Die meisten Messversuche beschränken sich auf die eindimensionalen Skalen nach Crowne und Marlow. Wie wichtig aber die Kontrolle weiterer Erklärungsfaktoren und Rahmenbedingungen für eine sinnvolle Messung und Interpretation des SD-Bias ist, zeigt die theoretische Weiterentwicklung des RC-Modells von Esser.

3.2.2 Modell nach Esser

In Deutschland wurde der Rational-Choice-Ansatz in der Soziologie vor allem von Hartmut Esser geprägt. Da er in seinen Arbeiten jedoch teilweise auf andere Begrifflichkeiten zurückgreift, soll an dieser Stelle nochmals kurz auf die Interpretation des allgemeinen RC-Ansatzes und seine Spezialform der SEU-Theorie nach Esser eingegangen werden¹⁹. Im zweiten Schritt erfolgt die Darstellung seines Modells zur Erklärung des Befragtenverhaltens.

In der klassischen SEU-Theorie wird von der Abwägung zwischen Kosten und Nutzen unter Berücksichtigung der Eintrittswahrscheinlichkeit gesprochen. Esser hingegen formuliert den Grundgedanken der Theorie wie folgt: „Personen wählen die ihnen vorstellbare Handlungsalternative, die am ehesten angesichts der vorfindbaren Situationsumstände bestimmte Ziele zu realisieren verspricht“ (Esser 1986a: 11). Abweichend von der klassischen SEU-Theorie verwendet Esser in seiner Definition die Begriffe Situation und Handlungsziele. Wie lassen sich diese

¹⁹ Die Interpretation des RC-Ansatzes nach Esser ist nicht kritiklos geblieben. Die Diskussion der Probleme und Lücken des Ansatzes reicht von älteren Arbeiten wie bei Opp et al. (1989) bis hin zu aktuellen Auseinandersetzungen (Kron 2004).

Begriffe mit dem klassischen SEU-Ansatz in Verbindung bringen? Für die Beantwortung dieser Frage ist es notwendig, die von Esser konzipierte Handlungstheorie genauer zu betrachten.

Die Handlungswahl selbst gliedert Esser in drei Schritte: Kognition, Evaluation und Selektion. Die Kognition ist von Erinnerungsprozessen des Akteurs und Assoziationsbildung geprägt (Esser 1991: 232). In dieser Phase kommt es zur Definition der Situation, aus der sich Handlungsalternativen und die sogenannten Ziele ableiten (Esser 1986a: 12). In der Phase der Evaluation wird dann eine Bewertung der einzelnen Handlungsalternativen vorgenommen. Dies erfolgt nach Esser durch die Verbindung der Handlungsziele mit den jeweiligen subjektiven Erwartungen, die als Wahrscheinlichkeiten interpretiert werden (Esser 1991: 232). Im Rahmen der Selektion wird im dritten Schritt auf Grundlage der berechneten SEU-Werte die Handlungsalternative mit dem größten subjektiven Nutzen ausgewählt (Esser 1990: 233).

Die Darstellung des RC-Modells von Esser macht deutlich, dass die hier verwendeten Formulierungen deutlich von vom ursprünglichen Rational-Choice-Ansatz abweichen. Anstelle der Begriffe Handlungsalternative, Handlungskonsequenz, Bewertung bzw. Handlungsnutzen und Realisierungswahrscheinlichkeit, spricht Esser von der Definition der Situation, Handlungszielen und Handlungstendenzen. Stellt Essers Modell damit im strengen Sinne noch ein RC-Modell dar und wenn ja, wie lassen sich beide Ansätze miteinander vereinbaren?

Zunächst ist die Verwendung des Begriffes Definition missverständlich. Als Definition der Situation bezeichnet Esser die subjektive Deutung von objektiven Merkmale, zur Reduktion der Komplexität der Umwelt (Esser 1999: 161). Damit stellt der Vorgang aber vielmehr eine Interpretation als nur eine Definition der Situation dar. Die Merkmale einer Situation müssen vom Akteur wahrgenommen und entsprechend evaluiert werden. Anschließend können Handlungsalternativen und ihre entsprechenden Konsequenzen abgeleitet werden.

Der Begriff Handlungsziel ist auf den ersten Blick kein Bestandteil der klassischen SEU-Theorie. Nach Esser ergeben sich die Ziele aus den Bedürfnissen, die der Akteur in einer bestimmten Situation verspürt. Über diese Bedürfnisse wiederum lassen sich konkrete Kosten und Nutzen ableiten. Damit sind nach Esser Handlungsziele als intendierte Konsequenzen zu verstehen, die sich durch eine Handlung ergeben. Hartmann geht in ihrer Auseinandersetzung mit Essers Modell ebenfalls auf die Problematik der Begrifflichkeiten näher ein. Sie vergleicht Handlungsziele mit perzipierten Handlungsfolgen²⁰. Beide Begriffe sind ihrer Meinung nach austauschbar, denn ein Handlungsziel erfasst sowohl das Erreichen von positiven als auch das Vermeiden von negativen Handlungsfolgen (Hartmann 1991: 127). Ein weiterer nicht ganz eindeutiger Begriff, den

²⁰ Folge ist hierbei als Synonym für Konsequenzen gemeint.

Esser verwendet, ist die Handlungstendenz. Er beschreibt damit das Produkt aus Bewertung und Wahrscheinlichkeit (Esser 1986a: 21). Aus seinen Arbeiten ist nicht zu ersichtlich, worin der Unterschied zwischen Handlungstendenz und Handlungsziel besteht.

Das RC-Modell von Esser entspricht trotz unterschiedlicher Begrifflichkeiten dem Grundgedanken des klassischen RC-Ansatzes. Zur besseren Vergleichbarkeit der Modelle sollen bei der folgenden Darstellung des Modells die Begrifflichkeiten der klassischen SEU-Theorie verwendet werden. Konkret bedeutet dies, dass die von Esser verwendeten Begriffe Handlungsziel und Handlungstendenz als Handlungskonsequenz interpretiert und benannt werden. Wie konkretisiert Esser nun diese allgemeine Handlungstheorie in Bezug auf das Antwortverhalten?

Die Anwendung des RC-Modells auf das Befragtenverhalten nach Esser

In einer Reihe von Veröffentlichungen hat Esser das allgemeine SEU-Modell im Hinblick auf das Antwortverhalten konkretisiert (1986a, 1986b, 1990, 1991). Esser beschreibt das Befragtenverhalten dahingehend,

„daß Befragte ihre Antworten „wählen“ und daß diese Wahl danach erfolgt, welche der (möglichen) Antwort-Alternativen für sie die (in einem weiten Sinne) günstigste Wahl darstellt. Diese Wahl hängt nach dieser theoretischen Darstellung davon ab, welche Vermutungen [...] die Befragten in Abhängigkeit der (sichtbaren oder erschlossenen) Merkmale der Befragungssituation über die Konsequenzen bestimmter Antworten haben; wie die Befragten die Situation in Bezug auf diese Konsequenzvermutungen einschätzen; und wie die Befragten diese Folgen vor dem Hintergrund ihrer Präferenzen bewerten. (Esser 1990: 232)“

Dem allgemeinen Modell folgend wird in der ersten Phase die Situation durch den Befragten interpretiert. Alltagstheorien werden durch sichtbare Merkmale des Interviewers, Rollenvorstellungen und den Kontext der Befragung aktiviert und so alle relevanten Handlungsalternativen und ihre Konsequenzen bestimmt. Bei den Handlungsalternativen handelt es sich in einem Interview mit geschlossenen Fragen um die Antwortoptionen, die der befragten Person vorgelegt werden, sowie die Möglichkeit des Interviewabbruchs. Mit jeder Antwortoption verbindet der Respondent Konsequenzen, d.h. Ereignisse, die auf seine Antwortabgabe folgen könnten. Im zweiten Schritt wird für jede dieser Konsequenzen die subjektive Eintrittswahrscheinlichkeit bestimmt und auf den Nutzen hin bewertet. Die Kalkulation der SEU-Werte und die anschließende Selektion stellen den zentralen Kern der Theorie dar, nach dessen Ergebnis die Antwortalternative mit dem höchsten subjektiven Nutzen ausgewählt wird (Esser 1991: 64f.). Welche Handlungskonsequenzen werden von Esser nun für die Interviewsituation konkret benannt?

In seinen Aufsätzen aus dem Jahr 1986 benennt Esser drei zentrale Handlungskonsequenzen, die im Rahmen einer Befragung zum Tragen kommen: das Bedürfnis nach persönlicher sowie kultureller Identität und die situationale Erwünschtheit. Die persönliche Identität bezeichnet die Bedeutung und Stärke des wahren Wertes der jeweiligen Einstellung und ihre Übereinstimmung mit den Antwortoptionen. Die kulturelle Identität basiert auf der Übereinstimmung der Antworten mit Rollenvorstellungen und Normen des Alltagsmilieus. Die situationale Erwünschtheit hingegen beschreibt generelle Bedürfnisse, die durch die Interviewsituation geprägt sind (Esser 1986a: 17f.). Seine weiteren Ausführungen beschränkt Esser auf die erste und dritte Handlungskonsequenz, da eine saubere Trennung zwischen persönlicher und kultureller Identität nicht möglich sei. Ausgehend von diesen Handlungszielen werden im Rahmen des Entscheidungsmodells Nutzen und Kosten abgeleitet. In jedem Interview stehen dem Befragten verschiedene Antwortalternativen (A1 und A2) zu Verfügung, die sich in diesem Beispiel auf die Antwortkategorien beziehen²¹. Jede Handlungsalternative ist über die verschiedenen Handlungskonsequenzen mit einer subjektiven Eintrittswahrscheinlichkeit verbunden (Esser 1986b: 322).

$$SEU(A_1) = p_{1T} * U_T + p_{1S} * U_S \quad (3.2)$$

$$SEU(A_2) = p_{2T} * U_T + p_{2S} * U_S$$

p_T = subjektive Wahrscheinlichkeit der Konsequenz „persönliche Identität“

U_T = Nutzen der Konsequenz „persönliche Identität“

p_S = subjektive Wahrscheinlichkeit der Konsequenz „soziale Anerkennung“

U_S = Nutzen der Konsequenz „soziale Anerkennung“

Im Zentrum der Theorie stehen zwei verschiedene Handlungstendenzen²²: die Tendenz eine „wahre“ Antwort zu geben und die Tendenz zur sozial erwünschten Antwort (Esser 1986b: 327). Die Tendenz zur „wahren“ Antwort ergibt sich aus dem Nutzen der Konsequenz „persönliche Identität“ (U_T) und der subjektiven Erwartung, dass die Einstellung innerhalb der Befragung auch geäußert werden kann²³. Die Tendenz zur sozial erwünschten Antwort hingegen wird aus der Intensität situationaler Bedürfnisse bspw. in Form des Anerkennungsbedürfnisses (U_S) und

21 Hierbei ist zu beachten, dass es in der Praxis in der Regel mindestens drei Optionen gibt. Die kürzeste Skala würde die Ablehnung und Zustimmung zu einer Aussage umfassen. Daneben hat der Respondent aber bei jeder Frage die Möglichkeit eine Antwort zu verweigern. Diese Non-Response Option stellt die dritte Handlungsalternative dar, die im Rahmen der hier vorgestellten Modelle aber nicht behandelt wird. Es ist davon auszugehen, dass das Ausweichen auf Non-Response Kategorien bei sensitiven Items durchaus ein relevantes Antwortverhalten darstellt. Die theoretischen Überlegungen zum SD-Bias konzentrieren sich jedoch ausschließlich auf die Erklärung von inhaltlichen Verzerrungen. Ein Modell zur Erklärung dieser Ausweichstrategie ist bisher nicht entwickelt worden

22 Die Handlungstendenzen beschreibt nach Esser das Produkt aus Ziel und Eintrittswahrscheinlichkeiten. Diese Definition ist damit gleichbedeutend mit den Handlungsnutzen.

23 Die Wahrscheinlichkeit ergibt sich vor allem durch die Stimuluseindeutigkeit, die sich aus präziser Fragestellung und den möglichen Antwortoptionen zusammensetzt.

der subjektiven Wahrscheinlichkeit von sozialen Konsequenzen gebildet (Esser 1986b: 327). Die Erwartung bestimmter sozialer Konsequenzen wiederum unterliegt nach Esser (1986a: 19) drei Bedingungen, die die Eintrittswahrscheinlichkeit beeinflussen:

- **Sichtbarkeit:** Die Befragten können soziale Konsequenzen nur wahrnehmen, wenn bestimmte Normen sichtbar sind.
- **Stereotypisierung:** Die Situationsmerkmale müssen mit unterschiedlichen Konsequenzenerwartungen verbunden sein.
- **Öffentlichkeit:** Die Antwortäußerung muss wahrgenommen werden können.

Für eine systematische Reaktion müssen alle drei Bedingungen gleichzeitig vorliegen. Während Esser die Sichtbarkeit und Stereotypisierung nicht weiter konkretisiert, lässt sich aus seinen Texten zumindest für die Öffentlichkeit auf eine dichotome Skalierung schließen. In seinem Aufsatz (1986a: 19) heißt es: „Wenn der Befragte annimmt, dass seine Reaktionen nicht öffentlich werden, dann sinkt entsprechend auch der Wert [der subjektiven Wahrscheinlichkeit].“ Der Formulierung entsprechend, handelt es sich um einen „entweder oder“ und nicht um einen „mehr oder weniger“ Zusammenhang.

Aus der Kombination der Konsequenzen in einer Befragung mit der subjektiven Eintrittswahrscheinlichkeit ergeben sich für jede Antwortalternative unterschiedliche SEU-Werte, die miteinander verglichen werden können. Der Befragte wählt in einem letzten Schritt die Alternative mit dem positivsten Gesamtwert aus. Je nachdem welcher der Handlungstendenzen der größere Gesamtnutzen zugeordnet wird, gibt der Befragte entweder eine „wahre“ Antwort oder äußert eine durch die soziale Erwünschtheit verzerrte Meinung. Betrachtet man die einzelnen Bedingungen genau, so wird jedoch deutlich, wie selten es nach der SEU-Theorie von Esser überhaupt zu einem SD-Bias kommen kann. Zum einen muss der Nutzen einer ehrlichen Antwort geringer gewertet werden, als der Nutzen der verzerrten Antwort. D.h. die Einstellungsstärke und damit die Verankerung der Meinung als Teil der persönlichen Identität, darf nicht zu hoch sein. Gleichzeitig muss der Befragte ein relativ hohes Anerkennungsbedürfnis in der Situation haben und soziale Konsequenzen (sowohl positive als auch negative) auf seine Antwort erwarten. Hierfür muss der Respondent nicht nur eine Norm wahrnehmen, sondern auch damit rechnen, dass der Interviewer auf die Normverletzung mit Konsequenzen reagiert. Letzteres ist nur zu erwarten, wenn eine gewisse Öffentlichkeit der Situation gegeben ist. Nur wenn alle diese Faktoren zusammenkommen, kann überhaupt mit Antwortverzerrungen gerechnet werden. Dabei ist zusätzlich zu beachten, dass auch dann nur eine Verzerrung zustande

kommt, wenn die situational induzierte Antwort nicht mit der „wahren“ Antwort übereinstimmt (Esser 1986b: 327).

Aus dem Modell von Esser lassen sich zusammenfassend zwei zentrale Handlungskonsequenzen für die Antwortverzerrung im Rahmen von sozialer Erwünschtheit ableiten: Die Förderung der persönlichen Identität und die soziale Anerkennung. Damit erhöht ein stark ausgeprägter persönlicher Bezug zum Befragungsthema in Verbindung mit einer hohen subjektiv erwarteten Wahrscheinlichkeit der adäquaten Meinungsäußerung die Kosten einer abweichenden Antwort und verringert den Nutzen und damit die Wahrscheinlichkeit eines SD-Bias. Dem gegenüber steht ein generelles Anerkennungsbedürfnis, das sich aus kulturellen oder situationalen Normen ergeben kann. In seinen Arbeiten diskutiert Esser, mit dem Hinweis, dass sich personale und kulturelle Identität nur schwer voneinander trennen lassen (Esser 1986a: 325), ausschließlich die situationale Erwünschtheit konkreter. Das vollständige theoretische Modell von Esser wurde bisher in noch keiner empirischen Studie vollständig überprüft.

Bewertung des Modells

Zusammenfassend lässt sich festhalten, dass Essers Modell zur Erklärung des Befragtenverhaltens auf Grundlage des RC-Ansatzes die Möglichkeit bietet, konkrete Einflussfaktoren und Bedingungen für den SD-Bias zu testen. Dennoch bleiben einige der verwendeten Begriffe äußerst vage und unpräzise. Dies zeigt sich deutlich bei der Benennung der drei Handlungskonsequenzen: persönliche Identität, kulturelle Identität und situationale Erwünschtheit. Zunächst bleibt festzuhalten, dass Esser diese drei Begriffe in den Aufsätzen nicht stringent verwendet. Während er in früheren Arbeiten (1986a, 1986b) die kulturelle Dimension eher der Identität zuordnet, spricht er später von kultureller und situationaler Erwünschtheit (Esser 1991: 69).

Es ist als problematisch zu bewerten, dass keiner der Bereiche tatsächlich eine Konsequenz, sondern vielmehr allgemeine Merkmale beschreibt, wie auch Hartmann bereits 1991 kritisierte (1991: 131). Hartmann versucht, die drei Determinanten zu konkretisieren: Die persönliche Identität erfasst demnach die Verankerung des Themas innerhalb der Persönlichkeit. Je stärker ausgeprägter eine Meinung ist, umso höhere Kosten verursacht eine abweichende Antwort. Hartmann bezeichnet diese Handlungskonsequenz als Vermeidung kostenverursachender Dissonanzen (1991: 131). Sowohl die kulturelle Identität als auch die situationale Erwünschtheit beziehen sich auf die Erlangung sozialer Anerkennung. Während sich Erstere auf allgemeine kulturelle Normen bezieht, werden bei der situationalen Erwünschtheit Merkmale der Situation, des Interviewers oder Dritter relevant (Hartmann 1991: 132). Damit lassen sich Essers Dimensionen nach Hartmann in drei Handlungskonsequenzen konkretisieren:

- Vermeidung kostenverursachender Dissonanzen (Ut),
- Soziale Anerkennung durch kulturelle/gesellschaftliche Normen und
- Soziale Anerkennung durch konkret angebbare Instanzen (Us).

Ein weiteres Problem ergibt sich aus den Voraussetzungen, die Esser benennt. Bedingungen außerhalb des Kosten-Nutzen-Kalküls und der Eintrittswahrscheinlichkeit sind nicht Bestandteil der SEU-Theorie, sondern vielmehr Hintergrundtheorien. Gerade für die Erklärung von Handlungen im Rahmen von sozialen Interaktionen, wie dem Interview, kann die soziale Wirklichkeit jedoch eine wichtige Größe darstellen. Der Befragte kann nur entsprechend einer kulturellen Norm handeln, die er auch wahrnimmt. Esser führt damit Prämissen ein, die notwendig sind, damit es überhaupt zu einer SEU konformen Abwägungen in Bezug auf sozial erwünschtes Antwortverhalten kommen kann. Die drei Bedingungen stellen bei Esser eine Vortheorie zur eigentlichen Handlungswahl dar, die es ermöglicht die soziale Wirklichkeit zu berücksichtigen. Nur bei Öffentlichkeit, Stereotypisierung und Sichtbarkeit kann die Handlungskonsequenz „Befriedigung des Anerkennungsbedürfnisses“ eintreten.

Problematisch ist hierbei, dass kulturelle Normen weitestgehend unberücksichtigt bleiben. Ist eine Befragungssituation anonym, kann es zu keiner Verzerrung kommen. Diese Schlussfolgerung bezieht sich jedoch ausschließlich auf das Verhältnis von Befragten zu Interviewer und die damit verbundenen Konsequenzen in der jeweiligen Situation. Fraglich ist hierbei, ob Anonymität einen Effekt auf die Wirkung gesellschaftlicher Normen besitzt. Folgt man der Konkretisierung durch Hartmann (s.o.), so wäre die Vermeidung von Dissonanz auch in anonymen Situationen denkbar. Für die Bewertung der subjektiven Wahrscheinlichkeit von Konsequenzen ist neben der Öffentlichkeit auch eine Sichtbarkeit und Stereotypisierung notwendig. Der Befragte muss eine bestimmte Norm wahrnehmen und damit eine Konsequenz verbinden. Beide Begriffe beschreiben das Konstrukt der Erwünschtheitswahrnehmung. Auch bei dieser Bedingung lässt sich jedoch nicht abschließend klären, ob sich Esser nur auf die Erwünschtheitswahrnehmung in der Situation (bspw. durch Merkmale oder Verhalten des Interviewers) bezieht oder ob diese Bedingung auch auf die kulturelle Erwünschtheit übertragbar ist. Während die Öffentlichkeit eindeutig eine situationale Bedingung darstellt, ist es denkbar, dass der Befragte sowohl gesellschaftliche als auch situationale Normen wahrnimmt und bewertet. Dabei ist es fraglich, welche Hierarchie zwischen beiden Erwünschtheitswahrnehmungen zu erwarten ist. Welcher Erwünschtheit würde ein Befragter folgen, wenn er mit gegensätzlichen Normen konfrontiert würde?

Der Schwerpunkt von Essers Ausführungen liegt auf dem Nutzen der Anerkennung. Die Einflussfaktoren im Rahmen der personalen Identität bleiben hingegen äußerst vage. Besonders problematisch ist dabei die subjektive Eintrittswahrscheinlichkeit, über welche der Nutzen der persönlichen Identität mit den Antwortoptionen verbunden wird. Sie ergibt sich nach Esser sowohl aus der Präzision der Frageformulierung wie auch aus der Passgenauigkeit der Antwortoptionen. Dabei wird jedoch nicht eindeutig klar, inwiefern sich aus diesen Faktoren eine Wahrscheinlichkeit ergeben soll. Die Präzision der Fragestellung ist für alle Antwortoptionen gleich. Hieraus lassen sich also keine variierenden Wahrscheinlichkeiten ableiten. Demnach bleibt nur die Passgenauigkeit, d.h., ob sich die eigene Einstellung in den Antwortvorgaben wiederfinden lässt. Anscheinend bezieht sich Esser hierbei auf die Passgenauigkeit der Antwortvorgaben zur wahren Einstellung. Dies erscheint paradox, da es ja gerade die wahre Einstellung ist, die der Befragte nicht preisgeben will. Der oft unpräzise Umgang mit Begriffen und Konstrukten erschwert eine Konkretisierung der einzelnen Faktoren. Gerade für die Operationalisierung und damit empirische Prüfung des Modells ist diese Konkretisierung jedoch notwendig. Durch die Ungenauigkeit einiger Bedingungen fällt eine Operationalisierung im Sinne von Esser schwer, da viel Raum für Interpretation bleibt.

Welche Ebene des SD-Bias versucht Esser nun, vor dem Hintergrund der in Kapitel 2.2.4 aufgestellten Kriterien einer Definition, in seinen Modellen zu erklären? Zunächst stellt sich die Frage, welche Bezugsgruppe für die Beurteilung der sozialen Norm herangezogen wird. Es ist nicht klar ob Esser zwischen kultureller und situationaler Erwünschtheit unterscheidet. Während bei der kulturellen Erwünschtheit gesamtgesellschaftliche Rollenerwartungen und Normen zur Beurteilung im Vordergrund stehen, spielen bei der situationalen Erwünschtheit die Merkmale und Verhaltensweisen des Interviewers oder dritter anwesender Personen eine Rolle. Bei seinen weiteren Ausführungen berücksichtigt Esser jedoch die erstere Form der Erwünschtheitswahrnehmung nicht weiter (vgl. Esser 1986a: 18). Dass es sich auf die Verzerrung durch anwesende Personen beschränkt, zeigt sich auch in den Rahmenbedingungen, die benannt werden, denn die Öffentlichkeit stellt einen in seinem Modell zentralen Faktor dar. „Erst bei Öffentlichkeit der Situation“, so Esser, „kann sich eine durch Sichtbarkeit, Stereotypisierung und situationale Interessen entstehende Erwünschtheitsneigung als Verhalten manifestieren“ (Esser 1986a: 19). Auf Grundlage dieser Einschränkung kann man davon ausgehen, dass sich das Modell von Esser nur auf Verzerrungen im Kontext von Interviewereffekten beschränkt.

Esser gibt keinen konkreten Hinweis auf den Geltungsbereich seines Modells in Bezug auf die inhaltliche Dimension. Fest steht, dass Esser Bewertungsfragen als Beispiel heranzieht. Ein weiterer Hinweis findet sich in seiner Diskussion über die Existenz des wahren Wertes in seinem Aufsatz „Können Befragte lügen?“ von 1986. Hierbei beschränkt er sich in der Diskussion auf die Stabilität von Einstellungen als Bestandteil der personalen Identität (Esser 1986a: 28ff.). Ob sein Modell vor diesem Hintergrund auch auf andere Frageformen wie bspw. berichtetes Verhalten oder Verhaltensintentionen anwendbar ist, bleibt fraglich.

Zur kognitiven Dimension (bewusste oder unbewusste Verzerrungen) finden sich keine eindeutigen Angaben in den vorliegenden Arbeiten. Es lässt sich nicht eindeutig sagen, ob Esser auch Formen der Selbsttäuschung unter die Begrifflichkeit des SD-Bias fasst. Ausgehend von dem allgemeinen RC-Ansatz können jedoch Rückschlüsse auf diese Dimensionen gezogen werden. Bereits bei dem Modell von Marlow und Crowne wurde der Unterschied zwischen Lerntheorie und Handlungstheorie hervorgehoben. Das Ergebnis, die Reaktion des Befragten, ist nach beiden Ansätzen identisch. Die Unterschiede sind eher auf der kognitiven Ebene zu verorten. Während in der Verhaltenstheorie die Abwägung in der Vergangenheit liegt und damit vor allem ritualisierte und automatisierte Prozesse beschrieben werden, findet bei der rationalen Wahl eine bewusst und aktuelle Abwägung der Handlungsalternativen statt. Basierend auf dieser Auffassung ist anzunehmen, dass sich Essers RC-Modell zur Erklärung sozial erwünschten Antwortverhaltens nur auf bewusste Formen der Verzerrung bezieht. Auf die zeitliche Dimension des SD-Bias, d.h., ob die Verzerrung im Sinne des SD-Bias die Einstellung auch über die Befragungssituation hinaus beeinflusst, geht Esser in seinen Aufsätzen nicht ein²⁴.

Es bleibt damit festzuhalten, dass die in Kapitel 2.2.4 aufgestellten Kriterien einer präzisen Definition bei Esser nicht erfüllt sind. Es finden sich einige Hinweise darauf, dass das Modell auf bewusste Verzerrungen durch Interviewereffekte von Bewertungsfragen beschränkt ist. Die zeitliche Komponente bleibt vollständig unbeachtet. Damit ist nicht eindeutig feststellbar, welche Ebene des SD-Bias sich durch sein Modell erklären lässt. Es ergibt sich ein Interpretationsspielraum, der dazu führt, dass der Ansatz für sehr verschiedene Studien herangezogen wird und die Ergebnisse miteinander verglichen werden, obwohl Zweifel an der Vergleichbarkeit angebracht sind, wie die Arbeiten von Stocke (2001, 2004), Skarbek-Kozietulska et al. (2012) und Wolter (2012) zeigen. Da Esser das Modell nie selbst empirisch überprüft hat, können zur Einschätzung nur die modifizierten Ansätze der drei genannten Autoren überprüft werden. Dass

²⁴ Es ist zu beachten, dass die bisherige Unterscheidung von RC- und Lerntheorie sich auf die Vergangenheit bzw. aktuelle Handlungssituation bezieht. Welche Folgen der Abwägungsprozess für die zukünftigen Entscheidungen hat, d.h. ob die soziale Norm sich langfristig auf die Einstellung auswirkt, bleibt hingegen offen.

die Studien²⁵ zu äußerst unterschiedlichen Ergebnissen kommen (besonders zwischen Stocké und Wolter), ist vor dem Hintergrund der eben diskutierten Schwächen und des daraus resultierenden unklaren Geltungsbereichs, wenig verwunderlich. Die Modifikationen des Ansatzes sind teilweise so grundlegend, dass es sinnvoll erscheint, die beiden Varianten im Rahmen der vorliegenden Arbeit als eigene Modelle vorzustellen.

3.2.3 Modell nach Stocké

Die theoretischen Arbeiten von Esser zur sozialen Erwünschtheit wurden von Stocké (2001, 2004) aufgegriffen und auf fremdenfeindliche Einstellungen angewendet. Stocké definiert den SD-Bias wie auch schon Esser als systematische Antwortverzerrung des wahren Wertes im Rahmen von Befragungen (Stocké 2004: 303). Im Mittelpunkt von Interviews stehen demnach die zwei Konsequenzen soziale Anerkennung und Authentizität. Kosten bzw. Nutzen dieser Konsequenzen werden zum einen über das Bedürfnis nach sozialer Anerkennung (U_S) und zum anderen über das Bedürfnis nach Authentizität (U_T) bestimmt. Während das Konzept des Bedürfnisses nach sozialer Anerkennung direkt von Esser übernommen wurde, ist zu vermuten, dass sich das Bedürfnis nach Authentizität aus der Vermeidung kostenverursachender Dissonanzen (siehe obige Diskussion Hartmann 1991) ergibt. Das Anerkennungsbedürfnis ist mit der möglichen Antwortoption, wie von Esser beschrieben, über eine subjektive Eintrittswahrscheinlichkeit verbunden. Letztere ergibt sich nach Stocké durch die Privatheit der Situation (p_A). Eine Handlung wie z.B. die Abgabe einer Antwort ist nur dann mit Konsequenzen verbunden, wenn sie in der sozialen Umgebung wahrgenommen wird (Stocké 2001: 5). Eine Voraussetzung dafür, dass der Befragte Anerkennung für eine bestimmte Antwortoption erhält, ist darüber hinaus die Erwünschtheitswahrnehmung. Während Esser diesen Faktor als Rahmenbedingung für subjektive Eintrittswahrscheinlichkeit bewertet, wird er im Modell von Stocké als Moderatorvariable interpretiert, die das Anerkennungsbedürfnis gewichtet (w_{EW}) (Stocké 2001: 5).

Das Bedürfnis nach Authentizität ergibt sich, aus dem generellen Motiv, Aspekte der persönlichen Identität ehrlich zu äußern. Die subjektive Eintrittswahrscheinlichkeit für diesen Nutzenfaktor wird durch die Zugänglichkeit einer Einstellung bestimmt (p_{ES}). Je stärker eine Einstellung in der Persönlichkeit verankert ist, umso größer sind die Kosten bei der Äußerung einer nicht authentischen Antwort. Auch dieses Handlungsziel wird von einer Moderatorvariable, nämlich der Eindeutigkeit und Passgenauigkeit der Antwortoptionen (w_P), beeinflusst (Stocké 2001: 5).

25 Die Artikel von Stocké basieren auf derselben Untersuchung. Der wesentliche Unterschied zwischen beiden Arbeiten besteht darin, dass in dem Artikel aus dem Jahr 2001 der Theorieabschnitt deutlich umfangreicher ausfällt.

Die Passgenauigkeit bezieht sich dabei auf die Übereinstimmung der eigenen Meinung mit den vorhandenen Antwortoptionen. Dieser Punkt stellt eine Modifikation des Modells von Esser dar. Bei Esser wird der Nutzen der Authentizität über die Zugänglichkeit der Einstellung und die subjektive Wahrscheinlichkeit durch die Stimuluseindeutigkeit definiert. Bei Stocké ergibt sich der Nutzen der Antwortoption nicht über die Zugänglichkeit, sondern über das Bedürfnis nach Authentizität. Die Einstellungsstärke bedingt nach Stocké die subjektive Wahrscheinlichkeit, die jedoch für alle Antwortoptionen gleich ist. Die unterschiedlichen Nutzenterme ergeben sich in seinem Modell durch die Passgenauigkeit der Antwortoptionen. Der Nutzen einer authentischen Antwort steigt mit der empfundenen Dissonanz bei einer Vermeidung der wahren Antwort. Die Berechnung der SEU-Werte der Antwortoptionen wurde auf Grundlage dieser Theorie von Stocké (2001: 6) wie folgt formalisiert:

$$SEU(A_1) = (p_A * w_{1,EW} * U_S) + (p_{ES} * w_{1,P} * U_T) \quad (3.3)$$

$$SEU(A_2) = (p_A * w_{2,EW} * U_S) + (p_{ES} * w_{2,P} * U_T)$$

p_A = Privatheit der Situation

U_S = Bedürfnis nach sozialer Anerkennung

w_{EW} = Erwünschtheitswahrnehmung

p_{ES} = Zugänglichkeit der Einstellung

U_T = Bedürfnis nach Authentizität

w_P = Eindeutigkeit und Passgenauigkeit der Antwortoptionen

Im Gegensatz zum klassischen RC-Ansatz geht Stocké jedoch nicht von einem deterministischen, sondern von einem probabilistischen Entscheidungsmodell aus. Die Wahrscheinlichkeit für die Antwortoption (A_1) ist damit die Funktion aus der Differenz des Handlungsnutzens bei der Antwortmöglichkeiten. Da sowohl Nutzen als auch subjektive Eintrittswahrscheinlichkeit für beide Antworten gleich sind, ergibt sich folgende Gleichung:

$$p(A_1) = f(\delta w_{1-2,EW} * p_A * U_S + \delta w_{1-2,P} * p_{ES} * U_T) \quad (3.4)$$

Exkurs: Beispiel zur Berechnung der Wahrscheinlichkeit eines SD-Bias

Gegeben sei eine Interviewsituation, in der ein Befragter die Möglichkeit hat, zwischen zwei Antwortoptionen zu wählen: Antwortoption 1 (A_1) entspricht dabei der wahren Meinung, Antwortoption 2 (A_2) würde von dieser abweichen und die sozial erwünschte Meinung darstellen. Der Befragte hat auf einer 10er Skala ein hohes Bedürfnis nach sozialer Anerkennung ($U_S=8$). Das Bedürfnis nach Authentizität ist hingegen auf derselben Skala nur schwach ausgeprägt ($U_T=2$)²⁶. Da es sich um ein Face-to-Face Interview handelt, ist die subjektive Wahrnehmung von sozialen Konsequenzen gleich 1 ($p_A=1$). Die Einstellung ist mit $p_{ES}=0.4$ eher schwach verankert. Diese vier Faktoren gehen nach Stocké bei jeder der

26 Die subjektive Wahrscheinlichkeit kann einen Wert zwischen 0 und 1 annehmen.

beiden Antwortoptionen mit dem gleichen Wert in die Berechnung des SEU-Wertes ein. Variiert wird nur die Moderatorvariable. Mit dem wahren Wert (A_1) verbindet der Befragte eine schwächere Erwünschtheitswahrnehmung als mit der sozial erwünschten Antwort (A_2). Die Passgenauigkeit der gewählten Antwort mit dem wahren Wert ist ebenfalls zwischen beiden Optionen unterschiedlich. Entschieden sich der Befragte für die wahre Antwort, liegt die Passgenauigkeit höher, als bei der Lüge (für beide Variablen liegt der Wertebereich zwischen -1 und +1, wobei +1 gleichbedeutend ist mit einer hohen Passgenauigkeit bzw. Erwünschtheitswahrnehmung). Die Werte werden anschließend in die allgemeine Formel eingeben und berechnet.

Formel

$$SEU(A_1) = (p_A * w_{1,EW} * U_S) + (p_{ES} * w_{1,P} * U_T)$$

$$SEU(A_2) = (p_A * w_{2,EW} * U_S) + (p_{ES} * w_{2,P} * U_T)$$

$$p(A_1) = f(\delta w_{1-2,EW} * p_A * U_S + \delta w_{1-2,P} * p_{ES} * U_T)$$

Anwendungsbeispiel

$$SEU(A_1) = (1 * w_{1,EW} * 8) + (0.4 * w_{1,P} * 2)$$

$$SEU(A_2) = (1 * w_{2,EW} * 8) + (0.4 * w_{2,P} * 2)$$

$$SEU(A_1) = (1 * (-0.7) * 8) + (0.4 * 0.8 * 2)$$

$$SEU(A_2) = (1 * 0.9 * 8) + (0.4 * (-0.4) * 2)$$

$$p(A_1) = \delta w_{1-2,EW} * 1 * 8 + \delta w_{1-2,P} * 0.4 * 2$$

$$p(A_1) = (-0.7 - 0.9) * 1 * 8 + (0.8 + 0.4) * 0.4 * 2$$

$$p(A_1) = -12.8 + 3.6$$

$$p(A_1) = -9.2$$

Der Gesamtnutzen variiert zwischen -10 und +10. Ein Wert von -9.2 bedeutet, dass für den Befragten eine hohe Wahrscheinlichkeit besteht, die zweite sozial erwünschte Antwortoption.

Ausgehend von diesem theoretischen Modell, leitet Stocké konkrete Erwartungen über den Einfluss der einzelnen Faktoren in Richtung des SD-Bias ab. Für das Handlungsziel Anerkennungsbedürfnis vermutet Stocké, dass die Erwünschtheitswahrnehmung die entscheidende Variable darstellt. Während das situationale Bedürfnis nach Anerkennung²⁷ und die Wahrschein-

27 Die doppelte Verwendung des Begriffes Anerkennungsbedürfnis ist in diesem Zusammenhang etwas verwirrend. Zum einen beschreibt der Begriff ein konkretes Handlungsziel innerhalb der Befragungssituation, zum anderen wird darunter ein Persönlichkeitskonstrukt verstanden, dass die Intensität des Bedürfnisses innerhalb der Situation messen soll.

lichkeit von Konsequenzen in Form der Anonymität für alle Handlungsalternativen gleich bleibt, sollte nur die Wahrnehmung der sozialen Norm für jede Antwortoption variieren. Je stärker die anderen Antwortoptionen im Vergleich zum wahren Wert als erwünscht wahrgenommen wird, desto größer ist nach Stockés Vermutung die Wahrscheinlichkeit eines SD-Bias. Darüber hinaus unterstellt er, dass in Befragungen, ohne Interviewer, die Erwünschtheitswahrnehmung keinen Effekt hat.

Stand der Forschung

Stocké (2004) prüfte in einer Studie sein Modell zur Erklärung sozial erwünschten Antwortverhaltens bezogen auf Vorurteile gegenüber Ausländern. Das Bedürfnis nach sozialer Anerkennung wurde dabei durch eine deutsche Kurzversion der Marlow-Crowne Skala (MC-Skala) erfasst. Die Privatheit der Antwortsituation wurde durch Zufallszuordnung der Befragten zu den beiden Alternativen Befragungsformen des persönlichen (CAPI) und schriftlichen (CASI) Interviews erfasst. Darüber hinaus wurde mit Blick auf die Notwendigkeiten der RC-Theorie zusätzlich die relative soziale Erwünschtheit erhoben. Hierfür mussten die Befragten nachgestellte Gesprächssituationen auf den Grad der Peinlichkeit hin einschätzen. Die Überprüfung der einzelnen Effekte auf die Einstellung zu Ausländern zeigte zunächst ein erstaunliches Ergebnis. Mit steigendem Anerkennungsbedürfnis nahmen, entgegen der theoretischen Erwartung, auch die geäußerten negativen Einstellungen gegenüber Ausländern zu. Der Einfluss des Befragungsmodus (Anonymität) war hingegen nicht signifikant. In einem zweiten Schritt wurden zusätzlich die multiplikativen Effekte berücksichtigt. Durch die Integration der Interaktionseffekte stellte sich die multiplikative Verknüpfung von Anerkennungsbedürfnis, Anonymität und relativer Erwünschtheit als stärkste Determinante für die Vorhersage des SD-Bias heraus. Personen mit einem hohen Anerkennungsbedürfnis und einer starken Normwahrnehmung orientierten sich nach der Interpretation von Stocké eher an der gesellschaftlichen Meinung zur erwünschten positiven Einstellung gegenüber Ausländern. Außerdem zeigten sich in dieser Gruppe signifikante und theoriekonforme Unterschiede im Antwortverhalten zwischen den Befragungsmodi. Stocké konnte damit zum einen Hinweise auf die Bestätigung des RC-Ansatzes finden, zum anderen machte er deutlich, dass sich eventuelle Fehlschlüsse in der Zusammenhangsstärke und Richtung der Variablen durch die Berücksichtigung von Interaktionseffekten aufklären lassen. Dass diese drei Faktoren für die korrekte Erfassung des SD-Bias jedoch nicht ausreichen, zeigte Stocké mit weiteren Analysen, deren Ergebnisse in Kapitel 3.2.6 im Rahmen des Frame-Selektion Modells dargestellt sind.

Bewertung des Modells

Die klassische SEU-Theorie besagt, dass sich der Handlungsnutzen einer Handlung aus der Summe der Einzelnutzen jeder Konsequenz ergibt, wobei sich die Summe aus der Bewertung der Konsequenz und ihrer subjektiven Eintrittswahrscheinlichkeit zusammensetzt. Moderatorvariablen wie bei Stocké gibt es in dieser Form in der klassischen SEU-Theorie nicht. Bei strenger Auslegung handelt es sich, schon aufgrund der Formalisierung, nicht mehr um ein SEU-Modell. Wie lässt sich Stocké im Einklang mit der SEU-Theorie interpretieren? Der Nutzen der Handlungskonsequenzen ergibt sich wie bei Esser aus dem Anerkennungsbedürfnis und dem Bedürfnis nach Authentizität. Beide Nutzenterme variieren nicht mit den Antwortoptionen, sondern stellen einen von der eigentlichen Handlung unabhängigen Faktor dar. Die unterschiedlichen Bewertungen der Optionen ergeben sich erst durch die variierenden subjektiven Wahrscheinlichkeiten der Konsequenzen. Stocké interpretiert diese subjektiven Wahrscheinlichkeiten jedoch ebenfalls als stabilen Wert zwischen den Antwortalternativen. Sowohl die Zugänglichkeit der Einstellung als auch die Anonymität ist immer gleich, egal wie der Respondent die an ihn gestellte Frage beantwortet. Bleibt man bei der klassischen SEU-Theorie, ergibt sich damit folgende Formalisierung des Modells:

$$SEU(A_1) = (p_A * U_S) + (p_{ES} * U_T) \quad (3.5)$$

$$SEU(A_2) = (p_A * U_S) + (p_{ES} * U_T)$$

- p_A = Privatheit der Situation
- U_S = Bedürfnis nach sozialer Anerkennung
- p_{ES} = Zugänglichkeit der Einstellung
- U_T = Bedürfnis nach Authentizität

Kürzt man die identischen Faktoren aus den Gleichungen ergeben sich für jede Antwortoption gleiche Nutzenterme. Eine Abwägung findet zwar statt, kommt jedoch zu dem Ergebnis, dass beide Alternativen gleich bedeutsam sind. Der Grundgedanke der Handlungstheorie als Entscheidungshilfe ist damit nicht mehr gegeben. Stocké löst dieses Problem, indem er eine zusätzliche Moderatorvariable in Form eines dritten Faktors in das Modell einfügt. Zwar sind Moderatorvariablen im ursprünglichen SEU-Modell nicht vorgesehen, es ist jedoch möglich, dass die, von Stocké beschriebene, Erwünschtheitswahrnehmung und die Passgenauigkeit einen Bestandteil der subjektiven Wahrscheinlichkeit darstellen. Die Wahrscheinlichkeit, dass das Bedürfnis nach sozialer Anerkennung befriedigt wird, steigt mit der Wahrnehmung der Erwünschtheit. Je stärker der Befragte eine bestimmte Einstellung als erwünscht wahrnimmt, umso eher erwartet er bei entsprechender Antwort eine positive Reaktion. Bei einer Einstel-

lung ohne eindeutige Erwünschtheit kann er sich dieser hingegen nicht sicher sein. Dasselbe gilt für die Passgenauigkeit. Die Wahrscheinlichkeit, die eigene Meinung zu äußern, steigt mit der Übereinstimmung von tatsächlicher Einstellung und vorhanden Antwortoptionen. Bei dieser Interpretation wären p_w die Erwünschtheitswahrnehmung und p_p die Passgenauigkeit. Beide Wahrscheinlichkeiten hängen jedoch von zusätzlichen Faktoren ab. Während die Erwünschtheitswahrnehmung von der Anonymität (w_A) beeinflusst ist, (denn in anonymen Situationen fällt die Anonymität deutlich geringer aus, als im direkten Kontakt mit einem Interviewer) wird die Passgenauigkeit von der Zugänglichkeit (w_{ES}) der Einstellung moderiert. Je stärker und eindeutiger die Einstellung eines Befragten ist, umso leichter fällt die Beurteilung der Passgenauigkeit der Antwortkategorien. Interpretiert man Stockés Ausführung wie beschrieben, ergibt sich folgende Gleichung.

$$SEU(A_1) = (p_{1,EW}(w_A) * U_S) + (p_{1,P}(w_{ES}) * U_T) \quad (3.6)$$

$$SEU(A_2) = (p_{2,EW}(w_A) * U_S) + (p_{2,P}(w_{ES}) * U_T)$$

p_{EW} = Erwünschtheitswahrnehmung

U_S = Bedürfnis nach sozialer Anerkennung

w_A = Privatheit der Situation

p_P = Eindeutigkeit und Passgenauigkeit der Antwortoptionen

U_T = Bedürfnis nach Authentizität

w_{ES} = Zugänglichkeit der Einstellung

Über diese Modifikationen der Zuordnungen der einzelnen Faktoren ist das Modell von Stocké mit der klassischen SEU-Theorie vereinbar. Ein Kritikpunkt bezieht sich auf die lückenhafte Konkretisierung des Handlungsziels Authentizität. Während Stocké für das Anerkennungsbedürfnis präzise Vorstellungen über die Wirkungsweise der einzelnen Faktoren benennt, werden die Variablen im Rahmen der Authentizität nur oberflächlich angesprochen. Diese fehlende Präzision erschwert vor allem die Operationalisierung der Konstrukte. Dieses Problem wird in Kapitel 4.2 nochmals aufgegriffen.

Welche Ebene des SD-Bias wird nun im Modell von Stocké betrachtet? Vor dem Hintergrund der in Kapitel 2.2.4 beschriebenen Kriterien einer Begriffsklärung ist zunächst nicht ersichtlich, was genau unter dem Konstrukt des wahren Wertes verstanden wird. Obwohl Stocké ebenso wie Esser die Reichweite seines Modells nicht auf eine bestimmte Frageform beschränkt, untersucht auch er die Erklärungskraft nur anhand von Bewertungsfragen. Gerade bei der Einstellungsmessung ist es jedoch notwendig, zwischen stabilen und instabilen Formen der Einstellung zu unterscheiden. Im Gegensatz zu Esser (1989a) findet sich bei Stocké (2001, 2004) keine Auseinandersetzung mit der Frage, was unter dem wahren Wert einer Bewertungsfrage zu ver-

stehen ist. Ebenfalls problematisch ist die Konkretisierung der kognitiven Ebene, auf der die Verzerrung stattfindet. Während Stocké in seinem Aufsatz aus dem Jahr 2001, in dem die Grundlagen seines theoretischen Modells gelegt wurden, keine Unterscheidung zwischen unbewusster und bewusster Verzerrung diskutiert wird, reduziert er drei Jahre später seine empirische Analyse auf die strategisch motivierte Fremdtäuschung. Die Selbsttäuschung hingegen bleibt unberücksichtigt²⁸. Dass Stockés Annahmen auf dem RC-Ansatz beruhen, kann, wie schon bei Esser, als zusätzlicher Hinweis für die Ausrichtung auf bewusste Formen einer Antwortverzerrung interpretiert werden. Davon ausgehend lässt sich vermuten, dass sich der Ansatz auf die Erklärung kurzfristiger Verzerrungen richtet.

Der letzte Bewertungsmaßstab bezieht sich auf die normative Bezugsgruppe. Während Esser zwischen dem Einfluss kultureller und situationaler Normen differenziert, benennt Stocké drei Quellen, die der Befragte zur normativen Orientierung nutzen kann: Die verbalen und nonverbalen Äußerungen des Interviewers, der stereotypische Hintergrund des Interviewers und gesellschaftliche Bewertungsstandards. Das Erklärungsmodell kann nach Stocké auf alle drei Normgruppen angewendet werden (2001: 4). Diese Einschätzung ist jedoch fraglich, denn seine Erklärungsfaktoren diskutiert Stocké ausschließlich im Hinblick auf Interviewereffekte. Die Interpretation ergibt sich vor allem aus dem vermuteten Zusammenhang zwischen Erwünschtheitswahrnehmung und Anonymität. Laut Stocké kann der Effekt einer sozialen Norm nur wirken, wenn in einer Befragungssituation ein Interviewer anwesend ist. Ohne Interviewer, so vermutet er, kann es zu keinem Effekt sozialer Normen kommen. Dies ist eine wichtige Einschränkung für den Geltungsbereich des Ansatzes, die eigentlich in der Definition der Antwortverzerrung hätte präzisiert werden müssen. Offen bleibt, wie Stocké seine restriktive Auslegung der Anonymität begründet. Im Grund genommen läuft diese Interpretation auf eine Dichotomisierung des Effekts hinaus, d.h. anonym und öffentlich²⁹. Weder in dem Modell von Esser noch in der Modifikation von Stocké findet sich eine Begründung für diese strenge Behauptung. Wenn das Modell tatsächlich auf alle Situationen übertragbar sein sollte und damit auch auf Befragungen ohne Interviewer, dann ist diese Interpretation der Anonymität nicht nachvollziehbar. Stockés Ausführungen wirken in diesem Punkt etwas widersprüchlich: Zum einen benennt er alternative Quellen für Normen, auch wenn eine direkte Bezugsperson fehlt, zum anderen schließt er eine Verzerrung ohne Interviewer aus. Die Konsequenzen dieses Problems werden

28 Fremd- und Selbsttäuschung sind zwei Dimensionen einer von Paulhus (1984) entwickelten SD-Skala, deren theoretische Grundlagen und Entwicklung in Kapitel 4.1.1.4 näher dargestellt sind.

29 Dies ergibt sich auch aus den erwarteten Zusammenhängen, die er in seiner Arbeit aus dem Jahr 2001 grafisch darstellt.

im Vergleich mit dem Modell von Skarbek-Kozietulska et al. (2012) deutlich, das im folgenden Kapitel besprochen werden soll.

3.2.4 Modell nach Skarbek-Kozietulska et al.

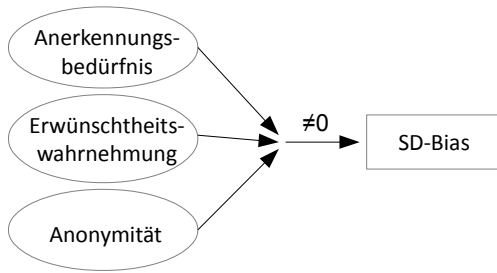
Skarbek-Kozietulska, Preisendörfer und Wolter untersuchten 2012 in einer Validierungsstudie die Determinanten korrekter Antworten in Befragungen. Sie definieren den SD-Bias als „begründetes Lügen“, dass sich von Antwortabweichungen aus Gründen wie Spaß oder Langeweile abgrenzt (Skarbek-Kozietulska et al. 2012: 6). Als heikle Frage, bei der es zu begründeten Lügen kommt, untersuchen die Autoren Fragen zu berichtendes delinquentes Verhalten. Da sich die theoretische Herleitung der Erklärungsfaktoren stark an den bereits vorgestellten theoretischen Arbeiten von Esser und Stocké orientiert, soll an dieser Stelle nur auf die Unterschiede der Modelle eingegangen werden.

Wie bei Stocké gehen Skarbek-Kozietulska et al. davon aus, dass der SD-Bias durch drei zentrale Determinanten beeinflusst wird: durch das Anerkennungsbedürfnis, durch die Erwünschtheitswahrnehmung und durch die Anonymität der Situation. Ihrer Ansicht nach ist die Anonymität auch als kontinuierliche Variable zu interpretieren. Die Wahrnehmung der sozialen Norm orientiert sich in einer anonymen Befragungssituation zudem nicht an dem Interviewer, sondern an einer vorgestellten Öffentlichkeit (Skarbek-Kozietulska et al. 2012: 9).³⁰ Ob und wie diese Konsequenzen von einer befragten Person in anonymen Situationen wahrgenommen werden können, erläutern Skarbek-Kozietulska et al. jedoch nicht. Vielmehr weisen sie darauf hin, dass andere Validierungsstudien einen additiven bzw. direkten Einfluss des Anerkennungsbedürfnisses und der Erwünschtheitswahrnehmung nachweisen konnten (Skarbek-Kozietulska et al. 2012: 9). Im Gegensatz zu Stocké (2004) gehen die Autoren aber nicht von einem strengen multiplikativen Modell aus: „Sieht man die drei genannten Bestimmungsfaktoren als kontinuierliche Variablen, erscheint die These vom ausschließlich interaktiven Zusammenwirken zweifellos relativ strikt. Zudem liegen empirische Befunde aufgrund von Validierungsdaten vor, die dafür sprechen, dass die Dimensionen SD-Neigung und SD-Belief unabhängig voneinander bzw. additiv auf das Missreporting wirken“ (Skarbek-Kozietulska et al. 2012: 9). So ist es nach ihrer Ansicht durchaus möglich, dass auch in anonymen Befragungssituationen ein SD-Bias auftreten kann.

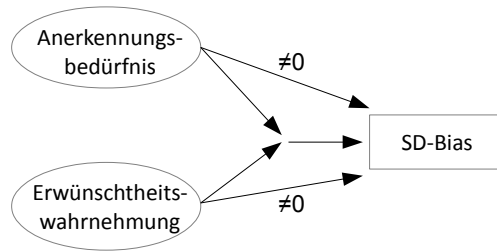
³⁰ Stocké argumentiert hingegen, dass sowohl die Erwünschtheitswahrnehmung als auch die subjektive Erwartung von Konsequenzen in anonymen Situationen nicht gegeben ist.

Abbildung 2: Vergleich des Erklärungsmodells von Stocké und Skarbek-Kozietulska et al.

Modell Stocké



Modell Skarbek-Kozietulska et al.



Quelle: Eigene Darstellung nach Stocké (2004) und Skarbek-Kozietulska et al. (2012)

Stand der Forschung

In einer Validierungsstudie überprüften Skarbek-Kozietulska, Preisendörfer und Wolter (2012) ihr Modell an einer schriftlichen Befragung, bei der sie Delinquenten über ihre strafrechtlichen Verurteilungen befragt wurden. Ziel der Studie war es, festzustellen, welche Faktoren das Lügen oder Gestehen einer Tat beeinflussen. Der Kostenfaktor der Antwortalternativen wurde durch den Grad der Deliktschwere nach einem externen Rating erfasst und zusätzlich zu den bereits besprochenen Faktoren im Rahmen der SEU-Theorie in das Modell integriert. Das Anerkennungsbedürfnis der Personen wurde über eine MC-Kurzskala sowie einem zusätzlichen Item aus dem Balanced Inventory of Desirable Responding (BIDR) gemessen. Für die Kontrolle des SD-Beliefs wurden die Personen zusätzlich nach der wahrgenommenen Verwerflichkeit delinquenten Verhaltens befragt. Da es sich um eine anonyme selbst administrierte schriftliche Befragung handelt, ist die dritte Bedingung für den SD-Bias, nicht gegeben. Die empirischen Befunde bestätigten ihre Annahme, dass sich die Erklärungsfaktoren des SD-Bias nicht multiplikativ, sondern direkt auf die abhängige Variable auswirken. So konnten sie trotz hoher von ihnen unterstellter Anonymität feststellen, dass 68% der Männer und nur 50% der Frauen bei der Frage nach strafrechtlichen Verurteilungen ehrlich antworteten. Im Gegensatz zu Stocké zeigte sich im multivariaten Modell zwar ein signifikanter Haupteffekt für die SD-Neigung und den SD-Belief jedoch kein signifikanter Interaktionseffekt. Die Autoren werten diese Ergebnisse als einen Beitrag zur Rehabilitierung des vielfach kritisierten Ansatzes von Marlow-Crowne, der einen SD-Bias ausschließlich auf der Basis direkter Effekte modelliert (Skarbek-Kozietulska et al. 2012: 19).

Bewertung des Modells

Das Erklärungsmodell von Skarbek-Kozietulska et al. orientiert sich an der Arbeit von Stocké (2004). Die Modifizierung des Modells bezüglich der Anonymität ist jedoch nur lückenhaft dar-

gestellt. Die Autoren gehen zwar auf die alternative Quelle der Erwünschtheitswahrnehmung ein, unterlassen es jedoch den bei Stocké (2001) zentralen Faktor der subjektiv erwarteten Konsequenz, für ihr Modell zu erläutern.

Aufgrund der zunächst großen Übereinstimmung von Skarbek-Kozietulska et al. und Stocké und der Orientierung beider an Esser sind sich auch die kognitiven und temporären Dimensionen des SD-Bias sehr ähnlich. Anders verhält sich dies in Bezug auf die inhaltliche Dimension. Während das Modell von Stocké auf Einstellungsfragen angewendet wird, überprüfen Skarbek-Kozietulska et al. ihren Ansatz anhand einer Wissensfrage. Wie in Kapitel 2.2.2 dargestellt, liegt der Hauptunterschied zwischen beiden Frageformen in ihrer Nähe zum wahren Wert. Aufgrund der sich hieraus ergebenden Konsequenzen ist es fraglich, ob die Einflussfaktoren für das sozial erwünschte Antwortverhalten für beide Formen identisch sind. Bei Betrachtung der Bezugsgruppen Dimension wird ein weiterer wichtiger Unterschied zwischen beiden Modellen deutlich. Das Modell von Stocké erklärt ausschließlich Verzerrungen durch Interviewereffekte. Skarbek-Kozietulska et al. fassen sie die Form der Verzerrung qua Definition weiter. Sie argumentieren, dass es auch in anonymen Situationen zu sozial erwünschtes Antwortverhalten kommen kann. Als normativer Bezugspunkt wird nicht der Interviewer, sondern eine vorgestellte Öffentlichkeit herangezogen. Beide Ansätze beziehen sich von den in Kapitel 2.2.4 aufgestellten Bewertungskriterien aus betrachtet auf unterschiedliche Formen des Phänomens soziale Erwünschtheit. Vor diesem Hintergrund überrascht es nicht, dass beide Studien nicht zu denselben Ergebnissen kommen. Allerdings bleibt dabei offen, ob die Ursache für diese Heterogenität in der inhaltlichen und/oder der Bezugsgruppen Dimension zu suchen ist.

3.2.5 Modell nach Wolter

Ebenfalls an einer Validierungsstudie zum Kriminalitätsverhalten untersuchte Wolter (2012) sein Modell zur Erklärung des sozial erwünschten Antwortverhaltens. Im Gegensatz zu Stocké und Skarbek-Kozietulska et al. findet sich in Wolters Arbeit eine umfangreichere Definition des zu untersuchenden Phänomens. Es wird auf Grundlage der klassischen Testtheorie zwischen dem Bedürfnis nach sozialer Anerkennung (SD-Neigung), der Erwünschtheitswahrnehmung (SD-Belief) und der sozial erwünschten Antwortverzerrung (SD-Bias) unterschieden (Wolter 2012: 39). Ausgehend von dieser differenzierten Betrachtung der Konstrukte wird das RC-Modell von Esser aufgearbeitet und erweitert. Im Zentrum seiner Ausführungen steht dabei die Frage, wie sich das Erklärungsmodell auf die Beantwortung heikler, dichotomer Verhaltensfra-

gen beantworten lässt. Hierzu modifiziert Wolter das Modell in vier grundlegenden Punkten, die im Folgenden kurz wiedergegeben werden sollen (siehe Wolter 2012: 135 ff.).

Erstens vereinfacht er das Modell von Esser, das sich auf Einstellungen bezieht, in Bezug auf die subjektive Wahrscheinlichkeit der Konsequenz Authentizität. Mag es bei Einstellungsfragen nicht immer eindeutig sein, ob die Antwortkategorien deckungsgleich mit der tatsächlichen Meinung des Befragten sind, ist es bei einer dichotomen Verhaltensabfrage laut Wolter keine Zweifel „mit welcher Antwort der Nutzen einer wahren Antwort realisiert wird“ (Wolter 2012: 135). Dies bedeutet, dass die subjektive Wahrscheinlichkeit für der Konsequenz Authentizität für die Antwortoptionen, die nicht der Wahrheit entspricht, gleich null ist ($p_{1T}=1$ und $p_{2T}=0$). Gleichzeitig unterstellt Wolter, dass basierend auf dieser Annahme für den Befragten auch eindeutig ist, welche der beiden Antwortoptionen die sozial erwünschte darstellt. Damit ergibt sich für die Wahrscheinlichkeit der Konsequenz „soziale Anerkennung“ folgende Formalisierung: $p_{1S}=0$ und $p_{2S}=1$.

Zweitens erscheint Wolter eine Erweiterung der Nutzenkomponenten für Verhaltensfragen, wie bspw. durch die kognitive Dissonanz, sinnvoll. Die dritte Erweiterung des ursprünglichen RC-Modells bezieht sich auf zusätzliche Kostenkomponente. Wolter argumentiert, dass mit einer sozial erwünschten Antwort auch immer die Gefahr der Entdeckung der Lüge einhergeht, die durch zusätzliche Kosten in Bezug auf die negative Reaktion des Gegenübers gekennzeichnet ist (Erwünschtheitsverlust). Die Erweiterung des Modells würde dann wie folgt formalisiert werden (Wolter 2012: 136):

$$SEU(A_{1,wahr}) = p_{1T} * U_T + p_{1S} * U_S \quad (3.7)$$

$$SEU(A_{2,falsch}) = p_{2T} * U_T + p_{2S} * U_S - p_a * C_L$$

p_T = subjektive Wahrscheinlichkeit der Konsequenz Authentizität

U_T = Nutzen der Konsequenz Authentizität

p_S = subjektive Wahrscheinlichkeit der Konsequenz soziale Anerkennung

U_S = Nutzen der Konsequenz soziale Anerkennung

p_a = subjektive Wahrscheinlichkeit des Aufdeckens

C_L = Kosten der Konsequenz Erwünschtheitsverlust

Im vierten und letzten Schritt versucht Wolter die Randomized-Response-Technik (RRT) in das Modell von Esser über die Veränderungen der Wahrscheinlichkeiten zu integrieren³¹. Da es sich bei dem Nutzen der Authentizität um ein intrinsisches Konstrukt handelt, ist die Eintrittswahrscheinlichkeit von der RRT nicht beeinflusst. Anders verhält es sich mit der Wahrschein-

31 Für nähere Erläuterungen zur Randomized-Response-Technik siehe Kapitel 4.1.2.1.1.

lichkeit der Konsequenz „soziale Anerkennung“. Im Extremfall wäre die Wahrscheinlichkeit hier für beide Antwortalternative (Lüge und Wahrheit) gleich null, da die Antwort nicht mehr von dem Interviewer wahrgenommen wird. Bei dieser Interpretation würde der Gesamtnutzen der wahren Antwort überwiegen. Nach Wolter ist jedoch damit zu rechnen, dass die RRT nicht automatisch beide Wahrscheinlichkeiten auf null setzt, denn aggregiert würde sich zumindest eine Prävalenz von 50% wahrer Antworten ergeben (Wolter 2012:138). Entscheidend ist, dass die Aufdeckungswahrscheinlichkeit mit der RRT gegen null tendiert (vorausgesetzt die Befragten haben den Einsatz der Technik verstanden und glauben die Instruktionen). Die Lüge zugunsten des Anerkennungsbedürfnisses ist damit im RRT-Modus für den Befragten sicherer. Diese Modellierung bietet eine Erklärung für die bisherige Inkonsistenz der Forschungsergebnisse: die RRT schützt durch ihre Anonymisierung nicht nur die Befragten, die die Wahrheit sagen wollen, sondern reduziert auch gleichzeitig die Anreize für solche Respondenten, die vorhaben eine sozial erwünschte Antwort abzugeben.

Stand der Forschung

Das Modell von Wolter wurde bisher nur von ihm selbst im Rahmen seiner Dissertation (2012) geprüft. Dafür wurde eine Validierungserhebung im Face-to-Face Design durchgeführt, bei der insgesamt 552 Personen vollständig befragt wurden, die für eine begangene Straftat verurteilt worden sind. Das Bedürfnis nach sozialer Anerkennung wurde über eine Kurzversion der MC-Skala gemessen. Zur Kontrolle der Erwünschtheitswahrnehmung wurden Kriminalitätsnormen und -einstellungen erfasst. Geschlecht, Alter, Migrationshintergrund, Deliktschwere und Bildungsabschluss stellen zusätzliche Kontrollvariablen dar (Wolter 2012:198 ff.). In der umfangreichen empirischen Analyse konnte Wolter (2012) zunächst zeigen, dass der RRT-Modus keinen Einfluss auf die Qualitätsindikatoren der Umfrage hatte. Insbesondere die Anonymitätsbedenken unterschieden sich nicht signifikant zwischen der anonymen und weniger anonymen Situation. Bezogen auf das Missreporting führte die RRT ebenfalls zu keiner Erhöhung der wahren Antworten. Hingegen wählten Befragte, die wegen eines schweren Delikts verurteilt wurden, häufiger die sozial erwünschte Antwortalternative. Hinsichtlich der RC-Erklärungsfaktoren zeigten die Analysen zwar, dass der Anreiz zur sozialen Erwünschtheit tatsächlich einen Einfluss auf das Antwortverhalten hat, die einzelnen Ergebnisse entsprachen jedoch nur selten den theoretischen Vermutungen. Entgegen der Lehrmeinung hatte das Anerkennungsbedürfnis keinen signifikanten Einfluss in der direkten Abfrage der heiklen Items, sondern ausschließlich im RRT-Modus. Wolter wertet dieses Ergebnis im Sinne seines Ansatzes. So ist es, wie oben beschrieben, durchaus möglich, dass die anonyme Situation im RRT-Modus nicht nur die negati-

ve, sondern auch die positiven Konsequenzen einer Lüge verringert. Wolter überprüfte zusätzlich die Interaktionshypothesen von Stocké (2004), die jedoch nicht an den empirischen Daten bestätigt werden konnten.

Bewertung des Modells

Die Arbeit von Wolter (2012) zählt zu den umfangreicheren theoretischen Auseinandersetzungen mit sozialer Erwünschtheit, die sich im deutschen Sprachraum finden lässt. Bei der Herleitung des Erklärungsmodells orientierte sich Wolter stark an dem RC-Modell von Esser. Im Gegensatz zu Stocké (2001, 2004) ist die Erweiterung theoriekonform vorgenommen worden. Diese Erweiterung stellt im Vergleich zu den bisherigen Modellen eine Besonderheit dar. Während Stocké die Anonymität der Situation in Form einer Moderationsvariable integriert, baut Wolter das experimentelle Design mit der RRT über die Wahrscheinlichkeiten ein. Das Modell wird anhand einer Validierungsstudie überprüft. Wolter argumentiert, dass es nur über diese externe Kontrolle der Antworten mittels Metadaten möglich ist, den SD-Bias zu untersuchen (Wolter 2012:118). Diese Behauptung ist jedoch durchaus fragwürdig, denn gerade bei retrospektiven Verhaltensabfragen können zusätzliche Effekte die Ergebnisse solcher Untersuchungen verzerren. So ist es denkbar, dass Erinnerungslücken, Verständnisprobleme oder Frageeffekte zu verzerrten Antworten unabhängig von sozial erwünschten Verhalten führen. Die Argumentation, dass nur über Validierungsstudien der wahre „wahre Wert“ zu ermitteln sei, ist nicht haltbar. Sicherlich ist der ermittelte Kontrollwert in solchen Erhebungen deutlich näher am „wahren Wert“ als es vielleicht bei Einstellungsmessungen möglich ist, aber dennoch sind eine Reihe von Verzerrungen denkbar (wie z.Bsp. Erinnerungsfehler).

Bezogen auf die Dimensionen, die in Kapitel 2.2.4 dargestellt wurden, zeigen sich große Übereinstimmungen mit dem Modell von Skarbek-Kozietulska et al.. Beide Modelle beziehen sich auf Verhaltensfragen. Da die Arbeit gleichermaßen auf dem RC-Modell von Esser basiert, ist auch die kognitive und temporäre Dimension sehr ähnlich. Aufgrund dieser großen Gemeinsamkeiten überraschte es nicht allzu sehr, dass auch die Ergebnisse der Modelle Übereinstimmungen aufweisen. Beide Analysen zeigten signifikante direkte Effekte des Anerkennungsbedürfnisses und der Erwünschtheitswahrnehmung. Die Interaktionseffekte von Stocké konnten hingegen nicht bestätigt werden. Dieser Vergleich der Ergebnisse von Stocké auf der einen und Skarbek-Kozietulska et al. sowie Wolter auf der anderen Seite gibt einen Hinweis darauf, wie groß die Relevanz der inhaltlichen Dimension für die Erklärungskraft des RC-Modells von Stocké ist, denn die Modelle weisen bezüglich dieser Dimension die größten Unterschiede auf. Tatsächlich scheint es, als würden bei der Verzerrung von Einstellungs- und Verhaltensfragen un-

terschiedliche Determinanten eine Rolle spielen. Um diese Vermutung zu überprüfen, ist es notwendig, das Modell von Stocké nochmals auf Einstellungsmessungen anzuwenden. Ansonsten ist es nicht möglich zu beurteilen, ob die schlechten Ergebnisse bei der Überprüfung des Modells von Stocké auf Fehlspezifikationen der Determinanten oder auf unterschiedliche Kontexteffekte (insbesondere die inhaltliche Dimension) zurückzuführen sind.

3.2.6 Erweitertes Frame-Selektion Modell nach Stocké

Neben den hier vorgestellten rationalen Grundmodellen findet sich ein weiterer Erklärungsansatz, der ebenfalls als Spezialfall des RC-Modells interpretiert werden kann: das Frame-Selektion-Modell (Grundlagen siehe Fazio 1990). Die Bewertung der ökonomischen Erklärungsfaktoren im Sinne der Kosten/Nutzen Relation können in diesem Modell je nach Rahmung der Situation von Akteuren unterschiedlich bewertet werden. Die Grundlage für die Anwendung des Frame-Selektion-Modells auf das Befragtenverhalten findet sich bei Esser (1990). In seinen Ausführungen deutet er eine Anwendung auf das Befragtenverhalten jedoch nur an. Theoretisch konkretisiert und empirisch überprüft wurde das Modell von Stocké und Wolter. Auf Grund der zur Verfügung stehende Daten zur Überprüfung der Modelle kann das Modell von Wolter jedoch nicht überprüft werden. Aus diesem Grund liegt der Schwerpunkt der Darstellung auf dem Modell von Stocké.

Die Erklärung des Befragtenverhaltens anhand des RC-Ansatzes bzw. der SEU-Theorie wurde häufig wegen zwei zentraler Aspekte kritisiert: Zum einen sei der RC-Ansatz nicht in der Lage den Sinn von Handlungen zu erfassen, zum anderen bleiben andere Ziele, die als die Nutzenmaximierung eine Rolle bei der Selektion der Handlungsalternativen spielen können, unberücksichtigt (Esser 1990: 239). Hartmut Esser begegnet diesen Kritikpunkten mit einer Erweiterung des klassischen RC-Modells über den Frame-Selektion-Ansatz. Das Modell der Frame-Selektion beschreibt einen dreistufigen Selektionsprozess: Zunächst selektiert der Akteur auf Grundlage von Alltagshypothesen ein mentales Modell für die jeweilige Situation (Frame) aus. Im zweiten Schritt wird für diesen Frame eine passende Handlungssequenz ausgewählt (Skript), dass in der Situation angewendet wird. Erst nach Ablauf dieser beiden Schritte erfolgt die eigentliche Auswahl der Handlung. Zwischen diesen drei Ebenen der Selektion werden verschiedene Formen der Informationsverarbeitung genutzt. Das Modell macht es möglich die Logik der Situation als eine erste Selektionsebene zu betrachten. Bei der Entscheidungsfindung wird damit berücksichtigt, dass es verschiedene Situationsdefinitionen gibt, zwischen denen gewählt werden kann. Grundlage der Theorie bilden zwei Annahmen. Zum einen geht dem Handeln von Perso-

nen immer eine Interpretation der Situation³² voraus. Diese Interpretation erfolgt durch die Wahl eines kulturellen Bezugsrahmens (Esser 2001: 259). Beim Framing werden beobachtbarer Objekte mit einem mentalen Modell der Situation verknüpft, das eine Richtlinie für den Akteur bietet. Das Handlungsmodell orientiert sich damit sowohl an Frames als auch an Skripten³³, die größtenteils auf sozialen Normen und Rollenvorstellungen basieren und durch Sozialisationsprozesse Bestandteile der Identität eines jeden Akteurs darstellen (Esser 2001: 261f.). Der jeweilige Modus bestimmt dabei die Entscheidungsstrategie, die angewendet wird. Esser betrachtet dabei vor allem zwei Extrempunkte: den reflexiv-kalkulierenden Modus (Rational Choice) und den automatisch-spontanen Modus (Esser 2001: 266). Während ein Akteur im ersten Modus nach rationalen Gesichtspunkten und der Abwägung von Kosten und Nutzen einer Situation seine Alternativen abwägt, erfolgt die Selektion nach dem automatisch-spontanen Modus über die unmittelbare Situationswahrnehmung (Kroneberg 2005: 347). Nach welcher Regel aber entscheidet der Akteur, ob die Selektion der Frames, Skripte und Handlungen nach dem rationalen oder spontanen Modus erfolgen? Für diese Verbindung aus Modell und Modus bedient sich Esser der SEU-Theorie³⁴. Da beide Logiken mit gegensätzlichen Kosten und Nutzen verbunden sind (Zeit und Genauigkeit) wird der Modus ausgewählt, der den größten Nutzen unter Berücksichtigung der Eintrittswahrscheinlichkeit aufweist. Der spontane Modus wird ausgewählt, wenn die Zugänglichkeit des mentalen Modells hoch ist und wenig Zeit zum Reflektieren zu Verfügung steht. Zum rationalen Modus kommt es eher, wenn die Reflektionskosten geringer eingeschätzt werden und der Nutzen einer umfangreichen Analyse hoch einzuschätzen ist (Kroneberg 2005: 347).

Der Ansatz des Frame-Selektion-Modells wurde von Esser (1990) nicht all zu ausführlich hin auf die Befragungssituation diskutiert. Nach Esser sind auch in einer Befragungssituation verschiedene Definitionen möglich. Neben dem Frame „Befragter im Interview“ der für Esser den Normalfall darstellt und einen Akteur beschreibt, der nach der Norm handelt, das Ziel der Befragung bestmöglich zu unterstützen, können bei einem Befragten durch bspw. geringe Vertraulichkeit auch weniger kooperative Rollenvorstellungen aktiviert werden. Neben dem guten Befragten kann also auch der Frame des Ja-Sagers oder der Tendenz zur sozial erwünschten Antwort aktiviert werden. In letzterem Fall ist das Anerkennungsbedürfnis das vorherrschende

32 In Kapitel 3.2.2 wurde bereits darauf hingewiesen, dass vor dem theoretischen Hintergrund die Formulierung „Interpretation der Situation“ anstatt des von Esser verwendeten Begriffs „Definition der Situation“ präziser erscheint.

33 Während Frames mentale Modelle von Situationen beschreiben, stellen Skripte die jeweiligen mentalen Modelle für die Handlungen in den jeweiligen Situationen dar (Kroneberg 2005: 246)

34 Esser diskutiert in Band 6 der „Speziellen Soziologie“ die Problematik, dass die SEU-Theorie formal eine Variante des RC-Ansatzes darstellt (Esser 2001: 269).

Ziel des Akteurs. Kosten würden sich bei Nichtwahl der SD Alternative in diesem Fall aus der negativen Reaktion des Interviewers und der damit verbundenen Abwertung der eigenen Person. Wählt der Befragte den kooperativen Frame aus, ist die Äußerung der „wahren Meinung“ das primäre Ziel. Die Kosten entstehen hierbei nicht über eventuelle Ablehnung, sondern vielmehr durch die Abweichung von dieser Handlungsprämisse (Esser 1990). In der Abhandlung wird an diesem Punkt jedoch nicht deutlich, welchen Selektionsmodus die Befragten nach welchen Gesichtspunkten wählen. Vielmehr werden zum Abschluss der theoretischen Überlegung Bedingungen genannt, unter denen Befragte vom Normalfall des „guten Befragten“ abweichen. Nach Esser muss es für den Frame der Antwortverzerrung eine transparente Situation geben. Das heißt, dem Befragten müssen die normativen Erwartungen in der Situation bekannt (Erwünschtheitswahrnehmung) und die Einstellung darf nicht zu stark ausgeprägt (Einstellungsstärke) sein. Erst wenn diese beiden Bedingungen erfüllt sind, ist eine Abweichung vom „wahren Wert“ nach Esser zu erwarten (Esser 1990).

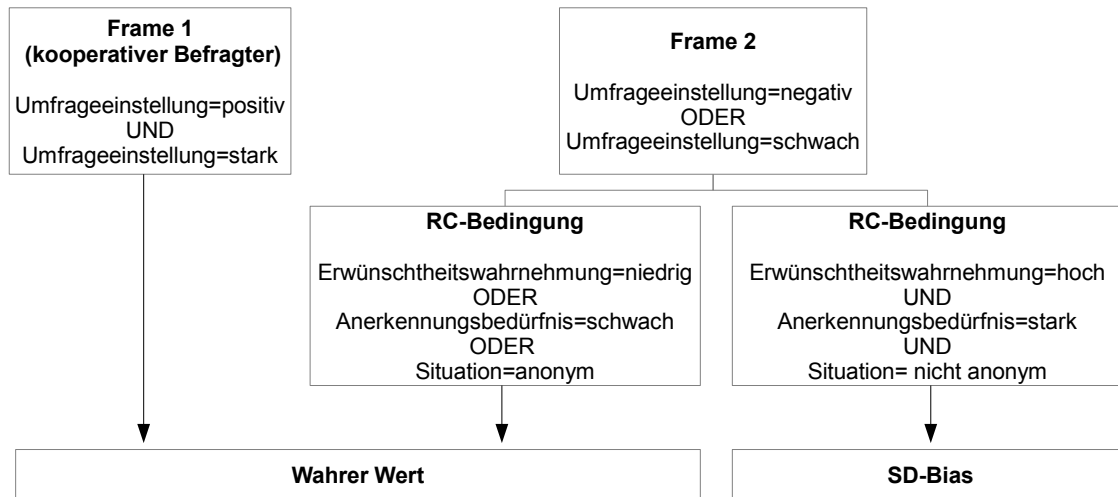
Die Entscheidung zwischen diesen beiden Frames läuft, wie bereits erwähnt, nach einem spezifischen Informationsverarbeitungsmodus ab. Entweder erfolgt die Entscheidung nach rationalen oder spontanen Kriterien. Zu den unterschiedlichen Varianten äußert sich Esser jedoch nicht. Vielmehr geht er davon aus, dass die Rolle des kooperativen Befragten die wahrscheinlichste Reaktion ist (Esser 1990: 244).

Stockés Framingmodell ist stark an den Vorarbeiten Essers orientiert und auch er sieht in der Befragtenrolle den wichtigsten Frame für die Erklärung sozial erwünschten Antwortverhaltens. Demnach unterstützen Respondenten die eine kooperative Befragtenrolle einnehmen das Ziel der Erhebung und antworten ehrlicher. Befragte mit konformistischer Rolle orientieren sich stärker in Richtung einer rationalen Abwägung. Insgesamt ist die kooperative Befragtenrolle im Vergleich zur konformistischen durch eine inhaltlich positive Einstellung zu Umfragen geprägt (Stocké 2004: 306). Ob der Befragte also eine kooperative oder konformistische Rolle einnimmt, hängt von der Umfrageeinstellung und der Stärke dieser Einstellung ab. Die kooperative Rolle kann durch unreflektierte Einstellungsaktivierungsprozesse bereits durch die freiwillige Teilnahme an einer Befragung aktiviert werden. In Anknüpfung an die Rollentheorie gehen die Befragten eine Selbstverpflichtung ein, in deren Rahmen sie versuchen, die Untersuchung durch möglichst wahre Antworten zu unterstützen (Stocké 2004: 306).

Generell kann sich eine positive Umfrageeinstellung gegenüber der Tendenz zur sozial erwünschten Antwort nur dann durchsetzen, wenn der Akteur über eine ausreichend starke Einstellung verfügt. Beide Bedingungen sind, wie in Abbildung 3 dargestellt, nach Stocké (2004)

multiplikativ miteinander verknüpft. Damit ist eine nicht rationale Antwort nur zu erwarten, wenn der Befragte über eine starke und positive Einstellung zur Umfrage verfügt (Stocké 2004: 306).

Abbildung 3: Einflussfaktoren des SD-Bias nach dem Frame-Selektion-Modell



Quelle: Eigene Darstellung nach Stocké (2004)

Esser bewertet die Wahrscheinlichkeit, dass systematische Antwortverzerrungen innerhalb eines Interviews entstehen im Rahmen des Framing-Ansatzes als äußerst gering. Seiner Meinung nach wenden die Befragten, wenn sie mit einer Interviewsituation konfrontiert werden, zunächst routiniert den Frame des kooperativen Befragten an. Erst wenn sozial erwünschtere Antwortalternativen hohe Gewinne versprechen und die Äußerung der wahren Meinung mit hohen Kosten verbunden ist, kann ein SD-Bias entstehen. Diese harten Bedingungen erschweren die systematische Kontrolle des SD-Bias in Interviews (Esser 1990: 242f.).

Eine umfangreiche Auseinandersetzung mit den Grundlagen des Framing-Modells sowie die Rolle der RRT findet sich bei Wolter (2012: 143 ff.). Da das Framing-Modell von Stocké aufgrund der fehlenden Berücksichtigung der Modusselektion kritisiert wird, stellte Wolter ein umfangreicheres Erklärungsmodell vor, das diese Schwäche ausgleichen sollte. Ebenso wie Stocké argumentiert Wolter, dass eine rationale Abwägung einer Antwort unter den Gesichtspunkten der sozialen Erwünschtheit für die Respondenten nur im reflexiv-kalkulierend Modus möglich ist. Bei der automatisch-spontanen Antwortabgabe hingegen spielt das Anerkennungsbedürfnis oder die Erwünschtheitswahrnehmung keine Rolle. Welche Faktoren beeinflussen nun, welcher Modus der Informationsverarbeitung ausgeführt wird? Wolter formuliert drei zentrale Hypothesen, welche die Modusselektion beschreiben. Nach der Reflexionsmodushypothese nimmt der Anteil der sozial erwünschten Antworten bei einer direkten Abfrage der sensitiven Items

zu, wenn der Respondent reflexiv-kalkulierend reagiert. Neben der direkten Frage beschäftigte sich Wolter zudem mit dem Einfluss der RRT. Hier ist zu vermuten, dass die besondere Erhebungsform sowohl bei ehrlichen als auch bei den gelogenen Antworten der reflexiv-kalkulierende Modus aktiviert wird, da beide Formen ein Reflektieren voraussetzen (Wolter 2012: 172). Die Interviewererfahrungshypothese hingegen bezieht sich auf den Einfluss der Interviewer, deren Verhalten den Framingprozess der Respondenten beeinflussen kann. Als dritten zusätzlichen Faktor für das Framing-Modell beschreibt Wolter über die Bildungshypothese. Die kognitiven Hürden, die die RRT erfordert, können von Befragten mit hoher Bildung leichter bewältigt werden (Wolter 2012: 174). Wolter erweitert das Framing-Modell zusammenfassend um drei Indikatoren: der Reflexionsmodus, die Interviewererfahrung und die Bildungshypothese. Darüber hinaus werden der Effekte von Drittpersonen, Dissonanzen und Incentives diskutiert.

Stand der Forschung

Getestet wurde das Framing-Modell von Stocké (2001, 2004), Skarbek-Kozietulska et al. (2012) und Wolter (2012). Stocké erfasste die generalisierte Umfrageeinstellung in Form von Fragen zur Nützlichkeit und Verlässlichkeit von Umfragen sowie die wahrgenommene Belastung der Befragungssituation. Neben der Umfrageeinstellung wurden die Verankerung der Umfrageeinstellung durch die Reaktionszeit zur Beantwortung der Fragen und die Umfrageerfahrung gemessen³⁵. Die multivariate Analyse zeigte, dass der, auf Basis des Framing-Modells vorhergesagte, Interaktionseffekt aus SD-Anreiz, der Umfrageeinstellung und der Latenzzeit einen signifikanten Beitrag zur Erklärung des Antwortverhaltens leistet. Laut Stocké (2004) wurde deutlich, dass sich der SD-Bias durch den SEU-Ansatz nicht korrekt vorhersagen lässt, wenn die Befragten eine positive und kognitiv stark verankerte Umfrageeinstellung haben.

Die Messung der Umfrageeinstellung und Umfrageerfahrung bei Skarbek-Kozietulska et al. orientierte sich stark an der Arbeit von Stocké. Im Rahmen der postalischen Befragung war eine Messung der Latenzzeit der einzelnen Fragen jedoch nicht möglich. Die Reaktionszeit wurde ersatzweise über die Tage gemessen, die die Befragten brauchten, um den Fragebogen ausgefüllt wieder zurückzuschicken. In der multivariaten Analyse konnte im Gegensatz zu Stocké jedoch kein Einfluss der Interaktionseffekte auf die Haupteffekte nachgewiesen werden. Neben den Erklärungsvariablen den RC-Ansatzes (SD-Neigung und SD-Belief) zeigte sich für die Autoren ein überraschend hochsignifikanter direkter Effekt der Antwortlatenzzeit. Entgegen der theoretischen

³⁵ Mit der Latenzzeit soll also nicht die Verankerung der erfragten Vorurteile, sondern die Einstellungsstärke der Umfrageeinstellung erfasst werden.

schen Vermutung war der Anteil der Falschangaben höher, je später der Fragebogen zurückgeschickt wurde.

Wolter konnte in seiner Analyse ebenfalls keinen signifikanten Einfluss der Umfrageeinstellung und -Verankerung (sowohl direkte als auch Interaktionseffekte) nachweisen. Der Reflexionsmodus, gemessen über die Gesamtdauer des Interviews, zeigte ebenfalls keinen relevanten Effekt auf das Befragtenverhalten. Bestätigen lies sich hingegen die Interviewerhypothese: Mit zunehmender Erfahrung des Interviewers nimmt die Validität der Daten bei der direkten Abfrage des Verhaltens zu. Bei der RRT lief der Effekt stattdessen genau in die andere Richtung. Der Bildungseffekt lies sich nur tendenziell für männliche Befragte bei der direkten Abfrage der sensitiven Items bestätigen (Wolter 2012: 173 ff.).

Bewertung des Modells

Die Schlussfolgerung von Esser, dass im Rahmen von Umfragen nur sehr selten mit einem SD-Bias gerechnet werden kann, ist nicht zwangsläufig plausibel. Die Stärke des SD-Bias dürfte in diesem Zusammenhang besonders von der Sensitivität der Frage abhängen. Je sensibler die Information ist, die erhoben werden soll, desto eher weichen die Respondenten von ihrer Rolle als kooperativer Befragter ab. Welche Reaktionen befragte Personen zeigen würden, wenn sie sich mit einem umfangreichen Erhebungsinstrument voller sensitiver Items konfrontiert sähen, ist völlig offen. Dass ein Großteil der Respondeten eine kooperativere Rolle zu Beginn eines Interviews einnimmt, ist zudem bisher nicht untersucht. Eine Erhebung von Stocké und Langenfeld (2003) zeigte, dass die Umfrageeinstellung der volljährigen Bevölkerung einer süddeutschen Großstadt zwar tendenziell positiv einzuschätzen ist, der Mittelwert der Antworten jedoch nur knapp über dem Mittelpunkt der Skala lag³⁶. Es bleibt fraglich, wie stark das kooperative Befragtenverhalten angesichts dieser Zahlen tatsächlich ist. Zusätzlich gibt es Vermutungen darüber, dass die steigende Zahl an Befragungen, mit denen die Respondenten konfrontiert werden, ein sogenanntes „over-surveying“ verursachen, welches geringere Teilnahmequoten zur Folge hat (siehe Dillman et al. 1974). Eine weitere Folge des „over-surveying“ könnte auch eine „Abstumpfung“ gegenüber Umfragen sein. Die Beantwortung von Fragebögen führt bei einigen Gruppen, wie bspw. Studierenden, nicht mehr zur Aufwertung der Persönlichkeit, sondern wird als lästige Pflicht wahrgenommen. In der Studie von Stocké und Langenfeld (2003) wird deutlich, dass Personen mit Interviewerfahrung durchschnittlich bereits an vier Erhebungen teilgenommen haben. In den letzten Jahren stieg insbesondere die Zahl an Onli-

³⁶ Bedenkt man, dass auch bei der Beantwortung der Fragen zur generellen Umfrageeinstellung Aspekte wie soziale Erwünschtheit einen Einfluss haben können, so ist davon auszugehen, dass die durchschnittliche Umfrageeinstellung noch etwas möglicherweise negativer ausfallen dürfte.

neumfragen erheblich an (ADM 2012: 12). Personen mit Internetanschluss werden regelmäßig mit Intercept-Befragungen³⁷ konfrontiert. Angesichts dieser Zahlen ist es fraglich, ob auch heute noch das Bild des kooperativen Befragten zutreffend ist. Darüber hinaus sind weitere Faktoren denkbar, die einen Einfluss neben der Häufigkeit von Befragungen ausüben können, wie z.B. öffentliche Aufmerksamkeit gegenüber dem Thema Datenschutz, die Einschätzung der Relevanz des genannten Befragungsziels oder auch nur die Belästigung durch Telefondirektmarketing.

Ebenfalls offen im Rahmen des Framing-Ansatzes bleibt, was genau die Respondenten unter einem kooperativen Befragten verstehen. Was verstehen die Respondenten als Unterstützung? Auch die Verzerrung der Antworten, um den vermuteten Umfragehintergrund zu unterstützen, ist im Rahmen des kooperativen Rollenverständnisses möglich. Dies zeigt sich beispielsweise an Umfragen zum Umweltverhalten. Hat der Befragte das Gefühl, dass die Umfrage das Ziel hat, auf Umweltprobleme aufmerksam zu machen und umweltfreundliche Maßnahmen zu fördern, kann ein besonders kooperativer Respondent seine persönliche Einstellung zu dem Thema verzerren, um diesem Ziel gerecht zu werden. In diesem Fall würde auch eine kooperative Befragtenrolle zu verzerrten Antworten im Rahmen der sozialen Erwünschtheit führen, weil der Respondent von seiner eigenen Meinung abweicht und die Antwortalternative auswählt, von der er denkt, dass sie die Ergebnisse der Umfrage am besten in eine Richtung beeinflussen, die übergeordneten Zielen entspricht.

Theoretisch stellt sich die Frage, womit die Unterscheidung zwischen Nutzentermen und Framing begründet wird. Der Nutzen der Anerkennung tritt in den Hintergrund, wenn die Einstellungsstärke hoch ist und der Befragte ein größeres Bedürfnis nach Authentizität aufweist. Die rationalen Faktoren spielen keine Rolle mehr, wenn der Respondent eine kooperative Befragtenrolle einnimmt. Sowohl argumentativ als auch formal ist der Unterschied zwischen den beiden Faktoren nicht ersichtlich. Der Erklärungszuwachs im Gegensatz zum RC-Modell ergibt sich durch die zusätzliche Berücksichtigung der Befragtenrolle über die Umfrageeinstellung und die Antwortlatenzzeit.

Die unvollständigkeit des Framing-Modells von Stocké wird auch von Wolter diskutiert (2012: 156). Seine Erweiterungen zur Erfassung der Modusselektion in Form von Antwortlatenz, Interviewererfahrung und Bildung zeigen, wie viele unterschiedliche Faktoren zusätzlich

³⁷ Bei den sogenannten Intercept-Befragungen werden ausschließlich Besucher einer bestimmten Internetseite kontaktiert. Ein Zufallsgenerator entscheidet darüber, bei welchem Besucher ein Pop-up Fenster mit der Befragung eingeblendet wird und bei welchem nicht.

eine Rolle bei der Beantwortung von Fragen spielen können. Daher ist es notwendig, bei künftigen Forschungsarbeiten diese Determinanten systematisch einer Prüfung zu unterziehen.

3.3 Zusammenfassung und Bewertung

Das Befragtenverhalten im Rahmen des RC-Ansatzes wurde in den letzten 50 Jahren der SD-Forschung von rudimentären bis hin zu äußerst umfangreichen Modellen theoretisch weiterentwickelt. Die ersten Anfänge zur Erklärung des sozial erwünschten Antwortverhaltens finden sich bei Crowne und Marlow (1964), die aufbauend auf der Lerntheorie einen zentralen Faktor für das Auftreten des SD-Bias beschreiben: Das Anerkennungsbedürfnis. Im deutschen Kontext wurde die theoretische Auseinandersetzung vor allem durch Hartmut Essers Arbeiten geprägt. Er entwickelte ein Handlungsmodell, bei dem sich der SD-Bias aus der Kombination von Zielen und der Wahrscheinlichkeit von Konsequenzen ergibt. Das Anerkennungsbedürfnis als Persönlichkeitskonstrukt kann seinem Ansatz nach nur zu einer Antwortverzerrung führen, wenn der Befragte eine Norm als sozial erwünscht wahrnimmt und sich nicht in einer völlig anonymen Situation befindet. Das entgegengesetzte Ziel zum Anerkennungsbedürfnis stellt der Wunsch nach Authentizität dar. In Kombination mit der Einstellungsstärke und der Passgenauigkeit der Antwortoptionen wirkt dieses Bedürfnis einem SD-Bias entgegen. Das theoretische Modell wurde von Stocké aufgegriffen und konkretisiert, wobei einer der zentralen Unterschiede in der Interpretation und Formalisierung der Nutzenterme besteht. Während Esser eine deterministische Selektion der Handlungsalternativen beschreibt, setzt Stocké sein Modell als stochastische Gleichung um. Skarbek-Kozietulska et al. stellen eine vierte Variante des Erklärungsmodells vor, das größtenteils auf dem Ansatz von Stocké beruht. Ihre Modifikation besteht im Wesentlichen darin, dass die strenge multiplikative Verknüpfung der Faktoren aufgegeben wird. Deutlich stärker modifiziert wird das Esser-Modell hingegen von Wolter. Er integriert zusätzliche Nutzen und Kostenterme in das Modell und modelliert über die subjektive Wahrscheinlichkeit außerdem einen Effekt der RRT.

Eine Spezialform des RC-Ansatzes zur Erklärung des Befragtenverhaltens wurde mit dem Framingansatz von Esser aufgegriffen und von Stocké konkretisiert. Im Frame-Selektion Modell werden die rationalen Erklärungsfaktoren durch die Befragtenrolle ergänzt. Über die Logik der Situation werden erstmals in diesem Zusammenhang Bedingungen genannt, unter denen Aspekte wie das Anerkennungsbedürfnis und die Erwünschtheitswahrnehmung keine Rolle bei der Auswahl der Antwortoptionen spielen. Wenn der Respondent in einer Befragungssituation

eine kooperative Rolle einnimmt, verstärkt die Logik der Situation den Nutzen „wahrer“ Antworten.

Es konnte gezeigt werden, dass alle Ansätze theoretische Schwächen haben, die bei der Umsetzung und Interpretation von Ergebnissen berücksichtigt werden müssen. Der unpräzise Umgang mit Begriffen und theoretische Lücken sind dabei besonders gravierend. Die eingeschränkte Vergleichbarkeit der Modelle zeigt sich besonders deutlich im Vergleich der Bewertungskriterien, die für eine präzise Definition des SD-Bias notwendig sind.

Im Kontrast zueinander stehen die Ergebnisse von Stocké und den Arbeiten von Wolter sowie Skarbek-Kozietulska et al.. In Bezug auf die Bewertung der gegensätzlichen Ergebnisse stellt sich die Bezugsgruppen und inhaltliche Dimension als problematisch heraus. Ein grundlegender Unterschied der Studien besteht im Hinblick auf die abhängigen Variablen, die über das jeweilige Modell erklärt werden soll. Während sich die Arbeit von Stocké (2004) auf Einstellungsfragen zur Ausländerfeindlichkeit bezieht, untersuchen Skarbek-Kozietulska et al. (2012) und Wolter (2012) die Effekte im Kontext mit berichteten delinquenten Verhalten. In Kapitel 2.2.2 wurde deutlich, dass beide Frageformen sich nicht auf dieselbe inhaltliche Dimension des SD-Bias beziehen. Während bei Einstellungsfragen möglicherweise kein stabiler wahrer Wert existiert und damit die sozial erwünschten Verzerrungen zu unterschiedlichen Phasen des kognitiven Prozesses greifen kann, unterstehen erinnerte Verhaltensfragen anderen Bedingungen.

Die unterschiedlichen Befragungsmodi haben zudem Konsequenzen für die Bezugsgruppen Dimension. In der anonymen schriftlichen Erhebung von Skarbek-Kozietulska et al. können soziale Normen nur über eine vorgestellte Öffentlichkeit wahrgenommen werden. Stocké entwickelte sein Erklärungsmodell jedoch ausschließlich für Interviewereffekte. Die unterschiedlichen Sichtweisen haben auch Folgen für den von Stocké (2004) vorgeschlagenen multiplikativen Zusammenhang der drei Erklärungsfaktoren. Ist eine der Voraussetzungen nicht erfüllt, würde der Respondent in diesem Fall den wahren Wert seiner Einstellung äußern. Skarbek-Kozietulska et al. gehen hingegen nicht von einem strengen multiplikativen Modell aus. Ihre Argumentation legt nahe, dass diese Annahme bei kontinuierlichen Variablen zu restriktiv ist. Ein Grund für diese Interpretation ist sicherlich die Tatsache, dass sie in ihrer Studie die Anonymität nicht variieren konnten. Bei einer strengen Auslegung hätten sie also für ihre Respondenten keinerlei Effekte im Sinne der sozialen Erwünschtheit feststellen dürfen. Ihre Ergebnisse zeigten jedoch eine klare Differenz zwischen tatsächlichen und berichteten Verhalten. Ob diese Verzerrung aber sozial erwünschtes Antwortverhalten darstellt und nicht viel mehr eine Folge von Erinnerungslücken oder unbewusster Selbsttäuschung ist, bleibt fraglich. Auch die Erfassung der

Erwünschtheitswahrnehmung müsste in beiden Varianten unterschiedlich erfolgen. Dennoch übertrugen Skarbek-Kozietulska et al. das Modell von Stocké auf ihren Untersuchungsgegenstand ohne die Problematik weitergehend zu diskutieren und verglichen die Ergebnisse vor dem gemeinsamen theoretischen Hintergrund miteinander. Deutlich konsequenter ist Wolter, der sein Erklärungsmodell speziell für dichotome Verhaltensfragen entwickelte.

Die Folgen dieses unpräzisen Umgangs mit den Begrifflichkeiten zeigen sich deutlich im heterogenen Stand der Forschung. Am häufigsten wird bis heute das Modell von Marlow und Crowne, dass nur das Anerkennungsbedürfnis als Erklärungsgröße beschreibt, verwendet. Neuere Forschungsergebnisse (Stocké 2004, Stocké/Hunkler 2004) zeigen aber, dass der Einsatz der Täuschungsskalen ohne gleichzeitige Kontrolle der Erwünschtheitswahrnehmung wenig empfehlenswert ist. Gerade beim Thema Vorurteile ist es möglich, dass sich gesellschaftliche Gruppen stark in ihrer Wahrnehmung von Tabus unterscheiden. Dabei ist es durchaus vorstellbar, dass Faktoren wie Alter, Bildung und politische Orientierung einen erheblichen Effekt auf die Erwünschtheitswahrnehmung haben. Gerade die Ergebnisse, die im Widerspruch zu den theoretischen Annahmen, umgekehrte Effekte oder fehlende Zusammenhänge aufzeigen, könnten sich teilweise durch diesen isolierten Einsatz der Skalen erklären lassen. Trotz neuerer Erkenntnisse werden weiterhin in Surveys wie aus dem GMF-Projekt, dem Allbus 2006, dem SOEP-Pretest 2005 und den Cella-Erhebungen die Skalen ohne weitere Kontrollfaktoren verwendet. In diesem Zusammenhang ist auffällig, dass es zu diesen frei zugänglichen Daten bisher keine veröffentlichten Studien zur Erklärungskraft des gemessenen Anerkennungsbedürfnisses gibt. Die Ergebnisse der Studie von Stocké (2004), können als Beleg für einen multiplikativen Effekt gewertet werden. Diese Ergebnisse konnten von Skarbek-Kozietulska et al. (2012) und Wolter (2012) trotz ähnlicher Vorgehensweise jedoch nicht bestätigen.

Die bisherigen empirischen Studien geben keine Möglichkeit zu beurteilen, welches dieser vier Modelle am besten den SD-Bias konzeptualisiert. Nichteinmal über die Stabilität der einzelnen Faktoren lassen sich eindeutige Aussagen treffen. Die Hauptursache für die heterogene Datenlage und die schlechte Vergleichbarkeit der Modelle liegt in den unterschiedlichen Rahmenbedingungen und Operationalisierungen. Eine Folge davon ist, dass Studien die sich auf ein und denselben Erklärungsansatz berufen, unterschiedliche Ebenen des SD-Bias betrachten und damit einer Vergleichbarkeit der Befunde nicht immer gegeben ist. Obwohl die beiden Validierungsstudien das Modell von Stocké nicht bestätigen konnten, ist dem Ansatz nicht grundsätzlich seine Erklärungskraft abzusprechen. Vielmehr ist aufgrund der Unterschiede der Studien nicht endgültig zu beurteilen, ob die fehlenden Effekte zulasten des Modells oder zulasten der

Differenzen in der inhaltlichen und/oder Bezugsgruppen Dimension gehen. Hierfür ist es sinnvoll, das Modell von Stocké auf andere Einstellungsfragen zu überprüfen³⁸. Aus diesem Grund soll im Rahmen dieser Arbeit die Erklärungskraft der Modelle von MC, Stocké und Skarbek-Kozietulska et al. an den Daten einer Erhebung überprüft werden. Eine Überprüfung des Modells von Wolter ist im empirischen Abschnitt dieser Arbeit, die auf Sekundäranalysen und einer Onlineerhebung beruht, aufgrund seines experimentellen Designs mit RRT nicht möglich.

³⁸ Offen bleibt in dieser Arbeit die Übertragbarkeit des Modells auf Verhaltens- und Wissensfragen. Die Überprüfung dieser beiden Frageformen bedarf aufgrund ihres Umfangs einer eigenständigen empirischen Arbeit.

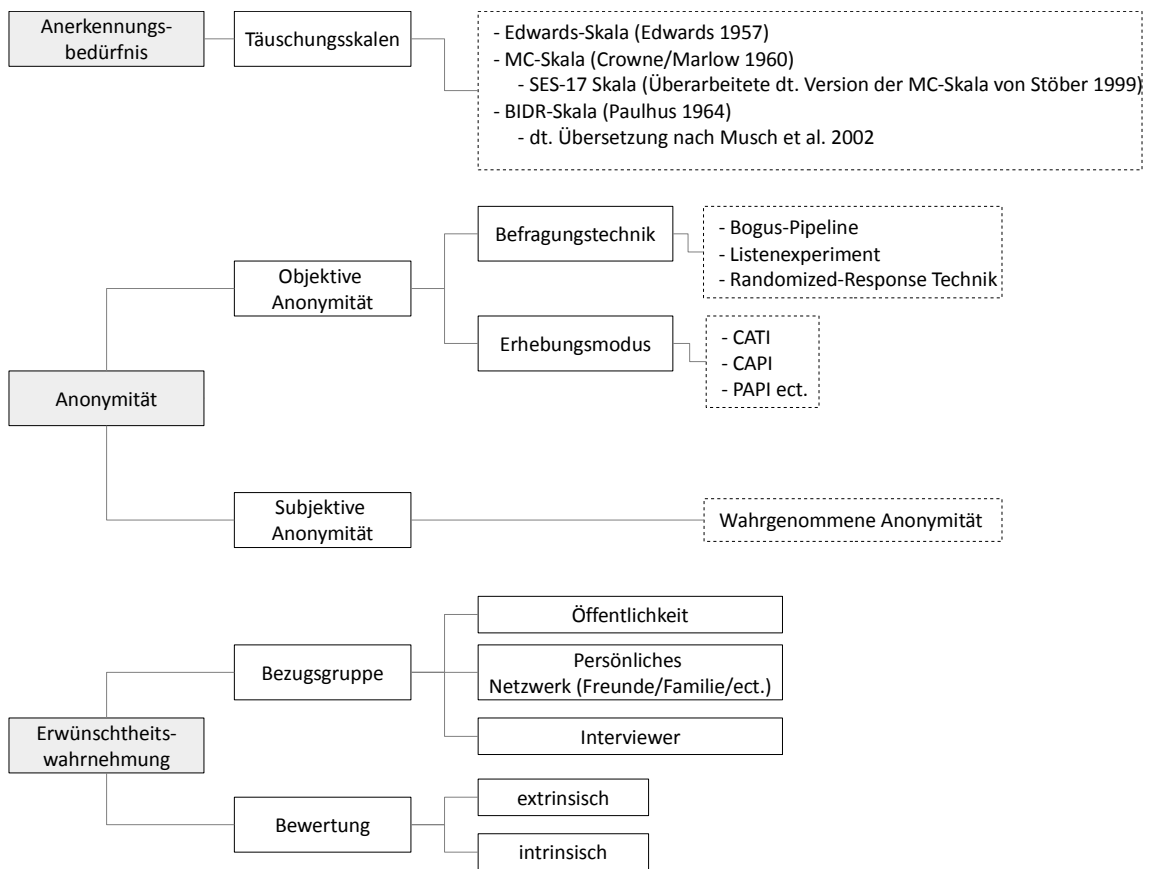
4 Operationalisierungsvarianten der Erklärungsfaktoren im Rahmen des RC-Ansatzes

In diesem Kapitel werden bisher eingesetzte Operationalisierungen der Konstrukte zur Erklärung des SD-Bias im Rahmen des Rational-Choice Ansatzes vorgestellt. In Kapitel 3 wurden drei Aspekte des Befragtenverhaltens dargestellt, aus denen sich Erklärungsfaktoren ergeben: die intendierte Handlungskonsequenz „soziale Anerkennung“ (Determinanten: Erwünschtheitswahrnehmung, Anerkennungsbedürfnis, Anonymität), die intendierte Konsequenz „persönliche Identität“ (Determinanten: Einstellungsstärke, Passgenauigkeit und Authentizität) sowie die kooperative Befragtenrolle aus dem Frame-Selektion-Modell. Leider gibt es bisher nur sehr wenige Studien, in denen präzise theoretische Modelle die Grundlage für empirische Analysen bilden. Forschungsarbeiten auf dem Gebiet des SD-Bias sind vielmehr stark empiristisch geprägt (siehe exemplarisch Banse/Gawronski 2003; Couper et al. 2003; Li/Bagger 2006). Eine Folge davon ist, dass in nur wenigen Studien mehrere Einflussfaktoren nebeneinander betrachtet werden. Vielmehr finden sich Forschungen zu dem isolierten Einfluss einzelner Determinanten, wie bspw. der Anonymität der Befragungssituation oder dem Anerkennungsbedürfnis. Im Folgenden werden deshalb auch solche Operationalisierungen vorgestellt, die nicht mit direktem theoretischen Bezug zum RC-Modell entwickelt wurden, gleichwohl aber als Messungen für die jeweiligen Konstrukte verwendet werden können. Anschließend an die Vorstellung erfolgt die kritische Diskussion der einzelnen Ansätze.

4.1 Erklärungsfaktoren der Handlungskonsequenz: soziale Anerkennung

Die in der bisherigen Forschung am umfangreichsten betrachtete Handlungskonsequenz stellt die soziale Anerkennung dar. Sie umfasst die Erklärungsfaktoren: Anerkennungsbedürfnis, Anonymität und Erwünschtheitswahrnehmung. Zu allen drei Faktoren wurden unterschiedliche Formen der Operationalisierung entwickelt (Abbildung 4), die im Folgenden dargestellt werden sollen.

Abbildung 4: Erklärungsfaktoren und Operationalisierungen der Handlungskonsequenz „soziale Anerkennung“



4.1.1 Anerkennungsbedürfnis

Die Forschung zur Messung des Anerkennungsbedürfnisses hat in den letzten 50 Jahren viele Skalen hervorgebracht. Bereits in den 60ern wurden Items entwickelt, die bis heute im Original oder in einer modifizierten Variante eingesetzt werden. Die Begriffsvielfalt für die Skalen zur Messung des Anerkennungsbedürfnisses ist groß. In der Literatur finden sich Bezeichnungen wie „Lügenskala“, „Täuschungsskala“ und „SD-Skala“. Alle diese Konstrukte berufen sich darauf, das Anerkennungsbedürfnis („need for social approval“) zu erfassen. Bei der Messung der Tendenz zur sozialen Erwünschtheit lassen sich grundlegend zwei verschiedene Ansätze unterscheiden. Zum einen ist es möglich, soziale Erwünschtheit als Persönlichkeitskonstrukt zu interpretieren. Bei der Erfassung im Rahmen einer Erhebung wird dabei versucht, die Tendenz der Befragten zu konformen Verhalten zu messen. Die Personen werden gebeten, eine persönliche Einschätzung über ihre eigene Tendenz zur sozialen Erwünschtheit zu geben. Zum anderen wird

bei der itembasierten Erhebung davon ausgegangen, dass es Verhaltensweisen gibt, die mehr oder weniger sozial erwünscht sind (DeMaio 1984: 258). Auf Grundlage dieser Unterscheidung werden den Befragten eine Reihe von sozial erwünschten aber sehr selten auftretenden bzw. sozial unerwünschten aber verbreiteten Verhaltensweisen vorgelegt. Diesen Verhaltensweisen muss der Respondent zustimmen oder sie ablehnen. Man geht davon aus, dass die Befragten im Interview dazu neigen sozial erwünschte Verhaltensweisen zu bejahen und sozial unerwünschte zu verneinen. Dabei hängt die Beantwortung von den Rahmenbedingungen (Anonymität, Wahrnehmung von Normen) der Situation ab. Aus diesem Grundgedanken leiten sich auch die Begriffe Lügen- und Täuschungsskala ab. Die Zustimmung zu dem Item „Ich sage immer die Wahrheit.“ wird deshalb als Täuschung zugunsten einer positiven Selbstdarstellung gewertet. Eine große Anzahl solcher Täuschungen geht also, so die theoretische Vermutung, mit einem hohen Anerkennungsbedürfnis einher. Welche Problematiken sich durch diesen Ansatz ergeben, wird im Anschluss an die Vorstellung der Operationalisierungsvarianten diskutiert. Im Folgenden sollen nun die wichtigsten Skalen, die auf Basis der Item-Eigenschaften entstanden sind, vorgestellt werden.

4.1.1.1 Edwards Skala

Die erste Skala, die mit dem Ziel entwickelt wurde, Erwünschtheitsverzerrungen zu messen, stammt von Edwards (1957). Er entwickelte eine Fragenbatterie, die sowohl die oben benannten itembasierten Messungen als auch Persönlichkeitskonstrukte beinhaltet. Grundlage für die Itemauswahl bildete das Minnesota Multiphasic Personality Inventory (MMPI), ein Persönlichkeitstest, der aus zehn klinischen Skalen besteht, die mit dem Zweck entwickelt wurden, psychiatrische Klassifikationen von Personen vorzunehmen. Der Schwerpunkt liegt auf der Unterscheidung zwischen „normalen“ Personen und solcher mit klinischer Störung (z.B. Schizophrenie) (Zimbardo/Gerrig 2008: 539). Edwards ließ in einer ersten Studie 1957 die verschiedenen Items dieser Skalen von Studierenden nach ihrer sozialen Erwünschtheit beurteilen, um auf Grundlage der Ergebnisse eine erste, 79 Items umfassende, Skala zu konstruieren. Von diesen 79 Items wurden in einer Anschlussstudie die 39 Items mit der größten Trennschärfe ausgewählt. Diese neue SD-Skala korrelierte sowohl positiv mit der Status- und Dominanzskala als auch negativ mit der Introversions- und Angstskala des MMPI (Reinecke 1991a: 87).

Zusammenfassend leitet sich aus der Operationalisierung der Skala von Edwards eine bestimmte Form des sozial erwünschten Antwortverhaltens ab: Personen mit hoher sozialer Erwünschtheit versuchen, bei der Beantwortung der Fragen einen guten Eindruck zu hinterlas-

sen. Als Resultat stimmen sie vor allem solchen Items zu, die in Übereinstimmung mit allgemeinen gesellschaftlichen Werten und Normen stehen (Reinecke 1991a: 90). Voraussetzung ist jedoch, dass die soziale Erwünschtheit „a fairly stable personality characteristic“ (Edwards 1957b: 39) darstellt.

Bewertung der Skala

Verschiedene Kritikpunkte lassen aber schon an der Gültigkeit der Edwards-Skala zweifeln. Während zum einen früh die Mehrdimensionalität der Skala mittels der Faktorenanalyse nachgewiesen (Messick 1960) und damit die von Edwards postulierte Eindimensionalität der Skala widerlegt wurde, stand zum anderen die fehlende Ausbalancierung der positiv und negativ formulierten Items in der Kritik. Als Reaktion entwickelte Edwards zwei Kontrollskalen, die sich jeweils nur aus positiv bzw. negativ formulierten Items zusammensetzten. Die Korrelation mit der Ursprungsskala machte deutlich, dass beide Sonderformen die Ergebnisse reproduzieren konnten. Edwards konnte damit eine Ja- oder Neinsagetendenz nicht nachweisen und schlussfolgerte aus diesen Ergebnissen die Stabilität seiner Skala (Reinecke 1991: 88). Die schwerwiegendste Kritik an der Skala bezieht sich jedoch auf die enge Verbindung mit den Persönlichkeitsskalen des MMPI. So lassen sich die hohen Korrelationen auch durch die große inhaltliche Nähe beider Konstrukte erklären. Demnach messe die SD-Skala nicht die Tendenz zu sozial erwünschten Antworten, sondern stelle vielmehr eine verkürzte Persönlichkeitsskala dar (Crowne/Marlow 1960).

4.1.1.2 Marlow Crowne Skala

Um die beschriebenen Schwachpunkte der Edwardsskala auszugleichen, wählten Marlow und Crowne (1960) einen anderen Ansatz zur Entwicklung ihrer Skala. Die grundlegenden Kritikpunkte an der Skala von Edwards veranlassten sie, ein neues Konzept zur verbesserten Messung der Erwünschtheitsverzerrung vorzustellen (DeMaio 1984:260). Ihr Hauptkritikpunkt bezog sich dabei auf die Basis der Itemauswahl. Demnach ist der MMPI vor allem für pathologische Persönlichkeitsstörungen konstruiert worden. Das hat zur Folge, dass hohe Werte auf der Skala nicht das Bedürfnis messen, sich besser darzustellen, sondern vielmehr das Vorkommen pathologischer Symptome in einer speziellen Population (Crowne/Marlow 1960:350).

Der Meinung von Crowne und Marlow nach kann die Messung des SD-Bias nicht über Persönlichkeitsitems erfolgen, da deren Beantwortung ebenfalls der Reflexion bestimmter gesellschaftlicher Normen unterliegt. Die Befragungssituation muss darüber hinaus als ein Wechselspiel zwischen Zielen und Erwartungen der Befragten betrachtet werden, in denen beide Fakto-

ren bei jeder Frage neu gegeneinander aufgewogen werden (Reinecke 1991a: 95). Für die entsprechende Umsetzung dieses Ansatzes in einer Skala wählten die Autoren verschiedene Items aus, die sowohl gesellschaftliche Anerkennung messen, als auch eine minimale pathologische Konsequenz bei einer sozial erwünscht oder auch sozial unerwünschten Antwort implizieren (Crowne/Marlow 1960: 350). Diese 50 Items bekamen 10 Mitarbeiter bzw. Studierenden des Departments für Psychologie an der Ohio State University zur Beurteilung vorgelegt. Ihre Aufgabe war es, die Aussagen auf ihre soziale Erwünschtheit bzw. Unerwünschtheit zu bewerten. Die 47 Items mit der stärksten Übereinstimmung in den Bewertungen wurden anschließend ausgewählt und zusammen mit der Edwardsskala auf ihren pathologischen Hintergrund geprüft. Anschließend wurde der Grad der Anpassung von zehn ausgewählten Personen auf einer fünf Punkte Skala eingeschätzt. Die Analyse belegte bei der Edwardsskala signifikant höhere Werte bei der Einschätzung der pathologischen Implikationen. Im dritten Schritt wählten die Autoren auf Grundlage einer Befragung von 76 Studierenden die Items mit der größten Trennschärfe aus. Das Ergebnis war die 33 Items umfassende MC-Skala zur Messung der Erwünschtheitsverzerrung. Die Korrelation mit den MMPI-Skalen zeigte weniger signifikante Zusammenhänge der MC-Skala als mit der ursprünglichen Edwardsskala, wodurch sich die Autoren in ihrem Ziel, eine Skala möglichst unabhängig von pathologischen Störungen zu entwickeln, bestätigt sahen. Lück und Timaeus (1969) übertrugen die Skala erstmals in Deutsche. Dafür übersetzten die Autoren zunächst alle 33 Items und legten sie anschließend 110 Studierenden zur Beurteilung vor. Zehn Items mit ungenügender Trennschärfe wurden in der anschließenden Analyse ausgeschlossen (Stöber 1999).

Beispielitems für die MC-Skala:

- Ich bin immer ein guter Zuhörer, egal, mit wem ich mich unterhalte.
- Ich bin immer bereit, meine eigenen Fehler zuzugeben.
- Manchmal versuche ich es jemandem heimzuzahlen, anstatt Gras darüber wachsen zu lassen.

Bewertung der Skala

Aber auch die MC-Skala wird sowohl in Bezug auf Validität als auch auf ihre inhaltliche Ausrichtung kritisiert. Edwards et al. (1962: 224) bemängeln, dass die MC-Skala eher eine Messung der Lügentendenz als der sozialen Erwünschtheit darstellt. Jost Reinecke stellte in „Interviewer- und Befragtenverhalten“ fest, dass beide Skalen sich auf unterschiedliche Reaktivitätseffekte und somit Dimensionen der Erwünschtheit beziehen (Reinecke 1991: 102). Somit ist eine Dis-

kussion über die „bessere Skala“, wie sie Edwards und Marlow/Crowne führten, nur begrenzt möglich.

Verschiedene Studien haben sich mit der Reliabilität und Validität (insbesondere der Konstruktvalidität) der Skala beschäftigt. Ein Versuch die Validität der Skala zu überprüfen findet sich bei Bradburn und Sudman (1979). Sie setzten die Skala in Bezug zu den sozio-ökonomischen Faktoren Alter, Bildung, Einkommen, Rasse, Wohnlage und Familienstand. Im Ergebnis zeigten sich die stärksten Zusammenhänge mit dem Alter und der Bildung, während das Einkommen und die Stadtgröße eine geringere Rolle spielten. In einer Reliabilitäts-Generalisierungs-Studie untersuchten Beretvas et al. (2002) verschiedene Einflussfaktoren und ihren Effekt auf die Reliabilität der Marlow-Crowne Skala. Als Datengrundlage dienten Studien aus vier großen Datenbanken, die zum Großteil aus Studierendenstichproben bestanden. In einem mixed-effect Modell kontrollierten die Autoren den Effekt des Alters, den Anteil der Männer sowie die Itemanzahl. Es zeigte sich, dass gerade das Geschlecht einen deutlichen Effekt auf die Reliabilität hat. Bei Frauen fand sich im Vergleich zu Männern eine stärkere interne Konsistenz. Jugendlichen weisen nur inakzeptable Reliabilitätswerte auf.

Die Struktur der MC-Skala untersuchte Barger (2002). An neun Versionen der MC-Skala (neben der Originalskala eine Reihe von Kurzskalen) überprüfte er die theoretisch vermutete eindimensionale Struktur mittels einer konfirmatorischen Faktorenanalyse. Alle Skalen zeigten nur eine geringe Passung. Außerdem unterschieden sich die Werte stark zwischen den zwei getesteten Studierendenstichproben. In einer anschließenden explorativen Faktorenanalyse versuchte Barger, der Struktur der MC-Skala näherzukommen. Die Analyse zeigte unübersichtliche 12 Faktoren der Gesamtskala. Als Schlussfolgerung stellt Barger die Anwendbarkeit der MC-Skala infrage und kritisiert den weiterhin bedenkenlosen Einsatz der Items.

Ebenfalls mit dem Ziel, die Konstruktvalidität der MC zu überprüfen, führten Leite und Beretvas (2005) an den bei einer Studierendenstichprobe erhobenen Items eine konfirmatorische Faktorenanalyse durch. Sie testeten dabei verschiedene theoretische Annahmen über die Faktorenstruktur der Skala. Eine zweidimensionale Struktur der MC-Skala wurde bspw. von Paulhus (1984) in Form von Fremd- und Selbsttäuschung als auch von Ramanaiah et al. (1977) getrennt nach positiven und negativen Formulierungen vorgeschlagen. Im Ergebnis zeigte sich weder die ein- noch die zweidimensionale Struktur der MC-Skala. Die Autoren vermuten hingegen, dass die MC-Skala ein mehrdimensionales Konstrukt darstellt. Auch die getesteten Kurzformen der Skala wiesen keine besseren Ergebnisse auf. Die Autoren empfehlen jedoch eine weitere Überprüfung der Skalen und fordern zusätzliche Forschungen auf dem Gebiet der Kon-

strukturalidität. Ihrer Meinung nach wäre es hilfreich, die MC-Skala, nicht wie von den Entwicklern vorgesehen, nur dichotom zu erheben, sondern auch mit einer ordinalen Antwortskala zu überprüfen.

4.1.1.3 SES-17 Skala

40 Jahre nach Entwicklung der MC-Skala stellte Stöber (1999) die Aktualität der Items infrage. Nach vier Jahrzehnten, so seine Argumentation, hat sich das Bild von sozial unerwünschten und erwünschten Verhaltensweisen in der Gesellschaft, besonders bei den jüngeren Generationen, stark geändert. Um die Skala zu aktualisieren, sollten die Items deshalb einer grundlegenden Überarbeitung unterzogen werden.

Nach demselben Schema, wie schon bei der Entwicklung der MC-Skala, beantworteten 44 Personen³⁹ die Frage, welches Verhalten zwar häufig vorkommt, aber sozial unerwünscht ist, bzw. welches Verhalten selten vorkommt, aber sozial erwünscht ist. Die rund 770 Nennungen wurden kategorisiert und zusammengefasst. Die daraus resultierenden 34 Items sowie die deutsche Übersetzung der MC-Skala beantworteten 91 Psychologiestudierende. Items mit geringer Trennschärfe wurden in der anschließenden Analyse ausgeschlossen, um die verbleibenden Aussagen in einer dritten Erhebung auf ihre Konsistenz und Validität zu überprüfen. Im Ergebnis präsentierte Stöber mit seiner SES-17 Skala eine neue Version der MC-Skala. Ob das Ziel dieser Studie, d.h. die Aktualisierung der Items der MC-Skala, wirklich erfolgreich war, wurde anschließend an 49 Studierenden getestet. Sie sollten die SES-17 Skala sowie die deutsche Version der MC-Skala erneut danach beurteilen, für, wie sozial erwünscht bzw. unerwünscht sie den Inhalt der einzelnen Items einschätzen. Alle getesteten Güterkriterien zeigten bessere Werte bei der SES-17 Skala als bei der MC-Skala (Stöber 1999: 3).

Beispielitems für die SES-17 Skala⁴⁰:

- Manchmal werfe ich Müll einfach in die Landschaft oder auf die Straße.
- Ich habe schon einmal geliehene Sachen nicht zurückgegeben.
- Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte.

39 Die Personen bestanden aus Studierenden, Teilnehmer einer Berufsberatung und dem Freundeskreis zweier Forschungspraktikanten.

40 Die Auswahl der Items ist auch identisch mit der SES-17 Kurzska.

4.1.1.4 BIDR-Skala

Alle bisher vorgestellten Skalen sollen das Anerkennungsbedürfnis ausschließlich eindimensional erfassen. Paulhus kritisierte dies als unzureichende Skalenkonstruktion in der Forschung zur sozialen Erwünschtheit und versuchte ein neues und umfangreicheres Konzept zu entwickeln (Paulhus 1984). Seine Überlegungen basieren auf der Tatsache, dass verschiedene Faktorenanalysen zu den bisher entwickelten Skalen eine zwei-faktorielle Trennung der Items aufzeigten (Damarin/Messick 1965, Sackeim/Gur 1978). Damarin und Messick interpretierten die zwei gefundenen Faktoren als Selbstachtung und propagandistische Verzerrung. Sackheim und Gur hingegen sahen in den Faktoren die Konstrukte Fremd- und Selbsttäuschung. Paulhus versucht in seinem Ansatz, diese beiden Neuinterpretationen miteinander zu verbinden. Er hielt es für notwendig, die eindimensionalen Skalen sowohl von Edwards als auch von Marlow und Crowne durch ein zweidimensionales Konstrukt, das Fremdtäuschung (FT) und Selbsttäuschung (ST) beinhaltet, zu ersetzen. Die Fremdtäuschung ist die bewusste Täuschung mit dem Ziel sich in einem möglichst positiven Bild darzustellen (Paulhus 2002: 55f). Bei dieser Dimension handelt es sich um eine Form der bewussten Verfälschung von Antworten. Erfasst wird dieselbe Persönlichkeitsdimension wie bei den klassischen Täuschungsskalen. Neben der bewussten Täuschung erfasst das Modell von Paulhus aber gleichzeitig die Dimension der Selbsttäuschung. Bei der Selbsttäuschung dienen unbewusste Verzerrungen der Antworten dem Ziel, das Selbstbild und Selbstwertgefühl zu schützen. Eine gewisse Tendenz zur Selbsttäuschung ist damit bei jeder Person zu finden und kennzeichnet ein psychisch gesundes Individuum (Musch et al. 2002: 121).

Paulhus entwickelte das Balanced Inventory of Social Desirability Responding (BIDR-Skala) 1984 auf Grundlage der Items von Sackeim und Gur (1978). Für die Beurteilung der Validität der Skala wurde die Edwards und Marlow-Crowne Skala herangezogen und von 86 Studierenden beantwortet. Mittels einer Faktorenanalyse konnte Paulhus belegen, dass die beiden Subskalen Fremd- und Selbsttäuschung als separate Faktoren interpretiert werden können (Musch et al. 2002: 122).

Musch et al. (2002) übertrugen die BIDR Skala ins deutsche. Dazu übersetzten sie die 40 Items und legten sie anschließend Schülern sowie Studierenden zur Beantwortung vor. Dabei verwendeten sie das siebenstufige Antwortformat der Originalskala. Die anschließende Faktorenanalyse zeigte, in Übereinstimmung mit dem englischen Original, die zwei Faktoren Fremd- und Selbsttäuschung. Die trennschärfsten Items mit mittlerer Schwierigkeit wurden für die

endgültige Skala ausgewählt, sodass zum Schluss 10 Items jeweils für das Konstrukt der Fremd- und der Selbsttäuschung übrig blieben. Drei weitere Kreuzvalidierungsstudien dienten der Überprüfung der Messung der neuen Skala.

Um die BIDR Skala auch in Bevölkerungsumfragen nutzen zu können, entwickelten Winkler et al. (2006) eine Kurzversion der deutschen Skala. Für einen Pretest im Rahmen des SOEP (Sozio-oekonomisches-Panel) wählten die Autoren 10 Items aus, um in einer anschließenden Faktorenanalyse für jedes Konstrukt drei Items zu bestimmen. Basierend auf rund 770 Fällen wurde außerdem die Reliabilität sowie interne und externe Validität der neuen Kurzsкала geprüft. Hierfür untersuchten Winkler et al. (2006) unter anderem die Korrelationen mit den Big-Five⁴¹ und der MC-Skala. Die Ergebnisse entsprachen den theoretischen Erwartungen, woraus die Autoren sowohl die Reliabilität als auch die Validität für bestätigt ansahen.

Beispielitems für die BIDR-Skala⁴²:

- Es ist schon mal vorgekommen, dass ich jemanden ausgenutzt habe. (Fremdtäuschung)
- Manchmal lüge ich, wenn ich muss. (Fremdtäuschung)
- Ich habe schon einmal zu viel Wechselgeld herausbekommen, ohne es der Verkäuferin zu sagen. (Fremdtäuschung)
- Ich bin mir meiner Urteile sehr sicher. (Selbsttäuschung)
- Der erste Eindruck, den ich von anderen Menschen gewinne, bewahrheitet sich meistens. (Selbsttäuschung)
- Ich weiß immer, warum ich etwas mag. (Selbsttäuschung)

Bewertung der Skala

Viele Studien beschäftigen sich mit der Validität der BIDR-Skala. Die Ergebnisse zeigen besonders in Bezug auf die faktorielle Struktur der Skala Probleme auf. Eine Metaanalyse von Li und Bagger (2006) machte deutlich, dass „social desirability did not create spurious effects on the relationships between personality measures and performance, nor did it function as a predictor of performance criteria“ (Li/Bagger 2006: 138).

41 Der Big-Five ist ein Persönlichkeitstest, der fünf zentrale Charaktereigenschaften erfasst: Extraversion, Offenheit für neue Erfahrungen, Verträglichkeit, Gewissenhaftigkeit und Neurotizismus.

42 Die Auswahl der Items ist auch identisch mit der BIDR-Kurzsкала, die in den GMF-Studien eingesetzt wurden.

Leite und Beretvas (2005) überprüften die BIDR-Skala an den Items einer Studierendenstichprobe mittels einer konfirmatorischen Faktorenanalyse. Es konnten keine zufriedenstellenden Ergebnisse in Bezug auf die Zweidimensionalität gefunden werden. Während die Selbsttäuschung einen alleinigen Faktor bildete, scheint die Fremdtäuschung eher zweidimensional angelegt zu sein.

Kroner und Weekes (1996) testeten die Skala an verschiedenen Stichproben von Strafgefangenen. Sie konnten mit einer konfirmatorischen Faktorenanalyse drei Dimensionen identifizieren: die von Paulhus (1984) vorgeschlagene Ebene der Fremdtäuschung sowie die Leugnung negativer Eigenschaften und die überzogene Selbstsicherheit.

In Deutschland führten Musch et al. (2002) eine Studie an Schülern durch, mit dem Ziel eine Kurzversion der BIDR-Skala zu entwickeln. Sie testeten die zweifaktorielle Struktur jedoch nicht mit einer konfirmatorischen, sondern über eine explorative Faktorenanalyse. Im Ergebnis berichten die Autoren die zwei von Paulhus postulierten Dimensionen. Die vollständige Analyse wird jedoch nicht berichtet. Es bleibt hier also offen, auf wie viele Faktoren die Items tatsächlich laden.

Wesentlich genauer gingen Li und Bagger (2007) in einer Reliabilitäts-Generalisierungs-Studie vor, bei der sie die Ergebnisse zur Reliabilität verschiedener Studien miteinander verglichen. Sie konnten eine hinreichende Reliabilität der Gesamtskala belegen. Die einzelnen Dimensionen, insbesondere die Selbsttäuschung, zeigten hingegen keine befriedigenden Werte. Eine der Kontrollvariablen stellte der Stichprobentyp dar. Wurde die Skala an einer Studierendenstichprobe getestet, zeigte sie bessere Reliabilitätswerte, als bei nicht Studierenden. Der Effekt war jedoch nicht signifikant. Einen deutlich positiven Einfluss hatte hingegen die Anzahl der genutzten Items. Weniger überraschend ist der Befund, dass je mehr Items der BIDR-Skala in einer Studie verwendet wurden, umso bessere Werte die Gesamtskala in der Reliabilitätsanalyse erreichte. Ein Test der BIDR über zwei Nationen (USA und Singapur) sowie individualistischen und kollektivistischen Wertvorstellungen führten Li und Reb (2009) durch. In einer konfirmatorischen Faktorenanalyse wurden keine Unterschiede zwischen den Werteinstellungen sichtbar. In Bezug auf die Länder konnten die Autoren unterschiedliche Strukturen der Skala nachweisen, die für eine schwache Äquivalenz der BIDR sprechen.

4.1.1.5 Bewertung

Zusammenfassend lässt sich festhalten, dass es eine Reihe verschiedener Skalen gibt, die zur Messung des Anerkennungsbedürfnisses eingesetzt werden. In Bezug auf ihre Verwendung sind mehrere Ebenen problematisch:

- der Einsatz der Skalen im Rahmen des Rational-Choice-Ansatzes,
- die verwendeten Begrifflichkeiten,
- die Reichweite und Aktualität,
- die theoretischen Strukturen der Skalen und
- die Interpretation der Items

Einsatz der Skalen im Rahmen des Rational-Choice-Ansatzes

Wie bereits erwähnt, lässt sich aus dem RC-Modell nach Esser ableiten, dass es bei der Messung auf das situationale Anerkennungsbedürfnis ankommt. Das spätere Modell von Stocké benennt diese Konkretisierung zwar nicht mehr, beruft sich jedoch auf die theoretischen Arbeiten von Esser. In verschiedenen Forschungsarbeiten wird die MC, SES-17 und BIDR Skala aber als Persönlichkeitskonstrukt interpretiert (siehe bspw. Banse/Gawronski 2003, Stocké 2004, Skarbak-Kozietulska et al. 2012). Bei genauer Betrachtung wird deutlich, dass die Messung keinesfalls unabhängig von der Situation stattfindet. Den Respondenten wird unterstellt, dass sie in Interviews zugunsten der positiven Selbstdarstellung und der sozialen Anerkennung sozial unerwünschte Verhaltensweisen verheimlichen. Diese Interpretation setzt jedoch zwei entscheidende Rahmenbedingungen voraus, die schon aus dem RC-Modell bekannt sind: Zum einen muss der Befragte eine soziale Norm in Bezug auf die Items wahrnehmen und zum anderen das Gefühl haben, dass seine Antworten vor dem Interviewer oder einer vorgestellten Öffentlichkeit wahrgenommen werden. Zugespitzt formuliert spielt also auch schon bei der Messung des Anerkennungsbedürfnisses über die vorgestellten Skalen die Erwünschtheitswahrnehmung und die Anonymität der Situation eine Rolle. Damit handelt es sich nicht mehr um die klassische Messung eines Persönlichkeitskonstruktes, sondern vielmehr um die Messung eines situationalen Bedürfnisses. Die Messung der Skalen unterliegt also denselben situationalen Einflüssen wie die Messung der sensitiven Items selbst. Dieser Zusammenhang wird bisher in Analysen jedoch nicht diskutiert.

Verwendete Begrifflichkeiten

Die zweite Problematik bezieht sich auf die verwendeten Begrifflichkeiten bei dem Einsatz der Skalen. Wie bereits erwähnt, finden sich in den einschlägigen Arbeiten Formulierungen wie Lügenskala, Täuschungsskala und SD-Skala, die zumeist synonym verwendet werden. Die Bezeichnungen Lüge und Täuschung beschreiben am besten den Grundgedanken der Skalen, denn den Befragten wird eine bewusste Manipulation zugunsten der Selbstdarstellung unterstellt. Die Entwickler der Skalen nehmen an, dass diese Täuschungstendenz mit dem Anerkennungsbedürfnis eng in Verbindung steht. Es ist also zu beachten, dass keine der Skalen eine direkte Messung dieses Bedürfnisses darstellt. Besonders irreführend ist die Bezeichnung SD-Skala. Sie unterstellt, dass mit den jeweiligen Items die soziale Erwünschtheit gemessen werden kann. Diese Interpretation geht jedoch weit über den Geltungsbereich der Skalen hinaus. Wie die theoretischen Arbeiten zu diesem Thema aufgezeigt haben, ist für die Messung der sozialen Erwünschtheit die Kontrolle weiterer Faktoren notwendig. Die vorgestellten Skalen messen hingegen nur die Tendenz eines Befragten zur Täuschung. Im Folgenden werden die Skalen deshalb ausschließlich als Täuschungsskalen bezeichnet.

Reichweite und Aktualität

Ebenfalls kritisch sind vor diesem Hintergrund die Stichproben zu bewerten, an denen die bisherigen Skalen entwickelt wurden. Für alle vorgestellten Skalen bildeten Studentstichproben die Grundlage für die normative Bewertung. Ob die Vorstellungen über sozial unerwünschte und erwünschte Verhaltensweisen dieser Gruppe jedoch auf die Bevölkerung übertragbar sind, ist nicht untersucht worden. Voraussetzung für das Funktionieren der Skala ist, dass die Items von den Befragten tatsächlich als sozial erwünscht oder sozial unerwünscht interpretiert werden. Die MC-Skala zum Beispiel basiert auf den Normvorstellungen der 50er Jahre. Es ist durchaus fraglich, ob die Items heute noch Aktualität besitzen (Kritik in Bezug auf die Aktualität der Items findet sich auch bei Stöber 1999).

Theoretische Strukturen der Skalen

Aus der Darstellung des Forschungsstandes ergeben sich zum Teil erhebliche Lücken in der Prüfung der Skalen. Zunächst ist festzuhalten, dass die Faktorenstruktur nur weniger Skalen in ihrer deutschen Version getestet wurde. Der Großteil der Studien stammt aus den USA. Da aber die Analyse von Li und Reb (2009) Hinweise dafür liefert, dass die Skalen nicht kritiklos von einem Land auf das andere übertragbar sind, bleibt ihre Anwendbarkeit in Deutschland weiterhin fraglich. Die wenigen deutschen Analysen wurden zudem nur auf Grundlage von ex-

plorativen Faktorenanalysen durchgeführt und nicht mit der angemesseneren aber restriktiveren konfirmatorischen Analyse überprüft. Besonders uneinheitlich sind die Ergebnisse für die MC-Skala, die anstatt einer Dimension in fast allen Studien eine mehrdimensionale Struktur bildet, dabei werden in allen Studien die Originalitems der Skala verwendet. Aus diesem Grund überarbeitete Stöber (1999) die Fragen grundlegend. Diese neue deutsche Version der MC-Skala wurde bisher jedoch nicht ausreichend auf ihre Faktorenstruktur hin überprüft. Erhebungen, in denen die MC oder SES-17 Skala eingesetzt wurden, beschränken sich zumeist auf die Angabe des Crombachs Alpha, als Maß für die Reliabilität bzw. interne Konsistenz (siehe bspw. Stocké 2004). Die Verwendung des Crombachs Alpha als Reliabilitätsmaß setzt jedoch bereits eine eindimensionale Skala voraus. Gerade diese Bedingung wurde in Bezug auf die MC-Skala jedoch häufig infrage gestellt. Die Überprüfung der Reliabilität mittels Crombach Alpha ist damit nicht ausreichend. Auch bezogen auf die BIDR-Skala fehlt bisher eine umfangreiche Prüfung der zweifaktoriellen Strukturen.

Auch die Kurzskalen sind bisher unzureichend getestet. Besonders im deutschen Kontext werden in verschiedenen Bevölkerungsumfragen (Allbus, GMF-Grad, GMF-Pretest, Cella, SOEP) immer wieder Kurzskalen der MC, SES-17 und BIDR Skala eingesetzt. Oft ist dabei nicht ersichtlich, nach welchen Kriterien die Items für die Kurzversion ausgewählt wurden. Auch findet sich für kaum eine der hier eingesetzten Varianten eine umfangreiche Überprüfung der Validität und Reliabilität. Um den Nutzen des Einsatzes dieser Skalen, unabhängig von ihrer Erklärungskraft, bewerten zu können, sollen im dritten Teil dieser Analyse häufig verwendete Kurzversionen der Täuschungsskalen überprüft werden.

Interpretation der Items

Ein weiterer wichtiger Punkt, der bei dem Einsatz der Skala berücksichtigt werden muss, ist die Interpretation der Antworten. Wie bereits erwähnt, werden die Antworten der Befragten nicht als Selbsteinschätzung, sondern als indirekter Hinweis auf eine Lüge gewertet. Das diese Sichtweise nicht unkritisch zu übernehmen ist, zeigt sich im Vergleich mit dem Konstrukt Selbstüberwachung⁴³. Die aus der Persönlichkeitspsychologie stammenden Items ähneln denen der Täuschungsskalen teilweise stark in ihrem Wortlaut.

„Manchmal lüge ich, wenn ich muss.“ (BIDR-Skala)

„Ich kann jemanden anlügen und ihm dabei mir ehrlicher Miene in die Augen sehen, wenn es einem guten Zweck dient“ (SÜ-Skala)

43 Sowohl das Konstrukt als auch die dazugehörige Skala werden in Kapitel 4.2.1.1 näher dargestellt.

Bei der BIDR-Skala wird eine Ablehnung der Aussage als Neigung zur Täuschung, bei der SÜ-Skala als ein Hinweis auf die positiv gesehene Fähigkeit zur Selbstüberwachung gewertet. Bisher wurden diese verschiedenen Interpretationen der Skalen jedoch nicht empirisch überprüft.

4.1.2 Anonymität

Die subjektive Wahrscheinlichkeit von Konsequenzen bei einer sozial unerwünschten Antwort ergibt sich nach den RC-Modellen unter anderem aus der Anonymität der Befragungssituation. Die Arbeiten, die sich mit dem experimentellen Vergleich von Antwortverhalten zwischen Gruppen mit unterschiedlichen Anonymisierungsgrad beschäftigen, stellen aufgrund ihres Umfangs einen eigenen Forschungsbereich dar. Die Analysen konzentrieren sich zumeist auf die Kontrolle der Anonymität durch objektive Messungen in Form von Experimenten in Bezug auf die Befragungstechnik oder den Erhebungsmodus. Der theoretische Hintergrund beschränkt sich bei diesen Analysen zumeist auf die Bedingungen von anonymen Situationen. Neben der objektiven Messung kann auch die subjektive Wahrnehmung der Privatheit erfasst werden. Beide Messungen wurden zwar nicht konkret für das RC-Modell entwickelt, der Grundgedanke kann jedoch auf das Konstrukt übertragen werden. Die Autoren der RC-Modelle (speziell Esser und Stocké) verweisen darauf, dass der Interviewer in anonymen Situationen nicht direkt auf die Antwort reagieren kann, da er entweder nicht als Person anwesend ist oder die Antwort nicht zugeordnet kann. Damit besteht für den Befragten weniger Gefahr, einer negativen Reaktion ausgesetzt zu sein. Die Kosten einer wahren Antwort entgegen der Erwünschtheitswahrnehmung sinken. Genau an diesem Punkt setzen auch die Messungen der Anonymität an, die im Folgenden dargestellt werden sollen.

4.1.2.1 Objektive Anonymität

In der Methodenforschung werden zwei verschiedene experimentelle Ansätze im Umgang mit dem SD-Bias differenziert. Entscheidend ist dabei, an welchem Punkt der Befragung durch den Forscher eingegriffen wird. Zum einen besteht, wie bereits vorgestellt, die Möglichkeit die individuelle Tendenz zur Antwortverzerrung parallel zu den inhaltlichen Konstrukten zu erheben, um anschließend in der Analyse den SD-Bias mathematisch zu kontrollieren. Zum anderen ist es möglich, bereits bei der Datenerhebung eine Verzerrung des wahren Wertes durch spezielle Befragungstechniken und Erhebungsdesigns zu vermeiden. Beide Methoden versuchen über Manipulationen oder Experimente, ein Gefühl von Anonymität bei den Befragten hervorzurufen.

4.1.2.1.1 Befragungstechniken

In der Methodenforschung wurden in den letzten 50 Jahren verschiedene Techniken zur Reduzierung des SD-Bias entwickelt. Eine vollständige Darstellung all dieser Techniken ist im Rahmen dieser Arbeit nicht notwendig. Dennoch soll ein kurzer Überblick über die drei wichtigsten Techniken: Bogous-Pipeline, Randomized Response (RRT) und das Listenexperiment geben werden. Da der Schwerpunkt des Kapitels Anonymität jedoch auf den Effekten durch Modi liegt, werden die drei Ansätze nur kurz vorgestellt und diskutiert.

Bogus-Pipeline

Die Bogus-Pipeline Technik wurde von Jones und Sigall im Jahre 1971 entwickelt. Ihr Ziel war es die Validitätsangst der Probanden zu nutzen, um durch das Vortäuschen einer Überprüfung der Angaben ehrlichere Antworten zu erhalten. Hierfür wurden Studierende in einem Experiment an eine spezielle Maschine angeschlossen. Den Befragten wurde erzählt, dass es sich bei dieser Maschine um einen Lügendetektor handelt, an den sie während der Befragung angeschlossen blieben. Tatsächlich handelte es sich jedoch um ein funktionsloses Gerät, das nur den Schein einer Überwachung vermitteln sollte. Im Ergebnis zeigte sich bei den Studierenden die an den „Lügendetektor“ angeschlossen waren eine niedrigere Tendenz zur positiven Selbstdarstellung. Diesen Effekt konnten auch Quigley-Fernandez und Tedeschi (1978) in einer Validierungsstudie bestätigen. In der SD-Forschung finden sich verschiedene Studien, die mithilfe der Bogus-Pipeline Technik versuchen sensible Informationen von den Befragten zu erhalten. Eine Metaanalyse zu diesem Thema konnte zeigen, dass durch die Bogus-Pipeline Technik Befragte signifikant häufiger sozial unerwünschte Einstellungen oder Verhaltensweisen äußerten (Roese & Jamieson 1993). Besonders bei sensiblen Fragen zum Verhalten, wie Rauchen (Bauman and Dent 1982; Murray et al. 1987) sowie Alkohol und Drogenkonsum (Tourangeau et al. 1997; Wish et al. 2000) konnten die Ergebnisse bestätigt werden.

Listenexperiment

Das Ziel des Listenexperimentes ist es, durch die besondere Abfrage des sensiblen Themas, ein Gefühl von Privatheit bei den Respondenten hervorzurufen. Hierfür wird der einen Hälfte von Befragten eine Liste mit Statements vorgelegt. Für diese nicht sensiblen Fragen sollen die Respondenten nun angeben über, wie viele, der Aussagen sie sich ärgern. Die andere Hälfte der Befragten erhält neben den neutralen Statements eine zusätzliche Aussage zu dem jeweils sensiblen Thema (Gosen et al. 2013). Die Respondenten werden also ausschließlich danach gefragt, wie viele der Aussagen sie wütend machen und nicht welche. Die Methode soll die Be-

denken der jeweiligen Person verringern und zu mehr ehrlichen Antworten führen (Streb et al. 2008: 81). Ein Beispiel für das Listenexperiment findet sich bei Kuklinski et al. (1997: 405):

“Now I’m going to read you three things that sometimes make people angry or upset. After I read all three, just tell me how many of them upset you. I don’t want to know which ones, just how many”.

1. the federal government increases the tax on gasoline;
2. professional athletes getting million-dollar salaries;
3. large corporations polluting the environment
- (4. black family moving in next door)⁴⁴

Doch wie lassen sich nun, trotz dieser anonymen Beantwortung, Rückschlüsse auf die sensiblen Fragen ziehen? Da alle Personen mittels Randomisierung auf die zwei Gruppen aufgeteilt wurden, sollte der Mittelwert der Items, die wütend machen, in beiden Gruppen gleich hoch sein. Ein höherer Mittelwert in der zweiten Gruppe müsste damit auf die sensitive Frage zurückzuführen sein. Die Differenz der Mittelwerte wird anschließend mit 100 multipliziert (Streb et al. 2008: 81). Folgendes Beispiel soll die Berechnung genauer verdeutlichen:

Mittelwert Gruppe 1: 2,5

Mittelwert Gruppe 2: 3,0

Differenz: 0,5 ($0,5 * 100 = 50$)

Bei diesem Beispiel wären 50 Prozent der Experimentalgruppe verärgert durch das sensitive Statement (Streb et al. 2008: 81). Diese Methode der Anonymisierung wurde bisher für verschiedene Aspekte bei denen ein SD-Bias bedürchtet wird in Bezug auf Rasse (Kuklinski et al. 1997; Sniderman/Carmines 1997; Gilens et al. 1998), Religion (Kane et al. 2004) und andere sensitive Themen angewendet (Dalton et al. 1994; Tsuchiya et al. 2007; Coutts/Jann 2008; Blair/Imai 2011).

Das Listenexperiment wird jedoch zunehmend kritisiert. Besonders neuere Studien stellen immer wieder die Validität der Technik infrage. Gosen et al. (2013) konnten in einer Studie zwei grundlegende Kritikpunkte nachweisen. Zum einen zeigten sich selbst dann zwischen beiden Gruppen signifikante Unterschiede im Mittelwert, wenn das zusätzliche Item nicht-sensitiv war und in einer direkten Befragung nur eine Zustimmung von 2% aufwies. Zum anderen variierten die Ergebnisse Test-Retest Reliabilität abhängig von der Sensitivität des Themas und der Anzahl

⁴⁴ Diese Aussage bekommt nur eine der beiden Befragtengruppen zur Auswahl.

der Items. Diese Ergebnisse zeigen, dass das Listenexperiment nicht unkritisch in der SD-Forschung eingesetzt werden darf.

Randomized-Response Technik

Ähnlich, wie bei dem Listenexperiment, wird versucht, mit der Randomized Response Technik (RRT) bei den Befragten ein Gefühl der Anonymität zu erzeugen. Ursprünglich wurde das Verfahren von Warner (1965) vorgestellt. Die Vertraulichkeit wird dadurch geschaffen, dass der Interviewer nicht weiß, auf welche Frage der Respondent antwortet. Bei der „forced response“ Methode wird per Zufallsmechanismus (z.B. Münzwurf) bestimmt, wie der Befragte antworten soll. Je nachdem, welche Kombination von Kopf oder Zahl bei zwei Würfeln erreicht wird, muss der Befragte dem sensitiven Item entweder zustimmen, es ablehnen oder die Wahrheit sagen. Bei der „unrelated question“ Variante hingegen wird über den Zufallsmechanismus entschieden, ob der Proband die heikle oder nicht-heikle Frage beantworten soll. Da die Wahrscheinlichkeiten der Zufallsmechanismen bekannt sind, lässt sich auf Grundlage der aggregierten Daten die Häufigkeit des sensitiven Verhaltens berechnen (Preisendörfer 2008: 11f.). In einer Metaanalyse bestätigten Lensvelt-Mulders et al. (2005), dass RRT die Häufigkeit von berichteten unerwünschten Verhalten signifikant höher war, als in einem klassischen Face-to-Face Interview. Dieses Ergebnis zeigte sich auch in Studien zum Drogenmissbrauch (Goodstadt/Gruson 1975), Abtreibung (Lara et al. 2004), Kindesmissbrauch (Zdep/Rhodes 1976), Kriminalität (Tracy/Fox 1981) und Sozialbetrug (van der Heijden et al. 2000). Dennoch gibt es immer wieder Kritik an dem Nutzen und der Umsetzbarkeit der Technik. In einer Studie von Preisendörfer (2008) konnte bei Studierenden kein sinnvoller Effekt der RRT festgestellt werden. Zum einen lagen die Probleme in der mangelhaften Umsetzung der Technik durch die Interviewer, zum anderen stellte sich vor allem das erzwungene Ja als problematisch heraus. Die Befragten neigten in solchen Fällen dazu, eine Diskussion mit dem Interviewer zu beginnen. Ähnliche fragliche Ergebnisse fanden auch Coutts und Jann (2008). Der Nutzen und Erfolg dieser Technik ist also durchaus umstritten. Einfacher für die Befragten scheint jedoch zumindest die unrelated-question Variante zu sein (Preisendörfer 2008: 13). Die im deutschen Sprachraum aktuellste Analyse der RRT findet sich bei Wolter (2012). Er konnte in seiner Validierungsstudie allerdings keine höhere Datenqualität durch den Einsatz der RRT bestätigen.

4.1.2.1.2 Erhebungsmodus

Neben den verschiedenen Befragungstechniken gibt es auch einen Deisgneffekt, mithilfe dessen sich die Anonymität in einem Interview an und abschalten lässt. Der Modus, in dem die

Befragung durchgeführt wird, kann genau diesen Effekt hervorrufen. Grundlage bildet die, in der Kommunikationsforschung häufig verwendete, Theorie der Kommunikationskanäle. Die systematische, kontrollierte und zielgerichtete Kommunikation, wird durch die jeweils gewählte Form der Kommunikation beeinflusst. Die Inhalte werden dabei über die sogenannten Kommunikationskanäle transportiert (Taddicken 2008: 93). Insgesamt lassen sich sechs verschiedene Kanäle (Sinne) voneinander unterscheiden:

Tabelle 1: Kommunikationskanäle

Kanal	Mittel
Textkanal	Kommunikation über den Text
Visueller Kanal	Kommunikation über Bilder, Videos, Layout, Mimik, Gestik, Körperhaltung, Blickkontakt und äußere Attribute
Auditiver Kanal	Kommunikation über den Gehörsinn
Taktiler Kanal	Kommunikation über Berührungen
Olfaktorischer Kanal	Kommunikation über den Geruchssinn
Gustatorischer Kanal	Kommunikation über den Geschmackssinn

Quelle: Eigene Darstellung nach Taddicken (2008)

Bei den persönlichen Kommunikationsformen müssen sich die Beteiligten für die Informationsweitergabe am selben Ort zur selben Zeit befinden. Dies trifft vor allem auf den taktilen, olfaktorischen und gustatorischen Kanal zu. Für den Textkanal ist die Anwesenheit eines Interviewers hingegen nicht notwendig. Der visuelle Kanal ist nur zum Teil auf den persönlichen Kontakt angewiesen. Für die visuelle Weitergabe von Informationen über Bilder, Videos und Layout ist, ähnlich wie der Textkanal, ein Interviewer nicht notwendig. Für die Vermittlung von Mimik, Gestik, Körperhaltung, Blickkontakt und äußere Attribute ist die persönliche Kommunikationsform stattdessen zwingend erforderlich. Ähnlich verhält es sich mit dem auditiven Kanal. Eine einseitige Kommunikation ist mittels Tonaufnahmen durchaus möglich, für ein wechselseitiges Gespräch benötigt man aber zumindest einen gemeinsamen Zeitpunkt. Doch in welchen Zusammenhängen stehen diese Kanäle nun mit der Anonymität?

Die Theorie der sozialen Präsenz von Short et al. (1976) zählt zu den klassischen kommunikationswissenschaftlichen Ansätzen. Die Autoren verstehen unter sozialer Präsenz „[...] the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships [...]“ (Short et al. 1976: 65 zitiert nach Taddicken 2008: 94). Definiert wird die Anonymität über die soziale Anwesenheit einer anderen Person. Diese Anwesenheit ist von zwei Einflussfaktoren gekennzeichnet. Zum einen hängt die Wahrnehmung von der Intensität der Kommunikation ab. Zum anderen hat die Anzahl der verwendeten Kommunikati-

onskanäle einen Effekt auf die wahrgenommene Anwesenheit. Je mehr Kommunikationskanäle in einem Interview zum Einsatz kommen, umso informationsreicher und persönlicher wird das Gespräch. Je weniger Kanäle eingesetzt werden, desto geringer ist das Wissen über den Kommunikationspartner. Dadurch sinkt nicht nur die Vertrautheit, sondern auch die Unmittelbarkeit der Situation (Taddicken 2008: 94). Welche Konsequenzen ergeben sich daraus für die verschiedenen Erhebungsmodi? Tabelle 2 gibt eine Übersicht über die momentan am häufigsten eingesetzten Befragungsmethoden und die jeweiligen Kommunikationskanäle, die durch diese Form der Erhebung angesprochen werden können. Der visuelle Kanal wird dabei nochmals in zwei Kategorien unterteilt. Der visuelle Kanal I bezieht sich auf starre und nicht im direkten Kontakt wahrgenommene Bilder, Videos und das Layout. Der visuelle Kanal II umfasst hingegen die Eindrücke, die in der direkten Interaktion mit Menschen, in Form von Mimik, Gestik, Körperhaltung und Blickkontakt, entstehen können.

Tabelle 2: Grad der Anonymität nach Befragungsmodus

Modi	Beschreibung	Textkanal	Visueller Kanal I	Visueller Kanal II	Auditiver Kanal	Taktiler/ Olfaktorischer/ Gustatorischer Kanal
Face to Face	Personal Interview (Aufzeichnungen auf Papier)	- (X)	- (X)	X	X ¹	X
CAPI	Computer Assisted Personal Interviewing (Aufzeichnung auf Computer)	- (X)	- (X)	X	X ¹	X
CATI	Computer Assisted Telephone Interviewing	-	-	-	X ¹	-
ACASI	Audio Computer-Assisted Self-Interviewing	X	X	-	X ²	-
ACAWI	Audio Computer Assisted Web Interviewing	X	X	-	X ²	-
PAPI	Paper And Pencil Interviewing (Selbstausfüller)	X	X	-	-	-
CASI	Computer Assisted Self Interviewing	X	X	-	-	-
CAWI	Computer Assisted Web Interviewing	X	X	-	-	-

Grad an Anonymität ↓

1) Wechselseitige Kommunikation

2) Nur einseitige Kommunikation möglich. Da Interviewer und Befragter nicht gleichzeitig anwesend sind, ist die Verständigen nur jeweils in eine Richtung umsetzbar. Der Interviewer kann so nicht individuell auf bspw. Nachfragen reagieren.

Quelle: Eigene Darstellung

Insgesamt lassen sich vier Gruppen von Modi am Grad der Anonymität unterscheiden. Bei dem Face to Face und CAPI Interview, die sich eigentlich nur durch die Form der Aufzeichnung durch den Interviewer unterscheiden, ist es möglich, die meisten Kommunikationskanäle anzusprechen. Folglich ist bei diesen beiden Formen der Erhebung, bei der sich die beteiligten Personen zur selben Zeit und am selben Ort aufhalten, der geringste Grad an Anonymität zu erwarten. Deutlich anonym ist hingegen das CATI Interview. Am Telefon wird zwar nur ein Kanal angesprochen, nämlich der auditive Kanal, da sich Befragter und Interviewer zwar nicht am selben Ort, aber zur selben Zeit aufhalten, ist eine wechselseitige Kommunikation möglich. Der In-

terviewer kann somit auf Nachfragen reagieren. Wichtig ist für den SD-Bezug außerdem, dass der Respondent in einem CATI Interview bei sozial unerwünschten Antworten mit einer sofortigen negativen Reaktion rechnen muss. Anders ist dies bei Erhebungsformen wie ACASI, T-ACASI und CAWI. Hier werden zwar zwei zusätzliche Kanäle (Textkanal und visueller Kanal I) angesprochen, dennoch ist davon auszugehen, dass die Befragungssituation deutlich anonym ist. Gerade der letzte Aspekt, nämlich die Angst vor der direkten Reaktion, fällt bei diesen Modi aus, da über die Technik nur eine einseitige Kommunikation möglich ist. Die letzte Modus Gruppe mit dem höchsten Grad an Anonymität umfasst PAPI, CASI und CAWI Interviews. Hier wird hauptsächlich über den Textkanal kommuniziert. Das CAWI-Interview stellt, da in der Regel kein persönlicher Kontakt zum Interviewer notwendig ist, eine Sonderform dar, denn die Online-Befragung ist durch ihre räumliche Ungebundenheit und Entzeitlichung gekennzeichnet. Durch diese Virtualisierung der Kommunikation ist sowohl für den Befragten als auch für den Interviewer/Forscher keine sensorische Erfahrung möglich (Taddicken 2008: 93). Dies trifft allerdings auch auf die postalische Befragung, eine Möglichkeit des PAPI Interviews, zu.

Bei dieser Darstellung der Modi ist zu beachten, dass es natürlich noch zusätzliche Sonderformen an Erhebungsmethoden gibt, die hier jedoch, aus Gründen der Übersichtlichkeit, außen vorgelassen wurden. So bietet sich durch moderne Technik mittlerweile die Möglichkeit, z.B. mittels Skype auch bei den CAWI-Interviews einen persönlichen Kontakt herzustellen.

Um die Effekte der Modi und ihrer Anonymitätsgrade auf den SD-Bias zu messen, werden klassischerweise parallele Befragungen in unterschiedlichen Erhebungsformen durchgeführt (siehe bspw. Stocké 2004). Ein etwas anderer Ansatz versucht die hohe Anonymität bei PAPI und CAWI Befragungen durch bestimmte Methoden zu verringern, um die Ergebnisse dann mit der anonymen Variante zu vergleichen. Um die Anonymität zu verringern, werden die Befragten bspw. gebeten, ihren Vor- und Nachnamen anzugeben. Bei Studierendenstichproben wird auch gerne darauf hingewiesen, dass die Antworten später in einem Seminar diskutiert werden.

Bewertung

Die Ergebnisse zu dem Einfluss des Modus auf den SD-Bias lassen jedoch kein klares Bild erkennen. Die, wie in Tabelle 2 vorgestellten, zu erwartenden Unterschiede zwischen den Erhebungsmethoden konnten für Zigaretten- und Drogenkonsum (Currivan et al. 2004; Newman et al. 2002; Tourangeau/Yan 2007), Schulnoten und Schulintegration (Heerweegh 2009; Kreuter et al. 2008), die BIDR-Skala (Joinson 1999), Fragen zur Gesundheit und dem Sexualverhalten (Tourangeau/Smith 1996) sowie zur Einstellung gegenüber Ausländern (Duffy et al. 2005; Stocké

2004) belegt werden. Es finden sich jedoch auch eine Reihe von Studien, die keine Effekte der Modi nachweisen konnten. Hancock und Flowers (2001) fanden keinen signifikanten Unterschied zwischen CAWI und PAPI Befragungen in Bezug auf die BIDR-Skala. Auch Couper et al. (2003) konnten keine Differenz zwischen CASI und ACASI bei der Beantwortung von Kirchgangshäufigkeit, Wahlverhalten und Drogenkonsum aufzeigen. Die umfangreichste Untersuchung zu den Modi Effekten findet sich jedoch von Rossiter (2009). In seiner Dissertation untersucht er die Unterschiede zwischen Face-to-Face, CATI, CAWI und PAPI Befragungen in Bezug auf die Beantwortung der BIDR-Skala. Es zeigte sich keinerlei signifikante Unterschiede zwischen den Modi. Der Autor erklärt dieses ungewöhnliche Ergebnis unter anderem mit seinem umfangreichen Design. Seiner Meinung nach wurde in bisherigen Studien bspw. der Fragereinfolgeeffekt nicht genügend berücksichtigt (Rossiter 2009: 121). Auch bei der Unterscheidung zwischen anonymen und nicht-anonymen Formen der PAPI und CAWI Befragungen finden sich keine einheitlichen Ergebnisse. Während Hancock/Flowers (2001) und Joinson (1999) keinen signifikanten Unterschied zwischen den Experimentalgruppen fanden, zeigten sich bei Couper et al. (2001), Sproull et al. (1996) und Mühlenfeld (2004) leichte Zusammenhänge.

Weiterhin ist fraglich, ob die Anonymität der einzelnen Erhebungsmodi tatsächlich so einheitlich wahrgenommen wird, wie die theoretischen Ansätze es unterstellen. Denkbar ist, dass einzelne Respondenten eine schriftliche Befragung als weniger anonym einschätzen als Befragte in einem Face-to-Face Interview⁴⁵. Dies ist besonders dann problematisch, wenn der Modus als einzige Operationalisierung für die Anonymität eingesetzt wird wie bei Stocké (2001, 2004). Der Grad der wahrgenommenen Privatheit wird in solchen Untersuchungen unterstellt ohne dass eine Überprüfung erfolgt.

4.1.2.2 Subjektive Anonymität

In dem vorangegangenen Kapitel wurde die Kontrolle des Anonymitätsgrades mittels des experimentellen Designs vorgestellt. Eine alternative Möglichkeit besteht darin, die wahrgenommene Anonymität direkt abzufragen. Diese Messung ist vor allem dann zusätzlich notwendig, wenn man davon ausgeht, dass sich die Wahrnehmung individuell zwischen den Personen unterscheidet. Während ein Befragter das Gefühl hat, bei einem CATI-Interview identifizierbar zu sein, kann ein anderer wiederum dieselbe Situation als sehr privat empfinden.

Im Zusammenhang mit sozialer Erwünschtheit gibt es kaum Studien, die sich mit dem Einfluss der subjektiven Anonymität beschäftigen. Vielmehr werden Unterschiede in der subjektiven

⁴⁵ Unveröffentlichte Daten eines Modi-Experimentes an der Universität Göttingen bestätigen diese Annahme.

ven Wahrnehmung bei den Modi-Experimenten unterstellt. Es ist jedoch fraglich, ob diese einseitige Kontrolle der Privatheit ausreichend ist. Dieselbe Befragungssituation kann von Personen aus den verschiedensten Gründen sehr unterschiedlich bewertet werden. Wie groß sind die erwarteten Unterschiede, zwischen einer Person die ein CATI Interview als sehr anonym bewertet und einer Person die eine CAWI Befragung als wenig anonym empfindet, tatsächlich? Aber gerade bei Erhebungen, die nicht durch ein Modusexperiment ergänzt werden, könnte die wahrgenommene Anonymität einen zusätzlichen Erklärungsbeitrag leisten.

In der Studie von Whelan und Meade (2009) wurde systematisch der Zusammenhang zwischen sensitiven Items und wahrgenommener Anonymität untersucht. Sie stellen fest, dass sich bisherige Studien nur auf die experimentellen Messungen beschränken. Sie kritisieren ebenfalls, dass bei dieser Messung individuelle Unterschiede in der Wahrnehmung unberücksichtigt bleiben, und führen Studien an, die diese neue Form der Operationalisierung bereits nutzen (Bates/Cox 2008; Whelan 2008). Bei der Definition des Konstrukts „wahrgenommene Anonymität“ beziehen sich die Autoren auf eine frühere Arbeit von Whelan und Tompson (2008), nach der sich drei Faktoren unterscheiden lassen:

„(a) the perception of the relative nonidentifiability of a respondent, (b) how immersed a respondent feels in a set of survey respondents, and (c) how likely it is that a respondent feels his or her personal information can be traced back to that individual for identification.“ (Whelan/Meade 2009: 2).

Ziel der Studie war es den Zusammenhang zwischen sensitiven Items und Anonymität in einem Web-Survey systematischer zu untersuchen. Im Ergebnis zeigte sich zunächst, dass die wahrgenommene Anonymität auch ohne die experimentelle Modifikation der Modi innerhalb eines Web-Survey variieren kann. In Bezug auf die Zusammenhgangsrichtung fanden die Autoren jedoch heraus, dass nicht die wahrgenommene Anonymität auf die sensitiven Fragen, sondern die sensitiven Fragen auf die Anonymität wirken. Je häufiger die befragten Studierenden zugaben bei Prüfungen oder ähnlichen zu betrügen, desto weniger anonym empfanden sie die Situation. Die Autoren begründen diesen Zusammenhang damit, dass erst das Befürworten negativer Aussagen dazu führt, dass sich die Befragten ihrer möglichen Identifizierbarkeit bewusst werden, besonders weil sie davon ausgehen, dass die Gruppe der Personen, die solch ein Verhalten zu gibt, kleiner ist. Dieses Ergebnis hätte zur Folge, dass nicht nur die Unterstellungen in Bezug auf die Anonymität bei Modi-Unterschieden fraglich sind, sondern auch, dass die wahrgenommene Anonymität eine völlig andere Zusammenhgangsrichtung aufweist, als bei der experimentellen Anonymität vermutet wird. Da es sich hierbei jedoch um die einzigen Studien in

diesem Zusammenhang handelt, müssen die Ergebnisse für eine korrekte Bewertung zunächst auf ihre Stabilität hin überprüft werden.

4.1.3 Erwünschtheitswahrnehmung

Ein zentraler Erkenntnisgewinn im Rahmen des Rational-Choice Ansatzes ist die Bedeutung der Erwünschtheitswahrnehmung. Schon Crowne und Marlow (1964) erwähnten in ihren Arbeiten, dass eine Voraussetzung für die Wirkung des Anerkennungsbedürfnisses ist, dass die Befragten die Antwortoptionen mit einer positiven oder negativen Konsequenz verknüpfen können. Eine notwendige Bedingung wurde bereits in Form der Anonymität der Befragung vorgestellt. Ein zweiter Faktor, der vor allem in dem Modell von Esser hervorgehoben wird, ist die Erwünschtheitswahrnehmung. Nur wenn der Befragte die Antwortoptionen mit einer sozialen Norm verknüpfen kann, d.h. er überhaupt eine erwünschte oder weniger erwünschte Antwort identifiziert, wird er in der Lage sein, auch entsprechend der Norm zu antworten.

Die Erwünschtheitswahrnehmung spielt also eine zentrale Rolle bei der korrekten Interpretation des SD-Bias. Eine sinnvolle Kontrolle der Verzerrung durch soziale Erwünschtheit ist nur möglich, wenn Informationen über die Wahrnehmung der sozialen Norm vorliegen, denn es ist durchaus möglich, dass sich die Stärke oder sogar die Richtung des SD-Bias zwischen verschiedenen Gruppen unterscheidet (Stocké/Hunkler 2004: 2). Unklar ist jedoch, wie diese Erwünschtheitswahrnehmung messbar ist. In der aktuellen Diskussion lassen sich zwei Dimensionen unterscheiden, anhand derer sich die Messungen einordnen lassen: Zum einen kann eine soziale Norm durch unterschiedliche Gruppen transportiert werden. Zum anderen ist es möglich, dass die Messung des Konstruktes durch verschiedene Bezugspunkte erfolgt. Beide Dimensionen sollen im Folgenden kurz dargestellt werden, um anschließend Probleme der bisherigen Operationalisierungen zu diskutieren.

4.1.3.1 Bezugsgruppen

Eine soziale Norm ist während eines Interviews über verschiedene Ebenen wahrnehmbar. Esser spricht in seinem RC-Modell die kulturelle und die situationale Erwünschtheit an. Während sich die erste Ebene auf die Beeinflussung des Respondenten durch allgemeine, gesellschaftliche Normen bezieht, werden in der situationalen Erwünschtheit die Normvorstellungen des Interviewers in der konkreten Situation relevant. Der Befragte kann aber auch, wie von Ajzen und Fishbein (1975) vorgeschlagen, durch die Einstellung wichtiger Dritte, also bspw. von Freunden

und Verwandten, beeinflusst werden. Folgende Fragebeispiele sollen die Unterschiede zwischen den drei Ebenen verdeutlichen:

- 1) „Die Äußerung von Vorurteilen gegenüber Juden wird von dem Interviewer als unangenehm und unpassend empfunden.“ (situationale Erwünschtheitswahrnehmung)
- 2) „Die Äußerung von Vorurteilen gegenüber Juden wird in meinem Freundeskreis als unangenehm und unpassend empfunden.“ (Erwünschtheitswahrnehmung einer sozialen Gruppe)
- 3) „Die Äußerung von Vorurteilen gegenüber Juden wird in der Öffentlichkeit als unangenehm und unpassend empfunden.“ (kulturelle Erwünschtheitswahrnehmung)

4.1.3.2 Persönliche und extrinsische Bewertung sozialer Erwünschtheit

Die zweite Dimension der Operationalisierung der Erwünschtheitswahrnehmung unterscheidet zwischen extrinsischer und persönlicher Evaluation. Stocké und Hunkler (2004) geben in ihrer Studie einen Überblick über die beiden Formen der Bewertungen. Wird das Konstrukt über die erste Form erhoben, müssen die Befragten angeben, für wie sozial erwünscht sie persönlich die betreffende Einstellung oder Verhaltensweise halten. Diese Angabe kann in der anschließenden Analyse als individuelle Information eingehen. Eine andere Möglichkeit besteht darin, die erhobene Bewertung über alle Befragten zu aggregieren, um so eine Information über das normative Klima der Stichprobe zu erhalten (Stocké/Hunkler 2004: 4).

Bei der extrinsisch evaluierten Bewertung werden die Respondenten hingegen gebeten eine Einschätzung darüber zu geben, für wie sozial erwünscht sie eine Einstellung oder Verhaltensweise in der Gesellschaft halten. Im Gegensatz zur ersten Methode werden sie also nicht um ihre persönliche Meinung gebeten, sondern um eine Einschätzung der gesellschaftlichen Sichtweise. Auch bei dieser Variante gibt es wieder die Möglichkeit die individuelle oder aggregierte Information in die Analyse einfließen zu lassen. Zwei Beispiele sollen die unterschiedlichen Operationalisierungen verdeutlichen:

Für die persönliche Bewertung sollen die Befragten für folgende Aussage den Grad der Zustimmung angeben:

- Mir ist es unangenehm in öffentlichen Situationen der Aussage: „Juden haben zu viel Einfluss in Deutschland“ zuzustimmen.

Bei der extrinsischen Bewertung würde der Fragetext hingegen wie folgt lauten:

- Ich bin der Meinung, dass die Aussage „Juden haben zu viel Einfluss in Deutschland“ öffentlich erwünscht ist.

4.1.3.3 Bewertung

Die Operationalisierungsvarianten der Erwünschtheitswahrnehmung werfen zwei zentrale Probleme auf. Auf Grundlage der drei verschiedenen Gruppen sowie der zwei Formen von Bewertungen ergibt sich für die Erwünschtheitswahrnehmung ein mehrdimensionales Konstrukt. Folgende Tabelle gibt einen Überblick darüber, wie ein und dieselbe Fragestellung für jede der Dimensionen lauten würde.

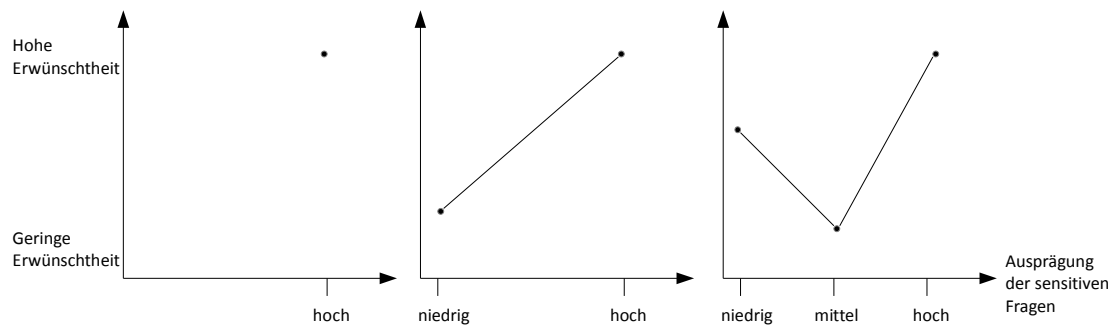
Tabelle 3: Dimensionen der Erwünschtheitswahrnehmung Beispiele

	extrinsisch	intrinsisch
Situationale Erwünschtheitswahrnehmung	Die Äußerung von Vorurteilen gegenüber Juden wird von dem Interviewer als unangenehm und unpassend empfunden.	Mir ist die Äußerung von Vorurteilen gegenüber Juden vor dem Interviewer unangenehm.
Erwünschtheitswahrnehmung einer sozialen Gruppe	Die Äußerung von Vorurteilen gegenüber Juden wird in meinem Freundeskreis als unangenehm und unpassend empfunden.	Mir ist die Äußerung von Vorurteilen gegenüber Juden vor meinen Freunden unangenehm.
Kulturelle Erwünschtheitswahrnehmung	Die Äußerung von Vorurteilen gegenüber Juden wird in der Öffentlichkeit als unangenehm und unpassend empfunden.	Mir ist die Äußerung von Vorurteilen gegenüber Juden in der Öffentlichkeit unangenehm.

Die Differenzierung zeigt, dass je nachdem welche Gruppe zur Normorientierung herangezogen wird und je nachdem, ob die Bewertung persönlich (intrinsisch) oder extrinsisch erfolgt, unterschiedliche Dimensionen angesprochen werden. Jede dieser Dimensionen kann eine andere Ebene des SD-Bias erfassen. Im Rahmen von anonymen Onlinebefragungen ist so bspw. zu erwarten, dass Interviewereffekte keinen Effekt auf die Beantwortung der Frage haben, weil keine Person anwesend ist, die eine Norm direkt vermittelt. Gesellschaftliche Normen könnten sich hingegen durchaus auch in Onlinebefragungen auf das Antwortverhalten auswirken. Die direkte Vergleichbarkeit der einzelnen Effekte ist dabei fraglich. Im Rahmen von Erhebungen ist es deshalb notwendig, dass die einzelnen Dimensionen reflektiert betrachtet werden, sodass eine präzise Einordnung der Ergebnisse im Rahmen der SD-Forschung möglich ist. Die oft widersprüchlichen Ergebnisse bisheriger Forschungsarbeiten lassen sich so bei einer genaueren Analyse vielleicht auf Unterschiede in der Operationalisierung dieses Konstruktes zurückführen.

Ein weiteres Problem im Kontext der Erwünschtheitswahrnehmung ist die Neutralitätsannahme. Üblicherweise wird die Erwünschtheitswahrnehmung in allen Dimensionen über die starke Ausprägung eines Merkmals mit einer bipolaren Skala und einem neutralen Mittelpunkt erhoben (Bsp.: „Für wie sozial unerwünscht halten Sie die Aussage: Juden in Deutschland haben zu viel Einfluss?“). Hiermit ist jedoch eine Neutralitätsannahme verbunden, die laut Stocké und Hunkler voraussetzt, dass eine schwache Merkmalsausprägung als neutral bewertet wird (Stocké/Hunkler 2004: 4). Es sind jedoch auch andere Erwünschtheitsprofile denkbar. Wird zusätzlich zu der Erwünschtheit der starken Ausprägung auch die Bewertung der schwachen Ausprägung erfasst, lassen sich monoton steigende oder fallende Profile identifizieren. Die Differenz beider Werte kann dann als Maß für die Stärke und Richtung verwendet werden. Aber auch hier wird immer noch ein linearer Verlauf unterstellt. Bei einem u-förmigen Profil wäre hingegen sowohl die schwache als auch die starke Ausprägung sozial erwünscht, während einzig die mittlere Ausprägung sozial unerwünscht ist (das Entsprechende gilt auch für umgekehrt u-förmige Profile). Um diese nicht-monotonen Verläufe zu erfassen, schlagen Stocké und Hunkler (2004) die zusätzliche Bewertung der mittleren Ausprägung vor.

Abbildung 5: Profile der Erwünschtheitswahrnehmung



Quelle: Eigene Darstellung nach Stocké und Hunkler (2004)

Ihre empirische Analyse dieser differenzierten Operationalisierungsmethoden zeigte, dass die bisher hauptsächlich verwendete Bewertung nur einer starken Merkmalsausprägung keine der sensitiven Fragen (Blutspende, Lebensglück, Fernsehkonsum, Alkoholismus) adäquat darstellen kann. Stattdessen eignete sich die Operationalisierung über spezifische Differenzwerte deutlich besser für die Abbildung der Erwünschtheitsunterschiede. Diese Ergebnisse zeigen, dass bei der Operationalisierung der Erwünschtheitswahrnehmung in zukünftigen Studien differenziertere Profile erhoben werden sollten, um eine über- oder unterinterpretation des SD-Bias zu vermeiden.

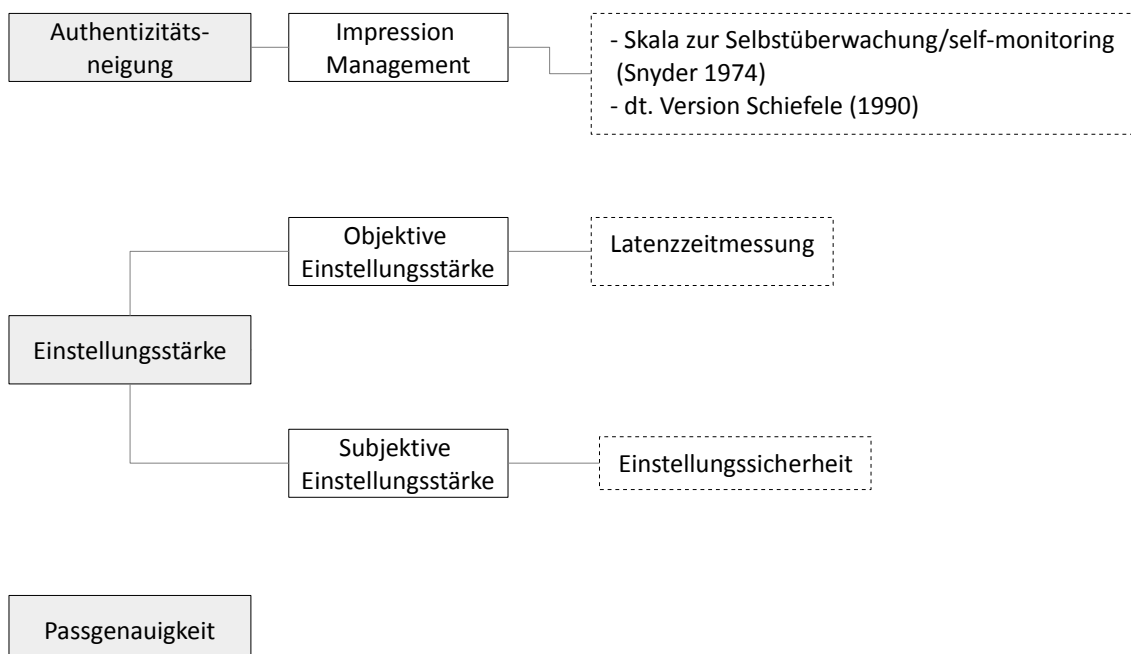
Eine besonders in der Persönlichkeitspsychologie angewendete Variante der Messung der Erwünschtheitswahrnehmung ist die Erfassung über faking good und faking bad Anweisungen.

Hierbei werden die Respondenten gebeten auf eine Frage die sozial akzeptierte Antwort (faking good) oder eben die sozial unerwünschte (faking bad) auszuwählen. Auf Grundlage dieser Daten können anschließend individuelle oder gruppenspezifische Differenzen gebildet werden. Diese Form der SD-Kontrolle wird vor allem in der Persönlichkeitsforschung eingesetzt (Hahn 2005; Rogers et al. 2003; Austin 1992).

4.2 Erklärungsfaktoren für die Handlungskonsequenz: Authentizität

Esser benennt in seinem Modell zur Erklärung des Befragtenverhaltens zwei zentrale Handlungskonsequenzen: soziale Anerkennung und Authentizität. Bei der Konsequenz Authentizität werden dem RC-Ansatz nach (siehe Modelle von Esser und Stocké in Kapitel 3.2.2 und 3.2.3) drei Erklärungsfaktoren abgeleitet: Das Bedürfnis nach Authentizität als Persönlichkeitsmerkmal, die Einstellungsstärke und die Passgenauigkeit der Antwortoptionen. Problematisch bei der Operationalisierung dieser Konstrukte ist jedoch, dass diese Faktoren weder in dem Modell von Esser noch in dem Modell von Stocké näher beschrieben werden. Beide Arbeiten konzentrieren sich in ihren Ausführungen auf das Bedürfnis nach Anerkennung. Während für diese Konstrukte konkrete Messungen vorgeschlagen werden, fällt die Präzisierung der Authentizität deutlich schwerer. Aus diesem Grund handelt es sich bei den folgenden Ausführungen auch nur um Operationalisierungsvorschläge (Übersicht siehe Abbildung 6) der Stand der Forschung bezüglich dieser Faktoren ist wesentlich weniger umfangreich als bei den zuvor behandelten Konstrukten. Ziel dieses Abschnittes ist es deshalb, mögliche Messungen zu vorschlagen sowie Problembereiche aufzuzeigen und zu diskutieren.

Abbildung 6: Erklärungsfaktoren und Operationalisierungen der Handlungskonsequenz „Authentizität“



4.2.1 Authentizität

Der Nutzenwert einer Antwortoption wird nach dem RC-Modell von Stocké unter anderem aus dem Bedürfnis nach Authentizität gebildet. Die Befragten unterscheiden sich in dem Bedürfnis innere Konflikte, die durch die Diskrepanz von Identität und Antwortabgabe entstehen können, zu vermeiden (Stocké 2001). Ist das Bedürfnis nach personaler Identität stark, kann am besten realisiert werden, wenn die gewählte Antwortoption identisch mit der als eigene Meinung wahrgenommenen Einstellung ist. Das Bedürfnis nach Authentizität gibt in diesem Zusammenhang den Grad der Befriedigung beim Ausdruck der authentischen Meinung wieder (Stocké 2001: 5). Aus der Impression-Management Theorie lässt sich eine Möglichkeit der Messung dieses Bedürfnisses nach Authentizität ableiten.

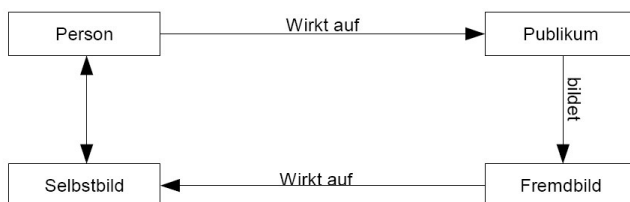
Die Grundlage für diese Theorie bilden die Arbeiten von James (1892), Cooley (1902) und Mead (1934), die im Rahmen der Entwicklung des symbolischen Interaktionismus die Idee eines flexiblen und dynamischen Selbst erstmals diskutierten. In den 50 und 60er Jahre untersuchte Goffman (1956, 1969) die motivationalen Grundlagen für diese Dynamik. Im Zentrum der Motive steht der Wunsch, der einzelnen Personen das Gesicht zu wahren und die soziale Interaktion aufrecht zu halten. Die Selbstpräsentation ist dabei das Resultat aus „hochstrukturierten und ritualisierten Interaktionen, die der Herstellung und Aufrechterhaltung des öffentlichen Selbst dienen“ (Hannover 1997: 10). Goffman vergleicht die soziale Interaktion dabei mit einem Theaterstück, in dem alle Beteiligten versuchen, ihr ehrliches Selbst darzustellen. Gleichzeitig ist diese Präsentation auf das Ziel abgestimmt, beim Publikum bestimmte Eindrücke zu erzeugen (Hannover 1997: 10f.). Goffmans Annahme über soziale Interaktionen findet sich noch heute in verschiedenen Theorien des Selbst wieder. Eine davon stellt das Impression-Management dar. Neben diesen soziologischen Ansätzen, gibt es auch ein sozialpsychologisches Erklärungskonzept, dass sich mit der Wirkung und Manipulation des Eindrucks von Personen auf ihre Umwelt beschäftigt.

Impression Management

Mummendey definiert das Impression Management in einem Satz: „Individuen kontrollieren (beeinflussen, steuern, manipulieren etc.) in sozialen Interaktionen den Eindruck, den sie auf andere Personen machen“ (Mummendey 1995:111). Im Mittelpunkt steht der strategische Einsatz von Sprache, Mimik, Gestik gegenüber der sozialen Umgebung. Doch warum verfälschen Personen ihre Signale? Nach Arkin (1980), der die Grundsteine der Impression-Management-Theorie legte, versuchen Menschen vor allem über negative Eigenschaften und Einstellungen

hinweg zu täuschen, um Ablehnung ihres Interaktionspartners zu vermeiden. Personen wählen diejenige Form der Selbstdarstellung, von der sie in der jeweiligen Situation die günstigsten Konsequenzen erwarten (Hannover 1997:12). Jellison und Arkin (1977) stellten fest, dass sich Menschen vor allem dann verstärkt in einem positiven Licht präsentieren, wenn sie über entsprechend direkte oder indirekte Signale eine Belohnung erhalten. Bei dieser Form der Selbstpräsentation steht die Wechselbeziehung zwischen Selbst- und Fremdbild im Vordergrund, d.h., „die Reaktionen der Interaktionspartner auf das jeweilige präsentierte Selbstbild beeinflussen das (weitere) Verhalten und die (weitere) Selbstpräsentation der betreffenden Person“ (Mummendey 1995:128).

Abbildung 7: Wechselbeziehungen zwischen Person und Publikum im Sinne der Impression Management Theorie



Quelle: Eigene Darstellung nach Mummendey (1995)

Im Rahmen der Impression Management Theorie gibt es verschiedene Taktiken, mithilfe derer die Personen versuchen, ihren Interaktionspartner zu beeinflussen. Mit der Taktik des Sich-Einschmeicheln stellte Jones (1964) die erste von einigen weiteren Strategien, die in der Literatur thematisiert werden, vor. Mittlerweile wird zwischen positiven Selbstdarstellungstechniken, wie der Eigenwerbung, die Signalisierung hoher Ansprüche, die Aufwertung über Kontakte oder das Herausstellen von Prestige, Macht, Offenheit, Attraktivität und negativen Selbstdarstellungstechniken in Form von Entschuldigungen, Understatement, Abwertungen und Bedrohungen unterschieden (Mummedey 1995: 141).

4.2.1.1 Selbstüberwachung

Das Konzept der Selbstüberwachung (SÜ) von Snyder (1974, 1987) stellt eine Taktik im Rahmen der Impression-Management-Theorie dar. „Unter Selbstüberwachung versteht man die Tendenz, das eigene Verhalten im Sinne sozialer Angemessenheit zu überwachen und bemüht zu sein, einen guten Eindruck von sich zu erzeugen“⁴⁶ (Schiefele 1990:1). Personen mit hoher

⁴⁶ Die Ähnlichkeit mit den gängigen Definitionen des SD-Bias ist bezogen auf die Selbstüberwachung sehr hoch. Der zentrale Unterschied zu den Täuschungsskalen, bei denen ebenfalls die Tendenz zur positiven Selbstdarstellung gemessen werden soll, besteht in Messung. Während bei den Täuschungsskalen die Tendenz indirekt über vom Forscher „unterstelltes Lügen“ gemessen wird, handelt

Selbstüberwachung sind aus Sorge um soziale Angemessenheit besonders empfindlich gegenüber dem Ausdruck und der Selbstdarstellung anderer Personen. Sie interpretieren die ihnen zur Verfügung stehenden Signale und versuchen dadurch zu erfahren, was in einer Situation angemessen ist, um dies als Leitlinien für die Überwachung und Steuerung der eigenen Selbstdarstellung zu verwenden (Snyder 1974: 528). Personen mit niedriger Selbstüberwachung sind hingegen weniger aufmerksam gegenüber den Signalen anderer. Sie haben eine geringere Motivation sich in einem positiven Licht darzustellen und streben nach einer Kongruenz zwischen ihrer Einstellung und ihrem Verhalten (Laux/Renner 2002; Schiefele 1990: 76; Ajzen et al. 1982). Damit zeichnet das Konstrukt zwei Extrempole der Persönlichkeit nach: Auf der einen Seite Personen mit hoher Selbstüberwachung, die dem typischen Selbstdarsteller entsprechen und auf der anderen Seite Personen mit geringer Selbstüberwachung, die sich durch ihr authentisches Verhalten auszeichnen (Laux/Renner 2002). Folgt man dieser Interpretation, stellt das Konstrukt der Selbstüberwachung eine Möglichkeit zur Messung der Authentizitätsneigung dar. Doch wo liegt der Unterschied zur Messung des Anerkennungsbedürfnisses, wenn eine hohe Selbstüberwachung mit einer starken Täuschungstendenz einhergeht?

Ein zentraler Unterschied zu dem Konzept des Anerkennungsbedürfnisses ist die Fähigkeit zur Selbststeuerung. Während Personen mit hohem Anerkennungsbedürfnis ebenso nach sozialer Anerkennung streben wie Personen mit hoher Selbstüberwachung, bedeutet dies nicht automatisch, dass sie auch die Fähigkeit besitzen die notwendigen Signale zu erkennen und gezielt umzusetzen. Diese Dimension wird von dem Konstrukt der Selbstüberwachung zusätzlich erfasst. Ein zweiter wichtiger Unterschied besteht in der Situationsabhängigkeit der Skalen. Während die Beantwortung der Items der Täuschungsskalen von der Normwahrnehmung und Anonymität der Situation abhängig ist, erfasst die Selbstüberwachung ein stabiles Persönlichkeitskonstrukt. Um diesen Unterschied zu verdeutlichen, soll zunächst die Skala vorgestellt werden.

Die Skala zur Selbstüberwachung wurde erstmals 1974 von Snyder vorgestellt. Seine ursprünglichen Items bezogen sich dabei auf fünf Aspekte:

1. Die soziale Angemessenheit der Selbstpräsentation,
2. das Heranziehen von sozialen Hinweisen um sich daran zu orientieren (POSÜ),
3. die Fähigkeit, die Selbstpräsentation zu kontrollieren und zu modifizieren (FZSÜ),
4. die Anwendung dieser Fähigkeit in bestimmten Situationen

es sich bei der Selbstüberwachung um eine Selbsteinschätzung der Respondenten.

5. und das Ausmaß, in dem die Selbstdarstellung in verschiedenen Situationen eingesetzt wird (SKSÜ).

Die 41 Items dieser fünf Aspekte wurden 192 Studierenden zur Beantwortung in Form einer Selbstbeschreibung vorgelegt. Ungefähr die Hälfte der Items stand bei Zustimmung die andere Hälfte hingegen bei Ablehnung für eine hohe Selbstüberwachung. In einer anschließenden Analyse wurden die 25 Items mit der größten Trennschärfe für eine Folgestudie ausgewählt. Dieser zweite Schritt diente der Überprüfung der Validität der Skala. Hierbei wurden 16 Mitglieder einer Studentengruppe an der Universität zu ihrem eigenen Verhalten befragt. Zusätzlich wurden sie gebeten, auch sechs ihrer Kommilitonen anhand einiger zentraler Selbstüberwachungsitems einzuschätzen. Es zeigte sich, dass Personen mit hoher Selbstüberwachung auch mehr Selbstüberwachungsmerkmale zugeschrieben wurden und umgekehrt. In einer weiteren Studie überprüft Snyder die Validität seiner Skala mittels professioneller Schauspieler, die eine Gruppe mit besonders hohen SÜ-Werten darstellen sollte. Niedrigere Werte, vermutete Snyder, sollten hingegen die Patienten einer psychiatrischen Anstalt aufweisen. In beiden Gruppen zeigten sich, im Vergleich zu der Studierendenstichprobe, die erwarteten signifikanten Unterschiede. In einem dritten Schritt überprüfte Snyder, ob Personen mit hohen SÜ-Werten tatsächlich besser in der Lage sind ihren Ausdruck und ihre Emotionen zu steuern. 53 Studierenden wurden gebeten am Telefon Sätze, die ein bestimmtes Gefühl vermitteln sollten, vorzulesen. Anschließend hörte sich eine Jury die Aufnahmen an und versuchte das Gefühl, das über die Stimme transportiert wurde, zu ermitteln. Der Grad der Übereinstimmung diente als Maß der Selbstkontrolle, die eine Dimension der Selbstüberwachung darstellt. Es zeigte sich, dass Personen mit hohen SÜ Werten besser dazu in der Lage waren, fremde Emotionen darzustellen. Ein weiteres Experiment ließ außerdem darauf schließen, dass diese Studierenden sich bei der Beantwortung von Persönlichkeitsfragen eher an dem Durchschnitt ihrer Kommilitonen orientierten. In verschiedenen Studien wurde die Selbstüberwachungsskala mit Skalen zur sozialen Erwünschtheit korreliert. Der Zusammenhang zwischen beiden Skalen war jedoch entweder sehr schwach (Snyder 1974: 534f.) oder gar nicht vorhanden (siehe Schiefele 1990: 81).

Bewertung

Die von Snyder (1974) a priori formulierten fünf Aspekte der Selbstüberwachung wurden in einer Reihe von Faktorenanalysen überprüft. Die Analysen konnten die Vermutung der fünf Faktoren nicht bestätigen. Eine Übersicht über die verschiedenen Studien findet sich bei Schiefele (1990: 81ff.), der die Skala auch erstmals ins Deutsche übersetzte. Er schlussfolgerte aus seiner Analyse, dass die SÜ-Skala kein eindimensionales Persönlichkeitskonstrukt abbildet, son-

dern dass ihr wahrscheinlich zwei Dimensionen zugrunde liegen. Die erste Dimension erfasst die Fähigkeit zur Selbstkontrolle, die zweite Dimension bezieht sich hingegen auf die tatsächliche Anwendung der Taktik (Personenorientierung und schauspielerische Fähigkeit) (Schiefele 1990: 91).

Wie sinnvoll ist nun die Verwendung dieser Skala zur Messung der Authentizität? Die Kombination von Motiv und Fähigkeit ist als Erweiterung zu den eher eindimensionalen Konstrukten der Täuschungsskalen zu sehen, gleichzeitig wird sie jedoch auch kritisch betrachtet. Laux und Renner (2002) bemängeln, dass bei Snyder bestimmte Persönlichkeitstypen innerhalb seiner Messung unberücksichtigt bleiben. So haben Personen mit hoher Selbstüberwachung eine schwache Motivation zur Authentizität und auf der anderen Seite eine starke Kompetenz zur Täuschung, während Personen mit niedriger Selbstüberwachung eine hohe Motivation und niedrige Kompetenz zur Selbstüberwachung aufweisen. Die Kombination hohe Motivation und ausgeprägte Kompetenz kommt hingegen in Synders Konstrukt nicht vor. Dabei ist es nach Laux und Renner (2002: 131) durchaus denkbar, dass Personen mit „hoher Selbstdarstellungskompetenz ihre besonderen Fähigkeiten nutzen könnten, um den Interaktionspartnern die eigenen Selbstbilder, Persönlichkeitseigenschaften, Einstellungen etc. möglichst genau zu vermitteln.“ Vor diesem Hintergrund der aktuellen Forschung erscheint es sinnvoll die Selbstüberwachung nicht als eindimensionales Konstrukt zu interpretieren. Vielmehr sollten die in den Analysen vorgefundenen Subskalen getrennt voneinander betrachtet werden. Schiefele (1990) konnte in seiner Analyse drei Faktoren identifizieren.

Zuvor wurde bereits angesprochen, dass es sich bei der Selbstüberwachung um ein Persönlichkeitskonstrukt handelt. Diese Auslegung spiegelt sich auch in den Items der Skala wieder. Anders als bei den Täuschungsskalen wird hier versucht, unabhängig von der Situation, Eigenschaften einer Person zu erfassen. Die Interpretation der Antworten beruht auf der Annahme, dass die Befragten dazu in der Lage sind, eine ehrliche und reflektierte Selbsteinschätzung abzugeben. Im Gegensatz dazu wird bei den Antworten der MC, SES-17 oder BIDR Skala den Befragten eine situationsabhängige Täuschung unterstellt. Während das Bedürfnis nach sozialer Anerkennung von Esser als situationsbezogenes Ziel definiert wird (Esser 1986a: 18), sind die Ausführungen in Bezug auf die Authentizität nicht so eindeutig. Das Anerkennungsbedürfnis wird zwar als Persönlichkeitskonstrukt beschrieben, gemessen wird das Konstrukt, wie in Kapitel 4.1.1 beschrieben, jedoch über eine situationsbezogene Skala. Es bleibt bei Esser und Stocké offen, ob die Authentizität der Theorie oder der vorgeschlagenen Operationalisierung der Autoren nach interpretiert werden soll. Geht man vom RC-Ansatz aus, so bestimmt die Definition

der Situation die vermuteten Handlungskonsequenzen und damit den Nutzen einer Option. Folgt man dieser Interpretation, müsste also auch das Bedürfnis nach Authentizität situationsbezogen gemessen werden. Im Sinne der hier vorgestellten Selbstüberwachungsskala stellt die Messung jedoch ein mehr oder weniger stabiles Persönlichkeitskonstrukt dar. Eine strenge Auslegung im Sinne des RC-Ansatz würde es hingegen notwendig machen, das Bedürfnis nach Authentizität abhängig von der Situation selbst zu erheben, in dem bspw. die Respondenten nach ihrem Verhalten während des Interviews befragt werden.

Ein Beispiel hierfür wäre folgendes Item aus der SÜ-Skala:

„Ich kann jemanden anlügen und ihm dabei mit ehrlicher Miene in die Augen sehen, wenn es einem guten Zweck dient.“

Auf die Situation bezogen, könnte das Item lauten:

„Ich kann den Interviewer anlügen und ihm dabei mit ehrlicher Miene in die Augen sehen, wenn es einem guten Zweck dient.“

Bisher wurde jedoch keine Skala zur Messung der situationalen Authentizität entwickelt. Aus diesem Grund ist es notwendig, auf die Messung des Erklärungsfaktors in Form eines Persönlichkeitskonstruktes zurückzugreifen. Ein weiterer wichtiger Punkt, der bei dem Einsatz der Skala berücksichtigt werden muss, ist die Interpretation der Antworten. Einzelne Items der Täuschungsskalen und der SÜ-Skala ähneln sich in ihrem Wortlaut stark (s.o.).

„Manchmal lüge ich, wenn ich muss.“ (BIDR-Skala)

„Ich kann jemanden anlügen und ihm dabei mir ehrlicher Miene in die Augen sehen, wenn es einem guten Zweck dient“ (SÜ-Skala)

Bei der BIDR-Skala wird eine Ablehnung der Aussage als Täuschung, bei der SÜ-Skala als ehrliche Antwort gewertet. Dieses Beispiel macht deutlich, dass viele der Items der SÜ-Skala gesellschaftlichen Normen unterliegen. Vor dem Hintergrund der sozialen Erwünschtheit ist es also durchaus denkbar, dass auch die SÜ-Skala wie eine Täuschungsskala zu interpretieren ist. Bisher wurden diese verschiedenen Auslegungen der Skala jedoch nicht empirisch überprüft.

4.2.2 Einstellungsstärke

Ein zentraler Bestandteil des Handlungsziels Authentizität ist die Verankerung der Einstellung innerhalb der Persönlichkeit des Befragten. Esser beschreibt diesen Faktor als „Bedeutung und Intensität der mit der Frage angezielten Einstellung für den Befragten selbst“ (Esser 1986a: 17).

Während Esser das Bedürfnis nach Authentizität als Handlungsziel beschreibt⁴⁷, interpretiert Stocké den Faktor als Wahrscheinlichkeit und nicht als Nutzen. Seiner Auslegung nach steigt die Wahrscheinlichkeit die persönliche Identität auszudrücken mit der Zugänglichkeit der Information im Gedächtnis des Befragten an (Stocké 2001: 5). Es ist hierbei jedoch zu beachten, dass die Zugänglichkeit, wie von Fazio (1995) beschrieben, nur eine von verschiedenen Indikatoren für die Einstellungsstärke darstellt. Die Zugänglichkeit einer Meinung bestimmt darüber, wie schnell die relevanten Informationen bezüglich einer Frage abgerufen, verarbeitet und geäußert werden können. Je größer die Zugänglichkeit einer Einstellung ist, umso weniger anfällig ist eine Person gegenüber anderen Einflüssen und umso höher die Wahrscheinlichkeit, dass die geäußerte Einstellung mit der von der Befragungssituation unbeeinflussten Meinung übereinstimmt. Die Zugänglichkeit stellt einen Indikator der Einstellungsstärke. Es wird angenommen, dass sich die Zugänglichkeit objektiv über die Latenzzeit messen lässt. Eine andere Möglichkeit besteht darin, die Einstellungsstärke als subjektive Einschätzung zu erfassen. Diese Operationalisierungsvarianten sollen im Folgenden ebenfalls dargestellt werden.

4.2.2.1 „objektive“ Einstellungsstärke

Ein objektives Maß (im Sinne von nicht-reaktiv) für die Zugänglichkeit ist die Latenzzeitmessung. Die Latenzzeit beschreibt den Zeitraum zwischen der Präsentation der Frage und der „Initialisierung der Reaktion einer Person auf den Stimulus“ (Meyerl/Urban 2008: 8). Die Interpretation der Antwortreaktionszeit als Maß für die Einstellungsstärke wird zumeist aus der Forschung zur Einstellung-Verhaltens-Konsistenz abgeleitet. In diesem Ansatz von Ajzen und Fishbein (Fishbein/Ajzen 1975; Ajzen 1991) wird zwischen geplanten und spontanen Verhalten unterschieden. Ein zentraler Bestandteil der Theorie ist dabei die Einstellungszugänglichkeit. Je zugänglicher eine Antwort ist, desto schneller erfolgt die Reaktion. Der Befragte muss nicht erst länger über seine Meinung bzw. über sein Verhalten nachdenken. Aus diesem Grund ist zu vermuten, dass bei hoher Zugänglichkeit auch seltener eine nachträgliche Editierung hin zu sozial erwünschteren Antworten stattfindet. Um diese Einstellungszugänglichkeit zu messen, gibt es zwei verschiedene Ansätze. Bei der meta-kognitiven Erfassung werden, wie im Kapitel zuvor beschrieben, die Befragten nach ihrer inhaltlichen Antwort hinsichtlich der Einschätzung ihrer Antwort in Bezug auf Sicherheit oder Reflektiertheit gebeten. Mit den operationalen Faktoren hingegen wird versucht, den Prozess der Antwortgenerierung direkt zu messen (Stocké 2003: 8). Urban und Mayerl (2007) erweiterten das klassische Modell des geplanten Verhaltens durch

47 Esser selbst nennt das Handlungsziel „persönliche Identität“, wie aber bereits dargestellt, beschreibt dieser Begriff kein konkretes Ziel eines Akteurs.

einen Modus, der sich durch die Informationsverarbeitung kennzeichnet. Als Maß für den Informationsverarbeitungsmodus wählten sie die Antwortlatenzzeit aus. Ihr Ansatz legt nahe, dass Personen mit kurzen Reaktionszeiten spontan reagieren und ihr Verhalten stärker durch die jeweilige Einstellung der Person sowie durch subjektiv wahrgenommene Normen bestimmt wird (Urban/Mayerl 2007: 695). Die Autoren halten jedoch fest, dass für letzteren Einflussfaktor auch gegenteilige Argumentationen zu finden sind. So argumentiert Terry et al. (2000), dass subjektive Normen überhaupt erst bei überlegten Antworten zum Tragen kommen. Empirisch konnten Urban/Mayerl (2007) keine signifikanten Unterschiede in der Wirkung von subjektiven Normen zwischen spontanen und überlegten Prozessen feststellen. Die Latenzzeit erwies sich damit zwar als Indikator für den Modus der Informationsverarbeitung im Sinne des Mode Modells von Fazio (1990), es gab aber keinen Hinweis auf die Wirkung der sozialen Norm.

Zwei grundlegende Transformationen dieser Reaktionszeit werden in der Literatur vorgeschlagen. Die inhaltliche Transformation orientiert sich an dem standardisierten Mittelwert jeder einzelnen Person (Stocké 2003: 11). Diese Berechnung trägt der Tatsache Rechnung, dass jeder Befragte eine individuelle Geschwindigkeit im Lesen und Verstehen hat. Um eine geeignete Vergleichsmessung zu erhalten, sind für die Berechnung dieser Basiszeit, Standardisierungsfragen notwendig. Die mittlere Antwortgeschwindigkeit wird dann von der Reaktionszeit der sozial sensiblen Frage abgezogen. Diese Differenz soll die Informationsverfügbarkeit widerspiegeln und gleichzeitig die Fehlervarianz auf Grundlage der itemspezifischen Unterschiede eliminieren (Fazio 1990). Eine andere Möglichkeit ist die mathematische Transformation der Reaktionszeiten. Mit Hilfe von Logarithmierung oder Wurzeltransformationen wird versucht, die schiefe Verteilung zu bereinigen (Stocké 2003: 11). In einem Vergleich der verschiedenen Transformationsmöglichkeiten konnte Stocké (2003) keinen signifikanten Unterschied zwischen den mathematischen Berechnungen feststellen. Nur die Differenzen zum standardisierten Mittelwert zeigte eine Verbesserung der Validität der Messung. Mayerl et al. (2005) verglichen zusätzlich verschiedene Berechnungsmethoden der Basisreaktionszeit. Es zeigte sich, dass der Residual-Score-Index⁴⁸ die besten Ergebnisse lieferte.

48 "The logic [...] is based on procedures of regression analysis. If the raw reaction time is regressed on the baseline speed, the residuals will indicate the proportion of time that is independent of baseline speed. This we call "Residual-Score". [...] The Residual-Score is the difference between the empirically observed reaction time (RT_{raw}) and the statistically expected reaction time (RT_{exp}) which is estimated as a linear function of baseline speed (BS)." (Mayerl et al. 2005: 5)

4.2.2.2 Subjektive Einstellungsstärke

In den meisten Studien ist es üblich, die Einstellungsstärke über die direkte Abfrage der Einstellungssicherheit zu erfassen (Bassili 1996: 638). Der einfachste Weg ist, nach der Beantwortung des sensitiven Items, die Respondenten zu fragen, wie sicher sie sich bei der Beantwortung waren (Krosnick/Schuman 1988 bei Bassili 1996). Eine andere Möglichkeit besteht im Vergleich der Einstellungsstärke mit einem anderen Thema. Diese eindimensionalen Abfragen können von der tatsächlichen Einstellungsstärke jedoch weit entfernt sein. Yankelovich et al. (1981) entwickelten aus diesem Grund einen komplexeren Index. Dieser Index erfasst vier Dimensionen der Einstellungsstärke. Die Befragten müssen sich dazu äußern:

- ob sie von dem Thema persönlich betroffen sind;
- oft darüber diskutieren;
- ihre Meinung beibehalten würden, egal was passiert und
- ob sie für eine feste Meinung zusätzliche Informationen benötigen (Bassili 1996: 638).

Diese subjektive Erfassung der Einstellungsstärke wird jedoch kritisiert. Bassili (1996:638) merkt an, „that meta-attitudinal properties of strength are typically neither represented with the attitude object in memory nor readily derived from the cognitive process that underlie attitudinal judgements“. Miller und Peterson (2004) widersprechen dieser Annahme. Sie führen an, dass die psychologische Forschung immer wieder gezeigt hat, dass die individuelle und subjektive Bewertung, sogar wenn sie nicht akkurat ist, einen bedeutenden Erklärungsbeitrag leisten kann. Miller und Peterson (2004) sind es auch, die eine neue Operationalisierung der Messung vorschlagen. Insgesamt unterschieden sie in ihrer theoretischen Arbeit vier Dimensionen: die Bedeutung, Zugänglichkeit, Ambivalenz und die Sicherheit der Einstellung. Die Tauglichkeit dieser neuen Form der Operationalisierung bedarf jedoch noch einer empirischen Überprüfung.

4.2.3 Passgenauigkeit

Die Passgenauigkeit der Antwortoptionen mit der tatsächlichen Einstellung der Befragten stellt den letzten Erklärungsfaktor im Rahmen des Handlungsziels „Bedürfnis nach Authentizität“ dar. Bei dem Modell von Esser finden sich folgende Konkretisierungen zu diesem Erklärungsfaktor: Bei der Stimuluseindeutigkeit handelt es sich um eine subjektive Erwartung, bei der die Präzision und Verständlichkeit des Frageinhalts und der Antwortvorgaben im Vorder-

grund steht (Esser 1986a: 20; Esser 1991: 70). Stocké interpretiert den Faktor hingegen als Moderatorvariable für den Handlungsnutzen, der von der wahrgenommenen Passgenauigkeit zwischen Inhalt der Antwortoptionen und der Vorstellung des Befragten über den wahren Wert abhängt (Stocké 2001: 5).

Doch wie lässt sich diese Passgenauigkeit bzw. Stimuluseindeutigkeit in Rahmen von Befragungen messen? Weder bei Esser noch bei Stocké finden sich Vorschläge zur Operationalisierung dieses Konstruktes. Während allgemein die Möglichkeit besteht, den Respondente nach seiner persönlichen Einschätzung der Passgenauigkeit zu befragen, ist diese Variante auf heikle Fragen nicht übertragbar, denn das Problem bei dieser Gruppe von Fragen ist ja gerade die Tatsache, dass viele der Befragten nicht ehrlich antworten. Wenn also im Nachhinein erfragt wird, ob der Inhalt der Frage und die Antwortoption passgenau zur eigenen Einstellung waren, kann nicht unterschieden werden, ob der Respondent die Passgenauigkeit für die wahre oder die verzerrte Antwort angibt.

Auch eine experimentelle Operationalisierung ist schwierig. Generell ist anzunehmen, dass mit dem Grad der Differenzierung der Antwortskala die Passgenauigkeit der Optionen zunimmt, da die Befragten eine größere Auswahl an Möglichkeiten zur Verfügung haben. Eine dichotomisierte Skala (Ja und Nein Option) bietet weniger Differenzierungsmöglichkeiten als bspw. eine 5-Punkte Skala (stimme zu bis stimme nicht zu). Die Passgenauigkeit nimmt bei dieser Operationalisierung allerdings auch nur für solche Personen zu, die keine Extremmeinung besitzen. Bei allen Befragten, die eine klare und eindeutig Meinung für oder gegen die Frage haben, bleibt die Erweiterung der Skala hingegen ohne Wirkung.

Bewertung

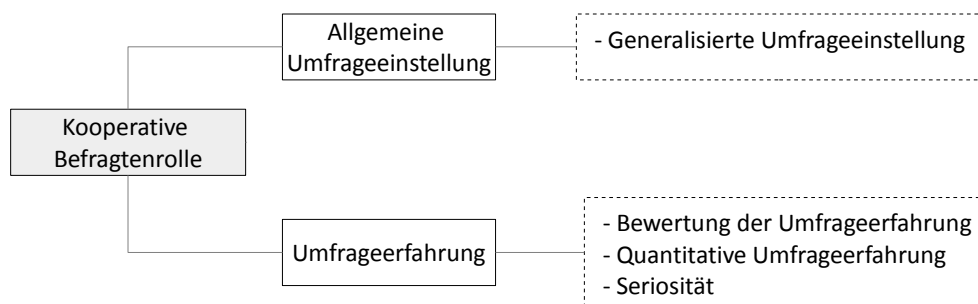
Zusammenfassend ist festzuhalten, dass eine sinnvolle Operationalisierung der Passgenauigkeit eines Items in Bezug auf heikle Fragen schwierig ist. Die Ausführungen von Esser und Stocké beschränken sich lediglich auf die theoretische Herleitung dieses Erklärungsfaktors, ohne dabei auf mögliche Messverfahren einzugehen. Vielmehr wird versucht die Erklärung eines Fehlers (SD-Bias) mit einem weiteren Erhebungsfehler (unpräzise Formulierung der Frage, unpassende Antwortoptionen) zu erfassen. Es ist deshalb fraglich, wie sinnvoll die Berücksichtigung dieses Erklärungsfaktors im Rahmen der Erklärungsmodelle überhaupt ist, wenn keine konkrete Möglichkeit zur Operationalisierung besteht. Dennoch stellt der Faktor im Rahmen der Theorie von Esser und Stocké ein wichtiges Konstrukt dar. Denn erst über die Stimuluseindeutigkeit wird der Nutzen der Handlungsziele mit einer subjektiven Wahrscheinlichkeit verbunden. Fällt die Passgenauigkeit weg, bleiben im Modell nur stabile Nutzenfaktoren (Bedürf-

nis nach Authentizität und Einstellungsstärke) bestehen. Der Nutzen wäre damit für jede Antwortoption gleich. Dieses Modell entspricht zwar dem Grundgedanken des RC-Ansatzes, die Anwendung ist jedoch nicht hilfreich, da der Nutzen für alle Antwortoptionen identisch ist. Vor diesem Hintergrund erscheint es notwendig, mögliche alternative Faktoren der subjektiven Erwartung aus der Theorie abzuleiten oder aber Operationalisierungsmöglichkeiten für die Passgenauigkeit zu entwickeln.

4.3 Erklärungsfaktoren des Frame-Selektion Modells

Im Rahmen des Ansatzes der Frame-Selektion wurde die kooperative Befragtenrolle als eine der Bedingungen formuliert, unter denen rationale Abwägungsgründe wie Anonymität und Anerkennungsbedürfnis außer Kraft gesetzt werden können. Stocké (2001, 2004) konkretisierte diesen Frame dahin gehend, dass sich der Grad der Kooperation eines Befragten aus seiner Einstellung zu Umfragen ableiten lässt. Je positiver die Umfrageeinstellung ist, desto weniger richtet sich der Befragte bei der Auswahl der Antwortoptionen nach den harten rationalen Faktoren, sondern tendiert eher dazu, die Umfrage durch Wiedergabe seiner „wahren“ Meinung zu unterstützen. In Abbildung 8 sind die Operationalisierungsvorschläge für die Erklärungsfaktoren des Framing-Modells dargestellt.

Abbildung 8: Erklärungsfaktoren und Operationalisierungen des Framing-Modells



4.3.1 Kooperative Befragtenrolle

Zur Messung der kooperativen Befragtenrolle schlägt Stocké (2001, 2004) die Umfrageeinstellung als Operationalisierung vor. Die Umfrageeinstellung bezeichnet die Bewertung vergangener Interviews sowie die generelle Einstellung zu Befragungen (Stocké/Langfeldt 2003:55). Verschiedene Studien zeigen, dass die Einstellung der Respondenten gegenüber der Umfrage einen erheblichen Einfluss auf die Datenqualität hat (Rogelberg et al. 2001; Erbslöh/Koch 1988; Stinchcombe et al. 1981). Stocké und Langfeldt (2003) unterscheiden zwischen zwei verschiedenen Dimensionen der Umfrageeinstellung: allgemeine Bestimmungsfaktoren und erfahrungsbasierte Determinanten.

4.3.1.1 Allgemeine Umfrageeinstellung

Die allgemeinen Bestimmungsfaktoren der Umfrageeinstellung unterliegen nach dem Rational-Choice Ansatz einer Kosten-Nutzen-Abwägung. Im Zentrum dieser Abwägung stehen meistens immaterielle Vor- und Nachteile der Befragung (Porst 1998: 3f.). Unterschiedliche Studien haben versucht die Kriterien, anhand derer die Umfragesituation bewertet wird, zu identifizieren. Eine der grundlegenden Arbeiten zu diesem Thema stammt von Groves, Singer und Corning (2000), die in ihrem leverage-salience Ansatz davon sprechen, „that a single survey design attribute will have different “leverages” on the cooperation decision for different persons“ (Groves et al. 2000: 306). Positive Aspekte können das Thema, die Seriosität, die gesellschaftliche Relevanz sowie das Publikationsinteresse sein. Der positive Nutzen von Umfragen kann also als Erstes über das Umfragethema bewertet werden. Personen, die zu dem Thema einer Umfrage eine persönliche Verbindung und somit auch ein gesteigertes Interesse haben, antworten nicht nur häufiger auf Fragen, ihre Antworten besitzen auch oft eine höhere Qualität als jene von Respondenten mit geringem Interesse (Goyder 1987; Groves et al. 2004).

Neben dem Thema spielt auch die wahrgenommene Seriosität der Umfrage eine entscheidende Rolle. Haben die Probanden das Gefühl, dass die Erhebung sinnvoll und professionell durchgeführt wird, antworten sie eher auf die Fragen. Empirische Hinweise auf diesen Zusammenhang finden sich bei Stocké und Becker (2004) sowie Stinchcombe et al. (1981). Die dritte Determinante für die positive Bewertung von Befragungen ist die gesellschaftliche Relevanz der Erhebung. Porst (1998) fragte in einer telefonischen Studie 288 Personen nach ihrem Wunschthema bei zukünftigen Befragungen. Rund 32% war das Thema egal, 19% hingegen wünschten sich Themen aus den Bereichen Politik, Wirtschaft und Gesellschaft. Als letzten Punkt ist das Publikationsinteresse zu nennen. Junge und gebildete Personen haben selbst ein größeres Interesse daran, Umfrageergebnisse in Zeitung und Fernsehen zu verfolgen (Stocké/Langfeldt 2003: 59). Es ist denkbar, dass solche Personen auch selbst eher dazu neigen, an Erhebungen teilzunehmen und die ihnen gestellten Fragen möglichst korrekt zu beantworten. Es lässt sich also festhalten, dass Respondenten, die Interesse an dem Befragungsthema haben und es für gesellschaftlich relevant einschätzen, das Erhebungsinstitut für seriös empfinden und selber an den Ergebnissen von Umfragen interessiert sind, sich in Befragungssituationen möglichst kooperativ verhalten. All diese positiven Anreize können dazu führen, dass trotz hohem Anerkennungsbedürfnis auch sensitive Fragen ehrlich und reflektiert beantwortet werden.

Es gibt jedoch immaterielle Nachteile von Umfragen, die diesen Effekt mindern können. Den positiven Aspekten von Erhebungen stehen die Kosten einer Teilnahme in Form von Zeitauf-

wand, Verletzung der Privatsphäre sowie der Datenschutz entgegen (für einen umfangreichen Überblick über die Belastungen im Rahmen einer Befragung, siehe Schnell (1997:166ff.)). Je länger eine Umfrage dauert, desto größer sind die Abbruchquoten (Goyder 1987: 46-48; Burchell/Marsh 1992: 235). Bei längeren Frageformulierungen lässt außerdem die Konzentration der Befragten nach, was zusätzlich zu Verzerrungen zulasten der Datenqualität führen kann. Die Angst vor Verletzung der Privatsphäre erhöht vor allem dann die Opportunitätskosten, wenn sensible Themenbereiche abgefragt werden. Auch hier ist davon auszugehen, dass mit steigender Sensibilität der gestellten Fragen auch die Kooperation der Befragten abnimmt. Das ist insbesondere dann der Fall, wenn ihre Einstellung oder ihr Verhalten nicht mit der sozial erwünschten Norm übereinstimmt. Der Datenschutz steht hingegen im engen Verhältnis zu der wahrgenommenen Professionalität des Erhebungsinstituts.

Insgesamt bewertet jeder Respondent die hier beschriebenen negativen und positiven Aspekte eines Interviews in Form einer Kosten-Nutzen-Abwägung. Aus dieser Abwägung ergibt sich eine generalisierte Umfrageeinstellung. Zahlreiche Studien haben Hinweise auf die positiven Effekte einer positiveren Umfrageeinstellung erbracht, dass z.B. Personen mit einer positiven generalisierten Umfrageeinstellung geringere Items-Nonresponse haben (Rogelberg et al. 2001), eher zu einer Teilnahme bereit sind (Stinchcombe et al. 1981), sowie eine höhere Wiederbefragungsbereitschaft besitzen (Schützenmeister 2002).

4.3.1.2 Umfrageerfahrung

Neben der allgemeinen Umfrageeinstellung spielen auch Determinanten eine Rolle, die auf den persönlichen Erfahrungen der Teilnehmer basieren. Es gibt Hinweise, dass die Erfahrungen die Personen während früherer Erhebungen gemacht haben, die generalisierte Einstellung gegenüber Umfragen prägen. Bradburn (1979) nennt vier zentrale Faktoren, die die individuelle erfahrungsbasierte Belastung beschreiben: die bisherigen Interviewlängen, der kognitive Aufwand, die Anzahl an durchgeführten Interviews und der psychische Stress. Wie bereits bei der generalisierten Umfrageeinstellung erwähnt, spielt die Dauer des Interviews eine entscheidende Rolle. Auch die Erfahrung mit bisherigen Befragungen und deren Dauer hat einen Effekt auf diesen Faktor. Das Unternehmen Forsa stellte bei einer Umfrage im Jahr 2000 fest, dass nur 14% der Befragten face-to-face Interviews bevorzugten. Die deutliche Präferenz von schriftlichen Interviews (40%) lässt sich unter anderem auf die Interviewdauer zurückführen (Stocké/Langfeldt 2003:61). In Bezug auf die Häufigkeit von Befragungen wird angenommen, dass negative Effekte in Form von „over-surveying“ auftreten können (Groves et al. 2004: 176).

Studien zu dem Zusammenhang mit der Teilnahmebereitschaft zeigten jedoch nur widersprüchliche Ergebnisse (Stocké/Langfeldt 2003: 61).

Der kognitive Aufwand der Respondenten in vorherigen Befragungssituationen wirkt, ebenso wie der wahrgenommene psychische Stress, auf die generalisierte Einstellung zu Umfragen. Eine umfangreiche Operationalisierung der hier vorgestellten Aspekte der Umfrageeinstellung findet sich bei Stocké und Langfeldt (2003) und sollen im Folgenden kurz vorgestellt werden:

- Generalisierte Umfrageeinstellung - Die Autoren entwickelten auf Grundlage verschiedener Studien (Dran/Hildreth 1995; Erbslöh/Koch 1988; Forsa 1993a, 1993b, 1996; Goyder 1986; Porst 1998 und Stinchcombe et al. 1981 alle bei Stocké/Langfeldt 2003) 16 Einstellungsfragen mit siebenstufiger Antwortskala.
- Bewertung der Umfrageerfahrung - Die Skala zur Bewertung der Erfahrung setzt sich aus sechs Items zusammen, die die Bereiche Belastung (zeitlicher und kognitiver Aufwand), Nutzen (Interesse und Spaß) sowie das erfahrene Frageverständnis (Unklarheiten und Sinn) erfassen.
- Quantitatives Ausmaß der bisherigen Umfrageerfahrung - Für die Erfassung der Häufigkeit der bereits durchgeführten Interviews, ist es möglich, die Befragten direkt nach der Anzahl der in den letzten 12 Monaten teilgenommenen Erhebungen zu fragen. Bei Stocké (2004) wird die bisherige Umfrageerfahrung auch als Maß für die Verankerung der Umfrageeinstellung interpretiert.
- Seriosität - Das Vertrauen in das Erhebungsinstitut kann auf zwei Wegen erfasst werden. Zum einen ist es möglich, die Befragten die Seriosität der bisherigen Institute einschätzen zu lassen. Zusätzlich kann das Vertrauen in das aktuelle Institut erhoben werden.

Bewertung

Zusammenfassend lässt sich festhalten, dass die vollständige Erfassung der Umfrageeinstellung ein komplexes Instrument benötigt. Die Unterscheidung von allgemeiner und erfahrungsbasierter Umfrageeinstellung, sowie die jeweiligen Unterkategorien in ihrer Gesamtheit zu erfassen, dürfte den Umfang eines normalen Fragebogen jedoch übersteigen, bedenkt man, dass die Umfrageeinstellung nur eine von vielen zu erhebenden Konstrukten im Rahmen des SD-Bias

ist. Deshalb ist es notwendig Kurzversionen der jeweiligen Skalen zu entwickeln, bzw. sich bei Erhebungen auf einzelne Aspekte der Umfrageeinstellung zu beschränken⁴⁹.

An diesem Punkt ist kritisch anzumerken, dass die Umfrageerfahrung, wie von Stocké (2004: 311) vorgeschlagen, allein kein Prädiktor für die Stärke der Einstellung sein kann. Viel Erfahrung in Bezug auf Umfragen bedeutet nicht automatisch, dass die Umfrageeinstellung besonders gefestigt ist. Denn gerade mit steigender Anzahl durchgeführter Interviews erhöht sich auch die Spannweite an positiven und negativen Erfahrungen. Eine Person, die nur an einem Interview teilgenommen hat und dieses als sehr belastend und negativ empfunden hat, könnte demnach eine viel stärkere Einstellung zu Umfragen haben als eine Person, die an vielen verschiedenen Befragungen mit positiven und negativen Erfahrungen teilgenommen hat.

49 In seiner Studie zur sozialen Erwünschtheit kontrollierte Stocké (2004) bspw. nur die generalisierte Einstellung sowie die Anzahl der bisherigen Interviews.

4.4 Zusammenfassung und Bewertung

Die in diesen Kapiteln vorgestellte Forschung zur Messung der Erklärungsfaktoren der RC-Modelle verdeutlichen, dass es bislang keine systematische Auseinandersetzung der Methodenforschung mit den Operationalisierungsvarianten gibt. Zum einen ist der Umgang mit den Begrifflichkeiten und den Operationalisierungen unpräzise. Gerade bei der Erwünschtheitswahrnehmung hat sich gezeigt, dass bei einem exakten Umgang mit der Operationalisierung verschiedene Ebenen des SD-Bias sichtbar werden, die bei der Analyse berücksichtigt werden müssen. Bei der Messung des Anerkennungsbedürfnisses in Form der Täuschungsskalen liegt die Gefahr vor allem in der Überinterpretation der Messung. Ausgehend von dem MC-Modell ist in vielen Forschungsarbeiten und aktuellen Surveys das Anerkennungsbedürfnis der einzige Erklärungsfaktor, der zur Kontrolle des SD-Bias eingesetzt wird. Entsprechend werden in diesem Zusammenhang die Skalen oft als SD-Skalen bezeichnet. Ohne Kenntnis der Richtung der Erwünschtheitswahrnehmung sowie wichtiger Rahmenbedingungen der Befragungssituation lassen sich die Analysen mit den Täuschungsskalen jedoch nur schwer interpretieren.

Aus der Darstellung des Forschungsstandes ergeben sich zum Teil erhebliche Lücken in der Prüfung der Täuschungsskalen. Der Großteil der Studien stammt aus den USA. Da aber die Analyse von Li und Reb (2009) Hinweise dafür liefert, dass die Skalen nicht kritiklos von einem Land auf das andere übertragbar sind, bleibt ihre Anwendbarkeit in Deutschland weiterhin fraglich, denn die Faktorenstruktur wurde für nur wenige Skalen in ihrer deutschen Version getestet. Die wenigen deutschen Analysen wurden zudem nur auf Grundlage explorativer Faktorenanalysen durchgeführt und nicht mit der strengeren konfirmatorischen Analyse überprüft. Besonders uneinheitlich sind die Ergebnisse für die MC-Skala, die anstatt einer Dimension in fast allen Studien eine mehrdimensionale Struktur bildet, dabei werden in allen Studien die Originalitems der Skala eingesetzt. Da die Entwicklung der Skala Jahrzehnte zurückliegt, ist es wenig verwunderlich, dass die Aktualität der Items infrage gestellt wurde (Stöber 1999). Aus diesem Grund überarbeitete Stöber (1999) die Fragen grundlegend. Diese neue deutsche Version der MC-Skala wurde bisher jedoch nicht ausreichend auf ihre Faktorenstruktur hin überprüft. Erhebungen, in denen die MC oder SES-17 Skala eingesetzt wurden, beschränken sich zumeist auf die Angabe von Crombachs Alpha, als Maß der internen Konsistenz (siehe bspw. Stocké 2004), das jedoch eine eindimensionale Skala voraussetzt (siehe obige Diskussion). Gerade diese Bedingung wurde in Bezug auf die MC-Skala jedoch häufig infrage gestellt. Die Überprüfung der Reliabilität mittels Crombachs Alpha ist damit nicht ausreichend. Auch bezogen auf die BIDR-Skala fehlt bisher eine Prüfung der zweifaktoriellen Struktur. Ziel der geplanten Analysen soll es

daher sein in einem ersten Schritt die Faktorenstruktur der wichtigsten Skalen zu überprüfen, um eine Aussage über die Reliabilität der Skalen treffen zu können.

Oft werden Ergebnisse verschiedener Täuschungsskalen miteinander verglichen und in Bezug gesetzt. Paulhus entwickelte, ausgehend von der Kritik der Mehrdimensionalität der MC-Skala, die BIDR-Skala. Aus seinen Erläuterungen lässt sich die Vermutung ableiten, dass die MC und damit auch die SES-17 Skala sowohl aus Items bezogen auf die Fremd- als auch auf die Selbsttäuschung bestehen. Zumindest für den deutschen Forschungsraum findet sich bisher keine Untersuchung, die diese Gemeinsamkeiten beider Skalen überprüft. Gleichzeitig wurde in Kapitel 4.2.1 eine Messung der Neigung zur Authentizität vorgeschlagen. Die SÜ-Skala erfasst, im Gegensatz zu den Täuschungsskalen, ein Persönlichkeitskonstrukt. Es dürfte daher keine Verbindung der SÜ-Skala zu den Täuschungsskalen geben, da sie unterschiedliche Dimensionen erfassen. Bisher gibt es aber keine Studie die alle hier vorgestellten Skalen (SES-17, FT, ST, SÜ) vergleichend überprüft. In der folgenden Analyse sollen deshalb im zweiten Schritt die gemeinsamen Dimensionen bzw. unterschiedlichen Strukturen der Skalen im Vergleich miteinander getestet werden.

Auch in Bezug auf die Messung des Anerkennungsbedürfnisses in Form von Kurzskalen finden sich Lücken. Besonders im deutschen Kontext werden in verschiedenen Bevölkerungsumfragen (Allbus, GMF-Grad, GMF-Pretest, Cella, SOEP) immer wieder Kurzskalen der MC, SES-17 und BIDR Skala eingesetzt. Oft ist dabei nicht ersichtlich, nach welchen Kriterien die Items für die Kurzversion ausgewählt wurden. Auch findet sich für kaum eine der hier eingesetzten Varianten eine umfangreiche Überprüfung der Validität und Reliabilität. Um den Nutzen des Einsatzes dieser Skalen, unabhängig von ihrer Erklärungskraft, bewerten zu können, sollen im dritten Teil dieser Analyse häufig verwendete Kurzversionen der Täuschungsskalen überprüft werden.

Eine weitere Forschungslücke betrifft die Vergleichbarkeit der Operationalisierungsvarianten. Bisher gibt es kaum Studien, die in ihren Analysen mehrere Messungen miteinander vergleichen und ihre Folgen diskutieren. Bisher ist nicht geklärt, ob die wahrgenommene Anonymität und die experimentelle Kontrolle des Modus tatsächlich ein und dasselbe Konstrukt messen und ob beide Operationalisierungen im Rahmen der Erklärung des SD-Bias funktional äquivalent sind.

Bezüglich der Erklärungsfaktoren des Handlungsziels Bedürfnis nach Authentizität fällt eine Konkretisierung der Operationalisierungsmöglichkeiten schwer. Die Ausführungen von Esser und Stocké beschränken sich hierbei auf kurze theoretische Ausführungen. Eine präzise Definition der Konstrukte oder konkrete Vorschläge für die Messung bleiben hingegen aus. Die in Ka-

pitel 4.2 diskutieren Operationalisierungsvarianten stellen deswegen nur unzureichende Versuche für eine Umsetzung im Rahmen von Befragungen dar. Vor allem wurden in Bezug auf die Konkretisierung der Faktoren Authentizitätsneigung und Passgenauigkeit Widersprüche und Ungenauigkeiten bei der praktischen Umsetzung der theoretischen Konstrukte deutlich, die in zukünftigen Arbeiten berücksichtigt werden müssen.

5 Empirische Analyse

5.1 Überblick über die Fragestellungen

Basierend auf den Forschungslücken, die in Kapitel 3 und 4 herausgearbeitet wurden, ergeben sich drei Schwerpunkte, die in der folgenden empirischen Analyse näher untersucht werden sollen:

1. Operationalisierung

Zunächst soll die Faktorenstruktur der wichtigsten Skalen überprüft werden, um eine Aussage über die Reliabilität der Skalen treffen zu können. Im zweiten Schritt werden die gemeinsamen Dimensionen bzw. unterschiedlichen Strukturen der Skalen im Vergleich miteinander getestet und im Anschluss häufig verwendete Kurzversionen der Täuschungsskalen überprüft werden. Da es bisher keine Erhebung gibt, bei der die relevanten Skalen unter denselben Rahmenbedingungen in ihrer vollständigen Form erhoben wurden, erfolgt die Prüfung auf der Basis einer Studierendenstichprobe, die an der Universität Göttingen erhoben wurde.

2. Erklärungskraft des MC-Modells

Für Deutschland ergibt sich die Forschungslücke, dass es zwar eine Reihe von Bevölkerungsumfragen gibt, die nach dem Prinzip von Marlow und Crowne nach Anerkennungsbedürfnis erheben, bisher jedoch kaum Veröffentlichungen zu diesen Daten zu finden sind. Aus diesen Gründen erscheint es notwendig, die bereits vorhanden Bevölkerungsdaten zu analysieren und die Erklärungskraft, unter Berücksichtigung der Kritik an der unsystematischen Forschung, des Modells von Marlow und Crowne zu überprüfen. Zentrales Kriterium für die Auswahl der Bevölkerungsumfragen war, dass sowohl Daten zu Täuschungsskalen als auch Vorurteilsitems erhoben wurden. Der Datenbestandskatalog der GESIS gibt bei der Stichwortsuche zum sozial erwünschten Antwortverhalten zwei Erhebungen an: die Cella Erhebungen sowie den Allbus 2006. Da in der Cella Befragung jedoch keine Vorurteilsitems enthalten sind, kommt dieser Datensatz nicht für die geplanten Analysen infrage. Neben dem Allbus als Datenquelle bietet auch das GMF-Projekt verschiedene Erhebungsdaten rund um das Thema Vorurteile. Zusätzlich zu menschenfeindlichen Syndromen wurden Täuschungsskalen jedoch nur in zwei der Erhebungen eingesetzt: Dem GMF-Graduiertensurvey sowie dem dazugehörigen Pretest. Im GMF-Pretest finden sich außerdem zusätzlich für einzelne Items zur Abfrage der Erwünschtheitswahrnehmung. Darüber hinaus bietet eine Onlineerhebung auf der Basis mehrerer Access Panels im

Auftrag der Universität Gießen die Möglichkeit das Modell an einer anonymen Webbefragung zu testen. Folgende Tabelle gibt eine Übersicht über die vorhandenen Surveys sowie der eingesetzten Täuschungsskalen und Vorurteilsitems.

Tabelle 4: Übersicht über die verwendeten Skalen⁵⁰ der einzelnen Surveys⁵¹

		Surveys			
		Allbus 2006	GMF-Graduierten-survey	GMF-Pretest	Web-sample
Täuschungs-skalen	MC-Skala	X			
	SES-17 Skala		X	X	X
	BIDR-Skala		X		X
Vorurteile	Antisemitismus	X	X	X	X
	Fremdenfeindlichkeit	X	X	X	
	Islamophobie		X	X	X
	Negative Einstellungen gegenüber Arbeitslosen		X	X	

In allen Surveys wurden nicht die vollständigen Skalen erhoben, sondern ausschließlich Kurzversionen, die teilweise nur drei Items umfassen. In keiner der vier Bevölkerungsumfragen wurden alle Items, die für die Überprüfung der RC-Modelle von Stocké und Skarbek-Kozietulska et al. notwendig sind.

3. Erklärungskraft der RC-Modelle im Vergleich

Auch die Modelle von Esser, Stocké, Skarbek-Kozietulska et al. und Wolter sind bisher nur unzureichend getestet. Da zudem zentrale Unterschiede in den Definitionen der sozialen Erwünschtheit bestehen, sind weder die Ergebnisse miteinander vergleichbar, noch lassen sich auf Grundlage der einzelnen Studien sichere Erkenntnisse über die Einflussfaktoren ableiten. Um die Modelle ausreichend zu testen, bedarf es einer größeren Anzahl an Analysen um die Reproduzierbarkeit der Ergebnisse zu überprüfen. Das Ziel der vorliegenden Arbeit ist es deshalb, an einem Datensatz die Modelle⁵² auf Einstellungsfragen am Beispiel der Vorurteile gegenüber Minderheiten anzuwenden, um die Stabilität sowie den Erklärungsbeitrag der Ansätze zu überprüfen. Der Schwerpunkt liegt hierbei jedoch auf den Arbeiten von Marlow-Crowne,

50 Die Liste der Items der verwendeten Skalen befindet sich in den Steckbriefen der jeweiligen Surveys im Anhang.

51 Die einzelnen Items der Skalen sind in Anhang nachzulesen.

52 Unter die beiden Modelle fällt auch ihre jeweilige Erweiterung über das Modell der Frame-Selektion, die hier ebenfalls überprüft werden.

Stocké und Skarbek-Kozietulska et al.. Das Modell von Wolter lässt sich auf Grund des experimentellen Designs in der empirischen Analyse nicht umsetzen und wird deshalb nicht überprüft. Da in den zuvor dargestellten Bevölkerungsumfragen die notwendigen Variablen nicht erhoben wurden, erfolgt die Prüfung der Erklärungsmodelle und ihr Vergleich ebenfalls an der Studierendenstichprobe der Universität Göttingen. Hier wurden sowohl die zentralen Täuschungsskalen als auch die Anonymität und die Erwünschtheitswahrnehmung für Antisemitismus, Fremdenfeindlichkeit, Islamophobie und negative Einstellungen gegenüber Langzeitarbeitslosen erfasst. Genauere Informationen zu den einzelnen Surveys mit Erhebungsdesign, Stichprobe und Items (Allbus 2006, GMF-Graduiertensurvey, GMF-Pretest, Websample und Studierendenstichprobe) sind aus Platzgründen im Anhang beigefügt.

Sensitive Einstellung als abhängige Variable: Vorurteile

Viele verschiedene Forschungszweige beschäftigen sich mit sensitiven Themen. Die Bandbreite reicht dabei von Gesundheitsfragen, über Vorurteile, bis hin zu Persönlichkeitstests im Rahmen von Einstellungsverfahren. Trotz ihrer Unterschiede haben alle Bereiche eine Gemeinsamkeit: Es wird angenommen, dass das sich nicht Befragten zu dem jeweiligen Thema nicht gerne äußern. Die unangenehme Situation entsteht dabei zum einen durch sehr intime Fragen (bspw. nach dem Sexualleben) oder Fragen, bei denen bestimmte Antworten nicht selten negative Reaktionen hervorrufen (bspw. delinquente Verhaltensweisen). Oft lässt sich nur schlecht einschätzen, wie viele der gegebenen Antworten der Interviewsituation entsprechend angepasst wurden und nicht das tatsächliche Verhalten oder den Ersten, ohne Rücksicht auf die Befragungssituation, Antwortimpuls abbilden.

Ein sensibles Thema, das häufig über Einstellungsfragen erfasst wird, sind negative Aussagen zu Minderheiten oder diskriminierte Gruppen. Gerade Vorurteile haben durchaus gesellschaftliche Relevanz, denn um gesellschaftliche Missstände bspw. in Form von Diskriminierungen aufzudecken, sind Informationen über die Häufigkeit von Vorurteilen in Bezug auf bestimmte Gruppen hilfreich. Gleichzeitig gelten solche Vorurteile gesellschaftlich als nicht sozial erwünscht. Deshalb wird vermutet, dass sie im Rahmen von Erhebungen tendenziell eher verschwiegen werden.

Ein Zusammenhang zwischen Vorurteilen und Erklärungsfaktoren ist nur zu erwarten, wenn hierbei tatsächlich sensitive Themen angesprochen werden. Der Umgang der Gesellschaft mit schwachen Gruppen ist in den letzten Jahrzehnten mehr und mehr in den Fokus der Öffentlichkeit gerückt. Die Forderung nach Gleichstellung und Gleichbehandlung ist mittlerweile in unserer pluralistisch, demokratischen Verfassung fest verankert. Diese Entwicklungen wurde nicht

zuletzt auch auf europäischer Ebene vorangetrieben. Die europäische Richtlinie zur Gleichbehandlung ohne Unterschied der Rasse oder ethnischen Herkunft bildet dabei eine der wichtigsten gesetzlichen Grundlagen (Richtlinie 2000/43/EG). In der Wissenschaft wird sich seit Langem intensiv mit dem Thema auseinandergesetzt. Bereits in den 50ern publizierte Adorno zusammen mit anderen Autoren einen Sammelband, der sich mit den Ursachen für faschistische Einstellungen auseinandersetzt (Adorno et al. 1950). Das Projekt „Gruppenbezogene Menschenfeindlichkeit“ wurde z.B. initiiert, um das Ausmaß und die Ursachen von Vorurteilen gegenüber verschiedenen Gruppen in Deutschland zu erfassen. Eine Folge dieser Entwicklung ist, dass Vorurteile gegenüber Minderheiten in unserer Gesellschaft nicht mehr als erwünscht angesehen und akzeptiert werden. Das Thema Diskriminierung und Ungleichbehandlung wird in Bezug auf viele verschiedene Gruppen (Ausländer, Frauen, Behinderte ect.) diskutiert. Diese Auseinandersetzung rückt damit mehr und mehr in das Bewusstsein des einzelnen Bürgers. Eine öffentliche Äußerung von negativen Stereotypen kann negative Reaktionen in der Umgebung hervorrufen. Dadurch hat sich das Thema Vorurteile in den letzten Jahrzehnten mehr und mehr zu einem sensitiven Thema entwickelt.

Im Zentrum der anschließenden Betrachtungen sollen deshalb vorurteilsbezogene Items stehen. Schwachen Gruppen und/oder Minderheiten, die Diskriminierung erfahren, sind äußerst vielfältig. Sie reichen von der Ablehnung Homosexueller über Ausländer, Obdachlose, Sinti und Roma, Juden, Behinderter, bis hin zu der Diskriminierung von Frauen. Einige Studien haben gezeigt, dass all diese Vorurteile durchaus miteinander zusammenhängen. Diese sogenannte gruppenbezogene Menschenfeindlichkeit hat ihren Ursprung in der Ablehnung von out-groups (Zick et al. 2008:364).

Ein weiterer Aspekt, der für die Vorurteilsforschung spricht, ist die Tatsache, dass Stocké sein Modell zur Erklärung am Beispiel der Fremdenfeindlichkeit getestet hat. Skarbek-Kozietulska et al. wendeten den Ansatz auf berichtetes Verhalten an und kamen zu anderen Ergebnissen. Mit einer Messung, die möglichst nah an Stockés Vorgehen anknüpft, ließe sich zum einen überprüfen, ob sich die Differenzen auf die inhaltliche Dimension zurückführen lassen und zum anderen, wie stabil sich das Modell von Stocké auch auf Vorurteile gegenüber anderen Minderheiten übertragen lässt.

In der folgenden empirischen Analyse wird in Kapitel 5.1 sowohl eine konformatorische als auch explorative Faktorenanalyse anhand der Daten einer Studierendenstichprobe vorgenommen, bei der die einzelnen SD-Skalen überprüft werden. Ziel von Kapitel 5.2 ist die Überprüfung der Erklärungskraft des Marlow-Crowne Modells zur Erklärung von Vorurteilen an unter-

schiedlichen deutschen Bevölkerungsumfragen. Im letzten Schritt werden in Kapitel 5.3 die drei zentralen Modelle von Marlow-Crowne, Stocké, Skarbek-Kozietulska et al. sowie einer eigenen Modellvariante an einer Studierendenstichprobe verglichen, um die Stabilität und den Nutzen der einzelnen Modelle in Bezug auf verschiedene Vorurteilsgruppen zu bewerten.

5.2 Strukturelle Stabilität und Dimensionen der Täuschungsskalen

Das Erklärungsmodell von Marlow und Crowne stellt den am häufigsten angewendeten Ansatz zur Erfassung des sozial erwünschten Antwortverhaltens in der Surveyforschung dar. In aktuellen Erhebungen wie dem Allbus 2006, den GMF-Survey, den Cella-Umfragen sowie in einzelnen Erhebungen (Duck/Hunsberger 1999, Rattazzi et al. 2003, Berinsky/Lavine 2007, Aosved et al. 2009, Stocké 2004) werden Täuschungsskalen eingesetzt, die teilweise vor über 40 Jahren entwickelt wurden. Trotz der jahrzehntelangen Verwendung der Skalen ist ihre Brauchbarkeit durchaus umstritten. Diese Kritik bezieht sich vor allem auf Aspekte der Validität und Reliabilität: die Aktualität, Entwicklung und Faktorenstruktur der Skalen (wie oben erwähnt). Den Grundstein für die Täuschungsskalen legte in den 50er Jahren Edwards (1957). Als Reaktion auf seine Skala entwickelten Crowne und Marlow (1960) ein Gegenmodell zur Messung des SD-Bias (vgl. Kapitel 3.2.1). Ein neues, zweidimensionales Konstrukt wurde erst 20 Jahre später durch Paulhus (1984) vorgestellt. All diese Skalen versuchen die Tendenz zur sozial erwünschten Antwort zu messen, in dem sie Verhaltensweisen abfragen, die zwar sozial sehr erwünscht sind, aber selten vorkommen oder die sozial unerwünscht sind, aber sehr häufig im Alltag zu finden sind. In der psychologischen Forschungstradition entstanden andere Messungen, die sich enger an Persönlichkeitskonstrukten orientieren. Die Skalen zur Selbstüberwachung (Snyder 1974) bildet einen Bereich der Persönlichkeit ab, denen ein Zusammenhang mit dem Anerkennungsbedürfnis nachgesagt wird und gleichzeitig, wie in Kapitel 4.2.1 beschrieben, als Messung der Authentizitätsneigung verwendet werden kann. Ziel der folgenden Analyse ist es die Struktur dieser Skalen, ihre Gemeinsamkeiten und Unterschiede sowie die Kurzskalen zu bewerten. Da bisher in keiner Bevölkerungsumfrage die aufgezählten Skalen gleichzeitig erhoben wurden, wird für die empirische Analyse eine Studierendenstichprobe der Universität Göttingen herangezogen.

5.2.1 Hypothesen

Um die Sinnhaftigkeit des Einsatzes der Skalen besser bewerten zu können, ist es notwendig, die theoretische Faktorenstruktur zu überprüfen. Aus dem dargestellten Ansatz zur Entwicklung der einzelnen Skalen (Kapitel 3.2) ergeben sich folgende Hypothesen für ihre Struktur. In der Literatur wird behauptet, dass die MC-Skala eine eindimensionale Struktur aufweist. Diese Vermutung soll anhand der Daten überprüft werden (Hypothese 1). Paulhus (1984) entwickelte die BIDR-Skala anhand einer zweidimensionalen Struktur, deren Überprüfung Gegenstand von Hypothese 2 ist. Die Items zur Selbstüberwachung hingegen werden in die drei Faktoren Fähig-

keit zur Selbstdarstellung, Selbstkonsistenz und Personen-Orientierung untergliedert (Hypothese 3).

Aus den Theorien lassen sich zwar Hinweise auf Gemeinsamkeiten zwischen den Skalen ableiten, strenge Vorhersage sind jedoch nicht möglich. Es ist denkbar, dass besonders die SES-17 und die BIDR-Skala stark zusammenhängen. Paulhus (1984) argumentierte auf der Basis verschiedener Faktorenanalyse sowie dem Wortlaut der Items, dass die MC-Skala sich aus beiden Subdimensionen der BIDR zusammensetzt. Da in der Version von Stöber jedoch mehr Items vorhanden sind, die sich auf das Konzept der Fremdtäuschung übertragen lassen, ist davon auszugehen, dass die Items der beiden Skalen stärker auf gemeinsame Faktoren laden (Hypothese 4). Die SÜ Skala erfasst ein Persönlichkeitskonstrukt und sollte keine gemeinsamen Faktoren mit den Täuschungsskalen (SES-17 und Fremdtäuschung) bilden (Hypothese 5).

Bezogen auf die Kurzversionen der SES-17 und BIDR Skala, die im Rahmen der zuvor erwähnten Bevölkerungsumfragen eingesetzt wurden, gibt es kaum Informationen zu den Kriterien für die Auswahl der einzelnen Items. Für eine effektive und sinnvolle Kurzskaala müsste es sich bei den ausgewählten Items jedoch um die Fragen mit den stärksten Faktorenladungen handeln, da sich nach diesen Kriterien möglichst trennscharfe Items auswählen lassen. Wenn die Items korrekt ausgewählt wurden, dann finden in den Kurzskaalen die ladungsstärksten Items (Hypothese 6).

5.2.2 Operationalisierung

Datengrundlage ist eine Erhebung unter Studierenden der Universität Göttingen (SST), in dem mehrere Täuschungsskalen gleichzeitig erhoben wurden. Dies ermöglicht es, die Skalen ohne Verzerrungen durch unterschiedliche Kontexte auf ihre Struktur, Gemeinsamkeiten und Unterschiede zu untersuchen. Natürlich unterliegt die Erhebung selbst speziellen Rahmenbedingungen. Durch die anonyme Situation der Onlinebefragung ist es möglich, dass der Mechanismus hinter den Täuschungsskalen sich geringer ausprägt als bspw. in Face-to-Face Erhebungen. Der Vorteil besteht jedoch darin, dass besonders diese wichtige Bedingung der Anonymität der Situation wie auch die Fragereihenfolge und der inhaltliche Kontext des Fragebogens für alle Skalen stabil ist. Problematisch ist jedoch die Verallgemeinbarkeit der selektiven Stichprobe. Wie bereits erwähnt, ist es fraglich, welche Unterschiede es zwischen Gesellschaftsgruppen in Bezug auf die Beantwortung der Fragen gibt. In diesem Zusammenhang soll die Studierendenstichprobe jedoch für den Test der theoretischen Konstrukte der Skalen verwendet werden. Da wie bereits erwähnt, fast alle Skalen ausschließlich an Studierenden entwickelt

wurden, ist davon auszugehen, dass sich die theoretischen Annahmen gerade bei dieser Gruppe am ehesten bestätigen lassen müssten.

In der Studie wurden die zwei Täuschungsskalen SES-17 und BIDR sowie die SÜ-Skala auf einer 4-stufigen Antwortskala erhoben. Die beiden Täuschungsskalen liegen in ihrer vollständigen Version vor, während für die SÜ-Skala eine Kurzskaala eingesetzt wurde, die auf den Faktorenladungen der von Schiefele (1990) durchgeführten Studie beruht. Alle Items wurden in der Befragung randomisiert. Folgende Tabelle gibt eine Übersicht über die einzelnen Items:

Tabelle 5: Itemübersicht⁵³ der Täuschungsskalen in dem SST Survey

Item	Kürzel
SES-17 Skala	
Manchmal werfe ich Müll einfach in die Landschaft oder Straße.	SES1
Eigene Fehler gebe ich stets offen zu und ertrage gelassen etwaige negative Konsequenzen.	SES2
Im Straßenverkehr nehme ich stets Rücksicht auf die anderen Verkehrsteilnehmer.	SES3
Ich habe schon einmal Drogen (Tabletten, Haschisch oder Ähnliches) konsumiert.	SES4
Ich akzeptiere alle anderen Meinungen, auch wenn sie mit meiner eigenen nicht übereinstimmen.	SES5
Meine Wut oder schlechte Laune lasse ich hin und wieder an unschuldigen oder schwächeren Leuten aus.	SES6
Ich habe schon einmal jemanden ausgenutzt oder übers Ohr gehauen.	SES7
In einem Gespräch lasse ich den anderen stets ausreden und höre ihm aufmerksam zu.	SES8
Ich zögere niemals, jemandem in einer Notlage beizustehen.	SES9
Wenn ich etwas versprochen habe, halte ich es ohne Wenn und Aber.	SES10
Ich lästere gelegentlich über andere hinter deren Rücken.	SES11
Ich würde niemals auf Kosten der Allgemeinheit leben.	SES12
Ich bleibe immer freundlich und zuvorkommend anderen Leuten gegenüber, auch wenn ich selbst gestresst bin.	SES13
Im Streit bleibe ich stets sachlich und objektiv.	SES14
Ich habe schon einmal geliehene Sachen nicht zurückgegeben.	SES15
Ich ernähre mich stets gesund.	SES16
Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte.	SES17
BIDR-Skala Selbsttäuschung	
Der erste Eindruck, den ich von anderen Menschen gewinne, bewahrheitet sich meistens.	ST1
Ich bin immer mir selber gegenüber ganz ehrlich gewesen.	ST2
Ich weiß immer, warum ich etwas mag.	ST3
Es fällt mir schwer, einen beunruhigenden Gedanken beiseite zu drängen.	ST4
Manchmal verpasse ich etwas, weil ich mich einfach nicht schnell genug entscheiden kann.	ST5
Ich bin ein vollkommen rational denkender Mensch.	ST6
Ich kann Kritik selten vertragen.	ST7
Ich bin mir meiner Urteile sehr sicher.	ST8

⁵³ Alle Items konnten auf einer 4-stufigen Skala von „stimme zu“ bis „stimme nicht zu“ beantwortet werden.

An meinen Fähigkeiten als Liebhaber habe ich schon gelegentlich gezweifelt.	ST9
Ich weiß immer die Gründe für meine Handlungen.	ST10

BIDR-Skala Fremdtäuschung

Manchmal lüge ich, wenn ich muss.	FT1
Es ist schon einmal vorgekommen, dass ich jemanden ausgenutzt habe.	FT2
Ich fluche niemals.	FT3
Manchmal zahle ich es lieber anderen heim, als dass ich vergebe und vergesse.	FT4
Ich habe schon einmal zuviel Wechselgeld herausbekommen, ohne es der Verkäuferin zu sagen.	FT5
Ich gebe grundsätzlich alles an, was ich zu verzollen habe.	FT6
Manchmal fahre ich schneller, als es erlaubt ist.	FT7
Ich habe Dinge getan, von denen ich anderen nichts erzähle.	FT8
Ich nehme niemals Dinge an mich, die mir nicht gehören.	FT9
Ich bin schon einmal wegen einer angeblichen Krankheit nicht zur Arbeit oder Schule gegangen.	FT10

Selbstüberwachungsskala – Fähigkeit zur Selbstüberwachung

Es fällt mir schwer, das Verhalten anderer Leute zu imitieren.	FZSÜ1
Ich wäre vermutlich ein(e) guter Schauspieler/in.	FZSÜ2
In einer Gruppe von Leuten bin ich selten im Mittelpunkt der Aufmerksamkeit.	FZSÜ3
Wenn Spiele eine schauspielerische Leistung erfordern, bin ich nie besonders gut.	FZSÜ4

Selbstüberwachungsskala – Personen Orientierung

Ich versuche nie, anderen zu gefallen.	POSÜ1
Auch wenn mir eine Sache keinen Spaß macht, gebe ich oft vor, mich gut zu unterhalten.	POSÜ2
Ich würde meine Auffassungen nie ändern, um Zuneigung von anderen zu erhalten.	POSÜ3
Um voranzukommen und beliebt zu sein, neige ich dazu, so zu sein, wie man es von mir erwartet.	POSÜ4

Selbstüberwachungsskala – Selbstkontrolle

Mein Verhalten ist für gewöhnlich der Ausdruck meiner tatsächlichen Gefühle.	SKSÜ1
Gegenüber unterschiedlichen Menschen verhalte ich mich oft so, als wäre ich nicht dieselbe Person.	SKSÜ2
Ich bin nicht immer die Person, als die ich erscheine.	SKSÜ3
Ich kann jemanden anlügen und dabei völlig ehrlich wirken.	SKSÜ4

5.2.3 Methode

Die Analyse gliedert sich in vier Schritten. Zunächst werden die Mittelwerte der Skalen sowie die Standardabweichungen für die SST-Stichprobe dargestellt. Im zweiten Schritt erfolgt die Überprüfung der Skalenstruktur mittels der Faktorenanalyse. Da es für die einzelnen Faktoren der Skalen eindeutige theoretische Herleitungen und Begründungen gibt, wird die Struktur der Skalen mit konfirmatorischen Faktorenanalysen geprüft. Die Analyse wird mit dem Programm

Mplus vorgenommen, das im Vergleich zu SPSS vielfältigere Möglichkeiten bei der Spezifikation der Grundeinstellungen bietet. Zur Schätzung der freien Parameter wird der ML-Schätzer verwendet. Voraussetzung für den ML-Schätzer ist das Vorliegen von metrischen Skalen mit Normalverteilung. Die Items in der SST-Stichprobe konnten auf einer vierstufigen Likertskala von „trifft voll und ganz zu“ bis „trifft überhaupt nicht zu“ beantwortet werden. Dieses Verfahren zur Messung persönlicher Einstellungen entspricht jedoch nur einem ordinalen Skalenniveau. In der Praxis wird diese Form der Skala aber schon allein aus praktischen Gründen wie eine metrische Messung verwendet. Dies zeigt sich auch bei der Bildung der Indizes der einzelnen Täuschungsskalen. Üblicherweise werden hier die Mittelwerte aus allen Items gebildet und dann als neue Gesamtskala in der Analyse eingesetzt. Diese Berechnung setzt jedoch eine metrische Skalierung voraus. Der Unterschied zwischen ordinalen und metrischen Skalen liegt in den messbaren Abständen. Während bei der ordinalen Skala die Antworten zwar in eine Reihenfolge gebracht werden können, liegt im Gegensatz zur metrischen Skala keine Information über die Abstände zwischen den einzelnen Kategorien vor. Während die konservative Interpretation aus diesem Grund von einer Nichtanwendbarkeit der ordinalen Skala in linearen Modellen spricht, sehen Befürworter den Einsatz nicht eingeschränkt. Die Abstände lassen sich, obwohl keine eindeutigen Informationen vorliegen, entweder schätzen oder es wird von einem annähernd gleichen Abstand zwischen den Rangfolgen ausgegangen (Kühnel 1993a: 32). Mit einer Monte-Carlo-Simulation konnte Kühnel (1993a) zeigen, dass auch wenn das Skalenniveau ignoriert wird, korrekte und fehlspezifizierte Modelle unterschieden werden können. Seiner Meinung nach ist es wichtiger in jedem Einzelfall neu darüber zu entscheiden, ob die Skala theoretisch auch als metrische Messung interpretiert werden kann. Im Falle der Messung von Persönlichkeitseigenschaften kann davon ausgegangen werden, dass es sich bei der vierstufigen Likertskala um eine ungenaue Messung einer metrischen Größe handelt.

Im dritten Schritt der Analyse werden mit einer explorativen Faktorenanalyse die gemeinsamen Dimensionen der Skalen untersucht. Die Extraktion der Faktoren erfolgt über das ML-Schätzverfahren mit anschließender Promaxrotation. Faktoren unter einem Eigenwert von 1 werden ausgeschlossen. Anschließend werden die Vermutungen zu den Gemeinsamkeiten und Unterschieden der Skalen mittels einer Korrelationsmatrix überprüft. Auf Grundlage der in Schritt zwei durchgeführten konfirmatorischen Faktorenanalyse soll im vierten Schritt die Qualität der eingesetzten Kurzskalen bewertet werden. Die Kurzskalen der SES-17 und BIDR Skala, die in den erwähnten Bevölkerungsumfragen eingesetzt wurden, umfassen jeweils drei Items. In Bezug auf den vorliegenden Datensatz wird die Stärke und Eindeutigkeit der Faktorenladun-

gen dieser Items im Vergleich zu den restlichen Items der Skalen bewertet, um eine Aussage über die Qualität der Auswahl treffen zu können.

5.2.4 Ergebnisse

Die Mittelwerte der einzelnen Täuschungsskalen zeigen im Vergleich, dass der Mittelwert zur sozialen Erwünschtheit bei der SES-17 Skala am höchsten ist. Fremd- und Selbsttäuschung liegen deutlich unter dem Mittelwert von 2.92. Auffällig ist, dass die Tendenz zur Selbsttäuschung in dieser Stichprobe höher ist, als der Hang zur Fremdtäuschung.

Tabelle 6: Mittelwerte der SD-Skalen für das SST-Survey

	SD-Skalen*			
	SES-17 Skala	BIDR-Skala		Selbstüberwachungs-skala
		Fremd-täuschung	Selbst-täuschung	
Mittelwerte ⁵⁴	2,92	2,18	2,43	2,41
Standardabweichung	0,35	0,46	0,36	0,41
N	664	659	660	652

* Skala der SD Items 1=geringes Anerkennungsbedürfnis bis 4=hohes Anerkennungsbedürfnis

Die erste Analyse bezieht sich auf die Fragestellung, ob sich die folgenden theoretischen Strukturen der einzelnen Skalen bestätigen lassen:

- SES-17 Skala: Eine Dimension
- BIDR Skala: Zwei Dimensionen (Fremd- und Selbsttäuschung) sowie beide Subskalen als eigenständige Skalen
- SÜ Skala: Drei Dimensionen (Selbstdarstellung, Selbstkonsistenz und Personen-Orientierung).

Tabelle 7 gibt einen Überblick über die Ergebnisse der konfirmatorischen Faktorenanalyse. Für die Beurteilung des Modellfits wird der Godness-of-Fit-Test sowie der Root Mean Square Error of Approximation (RMSEA) herangezogen. Bei Letzterem gilt ein Wert kleiner als 0.05 als Minimalvoraussetzung für einen ausreichenden Fit.

⁵⁴ Die Mittelwerte wurden auf Grundlage einer Mittelwertsfunktion gebildet. Im Fall von fehlenden Werten wurde der Mittelwert über die noch vorhandenen Informationen berechnet.

Tabelle 7: Ergebnisse der konfirmatorischen Faktorenanalyse

SD-Skalen	Itemzahl	Anzahl der Faktoren	N	χ^2	df	p-value	RMSEA	prob<0.5
BIDR-Skala	20	2	251	299,678	169	0,000	0,056	0,184
Fremd- täuschung	10	1	322	60,702	35	0,005	0,048	0,548
Selbst- täuschung	10	1	405	69,574	35	0,001	0,049	0,499
SES-17-Skala	17	1	433	345,408	119	0,000	0,066	0,001
Selbstüberwachungs- skala	12	3	384	137,344	51	0,000	0,066	0,022

Der Goodness-of-Fit-Test ist bei allen Modellen signifikant⁵⁵. D.h., dass in keinem der Fälle die gewählte Faktorenstruktur an den Daten bestätigt werden kann⁵⁶. Bezogen auf den RMSEA zeigen nur die Subskalen der BIDR-Skala annehmbare Werte⁵⁷. Beide Indizes haben zwar einen signifikanten Goodness-of-Fit-Test, der RMSEA-Wert liegt aber in einem Bereich unter 0.05. Damit entsprechen die Subskalen der BIDR-Skala und hier vor allem die Fremdtäuschung am ehesten der theoretisch vermuteten Struktur. Dieses Ergebnis ist jedoch keinesfalls zufriedenstellend, da der RMSEA-Wert nur den Minimaloraussetzungen entspricht und der Goodness-of-Fit-Test überhaupt keinen angemessenen Modellfit zeigt. Die Werte sind damit weit von einem guten Modellfit entfernt. Die SES-17 Skala sollte, entsprechend der Theorie ausschließlich auf einen Faktor laden. Auch hier finden sich, wie bereits erwähnt, problematische Werte in Bezug auf den Goodness-of-fit-Test. Im Gegensatz zur BIDR-Skala liegt jedoch auch der RMSEA-Wert deutlich über 0.05. Etwas besser sind die Werte der Modellfits hingegen für die SÜ-Skala, aber auch hier ist der Goodness-of-fit-Test signifikant und der RMSEA-Wert liegt über 0.05. Für die SES-17 und die SÜ-Skala kann die dimensionale Struktur (Hypothese 1 und 3) damit nicht bestätigt werden. Welche Items verursachen den schlechten Modellfit der SES-17 Skala? In Tabelle 8 sind alle Items der Skala und ihre Faktorenladungen dargestellt.

55 Beim Goodness-of-Fit-Test lautet die Nullhypothese, dass das vermutete Modell zu den Daten passt. Bei einem p-value Wert unter 0,05 wird die Nullhypothese entsprechend abgelehnt, da das Modell signifikant von der Datenstruktur abweicht.

56 Die Ergebnisse basieren auf dem Maximum-Likelihood-Schätzer. Zur Überprüfung der Robustheit der Ergebnisse wurde ebenfalls eine Analyse auf Basis des MLR-Schätzers durchgeführt (Ergebnisse hier nicht dargestellt). Auch ohne die Verteilungsannahme konnten die Faktorenstrukturen nicht bestätigt werden. Die standardisierten Residuen der Kovarianzen zeigen sehr hohe positive bzw. niedrige negative Werte, die ebenfalls auf Fehlspezifikationen der einzelnen Modelle hinweisen.

57 Kleine RMSEA-Werte sprechen für einen guten Modellfit. Ein Wert unter 0.05 wird als Minimalvoraussetzung für einen guten Modellfit angesehen.

Tabelle 8: Standardisierte Schätzer nach der konfirmatorischen Faktorenanalyse für die SES-17 Skala

Item	Itemformulierung	Faktorladung
SES1	Manchmal werfe ich Müll einfach in die Landschaft oder auf die Straße.	-0,249
SES2	Eigene Fehler gebe ich stets offen zu und ertrage gelassen etwaige negative Konsequenzen.	0,231
SES3	Im Straßenverkehr nehme ich stets Rücksicht auf die anderen Verkehrsteilnehmer.	0,277
SES4	Ich habe schon einmal Drogen (Tabletten, Haschisch oder ähnliches) konsumiert.	-0,392
SES5	Ich akzeptiere alle anderen Meinungen, auch wenn sie mit meiner eigenen nicht übereinstimmen.	0,198
SES6	Meine Wut oder schlechte Laune lasse ich hin und wieder an unschuldigen oder schwächeren Leuten aus.	-0,356
SES7	Ich habe schon einmal jemanden ausgenutzt oder übers Ohr gehauen.	-0,435
SES8	In einem Gespräch lasse ich den anderen stets ausreden und höre ihm aufmerksam zu.	0,229
SES9	Ich zögere niemals, jemandem in einer Notlage beizustehen.	0,205
SES10	Wenn ich etwas versprochen habe, halte ich es ohne Wenn und Aber.	0,159
SES11	Ich lästere gelegentlich über andere hinter deren Rücken.	-0,369
SES12	Ich würde niemals auf Kosten der Allgemeinheit leben.	0,138
SES13	Ich bleibe immer freundlich und zuvorkommend anderen Leuten gegenüber, auch wenn ich selbst gestresst bin.	0,308
SES14	Im Streit bleibe ich stets sachlich und objektiv.	0,328
SES15	Ich habe schon einmal geliehene Sachen nicht zurückgegeben.	-0,334
SES16	Ich ernähre mich stets gesund.	0,188
SES17	Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte.	-0,307

Eine Systematik für sehr schwache Items lässt sich nicht feststellen. Vielmehr scheinen bestimmte Fragen zusätzliche Dimensionen zu erfassen. Das Ergebnis für die Skalen fällt deutlich negativer aus als erwartet. Abgesehen von der BIDR-Skala hält keine theoretische Struktur der Indizes einer empirischen Überprüfung stand. Sowohl die einfaktorielle Lösung im Sinne der SES-17 Skala, als auch die dreifaktorielle Struktur der SÜ-Skala können nicht bestätigt werden. Hierbei ist zu erwähnen, dass auch die Reihenfolge der Items keinen Einfluss auf ihre Reliabilität zu haben scheint. Innerhalb der Erhebung bekamen alle Teilnehmer per Zufall zwei verschiedene Versionen des Fragebogens zur Beantwortung vorgelegt. In der ersten Version wurden alle Skalen sortiert abgefragt. In der zweiten Version wurden die einzelnen Items über alle Skalen randomisiert. Die Ergebnisse der konfirmatorischen Faktorenanalyse getrennt nach die-

sen beiden Gruppen, die an dieser Stelle nicht näher dargestellt werden, zeigten keine systematischen Unterschiede zwischen den Fragebogenvarianten.

Die Zusammenhänge der Indizes untereinander werden über eine Korrelationsmatrix von den Skalen analysiert. Es wurde die Vermutung aufgestellt (Hypothese 4), dass besonders die SES-17 und die FT-Skala Gemeinsamkeiten aufweisen, während die SÜ-Skala mit keinem der beiden Täuschungsskalen stark zusammenhängt.

Tabelle 9: Korrelationsmatrix für die Mittelwerte der SD-Skalen

	FT ¹	SES-17	ST ²	FZSÜ ³	POSÜ ⁴	SKSÜ ⁵	SÜ ⁶
FT	1,000	0,515	0,178	-0,247	-0,217	-0,333	-0,384
(N)	(659)	(659)	(658)	(648)	(635)	(647)	(652)
SES-17		1,000	0,234	-0,218	-0,221	-0,219	-0,330
(N)		(664)	(660)	(648)	(635)	(647)	(652)
ST			1,000	0,126	-0,279	n.s.*	n.s.
(N)			(660)	(648)	(635)	(647)	(652)
FZSÜ				1,000	0,076**	0,238	0,766
(N)				(648)	(632)	(631)	(648)
POSÜ					1,000	0,165	0,589
(N)						(648)	(635)
SKSÜ						1,000	0,618
(N)						(647)	(647)
SÜ							1,000

1) BIDR Skala – Fremdtäuschung

2) BIDR Skala – Selbsttäuschung

3) Selbstüberwachungsskala – Fähigkeit zur Selbstüberwachung

4) Selbstüberwachungsskala – Personen Orientierung

5) Selbstüberwachungsskala – Selbstkontrolle

6) Selbstüberwachungsskala – Gesamt

* Irrtumswahrscheinlichkeit über 5%

** Sign. 0.056

Die Ergebnisse der Analyse legen eine Bestätigung dieser Vermutungen nahe. Die SES-17 Skala korreliert sowohl mit der FT als auch mit ST Skala signifikant. Der Zusammenhang mit der FT-Skala (0.515) ist jedoch wesentlich stärker als mit der ST-Skala (0.178). In Bezug auf die Beziehung zwischen den Täuschungsskalen und der SÜ-Skala zeigt die Korrelationsmatrix ein differenzierteres Bild. Sowohl SES-17 als auch FT Skala weisen einen signifikanten negativen Zusammenhang mit der Gesamt- sowie den Subskalen auf. D.h. Im Fall der FT-Skala, je größer die Tendenz einer Person zur Fremdtäuschung ist, umso geringer fällt ihre Personen-Orientierung, Selbstkonsistenz und Fähigkeit zur Selbstdarstellung aus. Der Zusammenhang mit der Selbsttäuschung ist nicht so eindeutig. Während auch hier die Personen-Orientierung einen negati-

ven Koeffizienten aufweist, besteht zur Fähigkeit zur Selbstdarstellung ein positiver Zusammenhang. Befragte, die unbewusst und zum Selbstschutz ihre Antworten verzerren, schätzen sich als eher schlechte Schauspieler ein, denen es schwerfällt, das Verhalten anderer Leute zu imitieren. Diese gegensätzlichen Effekte sind möglicherweise auch der Grund für den nicht-signifikanten Effekt der Hauptskala. Die Subskala Selbstkonsistenz hingegen zeigt keinen signifikanten Zusammenhang zur ST-Skala. Entgegen den theoretischen Erwartungen (Hypothese 5) weist die SÜ-Persönlichkeitsskala also durchaus Gemeinsamkeiten mit den Täuschungsskalen auf. Darüber hinaus bestätigt die Analyse die Bedeutung der Trennung von FT und ST. Die unterschiedlichen Effekte mit den SÜ-Subskalen verdeutlichen, dass beide Konstrukte tatsächlich unterschiedliche Dimensionen erfassen. Zusätzlich zur Korrelationsmatrix lassen sich die Zusammenhänge der Skalen auch mittels einer explorativen Faktorenanalyse untersuchen.

Tabelle 10: Faktorenstruktur der explorativen Faktorenanalyse auf Basis des ML-Schätzers und der Promaxrotation

Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	Faktor 8
ST1	FT1	FZSÜ1	ST6	FT3	SES4	POSÜ4	SES9
ST2	FT2	FZSÜ2	ST7	FT7		SKSÜ2	SKSÜ1
ST3	FT4	FZSÜ3	SES2	SES3		SKSÜ3	
ST5	SES7	FZSÜ4	SES6	SES7			
ST6	SES17		SES13	SES11			
ST8			SES14	SKSÜ4			
ST10							

Es zeigt sich, dass die drei Gesamtskalen (inklusive Subskalen sieben Skalen) auf acht verschiedene Faktoren laden. Faktorenladungen wurden ab einer Stärke von 0.4 als relevant eingestuft⁵⁸. Der größte Faktor wird ausschließlich durch die Items der Selbsttäuschung gebildet. Auch bei der konfirmatorischen Analyse zählte dieses Konstrukt mit der FT-Skala zu den stabilsten Messungen. Den zweiten Faktor bildet ein Teil der Fremdtäuschungsskala, sowie zwei Items aus der SES-17 Skala. Ebenfalls eindeutig ist der dritte Faktor der Fähigkeit zur Selbstdarstellung sowie Faktor vier, der hauptsächlich aus Items der SES-17 Skala besteht. Alle anderen relevanten Faktoren setzten sich aus verschiedenen Skalen zusammen. Wie sind die Ergebnisse nun in Bezug auf die theoretischen Vermutungen zu interpretieren? Es wurde vermutet, dass die SES-17 Skala stärker mit der FT als mit der ST-Skala zusammenhängt. Tatsächlich bilden Teile der FT und SES-17 Skala gemeinsame Faktoren (Faktor zwei und fünf). Es lassen sich ebenfalls Gemeinsamkeiten zwischen der SES-17 und der ST-Skala finden (Faktor vier), diese fallen jedoch deutlich schwächer aus. Es scheint, als würde sich alle drei Täuschungsskalen aus zusätzlichen

⁵⁸ Faktorladungen von 0.4 und größer wurden berücksichtigt, weil es sich im Rahmen der explorativen Analysen zeigte, dass damit die theoretische Skalenstruktur näherungsweise am Besten reproduziert werden konnte.

Subskalen zusammensetzen, deren Konstrukte sich teilweise überschneiden. Faktor zwei scheint sich demnach vor allem auf den fairen Umgang mit den Mitmenschen zu beziehen, der sowohl von Items der FT als auch der SES-17 Skala erfasst wird. Faktor vier hingegen beschreibt eher den Umgang und die Kontrolle der eigenen Gefühle und setzt sich wiederum aus Items der SES-17 und der ST Skala zusammen. Faktor fünf ist stattdessen weniger eindeutig einer Interpretation zuzuordnen. Die FT und ST Items laden auf keinen gemeinsamen Faktor. Dieses Ergebnis spricht für die Vermutung, dass die SES-17 Skala ebenso wie die MC-Skala kein eindimensionales Konstrukt darstellt, sondern sich aus Aspekten der Fremd- und Selbsttäuschung zusammensetzt.

Die theoretischen Überlegungen lassen vermuten, dass die SÜ-Skalen ein Persönlichkeitskonstrukt messen, während die SES-17 und BIDR-Skala eher eine Täuschungstendenz erfassen. Entsprechend liegt die Vermutung nahe, dass die Items dieser beiden Skalengruppen keine gemeinsamen Dimensionen abbilden (Hypothese 5). Bis auf einige wenige Ausnahmen trifft dies zu. Nur in Faktor fünf und acht finden sich Gemeinsamkeiten. Bei genauerer Betrachtung der Frageformulierungen zeigt sich ein differenziertes Bild. Die Gemeinsamkeiten des SÜ Items mit den SES-17 und FT Items in Faktor fünf sind nachvollziehbar. Der Item lautete „Ich kann jemanden anlügen und dabei völlig ehrlich wirken“. Diese Aussage stammt zwar aus der Persönlichkeitsmessung, erfasst jedoch eine sozial unerwünschte Verhaltensweise, die relativ häufig vorkommt. Damit ist die Formulierung nah an den Items der beiden Täuschungsskalen (z.B. „Manchmal lüge ich, wenn ich muss“ oder „Ich fluche niemals“). In Bezug auf Faktor acht sind die Gemeinsamkeiten hingegen schwieriger nachzuvollziehen. Dennoch ist die Trennung zwischen den Täuschungs- und den Persönlichkeitsskalen relativ eindeutig.

Als Fazit zu Hypothese 4 und 5, bezogen auf die Gemeinsamkeiten und Unterschiede der Skalen, lässt sich festhalten, dass weder die Vermutungen zu den Unterschieden noch zu den Gemeinsamkeiten der Skalen eindeutig bestätigt werden können. Dennoch lassen sich Hinweise darauf finden, dass die SES-17 Skala stärker mit dem Konstrukt der Fremdtäuschung zusammenhängt, als mit der Selbsttäuschung. Generell sind jedoch beide Dimensionen über die SES-17 Skala erfasst. Dieses Ergebnis spricht zumindest auch gegen die gängige eindimensionale Interpretation der MC-Skala, die die Grundlage für die SES-17 Skala bildet, wie sie bspw. bei Stocké (2004) sowie Skarbek-Kozietulska et al. (2012) zu finden ist. Die SÜ Skala zeigt tatsächlich weniger Gemeinsamkeiten mit den Täuschungsskalen auf, als diese untereinander zusammenhängen. Dennoch sind die Konstrukte nicht völlig voneinander getrennt und weisen zumindest auf Item-Ebene Ähnlichkeiten auf.

Hypothese 6 bezog sich auf die Reliabilität der Kurzskalen, welche im Rahmen deutscher Bevölkerungsumfragen häufig eingesetzt werden. Die Auswahl der Items für die BIDR-Kurzskala basiert auf der Analyse von Winkler et al. 2006. Nach welchen Kriterien die Items für die SES-17 Kurzskalen ausgewählt wurden, gibt es hingegen keine Information. Bei der Bildung einer Kurzskala ist es jedoch üblich die Items, die am stärksten auf den jeweiligen Faktor laden, auszuwählen, um eine größtmögliche Trennschärfe zu erreichen. Fraglich ist nun, ob die in den Surveys verwendeten Items, auch bei der SST-Stichprobe die stärksten Effekte produzieren. Für die Beurteilung der Faktorenladung wird die konfirmatorische Faktorenanalyse herangezogen. Sortiert man die Items nach der Stärke ihrer Ladung auf den Faktor, so wird deutlich, dass die Kurzskalen keinesfalls nur aus den Items mit den höchsten Ladungen bestehen. Nur bei der Fremdtäuschung ist ein Item der Kurzskala überhaupt in dieser Stichprobe auch das Item mit der stärksten Ladung. Alle anderen Items bewegen sich eher im Mittelfeld. In Tabelle 11 sind alle Items sortiert nach ihrer Faktorenladung dargestellt. Die markierten Items kennzeichnen die Auswahl für die Kurzskalen.

Tabelle 11: Standardisierte Schätzer nach der konfirmatorischen Faktorenanalyse für die SD-Skalen der SST⁵⁹

SES-17 Skala		
Item	Itemformulierung	Standardisierte Faktorladung
SES7	Ich habe schon einmal jemanden ausgenutzt oder übers Ohr gehauen.	-0,435
SES4	Ich habe schon einmal Drogen (Tabletten, Haschisch oder ähnliches) konsumiert.	-0,392
SES11	Ich lästere gelegentlich über andere hinter deren Rücken.	-0,369
SES6	Meine Wut oder schlechte Laune lasse ich hin und wieder an unschuldigen oder schwächeren Leuten aus.	-0,356
SES15	Ich habe schon einmal geliehene Sachen nicht zurückgegeben.	-0,334
SES14	Im Streit bleibe ich stets sachlich und objektiv.	0,328
SES13	Ich bleibe immer freundlich und zuvorkommend anderen Leuten gegenüber, auch wenn ich selbst gestreift bin.	0,308
SES17	Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte.	-0,307
SES3	Im Straßenverkehr nehme ich stets Rücksicht auf die anderen Verkehrsteilnehmer.	0,277
SES1	Manchmal werfe ich Müll einfach in die Landschaft oder auf die Straße	-0,249
SES2	Eigene Fehler gebe ich stets offen zu und ertrage gelassen etwaige negative Konsequenzen.	0,231
SES8	In einem Gespräch lasse ich den anderen stets ausreden und höre ihm aufmerksam zu.	0,229
SES9	Ich zögere niemals, jemandem in einer Notlage beizustehen.	0,205
SES16	Ich ernähre mich stets gesund.	0,188
SES12	Ich würde niemals auf Kosten der Allgemeinheit leben.	0,138
SES5	Ich akzeptiere alle anderen Meinungen, auch wenn sie mit meiner eigenen nicht übereinstimmen.	0,198
SES10	Wenn ich etwas versprochen habe, halte ich es ohne Wenn und Aber.	0,159

⁵⁹ Die fett markierten Items stellen die Auswahl der Kurzskalen dar.

FT-Skala

Item	Itemformulierung	Standardisierte Faktorladung
FT2	Es ist schon einmal vorgekommen, daß ich jemanden ausgenutzt habe.	0,574
FT8	Ich habe Dinge getan, von denen ich anderen nichts erzähle.	0,486
FT10	Ich bin schon einmal wegen einer angeblichen Krankheit nicht zur Arbeit oder Schule gegangen.	0,471
FT1	Manchmal lüge ich, wenn ich muß.	0,457
FT4	Manchmal zahle ich es lieber anderen heim, als daß ich vergebe und vergesse.	0,404
FT7	Manchmal fahre ich schneller, als es erlaubt ist.	0,395
FT5	Ich habe schon einmal zuviel Wechselgeld herausbekommen, ohne es der Verkäuferin zu sagen.	0,329
FT6	Ich gebe grundsätzlich alles an, was ich zu verzollen habe.	-0,307
FT3	Ich fluche niemals.	-0,264

ST-Skala

Item	Itemformulierung	Standardisierte Faktorladung
ST10	Ich weiß nicht immer die Gründe für meine Handlungen.	0,565
ST2	Ich bin nicht immer mir selber gegenüber ganz ehrlich gewesen.	0,433
ST6	Ich bin ein vollkommen rational denkender Mensch.	0,412
ST8	Ich bin mir meiner Urteile sehr sicher.	0,388
ST5	Manchmal verpasse ich etwas, weil ich mich einfach nicht schnell genug entscheiden kann.	-0,318
ST9	An meinen Fähigkeiten als Liebhaber habe ich schon gelegentlich gezweifelt.	-0,245
ST1	Der erste Eindruck, den ich von anderen Menschen gewinne, bewahrheitet sich meistens.	0,175
ST4	Es fällt mir schwer, einen beunruhigenden Gedanken beiseite zu drängen.	-0,142
ST7	Ich kann Kritik selten vertragen.	-0,078
ST3	Ich weiß immer, warum ich etwas mag.	0,551

Das in der ST Kurzskaala enthaltene Item „Der erste Eindruck, den ich von anderen Menschen gewinne, bewahrheitet sich meistens.“ besitzt einen Koeffizienten unter 0.2. Die drei Items der SES-17 Skala liegen zwischen einem Wert von 0.25 und 0.33. Betrachtet man nun zusätzlich die Ergebnisse der explorativen Faktorenanalyse, so wird deutlich, dass die Items der Kurzskaalen auch nicht alle auf gemeinsame Faktoren laden. Das Item FT1 und FT2 der Fremdtäuschungs-skala laden zwar gemeinsam auf Faktor 2, für FT5 ist hingegen überhaupt keine ausreichend starke Ladung für einen der Faktoren zu finden. Die drei Items der SES-17 Kurzskaala weisen ebenfalls für keinen der Faktoren eine relevante Ladung auf. Nur bei der Selbsttäuschungsskala laden alle drei Items der Kurzskaala auf einen gemeinsamen Faktor. Generell ist die empirische Auswahl von Kurzitems für allgemein einsetzbare Skalen fragwürdig, da die Ergebnisse immer

zeit- und populationsabhängig sind. Die Auswahl „schwacher“ Items für eine Kurzskala könnte in Zusammenhangsanalysen fehlende oder falsche Effekte produzieren. Natürlich bezieht sich die hier vorgenommene Analyse auf eine Studentenstichprobe, während die Kurzskalen bei Bevölkerungsumfrage eingesetzt werden. Die hier vorliegenden schlechten Ergebnisse für die Kurzskalen sind damit zwar nicht uneingeschränkt verallgemeinerbar, legen jedoch die Vermutung nahe, dass diese Versionen der Skalen nicht kritiklos einsetzbar sind und einer umfangreicheren Überprüfung bedürfen.

5.2.5 Zusammenfassung und Bewertung

Im ersten Teil der Analyse zeigte die konfirmatorische Faktorenanalyse, dass sich für keine der getesteten Skalen die theoretisch vermutete Struktur vollständig bestätigen lässt. Nur die FT und ST Skala wiesen für ein Maß des Modelfits zumindest ausreichende Werte auf. Dieses Ergebnis ist besonders vor dem Hintergrund des bisherigen Einsatzes der Skalen verheerend, denn im Gegensatz zur internationalen Forschung gibt es für Deutschland bisher noch kaum Untersuchungen, die sich mit der theoretischen Faktorenstruktur der Skalen beschäftigen. Gerade in Bezug auf die MC und SES-17 Skala wird die Eindimensionalität unreflektiert vorausgesetzt. Wenn sich jedoch zeigt, dass die Skalen keinesfalls nur eine Täuschungsdimension erfassen, ist auch die Rechtfertigung der Nutzung der Skala über Crombachs Alpha nicht praktikabel. Sinnvoll wäre es jedoch für die Reproduzierbarkeit der Ergebnisse die Analyse nochmals für eine breitere Stichprobe durchzuführen.

Der zweite Teil der Analyse beschäftigte sich mit den Gemeinsamkeiten der Skalen untereinander. Mittels einer explorativen Faktorenanalyse und einer Korrelationsmatrix sollten Gemeinsamkeiten und Unterschiede der Skalen und Subskalen analysiert werden. Die Ergebnisse machen deutlich, dass die Items der SES-17 und der FT-Skala tatsächlich ähnliche Inhalte erfassen. Bei beiden Skalen scheint primär die Fremdtäuschung in Form von Lügen bei unangenehmen und sozial unerwünschten Themen im Vordergrund zu stehen. Es finden sich jedoch auch Gemeinsamkeiten zwischen Items der SES-17 und ST Skala, was, wie schon bei der konfirmatorischen Faktorenanalyse, auf eine Mehrdimensionalität der SES-17 Skala hinweist.

Besonders fraglich erscheinen bei der Analyse die Gemeinsamkeiten der Täuschungsskalen mit der Selbstüberwachung. Teilweise laden Items von beiden Skalentypen auf gemeinsame Faktoren. Betrachtet man jedoch den Wortlaut der betreffenden Fragen, so ist das Ergebnis nicht sonderlich überraschend. Viel schwieriger ist dadurch die Interpretation. Einige der Fragen innerhalb der Persönlichkeitsskala erfassen Verhaltensweisen, die zwar sozial unerwünscht

sind, aber relativ häufig vorkommen. Diese Items funktionieren also nach einem ähnlichen Prinzip wie die Items der Täuschungsskalen. Ihre Interpretation ist jedoch genau gegenteilig. Während bei den Täuschungsskalen die Ablehnung eines solchen Items als sozial erwünschte Antwort gewertet wird, ist bei den Persönlichkeitsskalen dieselbe Reaktion als ehrliche Einschätzung der eigenen Charakterstärken zu interpretieren. Gerade die SÜ Skala enthält viele Fragen, bei denen eine Zustimmung als negative Eigenschaft ausgelegt wird und damit die Beantwortung der Items selbst einer gewissen sozialen Erwünschtheit unterliegt.

Auch die Auswahl der Items für die deutschen Kurzskalen ist fragwürdig. Nur wenige der Items zeigten in Übereinstimmungen mit den Daten der Studierendenstichprobe ausreichend starke Effekte. Besonders bei der SES-17 Skala setzte sich die Kurzskala aus schwachen bis sehr schwachen Items zusammen. Dieses Ergebnis hat für den praktischen Einsatz der Skalen ungünstig. Im Rahmen von Untersuchungen werden generell nur schwache Effekte der Täuschungsskalen vermutet, da mit ihrer Hilfe ausschließlich ein Fehlerterm gemessen werden soll. Wenn bei dieser Messung auch noch zusätzlich schwache Skalen eingesetzt werden, können Effekte völlig verschwinden. Gerade wenn sich Kurzskalen nur aus drei Fragen zusammensetzen, ist die Auswahl dieser Items von besonderer Bedeutung. Aus praktischen Gründen ist es sicherlich nur selten möglich, im Rahmen von Surveys die Gesamtskalen der einzelnen Konstrukte zu erheben. Dennoch sollten die verwendeten Kurzskalen vor ihrem Einsatz ausreichend getestet sein. Einzig für die BIDR-Skala gibt es für den deutschen Raum eine getestete Kurzskala (Winkler et al. 2006).

Über die Analyse hinaus gibt es weitere Kritikpunkte an den Skalen. Zum einen liegt die Entwicklung vieler Skalen bereits über 30 Jahre zurück. Es ist fraglich, ob die Items trotz gesellschaftlicher Veränderungen weiterhin verwendet werden können oder einer Überarbeitung bedürfen. Die Notwendigkeit einer Aktualisierung war bereits für Stöber (1999) Ausgangspunkt seiner Kritik. Er überarbeitete die MC-Skala umfassend und passte sie an aktuelle Vorstellungen über sozial erwünschtes Verhalten an. Dennoch wird teilweise auf die Originalskala zurückgegriffen (wie bspw. Stocké 2004). Andere Skalen, wie die BIDR und SÜ wurden hingegen seit ihrer Entwicklung nicht aktualisiert. Besonders in der BIDR-Skala finden sich Items, an deren Aktualität bezweifelt werden darf⁶⁰.

Ein weiteres Manko ist die Personengruppe, mit deren Hilfe ein Großteil der Skalen entwickelt wurde. Edwards legte Studierenden eine Reihe von Items vor, die diese nach dem Grad

60 Bsp. „An meinen Fähigkeiten als Liebhaber habe ich schon gelegentlich gezweifelt.“ oder „Ich gebe grundsätzlich alles an, was ich zu verzollen habe.“

der sozialen Erwünschtheit beurteilen sollten. Diese Auswahl wurde dann erneut von Studierenden beantwortet, um die trennschärfsten Items zu isolieren. Die Marlow-Crowne Skala wurde an den Mitarbeitern und Studierenden des Departments für Psychologie an der Ohio State Universität entwickelt. Ebenso die Selbstüberwachungs-Items. Die zweifaktorielle Struktur der BIDR testete Paulhus (1984) ausschließlich an Studierendenstichproben. Nur die überarbeitete Version der MC-Skala von Stöber (1999) wurde neben Studierenden auch an den Teilnehmern einer Berufsberatung und dem Freundeskreis von Forschungspraktikanten getestet.

Dies bedeutet, dass alle Skalen mehr oder weniger an Studierenden entwickelt wurden. Diese selektive Stichprobe von hochgebildeten, jungen Personen bildet bei den Täuschungsskalen die Grundlage für die Beurteilung, welche Verhaltensweisen als sozial erwünscht oder sozial unerwünscht gelten. Da diese Skalen nicht nur im studentischen Kontext, sondern auch in repräsentativen Bevölkerungsumfragen eingesetzt werden, ist es fraglich, ob die Normvorstellungen dieser Gruppe überhaupt auf die Allgemeinheit übertragbar sind. Dieser Kritikpunkt gilt besonders für die Täuschungsskalen. Voraussetzung für ihre funktionale Äquivalenz ist, dass die aufgeführten Verhaltensweisen für alle Befragten ein ähnlicher Grad an Sensibilität besitzen. Es ist jedoch durchaus vorstellbar, dass sich die Normvorstellungen zwischen bestimmten Gruppen in der Gesellschaft stark unterscheiden (siehe Li/Reb 2009). Die Ergebnisse der vorliegenden Analyse sowie die zwei inhaltlichen Kritikpunkte stellen den Nutzen der Skalen im Rahmen von Surveys infrage. Zumindest aber wurde deutlich, dass die Skalen in Erhebungen nicht unreflektiert eingesetzt werden sollten.

5.3 Erklärungsbeitrag des Marlow-Crowne Modells innerhalb von Bevölkerungsumfragen

Ziel dieser Analyse ist es, die Erklärungskraft des Modells von Marlow und Crowne für deutsche Bevölkerungsumfragen bezogen auf sensitive Einstellungsfragen zu überprüfen. Die Anwendung erfolgt anhand verschiedener Vorurteile gegenüber Minderheiten, die sich in ihrem Grad an Sensitivität unterscheiden, um die Stabilität der Ergebnisse zu ermitteln.

5.3.1 Hypothesen

In Bezug auf die theoretischen Annahmen lassen sich verschiedene Hypothesen für die erwartete Erklärungskraft der einzelnen Skalen formulieren. Das Modell von Marlow und Crowne geht davon aus, dass zwischen den Täuschungsskalen und den Vorurteilen jeweils ein negativer Zusammenhang besteht, d.h. mit steigendem Anerkennungsbedürfnis nimmt die Zustimmung zu den Vorurteilsitems ab (Hypothese 1). Unter Berücksichtigung der Kritik an ihrem Modell in Kapitel 3.2.1 sowie an der bisherigen Messung des Anerkennungsbedürfnisses (Kapitel 4.1.1) lautet die Gegenhypothese, dass es zwischen dem Anerkennungsbedürfnis und den Vorurteilsitems keinen systematischen Zusammenhang gibt.

Im Rahmen des Erklärungsmodells von Marlow und Crowne wird keine genauere Erläuterung des Anerkennungsbedürfnisses vorgenommen. Es ist deshalb nicht eindeutig, ob sich Marlow und Crowne in ihrer Theorie nur auf die Fremd- oder auch auf die Selbsttäuschung beziehen. In ihrer eigens entwickelten Skala werden nach Paulhus (1984) beide Aspekte miteinander vermischt. Diese Vermischung gilt auch für die aktualisierte und überarbeitete deutsche Version der Skala (SES-17). Es ist fraglich, ob der Einsatz dieser zwei Skalen vor dem Hintergrund der neusten Ergebnisse in Bezug auf die Zweidimensionalität noch sinnvoll ist. Über einen Vergleich der Effekte der Fremd- und Selbsttäuschung sollen eventuelle Unterschiede zwischen beiden Skalen verdeutlicht werden und so ermittelt werden, welche der beiden Skalen dem Grundgedanken des Modells von Marlow und Crowne deutlicher entspricht. Dabei ist zu beachten, dass die SES-17 Skala in allen Surveys nur in einer Kurzversion eingesetzt wurde. Die hierbei verwendeten Items entsprechen inhaltlich eher der Fremd- als der Selbsttäuschung. Es ist also davon auszugehen, dass die SES-17 und die Fremdtäuschungsskala inhaltlich ähnliche Aspekte erklären und deshalb in ihren Effekten deutlich näher beieinanderliegen als mit der Selbsttäuschungsskala (Hypothese 2).

Sowohl die Auswahl der Skalen als auch die der Vorurteilsgruppen orientiert sich an den zur Verfügung stehenden Datensätzen. Um das Modell an mehreren Gruppen testen und vergleichen zu können, wurden vier Einstellungen ausgewählt, die in allen Befragungen erhoben wurden: negativen Einstellungen gegenüber Juden, Muslimen, Ausländern und Langzeitarbeitslosen. Wie bereits erwähnt, gibt es verschiedene Ansätze, die von einer Differenzierung zwischen den einzelnen Vorurteilen ausgehen. Duckitt und Sibley (2007) unterscheiden zwischen als minderwertig, gefährlich und dissidente wahrgenommene Gruppen. Während sich Langzeitarbeitslose und Ausländer aufgrund ihres geringen Status der „minderwertigen Gruppe“ zuordnen lassen, ist die Unterscheidung in Bezug auf Juden und Muslime nicht eindeutig. Zum einen werden Muslime teilweise als Bedrohung für die Sicherheit wahrgenommen (bspw. in Form von Terrorismus), zum anderen weichen sie durch ihre Religion auch von gesellschaftlichen Normen und Werten ab. Antisemitismus taucht hingegen in den Gruppierungen überhaupt nicht auf. Die Unterteilung von Duckitt und Sibley lässt sich also nur schlecht auf die vorliegenden Vorurteilsgruppen übertragen.

Neben dieser Differenzierung besteht eine zweite Möglichkeit darin, die verschiedenen Vorurteile nach ihrem Grad an Tabuisierung zu unterscheiden. Die negative Einstellung gegenüber Juden ist in Deutschland schon allein aus historischen Gründen ein besonders sensibles Thema. „Antisemitismus beschreibt die Ablehnung und Diskriminierung von Menschen aufgrund dessen, dass sie Juden sind und ihnen damit bestimmte Eigenschaften zugeschrieben werden“ (Heyder et al. 2005: 145). Es werden zwei Formen des Antisemitismus unterschieden. Während der klassische Antisemitismus auf negativen Stereotypen basiert, stellt der sekundäre Antisemitismus eine Verharmlosung der nationalsozialistischen Verbrechen in den Vordergrund (Heyder et al. 2005: 147f.). Gerade diese zwei Varianten der Vorurteile gegenüber Juden werden in Deutschland tabuisiert. Die Tabuisierung in Westdeutschland erfolgte direkt nach Ende des Zweiten Weltkrieges im Rahmen der Entnazifizierung. Symbolcharakter hatte dabei sowohl das Verbot der sozialistischen Reichspartei 1952 als auch die Ergänzung des Strafgesetzbuches um die Leugnung des Holocaust als Tatbestand der Volksverhetzung 1994 (Leibold/Kühnel 2009: 135f.). In der DDR fand die Aufarbeitung der Nazi-Verbrechen hauptsächlich unter Gesichtspunkten der Kapitalismuskritik statt. Während auch hier der Völkerverhass als Verbrechen rechtlich verankert wurde, zeigten sich die Grenzen der Aufarbeitung aber vor allem in der Ablehnung der Rückerstattung jüdischer Vermögen (Leibold/Kühnel 2009: 134). Aktuelle Studien zeigen, dass der klassische Antisemitismus auch nach der Wiedervereinigung weiterhin zurückgegangen ist (siehe Leibold/Kühnel 2009). Fraglich ist jedoch, ob dieser Rückgang tatsächlich die Verhältnisse in der Bundesrepublik widerspiegelt, oder die Folge einer zunehmenden öffentli-

chen Tabuisierung ist. Einen Hinweis hierauf bietet die theoretische Arbeit von Bergmann und Erb (1991), die im Rahmen der GMF-Projekte von Heyder et al. (2005) bestätigt wurde. Da der Antisemitismus einem Kommunikationstabu unterliegt, werden die negativen Einstellungen über eine Umwegkommunikation in Form von israelbezogener Kritik geäußert (Bergmann/Erb 1991). Diese Befunde zeigen, dass Vorurteile gegenüber Juden in Deutschland auch heute noch ein äußerst sensibles Thema darstellen.

Negative Stereotypen gegenüber Muslimen in Deutschland, in Form von Islamophobie stellen ein weiteres Vorurteil dar. Islamophobie bezeichnet „generelle ablehnende Einstellungen gegenüber muslimischen Personen und allen Glaubensrichtungen, Symbolen und religiösen Praktiken des Islam“ (Leibold/Kühnel 2003). Diese Ablehnung sollte dabei nicht nur im Kontext der allgemeinen Fremdenfeindlichkeit betrachtet werden. Der Anteil der Muslime in Deutschland ist mit der Einwanderungswelle der sechziger und siebziger Jahre deutlich gestiegen. Nach den katholischen und evangelischen Christen stellen die Muslime heute die drittgrößte Religionsgemeinschaft dar (Seidel 2003: 261). Obwohl die freie Religionsausübung durch das Grundgesetz geschützt wird, fordern Teile der Bevölkerung gerade für die Muslime eine Einschränkung dieser Rechte. Die Ursachen dieser großen Ablehnung des Islam reichen dabei von historischen Gründen (siehe Seidel 2003), über aktuelle Ereignisse wie den 11. September bis hin zu der größeren kulturellen Distanz. Obwohl die islamkritischen Diskussionen in den letzten Jahren zugenommen haben, zeigen aktuelle Studien im Vergleich mit anderen Minderheiten, dass kein deutliches Feindbild zu erkennen ist (Leibold/Kühnel 2003). Im Gegensatz zum Antisemitismus lässt sich daher vermuten, dass die Äußerung von Kritik gegenüber dem Islam keiner solchen strengen Tabuisierung unterliegt.

Neben diesen zwei religiösen Minderheiten spielen auch die Vorurteile gegenüber Ausländern in Deutschland eine große Rolle. Die Gewalttaten gegen Fremde hatten ihren dramatischen Höhepunkt in den 90er Jahren. Die Vorfälle von rechtsradikaler Propaganda und Volksverhetzung sind bis heute ansteigend (Babka von Gostomski et al. 2007: 102f.). Im Gegensatz zum Antisemitismus und der Islamophobie orientieren sich die negativen Stereotypen jedoch nicht an einem religiösen Merkmal, sondern vielmehr an der Staatszugehörigkeit und dem Herkunftsland.

Die Diskriminierung von Langzeitarbeitslosen ist im Vergleich zu den bisherigen Gruppen ein relativ neues Thema. Mit Einführung der Hartz IV Gesetz hat sich die soziale Lage ökonomisch schwacher Gruppen in Deutschland deutlich verschärft. Die Ökonomisierung des Sozialen führt zunehmend zu einer Desintegration von Personen, die nicht mehr in den Arbeitsmarkt einge-

bunden sind. Diskriminierende Äußerungen werden somit über ökonomische Prinzipien legitimiert (Heitmeyer/Endrikat 2008). Eine Gruppe, die von dieser Abwertungslogik besonders betroffen ist, stellen die Langzeitarbeitslosen dar. Diese Vorurteile werden mittlerweile von einem erheblichen Anteil (knapp 60%⁶¹) der Bevölkerung getragen (Heitmeyer/Endrikat 2008: 65ff.). Es muss davon ausgegangen werden, dass anders als bei den bisherigen Minderheiten die Diskriminierung von Langzeitarbeitslosen keiner öffentlichen Tabuisierung unterliegt. Bestimmte Interessengruppen fördern das negative Bild des Arbeitslosen sogar zusätzlich.

In Bezug auf die verschiedenen Vorurteile ergeben sich also, abgeleitet aus der unterschiedlichen Tabuisierung, folgende zu erwartende Unterschiede: Die öffentliche Tabuisierung des Antisemitismus in Deutschland ist aufgrund der besonderen historischen Verhältnisse wesentlich stärker verankert, als die Vorurteile gegenüber Langzeitarbeitslosen. Welchen Tabus die negativen Einstellungen gegenüber Muslimen und Ausländern unterliegen, hängt hingegen stark vom sozialen Kontext ab. Vorurteile gegenüber Ausländern und Muslimen werden jedoch häufiger geäußert als gegenüber Juden (Hypothese 3) (Heitmeyer 2002). Vor diesem Hintergrund lässt sich folgende dritte Hypothese ableiten: Der Zusammenhang zwischen dem Anerkennungsbedürfnis und den geäußerten Vorurteilen ist bei Antisemitismussmessungen am stärksten ausgeprägt. Am schwächsten ist der Effekt in Bezug auf Vorurteile gegenüber Langzeitarbeitslosen. Der Zusammenhang, zwischen den negativen Einstellungen gegenüber Muslimen und Ausländern rangiert zwischen diesen beiden Extrempunkten.

5.3.2 Operationalisierung

Es wurde bereits erwähnt, dass sich die Auswahl der Skalen und sensitiven Items an den zur Verfügung stehenden Surveys orientiert. Als Datengrundlage dienen die vorgestellten, deutschen Bevölkerungsumfragen. Aus Gründen der Lesbarkeit finden sich nähere Informationen zu den einzelnen Surveys (Fallzahl, Grundgesamtheit, Erhebungsmodi, Erhebungsjahr usw.) im Anhang.

Alle vier Surveys variieren in ihrem Grad an Anonymität, der Fragereifolge sowie der Grundgesamtheit. Es kann daher nur aufgezeigt werden, welchen Erklärungsbeitrag die Täuschungsskalen innerhalb der einzelnen Surveys leisten. Ein Vergleich der Effekte der Surveys untereinander ist hingegen nur schwer möglich. Da bei der theoretischen Entwicklung der Skalen keine Einschränkungen bezüglich der Anwendung und Wirksamkeit gemacht wurden, ist jedoch davon auszugehen, dass die vermuteten Zusammenhänge, auch unabhängig der vielen Unter-

⁶¹ Angaben beziehen sich auf Zustimmung zu der Aussage „Ich finde es empörend, wenn sich die Langzeitarbeitslosen auf Kosten der Gesellschaft ein bequemes Leben machen.“.

schiede zwischen den Surveys, überall auftreten. Sowohl für die Erhebung der Vorurteile als auch für die Täuschungsskalen liegen in allen Surveys nur Kurzskalen vor. Die jeweiligen Items sowie ihre Antwortoptionen sind aus Platzgründen im Anhang dargestellt.

Als sozio-ökonomische Kontrollvariablen werden zusätzlich das Alter, Geschlecht und der Bildungsstand berücksichtigt. Das Geschlecht ist binär mit 0 für männlich und 1 für weiblich über alle Surveys identisch codiert. Das Alter wurde als metrische Variable aus dem Geburtsjahr gebildet. Die Bildung wird über den höchsten Schulabschluss gemessen und als ordinale Variable in Form von Dummyvariablen in dem Modell kontrolliert. Die Kategorie „niedrige Bildung“ umfasst den Haupt- und Volksschulabschluss, den Abschluss an der polytechnischen Oberschule nach der 8. Klasse und Befragte ohne Schulabschluss. Eine „mittlere Bildung“ (Referenzgruppe) besitzen Befragte mit Realschulabschluss bzw. einen Abschluss an der polytechnischen Oberschule nach der 10. Klasse. Alle Abschlüsse über der mittleren Reife wurden der Kategorie „hohe Bildung“ zugeordnet.

5.3.3 Methode

Bei der Beziehung zwischen Vorurteilen und Täuschungsskalen handelt es sich um einen asymmetrischen Zusammenhang. Die theoretischen Annahmen gehen davon aus, dass die Ausprägung des Anerkennungsbedürfnisses auf die geäußerten negativen Stereotypen wirkt. Bei dem SD-Bias handelt es sich zudem nur um die Messung einer Verzerrung im Rahmen der Datenerhebung. Damit müsste die Erklärungskraft der Täuschungsskalen immer unter Berücksichtigung zusätzlicher erklärender Variablen (wie bspw. sozio-ökonomische Faktoren) erfolgen, denn nur so lässt sich einschätzen, wie relevant die Skala tatsächlich ist. Zunächst werden in der Analyse deshalb die Mittelwerte und Standardabweichungen für alle Indizes betrachtet, um einen Überblick über die Unterschiede zwischen den Skalen sowie zwischen den Surveys zu erhalten. Für die Prüfung der Hypothesen werden schrittweise Regressionen zwischen den Täuschungsskalen und den Vorurteilsitems betrachtet. Crowne und Marlow beschreiben in ihrem Modell einen einfachen linearen Effekt der Skala. Aus diesem Grund soll der Zusammenhang zwischen Täuschungsskalen und Vorurteilsitems über die lineare Regression ermittelt werden. Dabei wird zunächst das bivariate Modell berechnet. Im zweiten Schritt werden sozio-ökonomische Kontrollfaktoren (Alter, Geschlecht, Bildung) zur Überprüfung der Stabilität in das Modell integriert.

5.3.4 Ergebnisse

5.3.4.1 Deskriptive Ergebnisse

Im ersten Schritt der Analyse wird ein allgemeiner Überblick über die Zustimmung zu den Täuschungsskalen sowie den Vorurteilsitems gegeben. Grundlage für die Auswahl der Minderheitsgruppen war die Annahme, dass sich die vier Formen von Vorurteilen in dem Grad ihrer gesellschaftlichen Tabuisierung unterscheiden. Als Hinweis dafür können die Mittelwerte der vier Gruppen jeweils miteinander verglichen werden.

Tabelle 12: Mittelwerte der Täuschungs- und Vorurteilsskalen

		Täuschungsskalen				Vorurteilsskalen				
		Marlow-Crowne Skala (count) ¹	Marlow-Crowne Skala (mean) ¹	SES-17 ¹	BIDR-Fremd-täuschung ¹	BIDR-Selbst-täuschung ¹	Antisemitismus ²	Fremdenfeindlichkeit ²	Islamophobie ²	Ablehnung von Langzeitarbeitslosen ²
Allbus 2006	Mean	7,26	1,74				3,46	4,54		
	St.D	1,92	0,19				1,72	1,42		
	N	3421	3419				3199	3192		
	Range	0-10	1-2				1-7	1-7		
GMF Graduierten Survey	Mean			3,51	3,12	2,93	1,59	2,11	1,83	2,58
	St.D			0,56	0,67	0,64	0,79	0,91	0,88	0,95
	N			446	446	446	797	840	830	839
	Range			1-4	1-4	1-4	1-4	1-4	1-4	1-4
GMF Pretest	Mean			3,35			1,55	1,94	1,80	2,54
	St.D			0,42			0,72	0,88	0,84	0,91
	N			212			188	200	197	200
	Range			1-4			1-4	1-4	1-4	1-4
Websurveys	Mean			3,37	2,92	2,93	1,64		2,13	
	St.D			0,60	0,57	0,51	0,76		0,91	
	N			678	682	680	621		642	
	Range			1-4	1-4	1-4	1-4		1-4	

1 Je höher der Wert desto geringer das Anerkennungsbedürfnis.

2 Je höher der Wert desto mehr Vorurteile.

Zunächst zeigt sich, dass über alle Stichproben hinweg die Befragten bei der Kurzversion der SES-17 Skala signifikant das höchste Anerkennungsbedürfnis aufweisen (vgl. Tabelle 12). Die Zustimmung zu der sozial unerwünschten Verhaltensweise der SES-17 Items⁶² scheint den Befragten also wesentlich schwerer zu fallen, als die Bejahung der FT-Items⁶³, denn die Mittelwerte sind über alle Stichproben bei der SES-17 Skala am größten. Die Differenzen zwischen Fremd- und Selbsttäuschung fallen hingegen geringer aus. Bei den Websurveys ist der Unterschied zwischen beiden nicht mehr signifikant.

62 Manchmal werfe ich Müll einfach in die Landschaft oder Straße.

Ich habe schon einmal geliehene Sachen nicht zurückgegeben/behalten.

Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte.

63 Ich bin immer ehrlich zu anderen.

Ich habe schon einmal zu viel Wechselgeld herausbekommen, ohne es der Verkäuferin zu sagen.

Es ist schon einmal vorgekommen, dass ich jemanden ausgenutzt habe.

Neben dem Vergleich der Täuschungsskalen, sind die Differenzen zwischen den Vorurteilen jedoch zentraler. In allen Stichproben ist die Zustimmung zu den Antisemitismus-Items am geringsten. Am meisten Vorurteile werden hingegen, entsprechend der theoretischen Vermutung, in Bezug auf Langzeitarbeitslose geäußert. In den GMF Surveys liegt zwischen beiden Skalen eine Mittelwertdifferenz von 0.9. Die Vorurteile gegenüber Muslimen und Ausländern liegen zwischen diesen beiden Extrempunkten. Anders als zunächst vermutet, sind jedoch auch hier die Differenzen bei den GMF-Surveys signifikant. Die Befragten äußerten signifikant mehr Vorurteile gegenüber Ausländern, als gegenüber Muslimen.

Bezogen auf die Stichprobenunterschiede fällt vor allem auf, dass in dem anonymen Websurvey das Anerkennungsbedürfnis gemessen über die FT-Skala geringer ist, als im Graduierten-survey. Gleichzeitig ist die Zustimmung zu negativen Einstellungen gegenüber Juden und Muslimen im Schnitt höher als bei den persönlichen Befragungen, was für eine höhere Anonymität spricht. Zunächst lässt sich also festhalten, dass sich die Vermutung, über unterschiedliche Stufen der Tabuisierung, auf Grundlage der Mittelwertvergleiche nicht falsifizieren lässt. Vorurteile gegenüber Juden werden wesentlich seltener geäußert als gegenüber Muslimen, Ausländern und vor allem gegenüber Arbeitslosen.

5.3.4.2 Multivariate Ergebnisse

Im nächsten Schritt werden nun die Zusammenhänge zwischen den Täuschungs- und Vorurteilsskalen auf Basis von linearen Regressionen näher dargestellt, um die jeweiligen theoretischen Annahmen zu überprüfen. In Hypothese 1 wurde vermutet, dass mit steigendem Anerkennungsbedürfnis die geäußerten Vorurteile abnehmen. Entsprechend der Codierung der Skalen wird damit ein **negativer Zusammenhang** zwischen Täuschungsskalen und Vorurteilen unterstellt. Die Ergebnisse der Regression sind in Tabelle A1 bis A24 im Anhang dargestellt.

Als Erstes fällt auf, dass der Zusammenhang zwischen Täuschungsskalen und Vorurteilen nicht überall signifikant ist. Systematische Unterschiede lassen sich jedoch auf den ersten Blick nicht erkennen. Auch die Richtung des Zusammenhangs entspricht größtenteils nicht den theoretischen Erwartungen. Abgesehen von einigen Ausnahmen, besteht fast ausschließlich ein positiver Zusammenhang zwischen den Konstrukten. Inhaltlich bedeutet dies, dass mit steigendem Anerkennungsbedürfnis auch die geäußerten Vorurteile zunehmen⁶⁴. Personen, die dazu neigen sich in einem positiven Licht darzustellen und eine gewisse Täuschungstendenz aufweisen, stimmten häufiger antisemitischen, fremdenfeindlichen, islamophoben und den negativen

64 Diesen paradoxen Zusammenhang fand auch Stocké in seiner Erhebung (2001, 2004).

Aussagen gegenüber Arbeitslosen zu. Nur im Websample ist tatsächlich ein theoriekonformer Zusammenhang zwischen der SES-17 Skala und Antisemitismus zu finden. Mit einer empirischen Irrtumswahrscheinlichkeit von 7,7% findet sich erwartungsgemäß eine negative Beziehung zwischen FT-Skala und Fremdenfeindlichkeit. Insgesamt ist jedoch über alle Stichproben hinweg eher ein positiver Zusammenhang zwischen den Konstrukten zu verzeichnen. Hypothese 1 lässt sich damit an den vorliegenden Daten nicht bestätigen.

Hypothese 2 unterstellte, dass FT und SES-17 Skala im Gegensatz zur ST inhaltlich ähnliche Konstrukte erklären. In Bezug auf die Daten müssten beide Skalen damit nicht nur auf ähnliche Vorurteile wirken, sondern auch größere Gemeinsamkeiten in den Effektstärken aufweisen. Zunächst einmal sind die signifikanten Zusammenhänge so unsystematisch über die Skalen verteilt, dass innerhalb der Stichproben kein sinnvoller Vergleich der Effektstärken möglich ist. Nur in drei Fällen finden sich überhaupt signifikante Effekte bei mehr als nur einer Skala. In dem GMF-Graduierten Survey stehen sowohl die FT als auch die SES-17 Skala mit Fremdenfeindlichkeit im Zusammenhang. In Bezug auf die Islamophobieitems sind es hingegen Fremd- und Selbsttäuschung. In dem Websample findet sich ein schwacher Zusammenhang zwischen Antisemitismus und der ST Skala. Auch bei Berücksichtigung der Modi-Unterschiede zwischen den einzelnen Stichproben ergibt sich kein systematisches Bild. Entsprechend des theoretischen Hintergrundes wäre es denkbar, dass bei weniger anonymen Erhebungsmodi die FT und SES-17 Skala stärker wirken, da sie die Tendenz zur Täuschung gegenüber anwesenden Personen erfassen. Die ST-Skala hingegen müsste auch in den Onlinestichproben einen stabilen Effekt haben. Sowohl die Ergebnisse zum GMF-Survey als auch der Vergleich zu der Onlinestichprobe Websample bestätigen diese Annahme jedoch nicht. Zusammenfassend lässt sich festhalten, dass auch Hypothese 2 durch die Daten nicht bestätigt wird.

Hypothese 3 bezieht sich auf die Beziehungsmuster zwischen den einzelnen Minderheiten. Bei Antisemitismusitems ist aufgrund der stärkeren Tabuisierung von einem stärkeren Einfluss der Skalen auszugehen, als bei den Langzeitarbeitslosen. Welches Bild zeichnet sich nun in den Daten ab? Die meisten signifikanten Effekte auf einem 10%-Niveau lassen sich tatsächlich in Bezug auf die Fremdenfeindlichkeit finden. In vier von fünf möglichen Kombinationen liegt ein Zusammenhang vor. Die negative Einstellung gegenüber Arbeitslosen wird erwartungsgemäß am wenigsten von den Täuschungsskalen beeinflusst. Antisemitismus und Islamophobie bewegen sich mit jeweils drei von acht und vier von sieben signifikanten Effekten in ähnlichen Bereichen. Berücksichtigt man nur Koeffizienten mit einer Irrtumswahrscheinlichkeit von 5%, sind die Einflüsse der Skalen auf die Islamophobie wesentlich größer. Dies zeigt sich auch in den Effektstär-

ken. Während zwischen den Täuschungsskalen und dem Antisemitismus eher sehr schwache Effekte zu finden sind (0.053 und 0.082), zeigen sich stärkere Koeffizienten bei den anderen drei Vorurteilen⁶⁵. Hypothese 3 lässt sich damit nur teilweise bestätigen. Während Vorurteile gegenüber Langzeitarbeitslosen theoriekonform tatsächlich am schwächsten mit den Täuschungsskalen zusammenhängen, ist auch der Zusammenhang mit dem Antisemitismus im Vergleich zur Islamophobie und Fremdenfeindlichkeit relativ schwach. Die Täuschungsskalen haben die stärkste Erklärungskraft bezogen auf Vorurteile gegenüber Ausländern und Muslimen. Doch warum ist der Zusammenhang zwischen Antisemitismus und Täuschungsskalen im Verhältnis schwächer?

Eine mögliche Erklärung bietet auch hier der Grad der Tabuisierung. Zuvor wurde eine lineare Beziehung zwischen Tabuisierung und Effekt der Täuschungsskalen unterstellt. Es ist jedoch durchaus vorstellbar, dass gerade Antisemitismus in Deutschland ein solch sensibles Thema darstellt, dass auch Personen ohne großes Anerkennungsbedürfnis judenfeindlichen Aussagen nicht zustimmen. Negative Einstellungen würden dann nur noch über die von Bergmann und Erb (1991) beschriebene Umwegkommunikation, also in Form von Israelkritik, geäußert werden. Eine weitere Erklärung könnte in unterschiedlichem Nonresponse-Verhalten liegen. Personen, die negative Einstellungen gegenüber Juden haben, könnten bei hohem Anerkennungsbedürfnis häufiger auf Kategorien wie „weiß nicht“ oder „keine Angabe“ ausweichen.

Diese Annahmen lassen sich für die GMF-Pretest-Daten überprüfen. In den Tabellen 13-15 sind die Regressionen der offenen Messung des Antisemitismus sowie Items zum subtilen Antisemitismus und zur Israelkritik wiedergegeben. Zusätzlich wurden für die offenen Antisemitismusfragen die Häufigkeiten der fehlenden Angaben gezählt (Tabelle 16). Es zeigt sich, dass die Täuschungsskala in Bezug auf den subtilen Antisemitismus keinen Einfluss hat. Bei der Israelkritik und dem offenen Antisemitismus besteht hingegen ein theoriekonformer, wenn auch schwacher signifikanter Zusammenhang. Das heißt mit steigendem Anerkennungsbedürfnis, nimmt die Zustimmung zu der Aussage „Israel hat ein Recht auf Selbstverteidigung gegenüber den Palästinensern.“ zu. Am stärksten ist der Zusammenhang jedoch bei den Missing Values. Je höher das Anerkennungsbedürfnis ist, desto eher weichen Befragte auf die Nonresponse-Kategorien aus⁶⁶. Dieses Ergebnis entspricht der Vermutung, dass sich die schwachen Zusammen-

65 Eine Ausnahme stellt in diesem Zusammenhang die MC-Skala im Allbus 2006 dar. Hier sind die Effekte auf die Fremdenfeindlichkeit ebenfalls relativ schwach, aber dennoch stärker als auf Antisemitismus.

66 Bezogen auf die Nonresponse-Kategorien wurde die selbe Analyse auch für das GMF-Grad und das Websample durchgeführt. Die Ergebnisse werden an dieser Stelle nicht näher dargestellt. Die FT Skala zeigte in dem Websample bei einer Irrtumswahrscheinlichkeit von <10% einen signifikanten Effekt.

hänge zwischen Antisemitismus und den Täuschungsskalen auf die starke Tabuisierung zurückführen lassen. Dieser Befund weist darauf hin, dass nicht nur die inhaltlichen Antworten, sondern auch die Nonresponse-Kategorien bei der Auseinandersetzung mit sozialer Erwünschtheit berücksichtigt werden müssen. Bisher konzentriert sich die Forschung in diesem Bereich jedoch ausschließlich auf die inhaltlichen Kategorien. Es ist deshalb nicht nur fraglich, ob die theoretischen Modelle ebenso auf die Erklärung der Nonresponse Antworten übertragbar sind, sondern auch, ob die hier dargestellten Ergebnisse reproduzierbar bzw. auf andere Vorurteilsgruppen übertragen werden können.

Tabelle 13: Lineare Regression des offenen Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest)⁶⁷

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,947	0,429		0,000	2,628	0,451		0,000
SES-17	-0,119	0,128	-0,069	0,356	-0,240	0,136	-0,140	0,079
Geschlecht					-0,227	0,105	-0,150	0,033
Alter					0,003	0,003	0,084	0,293
Bildung niedrig					0,397	0,152	0,217	0,010
Bildung mittel								
Bildung hoch					-0,281	0,122	-0,193	0,022
korr. R ² (F-Test)	-0,001			0,356	0,137			0,000
N		182				182		

⁶⁷ Antisemitismus (1=keine Vorurteile, 4=viele Vorurteile)

SES-17 (1=hohes Anerkennungsbedürfnis, 4=niedriges Anerkennungsbedürfnis)

Geschlecht (0=männlich, 1=weiblich)

Bildung niedrig (Haupt- und Volksschulabschluss, Abschluss an der polytechnischen Oberschule nach der 8. Klasse und Befragte ohne Schulabschluss)

Bildung mittel (Realschulabschluss bzw. Abschluss an der polytechnischen Oberschule nach der 10. Klasse)

Bildung hoch (alle Abschlüsse über der mittleren Reife)

Tabelle 14: Lineare Regression des subtilen Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest)⁶⁸

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,818	0,537		0,001	2,490	0,601		0,000
SES-17	0,120	0,160	0,062	0,454	0,200	0,190	0,103	0,295
Geschlecht					-0,367	0,146	-0,208	0,013
Alter					-0,005	0,005	-0,103	0,302
Bildung niedrig					0,099	0,211	0,046	0,640
Bildung mittel								
Bildung hoch					-0,162	0,168	-0,095	0,336
korr. R ² (F-Test)	-0,003			0,454	0,029			0,105
N		147				147		

Tabelle 15: Lineare Regression der Israelkritik auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest)⁶⁹

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,204	0,521		0,000	3,115	0,575		0,000
SES-17	-0,363	0,156	-0,183	0,021	-0,299	0,176	-0,151	0,091
Geschlecht					0,155	0,130	0,093	0,236
Alter					-0,006	0,004	-0,126	0,161
Bildung niedrig					0,166	0,197	0,079	0,399
Bildung mittel								
Bildung hoch					-0,168	0,154	-0,103	0,277
korr. R ² (F-Test)	0,027			0,021	0,047			0,030
N		158				158		

Tabelle 16: Lineare Regression der Missing Values für den offenen Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne auf die Missing Values für den offenen Antisemitismus (GMF Graduierten Pretest)⁷⁰

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	-0,597	0,329		0,071	-0,272	0,372		0,465
SES-17	0,265	0,098	0,191	0,007	0,249	0,111	0,180	0,026
Geschlecht					-0,008	0,085	-0,007	0,925
Alter					-0,001	0,003	-0,027	0,733
Bildung niedrig					-0,200	0,120	-0,137	0,099
Bildung mittel								
Bildung hoch					-0,315	0,097	-0,270	0,001
korr. R ² (F-Test)	0,032			0,007	0,064			0,004
N		195				195		

68 Legende siehe Fn. 59

69 Legende siehe Fn. 59

70 Legende siehe Fn. 59

Zusammenfassend lässt sich festhalten, dass sich keine der theoretischen Hypothesen eindeutig durch die Daten bestätigen lässt. Die einzelnen Effekte sind unsystematisch oder nicht signifikant. Auch die Differenzen zwischen den Stichproben sind groß. Eine Ursache hierfür liegt sicherlich in den unterschiedlichen Designs der Erhebungen. Der Modus reicht von persönlichen Interviews (Allbus 2006), über CATI-Befragungen (GMF-Surveys) bis hin zu Onlinestichproben (Websample). Die Frage nach der Anonymität der Interviewsituation ist aber gerade bei sensiblen Fragen nicht unerheblich. Es kann davon ausgegangen werden, dass mit steigender Anonymität der Situation die Relevanz des Anerkennungsbedürfnisses abnimmt. Stocké (2004) interpretiert in seiner Studie die Anonymität sogar im Sinne einer Dichotomie. Dies würde bedeuten, dass in einer privaten Befragungssituation ohne Interviewer, wie es bei Onlineerhebungen der Fall ist, überhaupt keine Zusammenhänge zwischen Täuschungsskalen und Vorurteilen zu erwarten sind. Die Analyse zeigt jedoch eindeutig, dass auch bei den Onlinestichproben signifikante Einflüsse der Skalen zu verzeichnet werden. Die verschiedenen Grundgesamtheiten sowie Fragebögen und Erhebungsmodi schränken die Vergleichbarkeit der Studien untereinander dennoch ein. Es sollte geklärt werden, welchen Erklärungsbeitrag die Skalen in Bezug auf Vorurteile leisten, wenn sie wie bisher in den Studien bivariat oder als Kontrollvariable eingesetzt werden. Das Ergebnis der Analyse fällt nicht sehr überzeugend aus. Die meisten Zusammenhänge weisen entgegengesetzte Kausalrichtung auf. Die wenigen signifikanten Effekte sind höchstens schwach ausgeprägt. Es ist also nicht verwunderlich, dass zu diesen Daten bisher noch keine Studien veröffentlicht wurden. Doch was sind die Ursachen für die heterogenen Befunde?

Eine mögliche Erklärung bieten die Skalen selbst. Es ist fraglich, ob die eingesetzten Messungen auch für den deutschen Raum eine zuverlässige Messung des SD-Bias darstellen. Besonders die Auswahl der in den Kurzskalen verwendeten Items ist fraglich. Die Faktorenanalyse in Kapitel 5.1 zeigte, dass die Auswahl nicht zwingend die ladungsstärksten Items wiedergibt. Eine andere Erklärung bieten theoretischen Arbeiten von Stocké (2004) sowie Stocké und Hunkler (2004). Wie bereits erwähnt betont er in seinem Aufsatz die Bedeutung der Erwünschtheitswahrnehmung. Demnach besteht in der Gesellschaft keine einheitliche Wahrnehmung über das Tabu von Vorurteilen. Vielmehr kann sich die Tabuisierung je nach sozialem Hintergrund zwischen einzelnen Gruppen stark unterscheiden. Bei der isolierten Interpretation der Täuschungsskala wird jedoch die Prämisse eines gesellschaftlichen Konsens gesetzt. Gerade die umgekehrten Kausalrichtungen weisen jedoch darauf hin, dass diese Prämisse falsch sein könnte. Mit Hilfe des GMF-Pretest ist eine Überprüfung dieser Annahme möglich. Hier wurden für

die Antisemitismus-, Fremdenfeindlichkeit- und Islamophobieitems zusätzlich die Erwünschtheitswahrnehmung erhoben.

Tabelle 17: Mittelwerte der Erwünschtheitswahrnehmung für verschiedene Vorurteile (GMF Graduierten Pretest)

	Erwünschtheitswahrnehmung ¹		
	Anti-semitismus	Fremdenfeindlichkeit	Islamophobie
Mean	1,962	2,652	2,436
St.D.	0,837	0,898	0,873
N	184	194	194

1) 1=Vorurteil öffentlich gar nicht erwünscht, 4=Vorurteil öffentlich sehr erwünscht

Die Vorurteile gegenüber Juden werden insgesamt am sozial unerwünschtesten wahrgenommen. Auf einer Skala von eins bis vier liegt die Erwünschtheitswahrnehmung für die Fremdenfeindlichkeit und die Islamophobiefragen hingegen im Bereich der Mitte der Skala. Ein Großteil der Befragten bewertet die Äußerung von Vorurteilen gegenüber diesen Gruppen also durchaus als sehr öffentlich erwünscht. Bei diesen Vorurteilen ist entsprechend der theoretischen Vermutung ein umgekehrter Zusammenhang denkbar. Die Tabellen A25 bis A30 im Anhang geben die Regressionen für alle Befragten nach Erwünschtheitswahrnehmung getrennt wieder.

Die Regressionen über alle Befragten gleichermaßen zeigt den bereits bekannten Zusammenhang zwischen SES-17 Skala und der Fremdenfeindlichkeit, welcher der theoretischen Erwartungen widerspricht. Trennt man hingegen nach Personen, die Vorurteile als öffentlich erwünscht und öffentlich unerwünscht empfinden, so zeigt sich, dass dieser Effekt nur von einer Gruppe verursacht wird. Tatsächlich äußern Personen, die Vorurteile gegenüber Ausländern als öffentlich erwünscht empfinden, häufiger Vorurteile, je größer ihr Anerkennungsbedürfnis ist. Zu beachten ist außerdem, dass, auch wenn die Effekte nicht signifikant sind, zumindest die Vorzeichen für die beiden Untergruppen nun der theoretischen Vermutung von Stocké entsprechen. Mit Berücksichtigung der Erwünschtheitswahrnehmung lassen sich also durchaus fehlende und widersprüchliche Zusammenhänge aufklären.

5.3.5 Zusammenfassung und Bewertung

Insgesamt bleibt festzuhalten, dass der bisherige Einsatz der Skalen in Rahmen der Surveyforschung wenig sinnvoll ist. Die isolierte Erhebung von Täuschungsskalen ohne Kenntnisse über die wahrgenommene Erwünschtheit produzieren nicht interpretierbare und unsystematische Effekte. Es zeigt sich, dass selbst bei Vorurteilen gegenüber Gruppen mit sehr hoher Tabui-

sierung wie beim Antisemitismus, der Einsatz der Täuschungsskalen ohne fundierte theoretische Hintergründe zu widersprüchlichen Ergebnissen führt. Hinzu kommt, dass der Einsatz von nicht ausreichend getesteten Kurzversionen der Täuschungsskalen die so schon schwachen Zusammenhänge zusätzlich verschleiern. Vor diesem Hintergrund scheint es notwendig alternative Erklärungsmodelle, die über das einfache Modell von Marlow und Crowne hinausgehen, auf ihren Nutzen für die Surveyforschung hin zu untersuchen.

5.4 Anwendung der RC-Modelle

Die vorangegangenen Untersuchungen haben ein unbefriedigendes Ergebnis zu den bisherigen Einsatzes der Skalen zur Messung sozial erwünschter Antworttendenzen ergeben. Die Analysen der Bevölkerungsumfragen haben nur schwache und unsystematische Zusammenhänge zwischen den einzelnen Skalen und Vorurteilen aufgezeigt. Zudem entsprach die Zusammenhangsrichtung in vielen Fällen nicht der theoretischen Vermutung. Die Faktorenanalyse offenbarte Schwächen der Kurzskalen und Struktur der einzelnen Täuschungsskalen. Diese Kritikpunkte stellen die Anwendung des Erklärungsmodells von Marlow und Crowne in der Surveyforschung infrage. Ihrem Ansatz nach lässt sich sozial erwünschtes Antwortverhalten allein durch die Messung des Anerkennungsbedürfnisses kontrollieren. Weitergehende Forschungen haben jedoch gezeigt, dass gerade die Erwünschtheitswahrnehmung bei der korrekten Interpretation von Zusammenhängen von entscheidender Bedeutung ist. Bisher findet jedoch auf dem Gebiet der sozialen Erwünschtheit nur wenig theoriegeleitete Forschung statt. Eine Ausnahme für die deutsche Forschung stellen die Studien von Stocké (2004) und Skarbek-Kozietulska et al. (2012) dar, die auf Grundlage des RC-Modells von Esser komplexere Konstrukte zur Erklärung des SD-Bias einsetzen. Während Stocké (2004) seine theoretischen Vermutungen alle empirisch belegen konnte, zeigte sich bei Skarbek-Kozietulska et al. (2012) ein anderes Bild. Hier bestätigt sich der Interaktionseffekt für das Rational-Choice Modell nicht. Beide Studien beziehen sich jeweils auf unterschiedliche sensitive Fragen bei der Erklärung sozial erwünschten Antwortverhaltens. Während Stocké sein Modell auf Einstellungsfragen überträgt, wenden Skarbek-Kozietulska et al. ihre Analyse auf berichtetes delinquentes Verhalten an. Ein weiterer Unterschied besteht in der Interpretation der Erklärungsfaktoren, insbesondere der Anonymität. Während bei Stocké nur der Interviewereffekt relevant ist, argumentieren Skarbek-Kozietulska et al. indirekt, dass die Erwünschtheitswahrnehmung auch durch gesellschaftliche Normen in anonymen Befragungen zum Tragen kommt. Es ist denkbar, dass die Erwünschtheitswahrnehmung der beiden Gruppen (Interviewer und Gesellschaft) unterschiedliche Effekte hervorrufen. Es ist fraglich, ob sich die heterogenen Ergebnisse auf die Unterschiede in der inhaltlichen oder Bezugsgruppen Dimension zurückführen lassen und damit beide Modelle unterschiedliche Phänomene des sozial erwünschten Antwortverhaltens zu erklären versuchen. Aufgrund der großen Kontextunterschiede der einzelnen Untersuchungen ist es deshalb auch nicht möglich zu beurteilen, welches der beiden Modelle von Stocké und Skarbek-Kozietulska et al. den größeren Erklärungsbeitrag liefert und ob eines der beiden Modelle tatsächlich eine Verbesserung gegenüber dem Ansatz von Marlow und Crowne darstellt. In der folgenden Ana-

lyse soll daher an einer Studierendenstichprobe die Erklärungskraft aller drei Modelle miteinander verglichen werden. Um die Stabilität der Konstrukte zu untersuchen, werden die Modelle, wie schon im Kapitel zuvor, auf vier verschiedene Minderheiten übertragen. Vor dem Hintergrund der Kritik an den Täuschungsskalen in Kapitel 4.1.1 werden zusätzlich verschiedene Messungen des Anerkennungsbedürfnisses eingesetzt, um die Effekte der Operationalisierung zu überprüfen.

5.4.1 Eckpunkte der Modellvariation

Der Ansatz von Marlow und Crowne versucht, allein über das Anerkennungsbedürfnis, sozial erwünschtes Antwortverhalten zu erklären. Das Rational-Choice Modell von Stocké besagt, dass nur über den Interaktionseffekt aus Anerkennungsbedürfnis, fehlende Anonymität und Erwünschtheitswahrnehmung der SD-Bias kontrolliert werden kann. In dem Modell von Skarbek-Kozietulska et al. hingegen können alle drei Faktoren auch unabhängig voneinander einen Effekt auf das Antwortverhalten haben. Die Überprüfung der Erklärungskraft sowie der Stabilität der verschiedenen Modelle erfolgt in drei Schritten:

1. Variation der sensitiven Items

Stocké überprüfte die Erklärungskraft seines Modells ausschließlich bezüglich Items der Fremdenfeindlichkeit. Skarbek-Kozietulska et al. hingegen testeten ihre Annahmen auf berichtetes delinquentes Verhalten. In keiner der theoretischen Erläuterungen gibt es eine Einschränkung hinsichtlich des Themas der sensitiven Fragen. Demnach müssten alle Modelle auf unterschiedlichste Gebiete übertragbar sein. Die Problematik der verschiedenen Ebenen des SD-Bias (Kapitel 2.2) hat jedoch gezeigt, dass es notwendig ist, für die Vergleichbarkeit der Ergebnisse, sich bei der Analyse auf eine Frageform zu beschränken. Um die Ergebnisse von Stocké möglichst nah zu replizieren, wird dabei zunächst die Fremdenfeindlichkeit untersucht. Zusätzlich werden die bereits diskutierten Vorurteile gegenüber Muslimen, Juden und Langzeitarbeitslosen betrachtet. Die Vermutungen über die Stärke des Zusammenhangs zwischen den einzelnen Minderheiten folgen dabei den Hypothesen aus Kapitel 5.3.

2. Überprüfung der Anonymität

Zunächst ist festzuhalten, dass es sich bei den Daten um eine anonyme Befragung handelt. Folgt man streng dem Modell von Stocké, dürfte es bei solch einer Onlinebefragung zu keinerlei sozial erwünschtem Antwortverhalten kommen. Skarbek-Kozietulska et al. (2012) widersprechen dieser kategorialen Auslegung der Anonymität. Sie finden auch in ihrer postalischen Be-

fragung Erwünschtheitsverzerrungen. Bei ihrer Analyse wird die Anonymitätsvariable jedoch überhaupt nicht berücksichtigt. Fraglich ist, ob bei der Anonymität tatsächlich ausschließlich der Modus relevant ist⁷¹. Entscheidender in diesem Zusammenhang scheint vielmehr die individuelle wahrgenommene Anonymität. Dies könnte auch die Ergebnisse von Skarbek-Kozietulska et al. (2012) erklären. Es ist möglich, dass einige der Respondenten, die postalische Befragung eben nicht als anonyme Situation wahrgenommen haben, sondern gerade bedingt durch den offiziellen schriftlichen Weg eine Identifikation für möglich hielten. Für die vorliegende Onlineumfrage ergibt sich eine ähnliche Problematik. Das Argument, dass Onlineumfragen generell als anonym wahrgenommen werden, ist nicht immer zutreffend (Whelan/Meade 2009). So wurden die Studierenden zum einen in einer persönlichen E-Mail mit ihrem Nachnamen kontaktiert. Zum anderen gibt es keine Information darüber, unter welchen Bedingungen der Fragebogen ausgefüllt wurde. Wenn die Respondenten die Fragen beantworten, während bspw. Kommilitonen anwesend waren, so ist die Situation anonym wie eine persönliche Befragung.

Um auf die theoretische Problematik der Anonymität einzugehen, werden deshalb drei verschiedene Modelle überprüft. In dem ersten Modell orientiert sich an den Operationalisierung von Stocké. Da in der Erhebung der Modus nicht variiert wurde, soll auf Grundlage der subjektiven Anonymität eine dichotome Variable erstellt werden, die unterscheidet zwischen Befragten, die die Situation als völlig anonym einschätzten und Befragten, die sich nicht vollständig privat fühlten. Diese Messung stützt sich auf die berichteten Ergebnisse der Validierungsstudien, die auch in anonymen Befragungen Antwortverzerrungen nachweisen konnten. Das zweite Modell folgt der Interpretation von Skarbek-Kozietulska et al. in dem hier die Variable Anonymität nicht berücksichtigt wird. Der Interaktionseffekt ergibt sich hier allein aus dem Anerkennungsbedürfnis und der Erwünschtheitswahrnehmung. In einem dritten Modell werden beide Interpretationen miteinander verbunden. Hierfür wird die Variable Anonymität nicht wie bei Skarbek-Kozietulska et al. völlig ausgeschlossen, sondern als kategoriale Messung in Form der wahrgenommenen Anonymität in das Modell von Stocké integriert⁷².

3. Variation des Anerkennungsbedürfnisses

Als letzte Variation sollen verschiedene Messungen des Anerkennungsbedürfnisses überprüft werden. In Kapitel 5.1 wurde an denselben Datensatz bereits gezeigt, dass vor allem das theoretische Konstrukt der SES-17 Skala den Analysen nicht standhält. Sowohl Stocké als auch

71 Hinzu kommt, dass durch die Variation des Modus auch andere Effekte als nur die Anonymität auf den Befragten auswirken können.

72 Da die Autoren in ihrem Ansatz keine Aussage zum Einfluss der Anonymität treffen, ist das Modell von Marlow und Crowne in dieser Aufzählung nicht berücksichtigt.

Skarbak-Kozietulska et al. verwendeten Items auf Grundlage der MC-Skala, deren eindimensionale Faktorenstruktur ebenfalls zu bezweifeln ist. Die Modelle sollen deshalb neben einer vollständigen und einer verbesserten Form der SES-17 Skala auch für die Fremdtäuschungsskala überprüft werden. Da sich bei dem SD-Bias aus der hier gewählten Definition um eine bewusste Form der Täuschung handelt, wird die Erklärungskraft der Selbsttäuschungsskala nicht angewendet. Alle Skalen werden als Index aus den Mittelwerten der Items gebildet, wobei der Mittelwert auf Basis der vorhandenen Werte, ohne Ausschluss der Fälle die einzelne Missings beinhalteten, berechnet wurde.

Aus diesen Variationen ergibt sich eine Vielzahl von Modellen. Würde man alle Varianten miteinander in Verbindung setzen, müssten in der folgenden Analyse 36 Modelle besprochen werden. Um die Übersichtlichkeit der Ergebnisse zu gewährleisten, wird deshalb nur ein Ausschnitt der Modelle diskutiert. Im Anhang (A31 bis A46) sind alle Modelle vollständig dargestellt.

5.4.2 Operationalisierung

Im Folgenden wird die Operationalisierung der einzelnen Konstrukte näher erläutert. Als sensitive Fragen wurden vier verschiedene Vorurteilmessungen ausgewählt. Insgesamt wurden Vorurteilsitems gegenüber Juden, Muslimen, Ausländern und Langzeitarbeitslosen erhoben. Die Vorurteile gegenüber Ausländern waren auch in der Studie von Stocké (2004) abhängige Variable. Die zusätzliche Überprüfung von Antisemitismus, Islamophobie und negativen Einstellungen zu Arbeitslosen ermöglicht es die Modelle auch für solche Vorurteile mit unterschiedlichen Erwünschtheitsgrad und Sensitivität in der Bevölkerung zu überprüfen. Der Wertebereich liegt zwischen 1 (hohe Zustimmung zu Vorurteilen) bis 4 (keine Zustimmung zu Vorurteilen). Das Anerkennungsbedürfnis wurde über die Täuschungsskalen SES-17 und BIDR (beide in ihrer vollständigen deutschen Version) erfasst. Die Analysen in Kapitel 5.1 zeigten jedoch auf Grundlage desselben Datensatzes bereits, dass die eindimensionale Struktur der SES-17 Skala einer empirischen Überprüfung nicht standhält. Aus diesem Grund soll zusätzlich zu der vollständigen Skala eine validere Kurzversion eingesetzt werden. Für die Skala wurden die ladungsstärksten Items einer explorativen Faktorenanalyse ausgewählt. Die so entstandene Kurzsкала wurde anschließend mit der konfirmatorischen Faktorenanalyse die Reliabilität überprüft und optimiert. Der Modus innerhalb der Befragung konnte nicht variiert werden, daher wird die Privatheit der Antwortsituation über die wahrgenommene Anonymität in Form einer kategorialen Skala erhoben. Für das Modell von Stocké wird die Anonymität in anonyme und nicht-anonyme

Wahrnehmung dichotomisiert. Für die eigene Modellvariante wird die Anonymität hingegen als kategoriale Variable belassen. Die Erwünschtheitswahrnehmung zu den einzelnen Vorurteilsitems wurde direkt abgefragt. Die Studierenden mussten auf einer Fünfer-Skala für jedes Item einschätzen, ob sie die Aussage für gesellschaftlich erwünscht oder unerwünscht halten. Aus diesen Angaben lässt sich für jede Person ein individueller Wert der Erwünschtheitswahrnehmung berechnen.

Um auch den dreifachen Interaktionseffekt noch sinnvoll interpretieren zu können, wurde der Wertebereich nach Stocké angepasst⁷³. Der Wertebereich des Anerkennungsbedürfnisses liegt auf einer Skala zwischen 0 (kein Anerkennungsbedürfnis) und 1 (sehr starkes Anerkennungsbedürfnis). Die Erwünschtheitswahrnehmung kann einen Wert von -1 (negative Einstellungen sozial erwünscht) bis 1 (negative Einstellung sozial unerwünscht) annehmen. Die Anonymität wird dichotomisiert, wobei 0 für eine anonyme Beurteilung der Situation steht und 1 für die nicht-anonyme Wahrnehmung. Damit liegt der Wertebereich des SD-Anreizes (also des dreifachen Interaktionseffekts) zwischen -1 (starker Anreiz für Äußerung von Vorurteilen) und 1 (starker Anreiz für Verneinung von Vorurteilen) (vgl. Stocké 2004).

Schließlich wurde für das Framing Modell die Umfrageeinstellung sowie die Umfrageerfahrung als Maß für die Einstellungsstärke erfasst. Beide Operationalisierungen orientierten sich an den Messungen von Stocké und Langfeldt (2003). Während die Umfrageeinstellung als metrischer Index in die Analyse eingeht, wird die Umfrageerfahrung anhand der Anzahl der bisherigen teilgenommenen Umfragen dichotomisiert. Befragte, die an mehr als vier Umfragen teilgenommen hatten, wird dabei eine große Umfrageerfahrung zugesprochen. Gerade diese Messung stellte sich jedoch als problematisch heraus. Die Studierenden wurden gebeten, die Anzahl der durchgeführten Umfragen in den letzten 12 Monaten anzugeben. Aus den Kommentaren zur Umfrage sowie den offenen Angaben zeigte sich, dass sich viele Probleme und Unsicherheiten bei der Erinnerung und Benennung einer konkreten Zahl ergeben. Der Aufbau des Framing-Modells folgt den Beispiel von Stocké. Hierfür wird in den Regressionsmodell nicht mehr alle direkten Effekte des RC-Modells berücksichtigt, sondern ausschließlich der dreifach Interaktionseffekt als Indikator für den SD-Anreiz. Warum Stocké die Modelle nicht ineinander geschachtelt prüft, sondern diese Operationalisierung wählt, ist aus seinen Texten nicht ersichtlich. Um die Modelle jedoch möglichst genau nachprüfen zu können, folgt die Analyse dieser Vorgabe.

73 Eine genauere Übersicht über die Umcodierung findet sich in Tabelle A47 des Anhangs.

Die Messung der sozio-ökonomischen Kontrollvariablen (Alter, Geschlecht) folgt der Operationalisierung aus Kapitel 5.3.

5.4.3 Methode

Im ersten Schritt erfolgt die Anwendung der theoretischen Konstrukte von Marlow und Crowne (1964), Stocké (2004) und Skarbek-Kozietulska et al. (2012) sowie des kombinierten Modells⁷⁴. Die Auswahl der Analyseverfahren orientiert sich an den vorgestellten Studien, bei denen die Überprüfung der Hypothesen über ein multiples Regressionsmodell erfolgte. Folgende Tabelle gibt eine Übersicht, welche Variablen in den jeweiligen Modellen berücksichtigt werden.

Tabelle 18: Vergleich der RC-Modelle bezüglich der erklärenden Variablen

Variablen	Erklärungsmodelle			
	Marlow und Crowne	Stocké	Skarbek-Kozietulska et al.	Lischewski
Direkte Variablen				
Anerkennungsbedürfnis (AB)	X	X	X	X
Erwünschtheitswahrnehmung (EW)	-	X	X	X
Anonymität (A)	-	X (dichotom)	-	X (ordinal)
Interaktionseffekte				
EW*A	-	X	-	X
AB*EW	-	X	X	X
AB*A	-	X	-	X
EW*A*AB	-	X	-	X
Framing-Variablen	-	X	-	X

Abgesehen von dem MC-Modell, erfolgt in jedem Modell eine schrittweise Regression, bei der zunächst nur die direkten Variablen und anschließend die Interaktionseffekte berücksichtigt werden. Die Darstellung der Ergebnisse der Framing-Variablen erfolgt gesondert. Wie bereits erwähnt, sollen die Modelle in Bezug auf zwei Dimensionen variiert werden: die Vorurteilsitems sowie die Messung des Anerkennungsbedürfnisses. Jedes der vier Modelle wird also in verschiedenen Varianten getestet, die in folgender Tabelle dargestellt sind:

⁷⁴ Im Anhang (A48) sind nochmals die Operationalisierungen im Vergleich zusammenfassend dargestellt.

Tabelle 19: Operationalisierungsmöglichkeiten und Modellvarianten

Anerkennungs- bedürfnisses	Abhängige Variable			
	Fremden- feindlichkeit	Islamophobie	Antisemitismus	Negative Einstellungen gegenüber Arbeitslosen
SES-17 Skala	Variante 1	Variante 2	Variante 3	Variante 4
Kurzskala SES-17	Variante 5	Variante 6	Variante 7	Variante 8
BIDR-Skala Fremden- feindlichkeit	Variante 9	Variante 10	Variante 11	Variante 12

Für die 4 Modelle aus Tabelle 18 werden also jeweils 12 Varianten berechnet, woraus sich insgesamt 48 Regressionen ergeben. Bei allen Regressionsmodellen werden die Kontrollvariablen Geschlecht und Alter, wie bei Stocké (2004), linear berücksichtigt⁷⁵. Aus Gründen des Umfangs und der Lesbarkeit sollen nicht alle 48 Regressionen einzeln besprochen, sondern vielmehr zusammenfassend ein Bild über die Ergebnisse gegeben werden. Um die Ergebnisse systematisch darzustellen, werden im ersten Schritt die Varianten eins bis vier, im zweiten Schritt fünf bis 8 und zuletzt neun bis zwölf berichtet.

⁷⁵ Die Variable Bildung stellt auf Grund der Stichprobe eine Konstante da und wird deshalb nicht zusätzlich kontrolliert.

5.4.4 Ergebnisse

5.4.4.1 Deskriptive Ergebnisse

In Tabelle 20 sind zunächst die Mittelwerte und Standardabweichung der abhängigen Variablen sowie der zentralen Einflussfaktoren wiedergegeben.

Tabelle 20: Deskriptive Werte der abhängigen und unabhängigen Variablen

Variable	Skala	Mittelwert	Standardabweichung	N
Anerkennungsbedürfnis¹				
SES-17	0 bis 1	0,635	0,117	664
Kurzversion SES-17	0 bis 1	0,585	0,162	662
BIDR – Fremdtäuschung	0 bis 1	0,557	0,214	659
Vorurteilsmessungen²				
Antisemitismus	1 bis 4	3,877	0,366	665
Fremdenfeindlichkeit	1 bis 4	3,547	0,631	681
Islamophobie	1 bis 4	3,821	0,422	675
Abl. Von Langzeitarbeitslosen	1 bis 4	2,680	0,773	651
Erwünschtheitswahrnehmung³				
Antisemitismus	-1 bis +1	0,654	0,441	661
Fremdenfeindlichkeit	-1 bis +1	0,168	0,570	663
Islamophobie	-1 bis +1	0,365	0,539	671
Abl. Von Langzeitarbeitslosen	-1 bis +1	-0,126	0,568	672
Anonymität⁴				
Anonymität ⁴ (dichotom)	0 und 1	0,122	0,328	622
Anonymität ⁴ (ordinal)	0 bis 1	0,316	0,265	622
Umfrageeinstellung⁵				
Umfrageeinstellung ⁵	1 bis 4	2,851	0,502	611
Umfrageerfahrung⁶				
Umfrageerfahrung ⁶	0 und 1	0,262	0,440	602

1) 0 "niedriges Anerkennungsbedürfnis" 1 "hohes Anerkennungsbedürfnis"

2) 1 "hohe Zustimmung zu negativen Items" 4 "keine Zustimmung zu negativen Items"

3) -1 "negative Einstellung erwünscht" +1 "negative Einstellung nicht erwünscht"

4) 0 "hohe Anonymität" 1 "niedrige Anonymität"

5) 1 "negative Umfrageeinstellung" 4 "positive Umfrageeinstellung"

6) 0 "Teilnahme an höchstens 4 Befragungen" 1 "Teilnahme an mehr als 4 Befragungen"

Bei den Vorurteilsitems zeichnen sich auch für diese Stichprobe die erwarteten Unterschiede zwischen den Minderheiten ab. Vorurteile gegenüber Juden und Muslimen wurden von den Studierenden wesentlich seltener bejaht als die Fremdenfeindlichkeitsitems. Am negativsten ist die Einstellung gegenüber Arbeitslosen, die mit einem Mittelwert von 2.7 auch nur knapp über

der Mittelkategorie liegen. Sie stellen damit die Gruppe mit den meisten negative Bewertung dar. Diese Unterschiede zeigen sich auch in Hinblick auf die Tabuisierung. Die Langzeitarbeitslosen sind die einzige Gruppe, bei der die Studierenden eine negative Einstellung als erwünscht wahrnehmen. Die stärkste Erwünschtheitswahrnehmung ist erwartungsgemäß für den Antisemitismus zu verzeichnen, gefolgt von Vorurteilen gegenüber Muslimen und Ausländern. Der Mittelwert bei der Fremdenfeindlichkeit scheint mit 0.168 überraschend gering auszufallen. Die wahrgenommene Anonymität hat einen Mittelwert von 0.122. Die Onlinebefragung wurde damit größtenteils als anonym eingeschätzt. Deutlich wird hier aber auch, dass diese Einschätzung nicht alle Befragten teilen. Immerhin 12% der Studierenden schätzten die Situation als nicht anonym ein. Ein Grund könnte hierfür möglicherweise das persönliche Anschreiben per E-Mail sein, mit dem die Studierenden kontaktiert wurden. Die Umfrageeinstellung ist tendenziell eher positiv ausgeprägt. In Bezug auf die Umfrageerfahrung zeigt sich hingegen, dass 74% der Befragten an mehr als vier Befragungen in den letzten 12 Monaten teilgenommen haben.

5.4.4.2 Multivariate Ergebnisse

Die Ergebnisse zu den Modellen werden im Anhang Tabellen A31 bis A34 dargestellt.

5.4.4.2.1 Ergebnisse der RC-Modelle mit der SES-17 Skala als Messung des Anerkennungsbedürfnisses

Fremdenfeindlichkeit

In dem MC-Modell zur Erklärung der Fremdenfeindlichkeit findet sich ein schwacher positiver Effekt (9% Irrtumswahrscheinlichkeit) der SES-17 Skala d.h., mit steigendem Anerkennungsbedürfnis äußerten die Studierenden weniger Vorurteile gegenüber Ausländern. Die erklärte Varianz ist mit 0.7% jedoch sehr gering. Berücksichtigt man zusätzlich die Variablen Anonymität und/oder Erwünschtheitswahrnehmung, steigt die Erklärungskraft deutlich an. Die Hinzunahme der Interaktionseffekte bringt hingegen nicht in jedem Modell eine Verbesserung. Während sich die Erklärungskraft der Modelle von Stocké und dem kombinierten Modell mit den Interaktionsvariablen deutlich verbessern, nimmt das adj. R^2 bei dem Konstrukt von Skarbek-Kozietulska et al. ab. Die Erwünschtheitswahrnehmung ist dabei aber über alle Modelle hinweg die einzige signifikante Erklärungsgröße: Je stärker die negative Äußerung gegenüber Ausländern als sozial unerwünscht eingeschätzt wird, umso weniger äußern die Befragten beim gleichen Item selbst eine negative Einstellung. Anonymität und die SES-17 Skala haben hingegen in keinem Modell als direkte Variablen einen signifikanten Effekt. Unter Hinzunahme der Interaktionseffekte in den Modellen mit Anonymität wird der direkte Effekt der Erwünschtheitswahrneh-

mung etwas schwächer. Dafür ist nun auch die Anonymität signifikant. Allerdings widerspricht die Zusammenhgangsrichtung den theoretischen Vermutungen. Bisherige Studien zum Einfluss von Anonymität gehen davon aus, dass die Befragten mit steigender Anonymität eher dazu bereit sind, Vorurteile zu äußern, da die Angst vor negativen Konsequenzen sinkt. In der vorliegenden Untersuchung besteht zwischen Anonymität und Fremdenfeindlichkeit jedoch der gegenteiliger Zusammenhang: je anonymier die Onlinebefragung wahrgenommen wird, umso eher werden positive Einstellungen geäußert. Zusätzlich findet sich ein signifikanter Interaktionseffekt zwischen Anonymität und SES-17 Skala. Um den Mechanismus hinter diesem Zusammenhang verständlicher zu machen, wird der zweite Modellschritt getrennt nach dem Grad der Anonymität dargestellt.

Tabelle 21: Lineare Regression des RC-Grundmodells auf die Fremdenfeindlichkeit – nach Grad der Anonymität getrennt

RC-Variablen	Modell 1 - hohe Anonymität		Modell 2 - geringe Anonymität	
	Beta	Sign.	Beta	Sign.
SES-17 ¹	0,096	0,033	-0,273	0,043
Erwünschtheitswahrnehmung ¹	0,098	0,029	-0,093	0,492
Konstante	3,320	0,000	4,078	0,002
adj. R ² (F-Test)	0,029	0,001	0,032	0,212
N	498		62	

1) 0 "niedriges Anerkennungsbedürfnis" 1 "hohes Anerkennungsbedürfnis"

2) 1 "hohe Zustimmung zu negativen Items" 4 "keine Zustimmung zu negativen Items"

Es wird deutlich, dass sich die einzelnen Effekte nach der Anonymitätswahrnehmung unterscheiden (vgl. Tabelle 21). Wird die Situation als anonym wahrgenommen, sinkt mit steigendem Anerkennungsbedürfnis die Tendenz Vorurteilen zu zustimmen. Ist die Situation hingegen für einen Befragten weniger anonym, wird den negativ formulierten Items mit steigendem Anerkennungsbedürfnis eher zugestimmt. Die Erwünschtheitswahrnehmung hat hingegen überraschenderweise nur bei hoher Anonymität einen signifikanten Effekt. Theoretisch war jedoch die Vermutung, dass die Kenntnis gesellschaftlicher Normen vor allem in nicht anonymen Situationen zum Tragen kommt, da die Befragten hier stärker das Gefühl haben für eine Abweichung von den Normen sanktioniert zu werden.

Antisemitismus

Das MC-Modell kann keinen signifikanten Beitrag zur Erklärung von antisemitischen Einstellungen leisten. Bei den erweiterten Modellen ist auch hier wieder die Erwünschtheitswahrnehmung die wichtigste direkte Variable. Die größte Erklärungskraft weisen jedoch auch hier die Interaktionsmodelle mit den Anonymitätsvariablen auf. Die Modelle zur Erklärung antisemiti-

scher Einstellungen zeigen sie den nach Esser und Stocké zu erwartenden dreifachen Interaktionseffekt. Ein positiver Wert bedeutet, dass bei den Befragten ein starker RC-Anreiz für positive Aussagen besteht. Ein negativer Wert hingegen spricht für einen starken Anreiz für negative Aussagen. Demnach lässt sich der Interaktionseffekt wie folgt interpretieren: Je stärker der Anreiz für eine positive Einstellung gegenüber Juden ist, desto eher werden negative Einstellungen geäußert. Wie lässt sich dieser paradoxe Zusammenhang erklären? Sowohl die SES-17 Skala als auch die Anonymität weisen entgegen der theoretischen Vermutung ein negatives Vorzeichen auf.

Tabelle 22: Vorhersagewerte¹ der Regressionanalyse des RC-Modells nach Stocké auf den Antisemitismus

Erwünschtheitswahrnehmung	Anerkennungsbedürfnis niedrig		Anerkennungsbedürfnis hoch	
	anonym	nicht anonym	anonym	nicht anonym
positive Einstellung erwünscht	3,66	3,57	3,64	3,43
negative Einstellung erwünscht	3,47	3,51	3,46	3,55

1) 1= hohe Zustimmung zu Vorurteilen, 4= keine Zustimmung zu Vorurteilen

Tabelle 22 zeigt, die Vorhersagewerte der Vorurteilsäußerungen für acht verschiedene Kombinationen von Eigenschaften. Dabei zeigt sich deutlich der paradoxe Effekt der Anonymität. Der Theorie nach müssten Personen, die die Situation als nicht anonym einschätzten und einen starken Anreiz zur sozial erwünschten Antwort haben, in ihren Vorhersagewerten höher liegen, als Befragte, die die Umfrage als anonym wahrnahmen. Die Ergebnisse zeigen jedoch den gegenteiligen Effekt. Erstaunlicherweise äußerten sich Personen mit hohem Anerkennungsbedürfnis und einer positiven Erwünschtheitswahrnehmung in der nicht-anonymen Situation negativer (3,43) gegenüber Juden als wenn Vorurteile (negative Einstellung) als erwünscht wahrgenommen wurden (3,55).

Islamophobie

Bezüglich der Islamophobie ist das Ergebnis für die theoretischen Modelle am verheerendsten. In keinen der Modelle findet sich ein signifikanter Effekt. Auch der negative adjustierte Determinationskoeffizient zeigt, die geringe Erklärungskraft der einzelnen Faktoren. Die Ursache für diese schlechte Erklärungskraft der Modelle in Bezug auf Muslime sind unklar. Ein möglicher Grund könnte in der Zusammensetzung der Minderheit liegen. Diese Gruppe grenzt sich durch ihren religiösen Glauben ab, während Ausländer aufgrund ihrer Herkunft und ihres rechtlichen Status, bzw. die Arbeitslosen wegen ihrer ökonomischen Lage diskriminiert werden. Es

ist möglich, dass das RC Modell kein allgemeines Modell von Vorurteilen darstellt, sondern nur negativen Stereotypen auf Grundlage der ethnischen Herkunft erklärt. In Bezug auf den Antisemitismus ist die Erklärungskraft deutlich höher. Es ist möglich, dass die Gruppe der Juden aufgrund der historischen Ereignisse keine rein religiöse Minderheit darstellt, sondern hier noch zusätzlich andere Abgrenzungsmerkmale hinzukommen.

Langezeitarbeitslose

Auch die negative Einstellung gegenüber Langzeitarbeitslosen wird nur unzureichend durch die Modelle erklärt. Bei einer Irrtumswahrscheinlichkeit von 5% finden sich auch hier keine signifikanten Effekte. Betrachtet man doch tendenzielle Einflüsse auf einem 10% Niveau, zeigen sich interessante Muster. In dem Interaktionsmodell besitzt das Anerkennungsbedürfnis einen schwachen negativen Einfluss. Das heißt, je größer das Anerkennungsbedürfnis ist, umso eher werden negative Einstellungen gegenüber Arbeitslosen geäußert. Der Zusammenhang ist demnach auf den ersten Blick genau entgegengesetzt zur theoretischen Vermutung, die unterstellt, dass Vorurteile als sozial unerwünscht wahrgenommen werden. Es ist möglich, dass diese Voraussetzung bei der Gruppe der Langzeitarbeitslosen nicht gegeben ist. Des Weiteren finden sich schwache Interaktionseffekte zwischen Anonymität und Erwünschtheitswahrnehmung sowie dem dreifach Erklärungsfaktor. Allerdings ist auch hier die Zusammenhgangsrichtung entgegen der theoretischen Vermutung. Eine Ursache könnte sein, dass Langzeitarbeitslose ebenso wie Muslime keine ethnische Minderheit darstellen, sondern vielmehr eine aufgrund von ökonomischen Bedingungen benachteiligte Gruppe. Es ist durchaus möglich, dass der RC Ansatz auf diese Gruppen nicht übertragbar ist, da die Sensitivität des Themas wesentlich geringer und in einem Modell zur Erklärung des SD-Bias ins Leere läuft. 15% der Befragten finden negative Äußerungen gegenüber Arbeitslosen weder sozial erwünscht noch unerwünscht. Die größte Gruppe stellen jedoch jene Personen, die negative Einstellungen gegenüber Arbeitslosen für gesellschaftlich erwünscht halten (26%). Bei einer solchen Verteilung ist es durchaus nicht überraschend, dass die SES-17 Skala zu paradoxen Ergebnissen führt, da sie das Anerkennungsbedürfnis misst. Wenn jedoch in Bezug auf Arbeitslose keine Anerkennung oder Ablehnung zu erwarten ist, sollte die Skala auch keinen Beitrag zur Varianzaufklärung leisten. Erst der Interaktionseffekt aus Erwünschtheitswahrnehmung und Anonymität führt zur Verstärkung des eigentlich nicht signifikanten Effektes. Die SES-17 Skala beeinflusst die Einstellung gegenüber Langzeitarbeitslosen, wenn die Befragten die Situation als anonym wahrnehmen und negative Einstellungen für sozial erwünscht halten. Der überraschende Effekt der Anonymität wurde bereits diskutiert. Ihn außen vorgelassen, ist der negative Einfluss der Skala nicht mehr paradox. Es ist also

davon auszugehen, dass die negativen Effekte vor allem auf der „gedrehten“ Erwünschtheit beruhen, denn die Arbeitslosen sind die einzige Gruppe, bei der die negativen Äußerungen nicht tabuisiert, sondern anscheinend von Studierenden sogar als erwünscht wahrgenommen werden.

Fazit

Zusammenfassend lässt sich Folgendes festhalten: Insgesamt finden sich keine systematischen Ergebnisse, die für eine Bestätigung der einzelnen theoretischen Modelle sprechen. Sowohl das Modell von Marlow und Crowne, als auch von Stocké, Skarbek-Kozietulska et al. und die kombinierte Variante lieferten entweder nur bezüglich einer Vorurteilsgruppe signifikante Ergebnisse oder Effekte, die den theoretischen Vermutungen widersprachen. Das Modell von Marlow und Crowne zeigt dabei die geringste Erklärungskraft. Bei Hinzunahme der direkten Faktoren Anonymität und Erwünschtheitswahrnehmung verbessert sich der Modellfit deutlich. Die Ursache hierfür liegt hauptsächlich in dem Einfluss der Erwünschtheitswahrnehmung. Für Fremdenfeindlichkeit und Antisemitismus stellt dieser Faktor die stabilste Erklärungsgröße dar. Dies zeigt erneut, wie wichtig die Berücksichtigung der individuellen Wahrnehmung sozialer Normen ist, um das Anerkennungsbedürfnis und die Verzerrung durch den SD-Bias zu erklären. Die größte Erklärungskraft weisen allerdings die Interaktionsmodelle von Stocké und der kombinierten Variante auf, auch wenn nur wenige der Interaktionseffekte überhaupt signifikant sind. Dabei ist die erklärte Varianz in dem kombinierten Modell etwas größer. Die größere Varianz der Anonymitätsvariable scheint demnach eine Verbesserung darzustellen.

Vergleicht man die theoretischen Modelle untereinander, so wird deutlich, dass die Modelle mit Anonymität vorzuziehen sind. Sowohl mit der Operationalisierung der Anonymität als dichotome als auch als ordinale Variable können die Modelle die negative Einstellung gegenüber Minderheiten besser erklären. Schlechter schneidet das Modell von Skarbek-Kozietulska et al. ab. Einzig die Erwünschtheitswahrnehmung zeigte hier als direkter Effekt einen signifikanten Einfluss. Der Interaktionseffekt aus Anerkennungsbedürfnis und Erwünschtheitswahrnehmung war in keinem Modell relevant. Dieses Ergebnis deckt sich mit den Analysen von Skarbek-Kozietulska et al. (2012), denn auch hier fanden die Autoren keine Interaktionseffekte. Sie schlussfolgerten, dass nicht die Interaktionseffekte, sondern die direkten Einflüsse relevant für das Modell sind und verwiesen auf das Modell von Marlow und Crowne, dass auch nur direkte Effekte kennt. Skarbek-Kozietulska et al. berücksichtigen aus diesem Grund die Variable Anonymität überhaupt nicht. Diese Interpretation ist jedoch vor dem Hintergrund der hier dargestellten Ergebnisse fraglich.

Bei den vorliegenden Daten handelt es sich ebenfalls um eine anonyme Erhebung. Der theoretische Hintergrund besagt, dass die Privatheit der Situation umso größer ist, je weniger Kommunikationskanäle eingesetzt werden. Dies hat zur Folge, dass gerade bei nicht-persönlichen Interviews die Befragten weniger das Gefühl haben, dass ihre Antworten identifiziert und ihnen zugeordnet werden können. Außerdem fällt die direkte persönliche Reaktion auf die gegebene Antwort weg. Dies sollte zur Folge haben, dass die Hemmung sozial unerwünschte Antworten zu geben, geringer ist, als in persönlichen Interviews. Wie bereits erwähnt, wird diese Annahme bisher über experimentelle Designs in Form von Modi-Unterschieden überprüft. Hier bestätigen sich die Annahmen größtenteils. Die wahrgenommene Anonymität wurde bisher hingegen noch nicht berücksichtigt.

Tabelle 23: Prozentuale Verteilung der wahrgenommenen Anonymität in der SST-Stichprobe

Anonymität	
sehr anonym	28,30 %
	32,48 %
	27,01 %
	9,16 %
überhaupt nicht anonym	3,05 %

Die Onlinebefragung wurde tendenziell eher als anonyme Situation wahrgenommen (vgl. Tabelle 23). Es gibt jedoch einen nicht geringen Anteil an Befragten (12%), die, entgegen der theoretischen Vermutung, das Interview als nicht anonym einschätzen. Sicherlich bedeutet dies nicht automatisch, dass die angenommene Rangfolge der Modi in ihrem Grad an Anonymität fraglich ist. Es ist durchaus möglich, dass der Mittelwert über die Erhebungsmethoden tatsächlich der theoretischen Vermutung entspricht. Fraglich werden durch diese Differenzierung vielmehr die strengen Kategorien von Stocké (2004) als auch die nicht Berücksichtigung des Faktors durch Skarbek-Kozietulka et al..

Allerdings liefert die wahrgenommene Anonymität paradoxe Ergebnisse. Vor allem in Verbindung mit der Täuschungsskala zeichnet sich ein negativer Zusammenhang mit der Fremdenfeindlichkeit ab (siehe Anhang A31). Der Effekt ist dabei genau entgegengesetzt der theoretischen Vermutung.

Die Studie von Whelan und Maede (2009) bietet eine mögliche Erklärung. Sie untersuchten, wenn auch nur bivariat, den Zusammenhang zwischen sensitiven Items und der wahrgenommenen Erwünschtheit. Ihre Analysen zeigten, dass nicht die subjektive Anonymität auf die sensitiven Items wirkte, sondern die Beantwortung der sensitiven Fragen einen Einfluss auf die wahrgenommene Erwünschtheit hatte. Je häufiger unangenehme Eigenschaften und Verhal-

tensweisen von den Befragten bejaht wurden, umso weniger anonym wurde die Situation eingeschätzt. Die Autoren begründeten diesen Effekt damit, dass Personen die sozial Unerwünschtes preisgeben, sich ihrer Identifizierbarkeit bewusster werden und die Wahrscheinlichkeit höher einschätzen erkannt zu werden, da sie zu einer besonderen Gruppe zählen. Diese Interpretation hätte zur Folge, dass je mehr Vorurteile geäußert werden, der Grad an wahrgenommener Anonymität abnimmt. Dieser Zusammenhang deckt sich mit den in der vorliegenden Analyse gefundenen Ergebnissen. Während bei der Operationalisierung der Anonymität über Modus Unterschiede ein positiver Effekt vermutet werden kann, da die Messung experimentell und damit unabhängig vom Subjekt vorgenommen wird, ist bei der wahrgenommenen Anonymität der Einfluss eventuell negativ. Auch würde sich in diesem Fall die Messung nicht für das Rational Choice Modell eignen, da die Anonymität nicht mehr unabhängige, sondern abhängige Variable wäre. Da die Anonymität am Ende des Fragebogens, d.h. nach den verschiedenen Vorurteilsitems erhoben wurde, scheint diese Interpretation der gegenteiligen Kausalrichtung auch in der vorliegenden Studie zutreffend. Fraglich ist, ob die Abfrage der Anonymität zu Beginn des Fragebogens andere Effekte produzieren würde. Denn wenn nicht, würde das bedeuten, dass die subjektive Anonymität innerhalb des RC-Modells nicht anwendbar ist. Ohne weitere Forschung zu diesem Gebiet lässt sich jedoch nicht abschließend klären, ob die subjektive Anonymität tatsächlich entgegengesetzt zu dem Modi Effekt wirkt und ob sich dieses Ergebnis auch für persönliche Interviews wiederholen lässt.

Der Vergleich der einzelnen Vorurteilsgruppen untereinander zeigt, dass es auch hier große Unterschiede zwischen der Erklärungskraft der einzelnen Modelle gibt. Die größte erklärte Varianz findet sich, über die drei theoretischen Modelle hinweg, bei der Fremdenfeindlichkeit. Da mit dem Modell ausschließlich ein systematischer Fehler kontrolliert werden soll, ist der vergleichsweise geringe Prozentsatz nicht sonderlich überraschend. Im Gegensatz dazu sind die Modelle völlig ungeeignet die Vorurteile gegenüber Muslimen zu erklären. Den theoretischen Erwartungen im Sinne des dreifachen Interaktionseffektes entsprechen hingegen die Ergebnisse zum Antisemitismus, allerdings entgegen der vermuteten Zusammenhangsrichtung. Die negativen Einstellungen gegenüber Arbeitslosen lassen sich ebenfalls nur bedingt durch die Modelle erklären. Nur bei einer Irrtumswahrscheinlichkeit von 10% lassen sich hier Tendenz zu einem Antwortverhalten nach den Interaktionseffekt finden. Allerdings auch hier entgegen der theoretisch vermuteten Zusammenhangsrichtung. Was sind die Ursachen für das Scheitern der theoretischen Modelle?

5.4.4.2.2 Ergebnisse der RC-Modelle mit der überarbeiteten SES-17 Skala als Messung des Anerkennungsbedürfnisses

Eine mögliche Schwäche liegt in der Messung des Anerkennungsbedürfnisses. In Kapitel 5.3 wurde festgehalten, dass gerade die SES-17 Skala empirisch den theoretischen Erwartungen nicht standhält. Gerade die Eindimensionalität ist infrage zu stellen. Es ist also durchaus denkbar, dass die fehlenden signifikanten Effekte des Anerkennungsbedürfnisses auf der schlechten Reliabilität der Skala beruhen. Aus diesem Grund werden die Modelle mit einer überarbeiteten SES-17_k Skala berechnet. Die Items der Skala wurden auf Grundlage einer explorativen Faktorenanalyse ausgewählt und die Eindimensionalität anschließend zusätzlich mit der konfirmatorischen Analyse überprüft. Welche Verbesserungen lassen sich nun für die Modelle mit dieser verbesserten Skala festhalten?

Die Ergebnisse zu den Modellen werden im Anhang Tabellen A35 bis A38 dargestellt.

Fremdenfeindlichkeit

Bezüglich der Erklärung fremdenfeindlicher Einstellung besitzt die Kurzversion der SES-17_k Skala über alle Modelle einen signifikanten positiven Effekt. Mit dieser Skala weist nun auch das MC-Modell eine signifikante Erklärungskraft auf. Die Ergebnisse des Konstrukts von Skarbak-Kozietulska et al. verbessern sich ebenfalls deutlich. Der Anteil der erklärten Varianz des Modells von Stocké konnte von 2.7% auf 5.2% erhöht werden. Damit stellt das RC-Modell von Stocké bezogen auf Fremdenfeindlichkeit insgesamt das Modell mit der größten Erklärungskraft dar. Auch in der Originalstudie von Stocké (2004) wurde das Modell an Items der Fremdenfeindlichkeit getestet. Im Gegensatz zu Stockés Ergebnissen ist jedoch auch in diesem Modell der entscheidende dreifach-Interaktionseffekt nicht signifikant. Überraschend sind die Ergebnisse bezüglich des kombinierten Modells. Mit der Kurzversion der SES-17_k Skala sinkt der hier der zuvor noch signifikante Effekt aus Anonymität und Anerkennungsbedürfnis auf eine Irrtumswahrscheinlichkeit von 8.5%.

Antisemitismus

Bezogen auf den Antisemitismus zeigt sich keine Verbesserung der Modelle durch die neue Skala. Die Erklärungskraft der einzelnen Modelle verschlechtert sich sogar. Dies zeigt sich z.B. bei den Interaktionseffekten in der kombinierten Variante, die mit der Kurzversion der SES-17_k Skala nicht mehr signifikant sind.

Islamophobie

Mit der verkürzten Skala findet sich bei dem Modell von Stocké zur Islamophobie ein signifikanter dreifacher Interaktionseffekt. Die Erklärungskraft des Gesamtmodells ist jedoch nicht ausreichend ($F\text{-Test}=0.197$), um den Effekt interpretieren zu können. Bezüglich der anderen Modelle finden sich keine Verbesserungen.

Langzeitarbeitslose

Interessant ist hingegen die Veränderung bei den Modellen zu den Langzeitarbeitslosen. Bei der original SES-17 Skala zeigte das Anerkennungsbedürfnis einen schwachen negativen Effekt. Dieser negative Effekt bedeutet, dass mit steigendem Anerkennungsbedürfnis die negativen Äußerungen gegenüber Arbeitslosen zunehmen. Diese Zusammenhangsrichtung ist damit entgegen der theoretischen Vermutung. Mit der Kurzsкала findet sich hingegen ein signifikant positiver Effekt. Der direkte Einfluss des Anerkennungsbedürfnisses verschwindet jedoch, sobald die Interaktionseffekte berücksichtigt werden. Stattdessen ist nun bei dem Modell mit Anonymität der dreifach Interaktionseffekt theoriekonform signifikant. Der negative Effekt bedeutet jedoch, dass Befragte mit starkem Anreiz zur als erwünscht wahrgenommenen positiven Aussagen, häufiger negative Einstellungen gegenüber Arbeitslosen äußerten. In Tabelle 24 sind die einzelnen Vorhersagewerte berechnet.

Tabelle 24: Vorhersagewerte¹ der Regressionanalyse des RC-Modells nach Stocké auf negative Einstellungen gegenüber Langzeitarbeitslosen

Erwünschtheits- wahrnehmung	Anerkennungsbedürfnis niedrig		Anerkennungsbedürfnis hoch	
	anonym	nicht anonym	anonym	nicht anonym
positive Einstellung erwünscht	3,68	3,58	3,68	3,38
negative Einstellung erwünscht	3,81	3,87	3,92	3,97

1) 1= hohe Zustimmung zu Vorurteilen, 4= keine Zustimmung zu Vorurteilen

Die Vorhersagewerte zeigen, den paradoxen Effekt. Personen, die die negative Einstellung gegenüber Arbeitslosen für erwünscht hielten, äußern sich im Durchschnitt positiver als Befragte mit positiver Erwünschtheitswahrnehmung.

Fazit

Die größte Veränderung bezüglich der Erklärungskraft bringt die Kurzversion der SES-17_k Skala bei der Erklärung von Fremdenfeindlichkeit. Hier zeigt sich über alle Modelle zusammen mit der Erwünschtheitswahrnehmung ein signifikanter Effekt. Diese Verbesserung ist jedoch

nicht über alle Vorurteile stabil. Während die Ergebnisse bei den Modellen zur Islamophobie fast unverändert bleiben, verschlechtert sich die Erklärungskraft bezüglich des Antisemitismus sogar. Bei der Erklärung von negativen Einstellungen gegenüber Langzeitarbeitslosen produziert die neue Skala hingegen Effekte entgegengesetzt zur theoretischen Vermutung. Insgesamt machen die Ergebnisse auch hier deutlich, dass die einzelnen Modelle am besten bezüglich der Fremdenfeindlichkeit funktionieren. Der Vergleich mit den Analysen der vollständigen SES-17_k Skala zeigt, wie empfindlich die Modelle und insbesondere die Interaktionseffekte auf unterschiedliche Operationalisierungen reagieren.

5.4.4.2.3 Ergebnisse der RC-Modelle mit der FT-Skala als Messung des Anerkennungsbedürfnisses

Wie bereits thematisiert, entspricht die theoretische Struktur, d.h. die Eindimensionalität der SES-17 Skala nicht den empirischen Befunden. Als stabileres Messinstrument haben sich in Kapitel 5.1 die Subskalen der BIDR-Skala erwiesen. Während die Fremdtäuschungsskala ebenso wie die SES-17 und MC-Skala als Messung der Täuschungstendenz gegenüber anderen konzipiert wurde, soll die ST-Skala eher die Neigung zur Selbsttäuschung. Für beide Skalen sind daher unterschiedliche Effekte zu erwarten. Der Effekt der FT-Skala müsste den bereits berichteten ähneln. Aus diesem Grund werden die Modelle nochmals mit der FT-Skala als Messung des Anerkennungsbedürfnisses berechnet. Die Ergebnisse zu den Modellen werden im Anhang Tabellen A39 bis A42 dargestellt.

Wie auch schon bei der verkürzten SES-17 Skala besitzen die Modelle zur Erklärung der Fremdenfeindlichkeit die höchste Erklärungskraft. Das Anerkennungsbedürfnis weist in allen Modellen einen theoriekonformen positiven Zusammenhang auf. Die erklärte Varianz liegt jedoch deutlich unter den Modellen mit der Kurzversion der SES-17 Skala. Vor allem die Interaktionseffekte unterscheiden sich zwischen beiden Operationalisierungen. Während die verkürzte SES-17 Skala in Verbindung mit der dichotomisierten Messung der Anonymität einen signifikanten Interaktionseffekt in dem Modell von Stocké hat, verschwindet der Interaktionseffekt bei der FT-Skala. Dafür ist der Interaktionseffekt zwischen FT und ordinaler Anonymität in dem eigenen Modell signifikant. Diese Unterschiede zeigen deutlich, wie anfällig die theoretischen Modelle auf unterschiedliche Operationalisierungen reagieren. Die für die Theorie zentralen Interaktionseffekte hängen stark von der Messung der einzelnen Konstrukte ab und das, obwohl SES-17 und FT Skala angeblich beide die Täuschungstendenz erfassen und bei der Messung der Anonymität nur die Skalierung verändert wurde.

Mit der Messung über die Fremdtäuschungsskala fällt der signifikante Einfluss des dreifachen Interaktionseffektes in dem modifizierten Modell weg. Ob sich der Rational-Choice Ansatz bestätigen lässt oder nicht, hängt also auch hier wieder zentral von der Messung des Anerkennungsbedürfnisses ab. Bezogen auf die Islamophobie zeigt sich keine Verbesserung der Modelle durch die neue Skala.

Bei der Erklärung der negativen Einstellung gegenüber Arbeitslosen verschwinden bei dem Einsatz der FT-Skala die Interaktionseffekte. Stattdessen finden sich gerade bei dieser Minderheit plötzlich signifikante Einflüsse des Alters und des Geschlechts. Je älter die Befragten sind, umso weniger Vorurteile gegenüber Arbeitslosen werden geäußert. Gleichzeitig stimmten Frauen den negativen Stereotypen eher zu als Männer.

5.4.4.2.4 Ergebnisse der Frame-Selektion-Modelle mit der überarbeiteten SES17-Skala als Messung des Anerkennungsbedürfnisses

Stocké erweiterte wie in Kapitel 3.2.6 beschrieben sein Rationa-Choice Modell um die Framingvariable Befragtenrolle. Seiner Theorie zugrunde müsste ein dreifacher Interaktionseffekt aus Anreiz zur sozial erwünschten Antwort (=Anerkennungsbedürfnis*Erwünschtheitswahrnehmung*Anonymität), Umfrageeinstellung und Umfrageerfahrung einen signifikanten Erklärungsbeitrag in den Modellen leisten. Auf Grundlage der bisherigen Ergebnisse sollen im Folgenden nicht alle Modelle um die Framingvariablen erweitert werden. Das Modell von Skarbek-Kozietulska et al. zeigte die schlechtesten Ergebnisse und wird deshalb nicht mehr berücksichtigt. Hingegen wird sowohl das Modell von Stocké als auch das modifizierte Modell mit den zusätzlichen Variablen zur Umfrageeinstellung und Erfahrung für alle Vorurteilsgruppen dargestellt. Als Messung für das Anerkennungsbedürfnis wird ebenfalls auf Grundlage der bisherigen Ergebnisse die Kurzversion der SES-17 Skala ausgewählt. Die Ergebnisse der Modelle sind im Anhang Tabelle A43 bis A46 dargestellt.

Die Tabellen im Anhang (A43 bis A46) zeigt, dass die Framingvariablen nur in Bezug auf die Fremdenfeindlichkeit signifikante Ergebnisse vorweisen. Während die direkten Effekte keine Rolle bei der Erklärung der Einstellung gegenüber Ausländern spielen, stellt der Interaktionseffekt aus Umfrageeinstellung und SD-Anreiz den stärksten Faktor dar. Die Vorhersagewerte in Tabelle 25 und 26 verdeutlichen die Interpretation der Koeffizienten.

Tabelle 25: Vorhersagewerte¹ der Regressionanalyse des Frame-Slektion-Modells nach Stocké auf die Fremdenfeindlichkeit

Umfrageeinstellung	Anreiz zur negativen Aussage hoch	Anreiz zur negativen Aussage niedrig
Umfrageeinstellung positiv	4,07	3,98
Umfrageeinstellung negativ	4,24	4,07

1) 1=hohe Zustimmung zu Vorurteilen, 4=niedrige Zustimmung zu Vorurteilen

Tabelle 26: Vorhersagewerte¹ der Regressionanalyse des Frame-Slektion-Modells nach Stocké auf die Fremdenfeindlichkeit getrennt nach Umfrageerfahrung

Umfrageeinstellung	Anreiz zur negativen Aussage hoch		Anreiz zur negativen Aussage niedrig	
	Umfrageerfahrung niedrig	Umfrageerfahrung groß	Umfrageerfahrung niedrig	Umfrageerfahrung groß
Umfrageeinstellung positiv	4,12	4,04	4,09	3,88
Umfrageeinstellung negativ	4,18	4,26	4,11	4,01

1) 1=hohe Zustimmung zu Vorurteilen, 4=niedrige Zustimmung zu Vorurteilen

Der negative Interaktionseffekt der Umfrageeinstellung und des SD-Anreizes besagt, dass sich der Anreiz zur positiven Aussage gegenüber Ausländern verstärkt, wenn der Befragte eine negative Umfrageeinstellung besitzt. Dieser Zusammenhang ist theoriekonform, da Personen die den Nutzen von Umfragen als sehr hoch einschätzen, eher dazu neigen ihren ersten, ohne Rücksicht auf die Befragungssituation entstandenen, Antwortimpuls preis zugeben, auch wenn diese nicht sozial erwünscht ist. Überraschend in diesem Zusammenhang ist jedoch, dass sich derselbe Effekt auch bei Respondenten mit starkem Anreiz zur negativen Aussage zeigt. Hier äußerten Personen mit starkem Anreiz und positiver Umfrageeinstellung häufiger fremdenfeindliche Einstellungen, obwohl der Theorie nach, der Zusammenhang genau entgegengesetzt sein müsste. Ein ähnliches Paradox zeigt sich bei zusätzlicher Berücksichtigung der Umfrageerfahrung. Der Anreiz zur positiven Aussage müsste nach Stocké dann am niedrigsten sein, wenn die Umfrageeinstellung positiv ist und diese in Form von Umfrageerfahrung stark verankert ist. Diese Vermutung lässt sich bestätigen, denn in Tabelle 26 wird deutlich, dass der Vorhersagewert für diese Gruppe am geringsten ist. Bei dem Anreiz zur negativen Aussage müsste der Effekt genau entgegengesetzt sein d.h., bei derselben Gruppe müssten sich im Durchschnitt die positiveren Aussagen finden lassen. Dies trifft jedoch nicht zu. Es zeigt sich damit, dass sich das Framing Modell an den Daten nicht bestätigen lässt. Der dreifache Interaktionseffekt aus An-

reiz, Einstellung und Erfahrung hat keinen Einfluss auf die Fremdenfeindlichkeit⁷⁶. Bei den anderen Vorurteilsgruppen finden sich keinerlei signifikante Einflüsse der Variablen.

5.4.5 Zusammenfassung und Bewertung

Ziel dieses Kapitels war es, die theoretischen Modelle von Marlow/Crowne, Stocké und Skarbek-Kozietulska et al. sowie einer kombinierten Variante auf ihre Erklärungskraft und Stabilität hin zu testen. Hierfür wurden verschiedene Messungen des Anerkennungsbedürfnisses in Bezug auf unterschiedliche Vorurteilsgruppen angewendet. Die Ergebnisse für die theoretischen Modelle sind durchgehend unbefriedigend. Für keine Modellvariante konnten die theoretischen Erwartungen bestätigt werden. Am geringsten ist die Erklärungskraft des Modells von Marlow und Crowne. Die Variablen von Skarbek-Kozietulska et al. können zwar einen deutlich größeren Anteil der Varianz erklären, aber auch hier sind die Ergebnisse nicht zufriedenstellend. Interessant ist in diesem Zusammenhang, dass bei ihrem Modell, die Regression mit den direkten Variablen in der Mehrheit der Fälle tatsächlich, wie von den Autoren vermutet, die besten Ergebnisse liefert. Die Berücksichtigung des Interaktionseffektes konnte die Erklärungskraft nicht erhöhen. Betrachtet man das Modell isoliert von allen anderen Ergebnissen, könnte man damit der Schlussfolgerung der Autoren folgen. Wird hingegen die Anonymität als Variable berücksichtigt, zeigt sich der Nutzen der Interaktionsvariablen. Dabei ist zu berücksichtigen, dass nur in zwei der geschätzten Modelle der dreifache Interaktionseffekt (der nach Esser und Stocké die zentrale Erklärungsgröße darstellt) überhaupt ein signifikanter Faktor ist. In beiden Fällen entsprach die Richtung des Zusammenhangs jedoch nicht den theoretischen Vermutungen. Trotzdem scheint es auch in nicht-persönlichen Interviews sinnvoll zu sein die Anonymität als Variable in das Erklärungsmodell zu integrieren, denn die vorliegenden Ergebnisse haben deutlich gezeigt, dass die Anonymität deutlich zwischen den Respondenten variiert und die Variable selbst einen erheblichen Teil zur Erklärung des SD-Bias beitragen kann. Paradoxerweise lieferte die wahrgenommene Anonymität, entgegen der theoretischen Vermutung, jedoch genau umgekehrte Zusammenhänge. Es ist damit nicht abschließend geklärt, ob die wahrgenommene Anonymität tatsächlich als alternative Messung zu dem Modi-Experiment verwendet

76 Dieses Ergebnis ändert sich auch nicht, wenn das Modell anstelle der SES-17 Skala als Messung des Anerkennungsbedürfnisses mit der Fremdtäuschungs-Skala berechnet wird. Auch in diesem Modell hat nur der Interaktionseffekt aus Einstellung und Erfahrung einen signifikanten negativen Effekt. Auch alternative Operationalisierungen der Umfrageerfahrung mit verschiedenen Dichotomisierungen oder als metrische Variable, führten zu keiner Verbesserung des Modells.

werden kann. Die Ergebnisse weisen daraufhin, dass es entscheidend ist, zu welchem Zeitpunkt die Anonymität erfasst wird⁷⁷.

In Bezug auf das Anerkennungsbedürfnis liefert die stabilsten Ergebnisse die Kurzversion der SES-17 Skala, gefolgt von der FT-Skala. Bei Ersteren zeigen sich zusätzlich zu den direkten Effekten, signifikante Interaktionseffekte mit der Anonymität. Dieser Interaktionseffekt ist vor dem Hintergrund des RC-Ansatzes überraschend, da die Theorie eher für den Interaktionseffekt von Anerkennungsbedürfnis und Erwünschtheitswahrnehmung spricht. Der einzig stabile Einflussfaktor, für Fremdenfeindlichkeit und Antisemitismus, der sich über fast alle Modelle hinweg finden lässt, ist die Erwünschtheitswahrnehmung. Dieses Ergebnis zeigt noch einmal, besonders vor dem Hintergrund von Kapitel 5.3 und der bisherigen Umsetzung der sozialen Erwünschtheit im Rahmen von Bevölkerungsumfragen, wie wichtig die Kontrolle dieser Variable ist. Die Ergebnisse machten deutlich, dass die Unterschiede in der Erwünschtheitswahrnehmung in der homogenen Stichprobe sehr groß sind. Ohne Berücksichtigung dieser Differenzen erscheint die Interpretation der Täuschungsskalen nicht sinnvoll, was sich insgesamt am Bsp. der Langzeitarbeitslosen zeigt.

Die Unterschiede zwischen den Vorurteilsgruppen sind gravierend. Diese Unterschiede dürften sich jedoch der Theorie nach nicht ergeben, da das Modell eigentlich auf alle sensitive Einstellungsfragen übertragbar sein müsste. Während alle Modelle und Operationalisierungen bei der Erklärung islamophober Einstellungen scheitern, ist die Vorhersage für die Fremdenfeindlichkeit deutlich besser. Die Ergebnisse zu den Langzeitarbeitslosen und den Juden variieren hingegen sehr stark zwischen den Modellen und Messungen. Die Ursachen für diese Unterschiede zwischen Gruppen sind nicht eindeutig. Die Vermutung liegt nahe, dass sowohl die Merkmale der Vorurteilsgruppen als auch die Varianzen in der Sensitivität die Differenzen hervorrufen. Auch die Berücksichtigung von zusätzlichen Framingvariablen, Umfrageeinstellung und Umfrageerfahrung, bringt keine Verbesserung der Erklärungskraft.

Die Verallgemeinbarkeit der vorliegenden Ergebnisse werden durch zwei Aspekte stark limitiert: die Stichprobenzusammensetzung und die verwendeten Operationalisierungen. In der Stichprobe sind ausschließlich junge, hochgebildete Personen mit sozialwissenschaftlicher Orientierung vertreten. Die widersprüchlichen Ergebnisse könnten also auf diese selektive Zusammensetzung zurückzuführen sein. Stocké (2004) prüfte seinen Ansatz an einer wesentlich breiteren Stichprobe. Es liegt die Vermutung nahe, dass die Wirkung der Erklärungsgrößen nicht

⁷⁷ Aus diesem Grund könnten auch in der vorliegenden Arbeit die Analyse durch ein Endogenitätsproblem beeinflusst sein.

unabhängig von formaler Bildung ist. Gegen diese Einschränkung spricht jedoch, dass es bisher keine Einschränkungen zur Anwendbarkeit des RC Ansatzes gemacht wurden. Bei Zusammenhangsanalysen wird indes die Zusammensetzung der Stichprobe als nicht so bedeutsam angesehen, wie bei der Schätzung von Anteilen für größere Bevölkerungsschichten. Auch die Messungen wurden zum Teil an Studierendenstichproben entwickelt und sollte daher gerade bei dieser speziellen Gruppe am besten funktionieren.

Neben der Stichprobe muss auch die Messung der einzelnen Konstrukte als nicht unproblematisch angesehen werden. Während das Anerkennungsbedürfnis differenzierter erhoben wurde, ist die Erwünschtheitswahrnehmung vor dem Hintergrund aktueller Studien durchaus kritisch zu bewerten. Stocké und Hunkler (2004) konnten für sensitive Verhaltensfragen nachweisen, dass die angenommene Monotonie die hinter der hier verwendeten Skalierung steht, nicht immer zutreffend ist. Den Studierenden wurde in der vorliegenden Stichprobe eine negative Aussage präsentiert. Für diese sollten sie auf einer Fünfer-Skala einschätzen, für wie sozial erwünscht, bzw. unerwünscht sie die Aussage halten. Die Aussage, dass ein Vorurteil sozial unerwünscht ist, bedeutet, jedoch nicht im gleichen Zug, dass eine positive Aussage als sozial erwünscht gilt. Vielmehr erhält man bei dieser Operationalisierung nur eine Information für die negative Einstellung. Ebenfalls fraglich ist die Bezugsgruppe für die Erwünschtheitswahrnehmung. In dieser Erhebung wurde nach der gesellschaftlichen Einschätzung gefragt. Es ist aber möglich, dass die Erwünschtheitswahrnehmung in Bezug auf Freunde und Verwandte die zuverlässigere Messung gewesen wäre. Nichtsdestotrotz ist dieser Faktor der einzig stabile Prädiktor für das Antwortverhalten. Die paradoxen Ergebnisse bezüglich der Anonymität könnten sich auf die Operationalisierung zurückführen lassen. Die wahrgenommene Anonymität wurde im Rahmen des RC Ansatzes erstmals eingesetzt. Bisherige Erhebungen beschränken sich auf den experimentellen Vergleich von Modi und unterstellen dabei die Wahrnehmung. Die direkte Messung führt jedoch zu völlig anderen Ergebnissen und Zusammenhängen. Vergleichbare Analysen zeigen, dass die Zusammenhangsrichtung des Zusammenhangs bei der Messung der subjektiven Anonymität nicht eindeutig ist. Ohne weitere Forschung lässt sich jedoch nicht sagen, welche der beiden Operationalisierungsvarianten vorzuziehen ist, bzw. ob sich der Effekt der wahrgenommenen Anonymität auch in persönliche Interviews einstellen würde.

6 Zusammenfassung und Diskussion

Ehrlichkeit zählt bis heute zu den zentralen Werten in unserer Gesellschaft und spielt nicht nur in der Fremd-, sondern auch in der Selbstwahrnehmung von Personen eine wichtige Rolle. Wer würde nicht behaupten, den Anspruch an sich selbst zu haben, möglichst ehrlich durch das Leben zu gehen? Dennoch sehen wir uns oft mit Situationen konfrontiert, in denen die Ehrlichkeit mit anderen Werten kollidiert. Höflichkeit, Zurückhaltung, Harmoniebedürfnis oder einfach nur die Angst vor negativen Reaktionen sind dabei nur einige Beispiele. Diese Eigenheit zwischenmenschlicher Interaktion stellt im Rahmen sozialwissenschaftlicher Befragungen ein zentrales Problem der Methodenforschung dar. Während in alltäglichen Situationen eine Lüge hilft miteinander ohne ständige Konflikte auszukommen, stehen diese Handlungsziele in Befragungen dem Anspruch der Forscher an die Datenqualität entgegen. Gerade bei sensiblen politischen Themen wie Vorurteilen ist die fehlerfreie Schätzung ohne Verzerrungen in Richtung einer als sozial erwünscht bewerteten Antwort wichtig. Neben der rein methodischen Fragestellung ist daher das zentrale Ziel der Kontrolle des sozial erwünschten Antwortverhaltens, die unverzerrte Messung von Einstellungen, Merkmalen oder Verhaltensweisen.

In den letzten 50 Jahren hat sich ein breites Forschungsfeld zu diesem Thema entwickelt, das sowohl spezielle Befragungstechniken wie z.B. RandomisdeResponseTechnik oder das Bogus-Pipeline-Verfahren als auch komplexe Modelle zur nachträglichen Bereinigung der Daten umfasst. In der Breite der Forschungsarbeiten liegt jedoch auch das zentrale Problem: Viele Studien zum Themenfeld social desirability sind nur selten theoriegeleitet, die Ergebnisse äußerst heterogen und zum Teil unsystematisch. Ausgehend von dieser Problematik war das Ziel dieser Arbeit den theoretischen sowie empirischen Forschungsstand der Ursachen und Wirkungsweisen der sozialen Erwünschtheit im deutschen Sprachraum darzustellen und die zentralen Lücken sowie ihre Folgen zu konkretisieren. Als Fazit der vorliegenden Arbeit lassen sich einige Problemfelder bei der Erklärung sozial erwünschten Antwortverhaltens benennen. Schwächen bisheriger Studien betreffen vor allem die Begrifflichkeiten, die theoretischen Modellierungen und die Operationalisierungen der jeweiligen Konstrukte.

Begrifflichkeiten

In Kapitel 2 wurde die Definition des Phänomens „soziale Erwünschtheit“ erläutert. Die meisten Arbeiten, die sich mit der theoretischen Erklärung des SD-Bias im Rahmen des RC-Ansatzes beschäftigen, benennen als Grundlage ihrer Herleitungen die Testtheorie. Sozial er-

wünschte Antwortverzerrungen stellen demnach einen systematischen Fehler im Rahmen von Befragung zugunsten einer sozial erwünschten Norm dar. Diese Definition besitzt zwei Schwachstellen. Zum einen muss für eine genaue Beschreibung der sozial erwünschten Antwortverzerrung eine zusätzliche Definition darüber vorliegen, was genau unter einer sozial erwünschten Norm verstanden wird. Zum anderen ist die Existenz eines wahren Wertes nach der Testtheorie eine implizit verbundene Prämisse. Ob diese Bedingung jedoch für alle Fragetypen gegeben ist, bleibt fraglich. Für eine präzise Definition sind deshalb vier verschiedene Dimensionen notwendig: Zum einen ist es von Bedeutung, welche Bezugsgruppe für die Bewertung von sensitiven Fragen herangezogen wird (Bezugsgruppen Dimension). Zum anderen lässt sich das Phänomen auch nach seinem Ursprung (kognitive Dimension) und seiner Wirkungsweise (zeitliche Dimension) eingrenzen. Zusätzlich zeigen bisherige Ergebnisse, dass es ebenfalls von Bedeutung ist, auf welchen Sachverhalt sich der SD-Bias bezieht (inhaltliche Dimension). Dabei ist es besonders wichtig zwischen Verhaltens- sowie Einstellungsfragen zu differenzieren, da diese unterschiedliche Konsequenzen für die Existenz eines wahren Wertes zur Folge haben. Je nachdem wie diese vier Dimensionen interpretiert werden, ergeben sich viele verschiedene Ebenen des SD-Bias. Die bisherigen Analysen beschränken sich jedoch größtenteils auf die einfache Definition nach der Testtheorie, ohne den Geltungsbereich zu diskutieren. Dadurch ist es kaum möglich die Ergebnisse der Analysen sinnvoll miteinander zu vergleichen, da aus den Arbeiten nicht eindeutig hervorgeht, auf welche Phänomene im Rahmen des sozial erwünschten Antwortverhaltens sich die Autoren beziehen. Um die Ergebnisse sinnvoll zu interpretieren, ist es deshalb notwendig, möglichst präzise entsprechend der vier genannten Dimensionen eine Definition des SD-Bias vorzunehmen.

Theorie

Im Anschluss an die Definition wurde in Kapitel 3 der theoretische Forschungsstand zur Anwendung des Rational-Choice Modells auf das Befragtenverhalten dargestellt. Nach einer allgemeinen Einführung in den RC-Ansatz folgte die Auseinandersetzung mit sechs Erklärungsmodellen: das RC-Modell von Crowne und Marlow, Esser, Stocké, Skarbek-Kozietulska et al. und Wolter sowie das Framing-Modell von Esser und Stocké. Zunächst ist festzuhalten, dass es insgesamt nur sehr wenige Studien gibt, die sich überhaupt mit der theoretischen Basis des Phänomens beschäftigen. Das am häufigsten angewendete Modell stammt von Marlow und Crowne, die das Anerkennungsbedürfnis als einzige Einflussgröße benennen. Aktuelle Forschungen haben jedoch gezeigt, dass bei dieser Interpretation wichtige Rahmenbedingungen wie die Erwünschtheitswahrnehmung nicht berücksichtigt werden. Ein deutlich umfangreicheres Modell

findet sich bei Esser. Er leitet zwei zentrale Einflussfaktoren für die Erklärung sozial erwünschten Antwortverhaltens ab: Das Bedürfnis nach sozialer Anerkennung und das Bedürfnis nach persönlicher Identität. Beide Faktoren setzen sich entsprechend der Rational-Choice Theorie aus Nutzen und Wahrscheinlichkeiten zusammen. Das Bedürfnis nach sozialer Anerkennung ergibt sich aus dem Anerkennungsbedürfnis, der Anonymität und der Erwünschtheitswahrnehmung. Das Bedürfnis nach persönlicher Identität setzt sich hingegen aus dem Nutzen der persönlichen Identität und der Passgenauigkeit von Antwortoptionen zusammen. Bei der Präzisierung der letzteren Faktoren bleibt Esser sehr vage. Sein Ansatz wurde von Stocké aufgegriffen und weiterentwickelt. Sowohl Esser als auch Stocké gehen von einem strengen Interaktionseffekt zwischen den einzelnen Faktoren aus. Ein im Hinblick auf die multiplikative Verknüpfung modifiziertes Modell stellten Skarbek-Kozietulska et al. vor. Wolter hingegen erweiterte den Ansatz um weitere Kosten- und Nutzenfaktoren. Jedes der genannten Modelle besitzt theoretische Schwachstellen, die bisher nur unzureichend diskutiert sind.

Esser unterscheidet zwischen verschiedenen Typen von Normen. Während situative Normen vor allem von Interviewereffekten geprägt werden, beziehen sich kulturelle Normen eher auf die Gesellschaft. Während er die Anonymität als Bedingung für die situativen Normen benennt, konkretisiert er die Wirkung dieser Rahmenbedingung auf die kulturellen Normen nicht. Damit ist nicht klar, ob sich nach Esser auch in anonymen Befragungen ein SD-Bias zumindest aus gesellschaftlichen Erwünschtheitswahrnehmungen ergeben kann. Auch ist die Hierarchie zwischen den Normen nicht klar geregelt. Der größte Kritikpunkt bezieht sich jedoch auf die Ungenauigkeiten bei dem Bedürfnis der persönlichen Identität. Esser selbst testete sein theoretisches Modell nicht. Die Interpretation der Faktoren und ihre Operationalisierung im Sinne Essers ist dadurch erschwert.

Bei dem Modell von Stocké findet sich zumindest eine theoretische Konkretisierung dieses Bedürfnisses. Es setzt sich demnach aus dem Bedürfnis nach Authentizität, der Einstellungsstärke und der Passgenauigkeit der Antwortoptionen zusammen. Doch auch bei Stocké finden sich Lücken in der Theorie. Ohne dass er den Geltungsbereich begründet, beschränkt sich Stocké auf den SD-Bias im Rahmen von Interviewereffekten. Kulturelle Normen spielen bei ihm hingegen keine Rolle mehr. Ein weiterer Kritikpunkt ist die Erweiterung des RC-Modells von Esser durch Moderatorvariablen. Es ist fraglich, ob diese Variante des Modells überhaupt noch zu den RC-Modellen gezählt werden sollte. Auch die Operationalisierung der einzelnen Faktoren ist unpräzise. Stocké operationalisiert die Anonymität über einen dichotomen Modus-Effekt. Ob diese strikte Auslegung gerade unter der Bedingung des multiplikativen Zusammenspiels

sinnvoll ist, stellen Skarbek-Kozietulska et al. und auch Wolter infrage. Sie plädieren für ein Modell mit hauptsächlich direkten Faktoren und lassen in ihrem Aufsatz den strengen Interaktionseffekt fallen. Da ihre Analyse jedoch auf Verhaltensfragen basiert, während Stocké sein Modell für Einstellungsfragen entwickelte, war es auf Grundlage der bisherigen Studien nicht möglich eine Aussage über die Erklärungskraft des Modells von Stocké zu treffen. Aus diesem Grund wurden in der vorliegenden Arbeit die Modelle auf Einstellungen übertragen. Die Analysen der Modelle haben gezeigt, dass ein strenger multiplikativer Ansatz auch für die Erklärung von Antwortverzerrung bei Einstellungen nicht angemessen zu sein scheint, da der Faktor Anonymität entscheidend ist und auch in scheinbar anonymen Situationen nicht einfach ignoriert werden darf.

Der Stand der Forschung bezogen auf die theoretischen Modelle ist vor diesem Hintergrund unzureichend. Eine besondere Lücke stellt die fehlende Berücksichtigung des Bedürfnisses nach persönlicher Identität dar. Während die Faktoren des Bedürfnisses nach Anerkennung von Stocké, Skarbek-Kozietulska et al., Wolter und auch in der vorliegenden Arbeit bereits untersucht wurden, fehlt eine Analyse dieses zweiten Nutzenfaktors bisher.

Es konnte gezeigt werden, dass alle Ansätze theoretische Schwächen aufweisen, die bei der Umsetzung und Interpretation von Ergebnissen berücksichtigt werden müssen. Der unpräzise Umgang mit Begriffen und die Existenz theoretischer Lücken sind dabei besonders gravierend. Die eingeschränkte Vergleichbarkeit der Modelle von Stocké, Wolter und Skarbek-Kozietulska et al. verdeutlicht sich besonders deutlich in der Gegenüberstellung der Bewertungskriterien, die für eine präzise Definition des SD-Bias notwendig sind. Wenig Berücksichtigung fanden bisher die Folgen der inhaltlichen Dimension, die sich in den unterschiedlichen Frageformen der Untersuchungsgegenstände widerspiegelt. Durch die Variation des Erhebungsmodus ist zudem nicht klar, ob sich beide Autoren auf dieselbe Bezugsgruppen Dimension beziehen. So ist es nicht erstaunlich, dass Stocké im Gegensatz zu Skarbek-Kozietulska et al. und Wolter zu unterschiedlichen Ergebnissen kamen. Die heterogene Datenlage bietet weder die Möglichkeit zu beurteilen, welche der Erklärungsmodelle am besten den SD-Bias konzeptualisiert, noch können Aussagen über die Stabilität der Faktoren getroffen werden.

Operationalisierung

In Kapitel 4 wurden die Operationalisierungsvarianten der Erklärungsfaktoren im Rahmen des RC-Modells herausgearbeitet. Hier offenbart sich, dass bislang keine systematische Auseinandersetzung der Methodenforschung (zu diesem Thema) mit den Operationalisierungsvarian-

ten erfolgte. Für dieselben Konstrukte werden verschiedene Messungen eingesetzt ohne, dass eine Reflexion der Implikationen und deren Folgen stattfindet.

Gerade für das Anerkennungsbedürfnis entwickelte man in den letzten Jahrzehnten viele Täuschungsskalen, deren Gemeinsamkeiten und Unterschiede bisher nur unzureichend untersucht wurden. In Kapitel 5.1 der vorliegenden Arbeit wird beispielsweise aufgezeigt, dass die theoretischen Dimensionen und damit die Struktur der Skalen nicht selbstverständlich unterstellt werden kann. Die Annahme der Eindimensionalität hinsichtlich MC und der SES-17 Skala kann vor dem Hintergrund der empirischen Ergebnisse als nicht haltbar bezeichnet werden. In vielen der Analysen werden die Skalen dennoch, ohne vorherige Überprüfung der Faktorenstruktur, eingesetzt. Es besteht dadurch die Gefahr, dass die Ergebnisse, die man im Rahmen der RC-Modelle produzierte, nur Artefakte überholter Messungen darstellen.

Ein weiteres Problem bei der Operationalisierung ergibt sich bei der Erwünschtheitswahrnehmung. Bei der normativen Bewertung der Items können viele verschiedene Ebenen relevant werden. Bisher gibt es keine Forschungsarbeiten, in welcher die Beziehungen der verschiedenen Ebenen untersucht werden. Während in einigen Studien nach der normativen Bewertung einer fiktiven Gesellschaft gefragt wird, erfragen andere die Einstellung des Freundeskreises. Zusätzlich kann auch die Wahrnehmung über den Interviewer eine Rolle spielen. Welche Form der Erwünschtheitswahrnehmung im direkten Vergleich stärker wirkt und in welcher Beziehung sie jeweils zu anderen Erklärungsfaktoren stehen, blieb bisher unberücksichtigt.

Probleme, die bei der Messung der Anonymität auftreten, sind umfassender Untersuchungsgegenstand in Kapitel 4.1.2. Während in bisherigen Studien der Faktor über Modus Experimente gemessen wurde, ist hier im Rahmen der RC-Modelle erstmals die wahrgenommene Anonymität erhoben. Die paradoxen Ergebnisse zeigten jedoch, dass gerade diese Variable anfällig für Fragebogeneffekte ist. Wie schon Whelan und Meade (2009) vermuteten, ist es möglich, dass der Kausalzusammenhang entgegen der bisherigen Theorie verläuft. Auch zu diesem Faktor ist deshalb weiterer Forschungsbedarf gegeben. Am unklarsten ist die Operationalisierung der Konstrukte im Rahmen des Bedürfnisses nach Authentizität. Kapitel 4.2 verdeutlicht, dass die theoretischen Überlegungen sich nur schwer in der Praxis umsetzen lassen. Hinzu kommt, dass sowohl Esser als auch Stocké mit den Begrifflichkeiten teilweise unpräzise umgehen, sodass eine Interpretation zusätzlich erschwert wird.

Zusammenfassend lässt sich festhalten, dass die vorgestellten Analysen gezeigt haben, wie anfällig die Modelle auf unterschiedliche Operationalisierungen reagieren. In der bisherigen Forschung werden die Messungen noch viel zu sehr unreflektiert variiert. Dadurch ist die Ver-

gleichbarkeit der unterschiedlichen Ergebnisse nicht mehr gegeben. Um die Güte der theoretischen Modelle ausreichend bewerten zu können, ist es deshalb zunächst notwendig, die Folgen der unterschiedlichen Operationalisierungen zu testen. Der unpräzise Umgang mit den Täuschungsskalen, die mangelnde Berücksichtigung der verschiedenen Ebenen der Erwünschtheitswahrnehmung, sowie die bisher zu unklaren Messungen der Determinanten des Handlungsziels Authentizität stehen dabei im Vordergrund der Kritik. Insgesamt ist es vor dem Hintergrund unzureichender und zum Teil divergierender Operationalisierungen wenig überraschend, dass die SD-Forschung in den letzten 50 Jahren vor allem widersprüchliche Ergebnisse hervorgebracht hat.

Sinn und Unsinn der SD-Forschung

In der empirischen Analyse dieser Arbeit (Kapitel 5.3 und 5.4) wurde der empirische Forschungsstand zu den Modellen betrachtet. Die Analyse der bisherigen Kontrolle des SD-Bias im Rahmen von deutschen Surveys hat gezeigt, dass der Einsatz von Täuschungsskalen ohne Berücksichtigung weiterer Determinanten wenig sinnvoll ist. Die isolierte Erhebung von Täuschungsskalen ohne Kenntnisse der wahrgenommenen Erwünschtheit produziert Effekte, die widersprüchlich erscheinen und schlecht interpretierbar sind. Um bewerten zu können, welches der Erklärungsmodelle eingesetzt werden kann, um eine höhere Erklärungskraft und Stabilität zu erreichen, wurden die theoretischen Modelle von Marlow/Crowne, Stocké und Skarbak-Kozietulska et al. sowie eine kombinierte Variante auf Basis einer Studierendenstichprobe untersucht. Die Befunde lassen erkennen, dass keines der Modelle die gewünschte Stabilität erreichte und zum Teil sehr empfindlich auf kleine Veränderungen in der Operationalisierung reagierten.

Was bedeuten die Ergebnisse im Rahmen der vorliegenden Arbeit für die SD-Forschung? Zunächst bleibt festzuhalten, dass das Problem des sozial erwünschten Antwortverhaltens gerade bei der Erfassung von sensitiven Themen wie Vorurteilen nicht zu unterschätzen ist. Natürlich existiert dieses Phänomen und empirische Ergebnisse sollten stets mit Blick auf die Rahmenbedingungen einer Erhebung (wie bspw. ein durch Interviewer beeinträchtigtes Anonymitätsempfinden der Befragten) interpretiert werden. Ob es jedoch sinnvoll ist, nach den bisherigen Modellen zusätzliche Variablen zu berücksichtigen, mit denen sich der SD-Bias nachträglich berechnen lässt, bleibt fraglich, da der Aufwand nicht im Verhältnis zum Nutzen steht.

Die angesprochenen Probleme in Bezug auf Begrifflichkeiten, Theorien und Operationalisierungen machen deutlich, wie wenig belastbare Erkenntnisse trotz jahrzehntelanger Forschung auf dem Gebiet der sozialen Erwünschtheit bisher vorliegen. Zwar gibt es eine Vielzahl von em-

pirischen Untersuchungen, deren Ergebnisse sind jedoch aufgrund der diskutierten Schwächen kaum miteinander zu vergleichen. Gerade die Defizite im theoretischen Bereich führen zu wenig systematischen Untersuchungen. Die Befunde aus diesen Forschungsarbeiten sind vor dem Hintergrund aktueller Erkenntnisse kaum sinnvoll zu interpretieren. Folgt man den theoretisch am weitesten entwickelten Ansätzen, so müssten, um die Verzerrung durch sozial erwünschtes Antwortverhalten korrekt zu erfassen, eine Reihe von Konstrukten gemessen werden. Im Rahmen einer einzelnen Erhebung bedeutet dies eine große Anzahl an Items. Für Befragungen, die nicht ausschließlich auf Methodenforschung fokussiert sind, ist die Kontrolle des SD-Bias aufgrund des gegebenen Umfangs an zusätzlichen Items unpraktisch. Die Abwägung zwischen längeren Befragungen mit korrekter Kontrolle des SD-Bias und kürzeren, die gleichzeitig geringer Kosten verursachen und niedrigere Verweigerungs- bzw. Abbruchquoten nach sich ziehen, fällt ausgesprochen leicht. In Anbetracht der dargelegten Mängel, Schwächen und Defizite im Bereich der wissenschaftlichen Auseinandersetzung mit sozial erwünschtem Antwortverhalten ist der Aufwand nicht zu rechtfertigen. Für die Methodenforschung ist die korrekte Messung des SD-Bias sicherlich ein interessantes methodisches Problem, ob die dabei gewonnen Erkenntnisse sich jemals in der Umfrageforschung sinnvoll umsetzen lassen bleibt abzuwarten.

Sowohl in der Aufarbeitung des theoretischen Forschungsstandes als auch bei der anschließenden empirischen Analyse der vorliegenden Arbeit wurden wichtige Komponenten der sozialen Erwünschtheit dargestellt. Gerade die Erwünschtheitswahrnehmung scheint eine zentrale Rolle bei der Interpretation des SD-Bias zu spielen. Es hat sich jedoch auch gezeigt, dass die bisherige Aufzählung der Erklärungsfaktoren längst nicht abschließend ist. Aber auch die Wirkung der Erklärungsmodelle in Bezug auf unterschiedliche sensitive Fragen (Verhalten oder Einstellung) ist bisher ungeklärt. Für eine Beurteilung des Nutzens umfangreicher SD-Modelle im Rahmen von inhaltlichen Fragestellungen wäre es ebenfalls notwendig, die Erklärungskraft der Modelle neben inhaltlichen Faktoren zu untersuchen. Wie groß ist der Anteil der erklärten Varianz der SD-Modelle in Bezug auf Vorurteile, wenn inhaltliche Konstrukte wie die „Soziale Dominanz Orientierung“ oder die „Autoritäre Persönlichkeit“ zusätzlich berücksichtigt werden? Um das Zusammenspiel der einzelnen Determinanten und die Folgen ihrer Messung korrekt einschätzen zu können, wäre ein systematisches Forschungsprogramm notwendig.

Literatur

- Adorno, T.W., 1950: *The Authoritarian personality*. New York: Norton.
- Ajzen, I., 1991: *The Theory of Planned Behavior. Some Unresolved Issues*. *Organizational Behavior and Human Decision Processes* 50: 179–211.
- Ajzen, I. & Fishbein M., 1975: *Belief, Attitude, Intention, and Behavior. An Introduction to Theory and Research*. MA: Addison-Wesley.
- Ajzen, I., Timko C & White J.B., 1982: Self-monitoring and the attitude-behavior relation. *Journal of Personality and Social Psychology* 42: 426–435.
- Aosved, A.C., P.J. Long & E.K. Voller, 2009: Measuring Sexism, Racism, Sexual Prejudice, Ageism, Classism, and Religious Intolerance: The Intolerant Schema Measure. *Journal of Applied Social Psychology* 39: 2321–2354.
- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., 2010: *Jahresbericht 2010*.
- Arkin, R.M., 1980: Self-presentation in: D. Wegener & R. Vallacher (Hrsg.), *The self in social psychology*. London, New York: Oxford University Press.
- Austin, J.S., 1992: The Detection of Fake Good and Fake Bad on the Mmpi-2. *Educational and Psychological Measurement* 52: 669–674.
- Babka Gostomski, C. von, B. Küpper & W. Heitmeyer, 2007: Fremdenfeindlichkeit in den Bundesländern. Die schwierige Lage in Ostdeutschland. S. 102–128 in: W. Heitmeyer (Hrsg.), *Deutsche Zustände. Folge 5*. Frankfurt am Main: Suhrkamp.
- Banse, R. & B. Gawronski, 2003: Die Skala Motivation zu vorurteilsfreiem Verhalten. Psychometrische Eigenschaften und Validität. *Diagnostica* 49: 4–13.
- Barger, S.D., 2002: The Marlowe-Crowne Affair: Short Forms, Psychometric Structure, and Social Desirability. *Journal of Personality Assessment* 79: 286–305.
- Bassili, J.N., 1996: Meta-judgmental versus operative indexes of psychological attributes. The case of measures of attitude strength. *Journal of Personality and Social Psychology* 71: 637–653.
- Bates, S. & J. Cox, 2008: The impact of computers versus paper-pencil survey, and individual versus group administration, on self-reports of sensitive behavior. *Computers in Human Behavior* 24: 903–916.
- Batson, C.D., S.J. Naifeh & S. Pate, 1978: Social Desirability, Religious Orientation, and Racial Prejudice. *Journal for the Scientific Study of Religion* 17: 31.
- Bauman, K.E. & C.W. Dent, 1982: Influence of an objective measure on self-reports of behavior. *Journal of Applied Psychology* 67: 623–628.
- Beretvas, S.N., J.L. Meyers & W.L. Leite, 2002: A Reliability Generalization Study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement* 62: 570–589.
- Bergmann, W. & R. Erb, 1986: *Kommunikationslatenz, Moral und Öffentliche Meinung. Theoretische Überlegungen zum Antisemitismus in der Bundesrepublik Deutschland*. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38: 223–246.
- Berinsky, A.J., 2004: Can We Talk? Self-Presentation and the Survey Response. *Political Psychology* 25: 643–659.
- Berinsky, A.J. & H. Lavine, 2007: *Self-Monitoring and Political Attitudes*.

- Blair, G. & K. Imai, 2012: Statistical Analysis of List Experiments. *Political Analysis* 20: 47–77.
- Bradburn, N.M., S. Sudman & E. Blair, 1979: Improving interview method and questionnaire design. San Francisco: Jossey-Bass.
- Bradburn, N. & S.W.B. Sudmann (Hrsg.), 2004: Asking Questions. The Definitive Guide to Questionnaire Design- For Market Research, Political Polls, and Social and Health Questionnaires, Revised Edition. San Francisco: Jossey-Bass.
- Burchell, B. & C. Marsch, 1992: The Effect of Questionnaire Length on Survey Response. *Quality & Quantity* 26: 233–244.
- Cooley, C., 1902: Human nature and the social order. New York: Scribners.
- Couper, M.P., E. Singer & R. Tourangeau, 2003: Understanding the Effects of Audio-CASI on Self-Reports of Sensitive Behavior. *Public Opinion Quarterly* 67: 385–395.
- Couper, M.P., R. Tourangeau & D.M. Steiger (Hrsg.), 2001: Social Presence in Web Surveys. SIGCHI conference on Human factors in computing systems, New York.
- Coutts, E. & B. Jann, 2008: Sensitive Questions in Online Surveys. Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). ETH Zurich Sociology Working Paper No. 3.
- Crowne, D.P. & D. Marlowe, 1964: The Approval Motive. New York: John Wiley & Sons.
- Currivan, D.B., A.L. Nyman, C.F. Turner & L. Biener, 2004: Does Telephone Audio Computer-Assisted Self-Interviewing Improve the Accuracy of Prevalence Estimates of Youth Smoking? Evidence from the Umass Tobacco Study. *Public Opinion Quarterly* 68: 542–564.
- Dalton, D.R., J. Wimbush & C. Daily, 1994: Using the Unmatched Count Technique (UCT) to Estimate Base Rates for Sensitive Behavior. *Personnel Psychology* 47: 817–829.
- Damarin, F. & S. Messick, 1965: Response styles as personality variables. A theoretical integration of multivariate research. Princeton, New York: Educational Testin Service.
- DeMaio, T.J., 1984: Social Desirability and Survey Measurement: A Review. S. 257–282 in: C. Turner, E. Martin & National Research Council . Panel on Survey Measurement of Subjective Phenomena (Hrsg.), *Surveying Subjective Phenomena*: Russell Sage Foundation.
- Diekmann, A., 2010: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Hamburg: Rowohlt Taschenbuch Verlag.
- Diekmann, A. & T. Voss (Hrsg.), 2004: Rational Choice Theorie in den Sozialwissenschaften. München: Oldenbourg.
- Dillman, D.A., E.H. Carpenter, J.A. Christenson & R.M. Brooks, 1974: Increasing Mail Questionnaire Response: A Four State Comparison. *American Sociological Review* 39: 744–756.
- Dran, E. & A. Hildreth, 1995: What the Public Thinks About How we Know What it is Thinking. *International Journal of Public Opinion Research* 7: 128–144.
- Duck, R.J. & B. Hunsberger, 1999: Religious Orientation and Prejudice: The Role of Religious Proscription, Right-Wing. *International Journal for the Psychology of Religion* 9: 157–179.
- Duckitt, J. & C.G. Sibley, 2007: Wing authoritarianism, social dominance orientation and the dimensions of generalized prejudice. *European Journal of Personality* 21: 113–130.

- Duffy, B., K. Smith, G. Terhanian & J. Bremer, 2005: Comparing data from online and face-to-face surveys. *International Journal of Market Research* 47: 615–639.
- Edwards, A.L., 1967: The social desirability variable. A broad statement. S. 32–47 in: I.A. Berg (Hrsg.), *Response set in personality assesment*. Chicago: Aldine.
- Erbslöh, B. & A. Koch, 1988: Die Non-Response-Studie zum Allbus 1986.: Problemstellung, Design, erste Ergebnisse. *ZUMA-Nachrichten* 22: 29–44.
- Esser, H., 1986a: Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. *ZUMA-Arbeitsbericht*.
- Esser, H., 1986b: Aus dem Leben der Forschung: Können Befragte lügen? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38: 314–336.
- Esser, H., 1990: "Habits", "Frames" und "Rational Choice". Die Reichweite von Theorien der rationalen Wahl (am Beispiel der Erklärung des Befragtenverhaltens). *Zeitschrift für Soziologie* 19: 231–247.
- Esser, H., 1991: Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als „rational choice“ in: R. Wittenberg (Hrsg.), *Person – Situation– Institution – Kultur*. Günter Büschges zum 65. Geburtstag. Berlin: Duncker& Humblot.
- Esser, H., 1999: *Soziologie. Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt/New York: Campus Verlag.
- Fazio, R.: Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. S. 247–282.
- Fazio, R.H.: Multiple Processes by which Attitudes guide Behavior: The MODE Model as an Integrative Framework. *Advances in Experimental Social Psychology* 23: 75–109.
- Fishbein M. & I. Ajzen, 1975: *Belief, Attitude, Intention and Behavior. An Introduction to Theory and Research*. Reading: Addison-Wesley.
- Forsa, 1993: *Umfragen in Deutschland. Erfahrungen und Meinungen der Bürger. Teil I*. *Context* 8: 1–7.
- Forsa, 1993: *Umfragen in Deutschland. Erfahrungen und Meinungen der Bürger. Teil II*. *Context* 8: 1–10.
- Forsa, 1996: *Umfragen über Umfragen. Ein Jahrzehnt im Überblick*. *Context* 14: 1–15.
- Gebhardt, R., A. Klein, M. Meier & W. Alban (Hrsg.), 2012: *Antisemitismus in der Einwanderungsgesellschaft. Beiträge zur kritischen Bildungsarbeit*. Weinheim: Beltz Juventa.
- Gilens, M., P.M. Sniderman & J.H. Kuklinski, 1998: Affirmative Action and the Politics of Realignment. *British Journal of Political Science* 28: 159–183.
- Goffman, E., 1959: *The presentation of self in everyday life*. New York: Doubleday.
- Goffman, E., 1969: *Wir alle spielen Theater. Die Selbstdarstellung im Alltag*. München [u.a.]: Piper.
- Goodstadt, M.S. & V. Gruson, 1975: The randomized response technique. A test of drug use. *Journal of the American Statistical Association* 70: 814–818.
- Gosen, S., P. Schmidt, S. Thörner & J. Leibold, 2013: *Is the List Experiment Doing its Job? Inconclusive Evidence*. Unveröffentlichtes Manuskript.
- Goyder, J., 1986: Surveys on Surveys. Limitations and Potentialities. *Public Opinion Quarterly* 50: 27–41.

- Goyder, J., 1987: *The Silent Minority. Nonrespondents on Sample Surveys*. Cambridge: Polity Press.
- Groves, R.M., 2004: *Survey errors and survey costs*. Hoboken, N.J: Wiley-Interscience.
- Groves, R.M., S. Presser & S. Dipko, 2004: The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly* 68: 2–31.
- Groves, R., E. Singer & A. Corning, 2000: Leverage-Saliency Theory of Survey Participation. Description and an Illustration. *Public Opinion Quarterly* 64: 299–308.
- Gulliksen, H., 1950: *Theory of mental tests*. New York: Wiley.
- Hahn, J., 2005: Faking Bad and Faking Good by College Students on the Korean MMPI–2. *Journal of Personality Assessment* 85: 65–73.
- Hancock, D.R. & C.P. Flowers, 2001: Comparing Social Desirability Responding World Wide Web and Paper-Administered Surveys. *Educational Technology Research and Development* 49: 5–13.
- Hannover, B., 1997: *Das dynamische Selbst. Die Kontextabhängigkeit selbstbezogenen Wissens*. Bern, Seattle: Huber.
- Hartmann, P., 1991: *Wunsch und Wirklichkeit: Theorie und Empirie sozialer Erwünschtheit*: Deutscher Universitäts-Verlag.
- Heerwegh, D., 2009: Mode Differences Between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects. *International Journal of Public Opinion Research* 21: 111–121.
- Heerwig, J. & B. McCabe, 2009: Education and Social Desirability Bias. The Case of a Black Presidential Candidate. *Social Science Quarterly* 90: 674–686.
- Heitmeyer, W., 2002: Gruppenbezogene Menschenfeindlichkeit. Die theoretische Konzeption und erste empirische Ergebnisse. S. 15–36 in: W. Heitmeyer (Hrsg.), *Deutsche Zustände. Folge 1*. Frankfurt am Main: Suhrkamp.
- Heitmeyer, W. & K. Endrikat, 2008: Die Ökonomisierung des Sozialen. Folgen für "Überflüssige" und "Nutzlose". S. 55–72 in: W. Heitmeyer (Hrsg.), *Deutsche Zustände. Folge 6*. Frankfurt am Main: Suhrkamp.
- Heyder, A., J. Iser & P. Schmidt, 2005: Israelkritik oder Antisemitismus? Meinungsbildung zwischen Öffentlichkeit, Medien und Tabus. S. 144–165 in: W. Heitmeyer (Hrsg.), *Deutsche Zustände. Folge 3*. Frankfurt am Main: Suhrkamp.
- Holtgraves, T., 2004: Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding. *Personality and Social Psychology Bulletin* 30: 161–172.
- Hülshager, U.R., F.H. Spinath, A. Küppers & S. Etzel, 2004: Experimentelle Untersuchung zweier Methoden zur Reduzierung Sozialer Erwünschtheit in einem computergestützten eignungsdiagnostischen Testverfahren. *Zeitschrift für Personalpsychologie* 3: 24–33.
- James, W., 1892: *Psychology: Briefer course* in: F. Burkhardt, F. Bowers & I. Skrupskelis (Hrsg.), *The works of William James*. Cambridge: Harvard University Press.
- Jellison, J. & R. Arkin: Social comparison of abilities: A self-presentation approach to decision making in groups. S. 235–257.
- Joinson, A., 1999: Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers* 31: 433–438.
- Jones, E., 1964: *Ingratiation: A social psychological analysis*. New York: Appleton-Century-Crofts.

- Jones, E.E. & H. Sigall, 1971: The Bogus Pipeline. A New Paradigm for Measuring Affect and Attitude. *Psychological Bulletin* 76: 349–364.
- Jungbauer-Gans, M. & P. Kriwy, 2004: Bildung und Gesundheitsvorsorge. Die Impfentscheidung. S. 331–348 in: A. Diekmann & T. Voss (Hrsg.), *Rational Choice Theorie in den Sozialwissenschaften*. München: Oldenbourg.
- Kalter, F., 1997: Wohnortwechsel in Deutschland. Ein Beitrag zur Migrationssoziologie und zur empirischen Anwendung von Rational-Choice-Modellen. Opladen: Leske + Budrich.
- Kane, J.G., S.C. Craig & K.D. Wald, 2004: Religion and Presidential Politics in Florida: A List Experiment*. *Social Science Quarterly* 85: 281–293.
- Köllisch, T. & D. Oberwittler, 2004: Wie ehrlich berichten männliche Jugendliche über ihr delinquentes Verhalten? *KzfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 56: 708–735.
- Krampen, G., 2005: Psychology of control and behavior. S. 97–115 in: W. Greve, K. Rothermund & D. Wentura (Hrsg.), *The adaptive self. Personal continuity and international self-development*. Cambridge: Hogrefe & Huber Publishers.
- Kreuter, F., S. Presser & R. Tourangeau, 2008: Social Desirability Bias in CATI, IVR, and Web Surveys. The Effect of Mode and Question Sensitivity. *Public Opinion Quarterly* 72: 847–865.
- Kroneberg, C., 2005: Die Definition der Situation und die variable Rationalität der Akteure. Ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie* 34: 344–363.
- Kroner, D.G. & J.R. Weekes, 1996: Balanced Inventory of Desirable Responding: Factor Structure, Reliability, and Validity with an Offender Sample. *Personality and Individual Differences* 21: 323–333.
- Krosnick, J.A. & H. Schuman, 1988: Attitude intensity, importance, and certainty and susceptibility to response effects. *Journal of Personality and Social Psychology* 54: 940–952.
- Kühnel, S.M., 1993a: Lassen sich ordinale Daten mit linearen Strukturgleichungsmodellen analysieren? *ZA-Information* 33: 29–51.
- Kühnel, S.M., 1993b: Zwischen Boykott und Kooperation. Teilnahmeabsichten und Teilnahmeverhalten bei der Volkszählung 1987. Frankfurt/M.: Peter Lang.
- Kuklinski, J.H., M.D. Cobb & M. Gilens, 1997: Racial Attitudes and the “New South”. *The Journal of Politics* 59: 323.
- Kunz, V., 2004: *Rational Choice*. Frankfurt/New York: Campus Verlag.
- Lambert, K.D., 2011: *Handlungstheorie zwischen Kommunitarismus und Rational Choice*: Grin Verlag.
- Lara, D., J. Strickler, C.D. Olavarrieta & C. Ellertson, 2004: Measuring Induced Abortion in Mexico: A Comparison of Four Methodologies. *Sociological Methods & Research* 32: 529–558.
- Laux, L. & K.-H. Renner, 2002: Self-Monitoring und Authentizität: Die verkannten Selbstdarsteller. *Zeitschrift für Differentielle und Diagnostische Psychologie* 23: 129–148.
- Leibold, J. & S.M. Kühnel, 2011: Einigkeit in der Schuldabwehr. Die Entwicklung antisemitischer Einstellungen in Deutschland nach 1989. S. 131–151 in: W. Heitmeyer (Hrsg.), *Deutsche Zustände*. Folge 10. Frankfurt am Main: Suhrkamp.

- Leibold, J. & S.M. Kühnel, 2002: Islamophobie. Sensible Aufmerksamkeit für spannungsreiche Anzeichen. S. 100–119 in: W. Heitmeyer (Hrsg.), *Deutsche Zustände. Folge 1*. Frankfurt am Main: Suhrkamp.
- Leite, W.L. & S.N. Beretvas, 2005: Validation of Scores on the Marlowe-Crowne Social Desirability Scale and the Balanced Inventory of Desirable Responding. *Educational and Psychological Measurement* 65: 140–154.
- Lensvelt-Mulders, G.J.L.M., 2005: Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation. *Sociological Methods & Research* 33: 319–348.
- Li, A. & J. Bagger, 2006: Using the BIDR to Distinguish the Effects of Impression Management and Self-Deception on the Criterion Validity of Personality Measures: A Meta-Analysis. *International Journal of Selection and Assessment* 14: 131–141.
- Li, A. & J. Reb, 2009: A Cross-Nations, Cross-Cultures, and Cross-Conditions Analysis on the Equivalence of the Balanced Inventory of Desirable Responding. *Journal of Cross-Cultural Psychology* 40: 214–233.
- Lück, H.E. & Tomaeus E., 1969: Skalen zur Messung Manifester Angst (MAS) und sozialer Wünschbarkeit (SDS-E und SDS-CM). *Diagnostica* 15: 134–141.
- Mayerl, J., 2005: Controlling The Baseline Speed Of Respondents: An Empirical Evaluation Of Data Treatment Methods Of Response Latencies in: C. van Dijkum, J. Blasius & B. van Hilton (Hrsg.), *Recent Developments and Applications in Social Research Methodology. Proceedings of the Sixth International Conference on Logic and Methodology*. Leverkusen-Opladen: Barbara Budrich.
- Mayerl, J. & Urban Dieter, 2008: *Antwortreaktionszeiten in Survey-Analysen. Messung, Auswertung und Anwendung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mead, G., 1934: *Mind, self and society from the standpoint of a social behaviorist*. Chicago: University of Chicago Press.
- Mehlkop, G. & R. Neumann, 2012: Die Erklärung der Zustimmung zu familienpolitischen Umverteilungsmaßnahmen. Evidenzen für das Wechselspiel von rationalen Erklärungsansätzen und der Bedeutung von Einstellungen zur Familie. *Zeitschrift für Soziologie* 41: 182–206.
- Messick, S., 1960: Dimensions of Social Desirability. *Journal of Consulting Psychology* 24: 279–287.
- Miller, J.M. & D.A.M. Peterson, 2004: Theoretical and Empirical Implications of Attitude Strength. *The Journal of Politics* 66.
- Mühlenfeld, H.-U., 2004: *Der Mensch in der Online-Kommunikation. Zum Einfluss webbasierter, audiovisueller Fernkommunikation auf das Verhalten von Befragten*. Wiesbaden.
- Mummendey, H.D., 1995: *Psychologie der Selbstdarstellung*. Göttingen, Seattle: Hofgreffe-Verlag für Psychologie.
- Murray, D., C. O'Connell, L. Schmid & C. Perry, 1987: The validity of smoking self-reports by adolescents. A reexamination of the bogus pipeline procedure. *Addictive Behaviors* 12: 7–15.
- Musch, J., R. Brockhaus & A. Bröder, 2002: Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit. *Diagnostica* 48: 121–129.
- Newman, J.C., D.C. Des Jarlais, C.F. Turner, J. Gribble, P. Cooley & D. Paone, 2002: The Differential Effects of Face-to-Face and Computer Interview Modes. *American Journal of Public Health* 92: 294–297.

- Paulhus, D.L., 1984: Two-component models of socially desirable responding. *Journal of Personality and Social Psychology* 46: 598–609.
- Paulhus, D.L., 2002: Socially desirable responding: The evolution of a construct. S. 49–69 in: H.I. Braun, D.N. Jackson, D.E. Wiley & S. Messick (Hrsg.), *Role of Constructs in Psychological and Educational Measurement*. Mahwah, NJ: L. Erlbaum.
- Podsakoff, P., S. MacKenzie, J. Lee & N. Podsakoff, 2003: Common method biases in behavioral research. A critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88: 879–903.
- Porst, R., 1998: Erfahrung mit und Bewertung von Umfragen. Was unsere Befragten über Umfragen denken. *ZUMA-Arbeitsbericht* 3: 1–14.
- Preisendörfer, P., 2008: Heikle Fragen in mündlichen Interviews: Ergebnisse einer Methodenstudie im studentischen Milieu. *ETH Zurich Sociology Working Paper*.
- Preisendörfer, P. & U. Liebe, 2011: Umweltsoziologie und Rational-Choice-Theorie. S. 221–239 in: M. Groß (Hrsg.), *Handbuch Umweltsoziologie*. Wiesbaden: Verlag für Sozialwissenschaften.
- Quigley-Fernandez, B. & J. T. Tedeschi, 1978: The Bogus Pipeline as Lie Detector. Two Validity Studies. *Journal of Personality and Social Psychology* 36: 247–256.
- Ramanaiah, N.V., T. Schill & L.S. Leung, 1977: A test of the hypothesis about the two-dimensional nature of the Marlowe-Crowne Social Desirability Scale. *Journal of Research in Personality* 11: 251–259.
- Rattazzi, M., A. Maria & V. Chiara, 2003: Social Desirability of Subtle and Blatant Prejudice Scales. *Psychological Reports* 92.
- Reinecke, J., 1991a: Interviewer- und Befragtenverhalten: theoretische Ansätze und methodische Konzepte. Opladen: Westdeutscher Verlag.
- Reinecke, J., 1991b: Intervieweffekte und soziale Erwünschtheit. Theorie, Modell und empirische Ergebnisse. *Journal für Sozialforschung* 31: 293–320.
- Roese, N. & D. Jamieson, 1993: Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin* 114: 363–375.
- Rogelberg, S., G. Fisher, D. Maynard, M. Hakel & M. Horvath, 2001: Attitudes toward surveys: Development of a measure and its relationship to respondent behavior. *Organizational Research Methods* 4: 3–25.
- Rogers, R., K. Sewell, M. Martin & M. Vitacco, 2003: Detection of feigned mental disorders - A meta-analysis of the MMPI-2 and malingering. *Assessment* 10: 160–177.
- Rossiter, J.C., 2009: A Comparison of Social Desirability Bias among Four Widely Used Methods of Data Collection as Measured by the Impression Management Subscale of the Balance Inventory of Desirable Responding.
- Rotter, J., 1960: Some implications of a social learning theory for the prediction of goal directed behavior from testing procedures. *Psychological Review* 67: 301–316.
- Sackheim, H. & R. Gur, 1978: Self-deception, selfconfrontation, and consciousness. S. 139–197 in: G. Schwartz & D. Shapiro (Hrsg.), *Consciousness and Self-Regulation. Advances in Research and Theory*. New York: Plenum Press.
- Schiefele, U., 1990: Einstellung, Selbstkonsistenz und Verhalten. Göttingen: Verlag für Psychologie Hogrefe.
- Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske+Budrich.

- Schumann, S., 2010: Repräsentative Umfrage: Praxisorientierte Einführung in empirische Methoden und statistische Analyseverfahren: Oldenbourg Wissenschaftsverlag.
- Schützenmeister, F.: Die Bereitschaft, sich wieder befragt zu lassen, in postalischen Erhebungen. *ZfS - Zeitschrift für Soziologie*; *ZfS*, Jg. 31, Heft 2 (2002); 138-154.
- Seidel, E., 2002: Die schwierige Balance zwischen Islamkritik und Islamophobie. S. 261–282 in: W. Heitmeyer (Hrsg.), *Deutsche Zustände*. Folge 1. Frankfurt am Main: Suhrkamp.
- Short, J., E. Williams & B. Christie, 1976: *The Social Psychology of Telecommunications*. London.
- Skarbek-Kozietulska, A., P. Preisendörfer & F. Wolter, 2012: Leugnen oder gestehen? Bestimmungsfaktoren wahrer Antworten in Befragungen. *Zeitschrift für Soziologie* 41: 5–23.
- Sniderman, P.M. & E.G. Carmines, 1997: Reaching beyond Race. *PS: Political Science and Politics* 30: 466.
- Snyder, M., 1974: Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology* 30: 526–537.
- Snyder, M., 1987: *Public appearances/private realities: The psychology of self-monitoring*. New York: Freeman.
- Sproull, L., M. Subramani, S. Kiesler, J. Walker & K. Waters, 1996: When the Interface Is a Face. *Human-Computer Interaction* 11: 97–124.
- Stinchcombe, A., C. Jones & P. Sheatsley, 1981: Nonresponse bias for attitude questions. *Public Opinion Quarterly* 45: 359–375.
- Stöber, J., 1999: Die Soziale-Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität. *Diagnostica* 45: 173–177.
- Stocké, V., 2001: Socially Desirable Response Behavior as Rational Choice: The Case of Attitudes Towards Foreigners. Sonderforschungsbereich 504 Arbeitspapier No.01-22.
- Stocké, V., 2003: Informationsverfügbarkeit und Response-Effects: Die Prognose von Einflüssen unterschiedlich kategorisierter Antwortskalen durch Antwortsicherheiten und Antwortlatenzen. Sonderforschungsbereich 504 03-25. Mannheim.
- Stocké, V., 2004: Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie* 33: 303–320.
- Stocké, V. & B. Becker, 2004: Determinanten und Konsequenzen der Umfrageeinstellung. Bewertungsdimensionen unterschiedlicher Umfragesponsoren und die Antwortbereitschaft der Befragten. *ZUMA-Arbeitsbericht* 28: 89–116.
- Stocké, V. & C. Hunkler, 2004: Die angemessene Erfassung der Stärke und Richtung von Anreizen durch soziale Erwünschtheit. Sonderforschungsbereich 504 04-16. Mannheim.
- Stocké, V. & B. Langfeldt, 2003: Umfrageeinstellung und Umfrageerfahrung. Die relative Bedeutung unterschiedlicher Aspekte der Interviewerfahrung für die generalisierte Umfrageeinstellung. Sonderforschungsbereich 504 03-34. Mannheim.

- Streb, M.J., B. Burrell, B. Frederick & M.A. Genovese, 2008: Social Desirability Effects and Support For a Female American President. *Public Opinion Quarterly* 72: 76–89.
- Taddicken, M., 2008: Die Bedeutung von Methodeneffekten der Online-Befragung in der empirischen Sozialforschung. Zusammenhänge zwischen computervermittelter Kommunikation und erreichbarer Datengüte. S. 91–107 in: N. Jakob, H. Schoen & T. Zerback (Hrsg.), *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung*. Wiebaden: VS Verlag für Sozialwissenschaften.
- Taylor, J.B., 1959: Social Desirability and MMPI performance. the individual case. *Journal of Consulting Psychology* 23: 514–517.
- Tedeschi, J.T., S. Lindskold & P. Rosenfeld, 1985: *Introduction to social psychology*. St. Paul, Minn: West Pub. Co.
- Terry, D.J., M.A. Hogg & B.M. McKimmie, 2000: Attitude-behaviour relations: The role of in-group norms and mode of behavioural decision-making. *British Journal of Social Psychology* 39: 337–361.
- Thomas, W.I. & D.S. Thomas, 1928: *The Child in America. Behavior Problems and Programs*. New York: Knopf.
- Tourangeau, R., L.J. Rips & K.A. Rasinski, 2000: *The psychology of survey response*. Cambridge, U.K., , New York: Cambridge University Press.
- Tourangeau, R. & T.W. Smith, 1996: Asking Sensitive Questions. The Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly* 60: 275–304.
- Tourangeau, R., T.W. Smith & K.A. Rasinski, 1997: Motivation to Report Sensitive Behaviors on Surveys: Evidence From a Bogus Pipeline Experiment¹. *Journal of Applied Social Psychology* 27: 209–222.
- Tourangeau, R. & T. Yan, 2007: Sensitive Questions in Surveys. *Psychological Bulletin* 133: 859–883.
- Tracy, P.E. & J.A. Fox, 1981: The Validity of Randomized Response for Sensitive Measurements. *American Sociological Review* 46: 187.
- Tsuchiya, T., Y. Hirai & S. Ono, 2007: A Study of the Properties of the Item Count Technique. *Public Opinion Quarterly* 71: 253–272.
- Urban Dieter & J. Mayerl, 2007: Antwortlatenzzeiten in der Survey-basierten Verhaltensforschung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59: 692–713.
- van der Heijden, P.G.M, G. van Gils, J. Bouts & J. Hox, 2000: A Comparison of Randomized Response, Computer-Assisted Self-Interview, and Face-to-Face Direct Questioning: Eliciting Sensitive Information in the Context of Welfare and Unemployment Benefit. *Sociological Methods & Research* 28: 505–537.
- Vogt, W., 1999: *Dictionary of statistics and methodology. A nontechnical guide for the social science*. CA: Sage Publications.
- Warner, S.L., 1965: Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association* 60: 63.
- Whelan, T.J., 2008: Antecedents of Anonymity Perceptions in Web-based Surveys.
- Whelan, T.J. & A.W. Meade, 2009: Examining DIF in Perceived Anonymity When Sensitive Items are Endorsed.

- Whelan, T.J. & L. Thompson, 2009: Development and validation of a scale for measuring respondents' perceptions of anonymity in survey research. North Carolina State University.
- Wilson, T.D., S.J. LaFleur & E.D. Anderson, 1996: The Validity and Consequences of Verbal Reports of Attitudes. S. 91–114 in: N. Schwarz & S. Sudman (Hrsg.), Answering questions. Methodology for determining cognitive and communicative processes in survey research. San Francisco: Jossey-Bass Publishers.
- Winkler, N., M. Kroh & Spiess Martin, 2006: Entwicklung einer deutschen Kurzsкала zur zweidimensionalen Messung von sozialer Erwünschtheit. Discussion Papers 579. Berlin.
- Wish, E., G. Yacoubian, D. Perez & J. Choyka, 2000: Test first and ask questions afterwards. Improving the validity of self-report drug use by arrestees. San Francisco.
- Wiswede, G., 2007: Einführung in die Wirtschaftspsychologie. München: UTB.
- Wolter, F., 2012: Heikle Fragen in Interviews. Eine Validierung der Randomized Response-Technik. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Yankelovich, D., F. Skelly & White A., 1981: The mushiness index. A refinement in public policy polling techniques. New York: Yankelovich, Clancey, Shulman.
- Zdep, S.M. & I.N. Rhodes, 1976: Making the randomized response technique work. Public Opinion Quarterly 40: 531–537.
- Zick, A., C. Wolf, B. Kuepper, E. Davidov, P. Schmidt & W. Heitmeyer, 2008: The syndrome of group-focused enmity. The interrelation of prejudices tested with multiple cross-sectional and panel data. Journal of Social Issues 64: 363–383.
- Zimbardo, P.G., R.J. Gerrig & R. Graf, 2008: Psychologie. München, Boston [u.a.]: Pearson Studium.

Anhang

Allbus 2006

Name	Allgemeine Bevölkerungsumfrage der Sozialwissenschaften 2006
Abkürzung	Allbus 2006
Erhebungsinstitut	Infratest im Auftrag von GESIS
Erhebungszeitraum	18.3.2006 - 21.08.2006
Grundgesamtheit	Alle erwachsenen Personen (Deutsche und Ausländer), die in der Bundesrepublik Deutschland (West und Ost) in Privathaushalten wohnen. Ausländische Personen wurden nur dann befragt, wenn das Interview in deutscher Sprache durchgeführt werden konnte.
Stichprobengröße	8424
Fallzahl	3421
Erhebungsmodus	Mündliche Interviews mit vollstrukturiertem Fragebogen (CAPI) und CASI-Befragung als „drop-off“ (ISSP)
Fragebogeninhalt	<ul style="list-style-type: none">- Schwerpunktthema: Einstellung gegenüber ethnischen Gruppen- Replikationsfragen: Familie, politische Einstellungen, subjektive Befindlichkeiten und Lagebeurteilung- Demografiefragen

Verwendete Skalen

Skala	Item
Marlow-Crowne Skala	Ich bin stets höflich, selbst zu Leuten, die ich abstoßend finde.
	Ich habe noch nie absichtlich etwas gesagt, um die Gefühle anderer zu verletzen.
	Manchmal bin ich ärgerlich auf Leute, die mich um einen Gefallen bitten.
	Vor einer Wahl informiere ich mich gründlich über die Eignung der verschiedenen Kandidaten.
	Manchmal bin ich beleidigt, wenn es nicht nach meinem Willen geht.
	Ich bin stets ein guter Zuhörer, gleichgültig, wer mein Gesprächspartner ist.
	Ich kann mich daran erinnern, dass ich schon einmal krank gespielt habe, um eine Pflicht zu umgehen.
	Bei Gelegenheit habe ich schon einmal jemanden ausgenutzt.
	Wenn ich einen Fehler gemacht habe, bin ich stets bereit, das zuzugeben.
	Ich halte mich immer selber an Grundsätze, deren Befolgung ich von anderen erwarte.

Antisemitismus

Juden haben auf der Welt zu viel Einfluss.

Durch ihr Verhalten sind die Juden an ihren Verfolgungen nicht ganz unschuldig.

Viele Juden versuchen, aus der Vergangenheit des

Dritten Reiches heute ihren Vorteil zu ziehen und die Deutschen dafür zahlen zu lassen.

Fremdenfeindlichkeit

Die in Deutschland lebenden Ausländer sollten ihren Lebensstil ein bisschen besser an den der Deutschen anpassen.

Wenn Arbeitsplätze knapp werden, sollte man die in Deutschland lebenden Ausländer wieder in ihre Heimat zurückschicken.

Quelle:

ZUMA – Methodenbericht 2007/09

Titel: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 2006

Autoren: Wasmer, Martina; Scholz, Evi; Blohm, Michael

Studierendenstichprobe der Universität Göttingen

Name	Studierendenstichprobe der Universität Göttingen
Verwendete Abkürzung	SSP
Erhebungsinstitute	Universität Göttingen
Erhebungszeitraum	19.11.2012-3.12.2012
Grundgesamtheit	Studierende des Studiums der Sozialwissenschaften im 1. bis 3. Semester
Fallzahl	908
Erhebungsmodus	CAWI Interview
Fragebogeninhalt	- Täuschungs- und Persönlichkeitsskalen - Vorurteile gegen Muslime, Juden, Ausländer und Langzeitarbeitslose

Verwendete Skalen

Skala	Item	Kürzel
SES-17	Manchmal werfe ich Müll einfach in die Landschaft oder Straße.	SES1
	Eigene Fehler gebe ich stets offen zu und ertrage gelassen etwaige negative Konsequenzen.	SES2
	Im Straßenverkehr nehme ich stets Rücksicht auf die anderen Verkehrsteilnehmer.	SES3
	Ich habe schon einmal Drogen (Tabletten, Haschisch oder ähnliches) konsumiert.	SES4
	Ich akzeptiere alle anderen Meinungen, auch wenn sie mit meiner eigenen nicht übereinstimmen.	SES5
	Meine Wut oder schlechte Laune lasse ich hin und wieder an unschuldigen oder schwächeren Leuten aus.	SES6
	Ich habe schon einmal jemanden ausgenutzt oder übers Ohr gehauen.	SES7
	In einem Gespräch lasse ich den anderen stets ausreden und höre ihm aufmerksam zu.	SES8
	Ich zögere niemals, jemandem in einer Notlage beizustehen.	SES9
	Wenn ich etwas versprochen habe, halte ich es ohne Wenn und Aber.	SES10
	Ich lästere gelegentlich über andere hinter deren Rücken.	SES11
	Ich würde niemals auf Kosten der Allgemeinheit leben.	SES12
	Ich bleibe immer freundlich und zuvorkommend anderen Leuten gegenüber,	SES13

	auch wenn ich selbst gestresst bin.	
	Im Streit bleibe ich stets sachlich und objektiv.	SES14
	Ich habe schon einmal geliehene Sachen nicht zurückgegeben.	SES15
	Ich ernähre mich stets gesund.	SES16
	Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte.	SES17
BIDR - Selbst- täuschung	Der erste Eindruck, den ich von anderen Menschen gewinne, bewahrheitet sich meistens.	ST1
	Ich bin immer mir selber gegenüber ganz ehrlich gewesen.	ST2
	Ich weiß immer, warum ich etwas mag.	ST3
	Es fällt mir schwer, einen beunruhigenden Gedanken beiseite zu drängen.	ST4
	Manchmal verpasse ich etwas, weil ich mich einfach nicht schnell genug entscheiden kann.	ST5
	Ich bin ein vollkommen rational denkender Mensch.	ST6
	Ich kann Kritik selten vertragen.	ST7
	Ich bin mir meiner Urteile sehr sicher.	ST8
	An meinen Fähigkeiten als Liebhaber habe ich schon gelegentlich gezweifelt.	ST9
	Ich weiß immer die Gründe für meine Handlungen.	ST10
BIDR - Fremd- täuschung	Manchmal lüge ich, wenn ich muss.	FT1
	Es ist schon einmal vorgekommen, dass ich jemanden ausgenutzt habe.	FT2
	Ich fluche niemals.	FT3
	Manchmal zahle ich es lieber anderen heim, als dass ich vergebe und vergesse.	FT4
	Ich habe schon einmal zuviel Wechselgeld herausbekommen, ohne es der Verkäuferin zu sagen.	FT5
	Ich gebe grundsätzlich alles an, was ich zu verzollen habe.	FT6
	Manchmal fahre ich schneller, als es erlaubt ist.	FT7
	Ich habe Dinge getan, von denen ich anderen nichts erzähle.	FT8
	Ich nehme niemals Dinge an mich, die mir nicht gehören.	FT9
	Ich bin schon einmal wegen einer angeblichen Krankheit nicht zur Arbeit oder Schule gegangen.	FT10
Selbstüber- wachung – Fähigkeit zur SÜ	Es fällt mir schwer, das Verhalten anderer Leute zu imitieren.	FZSÜ1
	Ich wäre vermutlich ein(e) guter Schauspieler/in.	FZSÜ2
	In einer Gruppe von Leuten bin ich selten im Mittelpunkt der Aufmerksamkeit.	FZSÜ3

	Wenn Spiele eine schauspielerische Leistung erfordern, bin ich nie besonders gut.	FZSÜ4
Selbstüberwachung – Personen Orientierung	Ich versuche nie, anderen zu gefallen.	POSÜ1
	Auch wenn mir eine Sache keinen Spaß macht, gebe ich oft vor, mich gut zu unterhalten.	POSÜ2
	Ich würde meine Auffassungen nie ändern, um Zuneigung von anderen zu erhalten.	POSÜ3
	Um voranzukommen und beliebt zu sein, neige ich dazu, so zu sein, wie man es von mir erwartet.	POSÜ4
Selbstüberwachung – Selbstkontrolle	Mein Verhalten ist für gewöhnlich der Ausdruck meiner tatsächlichen Gefühle.	SKSÜ1
	Gegenüber unterschiedlichen Menschen verhalte ich mich oft so, als wäre ich nicht die selbe Person.	SKSÜ2
	Ich bin nicht immer die Person, als die ich erscheine.	SKSÜ3
	Ich kann jemanden anlügen und dabei völlig ehrlich wirken.	SKSÜ4
Erwünschtheitswahrnehmung	Die meisten Langzeitarbeitslosen sind nicht wirklich daran interessiert, einen Job zu finden.	
	Ich finde es empörend, wenn sich die Langzeitarbeitslosen auf Kosten der Gesellschaft ein bequemes Leben machen.	
	Der Islam sollte in Deutschland verboten sein.	
	Es sollte besser gar keine Muslime in Deutschland geben.	
	Wenn Arbeitsplätze knapp werden, sollte man die in Deutschland lebenden Ausländer wieder in ihre Heimat zurückschicken.	
	Es leben zu viele Ausländer in Deutschland.	
	Juden haben in Deutschland zu viel Einfluss.	
	Durch ihr Verhalten sind die Juden an ihren Verfolgungen mitschuldig.	
Umfrageeinstellung	Umfragen sind sehr wichtig für Wissenschaft, Politik und Wirtschaft.	
	Umfragen bringen Abwechslung und sind interessant.	
	Die Ergebnisse von Umfragen sind in den meisten Fällen richtig.	
	Die Teilnehmer an Umfragen sagen meistens ihre wahre Meinung.	
Umfrageerfahrung	An wie vielen Umfragen haben Sie in den letzten 12 Monaten teilgenommen?	
Anonymität	Wie anonym (sicher und unbeobachtet) haben Sie die Befragungssituation empfunden?	
Antisemitismus	Juden haben in Deutschland zuviel Einfluss.	
	Durch ihr Verhalten sind die Juden an ihren Verfolgungen mitschuldig.	

**Fremden-
feindlichkeit** Wenn Arbeitsplätze knapp werden, sollte man die in Deutschland lebenden
Ausländer wieder in ihre Heimat zurückschicken.

Es leben zu viele Ausländer in Deutschland.

**Islamo-
phobie** Der Islam sollte in Deutschland verboten sein.

Es sollte besser gar keine Muslime in Deutschland geben.

**Langzeit-
arbeitslose** Die meisten Langzeitarbeitslosen sind nicht wirklich daran interessiert, einen Job
zu finden.

Ich finde es empörend, dass sich die Langzeitarbeitslosen auf Kosten der
Gesellschaft ein bequemes Leben machen.

Studierendenpanel der Universität Göttingen zum Thema G8/G9

Name	Studierendenpanel der Universität Göttingen Sozialwissenschaften und Chemie
Abkürzung	G8G9
Erhebungsinstitut	Universität Göttingen
Erhebungszeitraum	6.2.2012-22.2.2012
Grundgesamtheit	Teilnehmer einer Einführungsveranstaltung im Rahmen des Studiums der Sozialwissenschaften bzw. Chemie
Fallzahl	317
Erhebungsmodus	CAWI Interview
Fragebogeninhalt	- Studienbedingungen - Leistungsdruck - Bewertung des Unterschieds zwischen 8 und 9 Jahren Gymnasium

Verwendete Skalen

Skala	Item
SES-17	Ich habe schon einmal geliehene Sachen behalten.
	Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte.
	Manchmal werfe ich Müll einfach in die Landschaft oder auf die Straße.
BIDR - Selbsttäuschung	Mein erster Eindruck von Menschen stellt sich gewöhnlich als richtig heraus.
	Ich bin mir oft unsicher in meinem Urteil.
	Ich weiß immer genau, wieso ich etwas mag.
BIDR - Fremdtäuschung	Ich habe schon mal zu viel Wechselgeld zurückbekommen und es behalten.
	Ich bin immer ehrlich zu anderen.
	Ich habe gelegentlich mal jemanden ausgenutzt.
Antisemitismus	Juden haben in Deutschland zuviel Einfluss.
	Durch ihr Verhalten sind die Juden an ihren Verfolgungen mitschuldig.
Fremdenfeindlichkeit	Wenn Arbeitsplätze knapp werden, sollte man die in Deutschland lebenden Ausländer wieder in ihre Heimat zurückschicken.
	Es leben zu viele Ausländer in Deutschland.

Gruppenbezogene Menschenfeindlichkeit - Graduiertensurvey

Name	Gruppenbezogene Menschenfeindlichkeit Graduiertensurvey
Abkürzung	GMF Grad
Erhebungsinstitut	Universität Göttingen
Erhebungszeitraum	2009/2010
Grundgesamtheit	Repräsentative Bevölkerungsumfrage
Fallzahl	843
Erhebungsmodus	CATI Interview
Fragebogeninhalt	- Vorurteile gegenüber verschiedenen Minderheiten (Obdachlose, Langzeitarbeitslose, Muslime, Juden, Ausländer, Homosexuelle)

Verwendete Skalen

Skala	Item
SES-17	Ich habe schon einmal geliehene Sachen nicht zurückgegeben.
	Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte.
	Manchmal werfe ich Müll einfach in die Landschaft oder auf die Straße.
	Ich akzeptiere alle anderen Meinungen, auch wenn sie mit meiner eigenen nicht übereinstimmen.
	Ich zögere niemals, jemandem in einer Notlage beizustehen.
	Wenn ich etwas versprochen habe, halte ich es ohne Wenn und Aber.
	Ich lästere gelegentlich über andere hinter deren Rücken.
Antisemitismus	Juden haben in Deutschland zuviel Einfluss.
	Durch ihr Verhalten sind die Juden an ihren Verfolgungen mitschuldig.
Fremdenfeindlichkeit	Wenn Arbeitsplätze knapp werden, sollte man die in Deutschland lebenden Ausländer wieder in ihre Heimat zurückschicken.
	Es leben zu viele Ausländer in Deutschland.
Islamophobie	Muslimen sollte die Zuwanderung nach Deutschland untersagt werden.
	Muslimen sollte die Zuwanderung nach Deutschland untersagt werden.
Langzeitarbeitslose	Die meisten Langzeitarbeitslosen sind nicht wirklich daran interessiert, einen Job zu finden.
	Ich finde es empörend, wenn sich die Langzeitarbeitslosen auf Kosten der Gesellschaft ein bequemes Leben machen.

Gruppenbezogene Menschenfeindlichkeit – Pretest zum Graduiertensurvey

Name	Gruppenbezogene Menschenfeindlichkeit Graduiertensurvey Pretest
Abkürzung	GMF Pretest
Erhebungsinstitut	Universität Göttingen
Erhebungszeitraum	2009
Grundgesamtheit	Repräsentative Bevölkerungsumfrage
Fallzahl	228
Erhebungsmodus	CATI Interview
Fragebogeninhalt	- Vorurteile gegenüber verschiedenen Minderheiten (Obdachlose, Langzeitarbeitslose, Muslime, Juden, Ausländer, Homosexuelle)

Verwendete Skalen

Skala	Item
SES-17	Ich habe schon einmal geliehene Sachen nicht zurückgegeben.
	Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte.
	Manchmal werfe ich Müll einfach in die Landschaft oder auf die Straße.
	Ich akzeptiere alle anderen Meinungen, auch wenn sie mit meiner eigenen nicht übereinstimmen.
	Ich zögere niemals, jemandem in einer Notlage beizustehen.
	Wenn ich etwas versprochen habe, halte ich es ohne Wenn und Aber.
Ich lästere gelegentlich über andere hinter deren Rücken.	
Antisemitismus	Juden haben in Deutschland zuviel Einfluss.
	Durch ihr Verhalten sind die Juden an ihren Verfolgungen mitschuldig.
Fremdenfeindlichkeit	Wenn Arbeitsplätze knapp werden, sollte man die in Deutschland lebenden Ausländer wieder in ihre Heimat zurückschicken.
	Es leben zu viele Ausländer in Deutschland.
Islamophobie	Muslimen sollte die Zuwanderung nach Deutschland untersagt werden.
	Muslimen sollte die Zuwanderung nach Deutschland untersagt werden.
Langzeitarbeitslose	Die meisten Langzeitarbeitslosen sind nicht wirklich daran interessiert, einen Job zu finden.
	Ich finde es empörend, wenn sich die Langzeitarbeitslosen auf Kosten der Gesellschaft ein bequemeres Leben machen.

Erwünschtheitswahrnehmung Durch die vielen Muslime hier fühle ich mich manchmal wie ein Fremder im eigenen Land.

Muslimen sollte die Zuwanderung nach Deutschland untersagt werden.

Juden haben in Deutschland zu viel Einfluss.

Durch ihr Verhalten sind die Juden an ihren Verfolgungen mitschuldig.

Es leben zu viele Ausländer in Deutschland.

Wenn Arbeitsplätze knapp werden, sollte man die in Deutschland lebenden Ausländer wieder in ihre Heimat zurückschicken.

Subtiler Antisemitismus Wie oft haben Sie Sympathie für die hier lebenden Juden empfunden?

Wie oft haben Sie Bewunderung für die hier lebenden Juden empfunden?

Israelkritik Israel hat ein Recht auf Selbstverteidigung gegenüber den Palästinensern.

Access Onlinestichprobe

Name	Access Panel Stichprobe
Abkürzung	Websample
Erhebungsinstitute	Toluna, Webfrager, Respondi und Promio im Auftrag der Universität Göttingen
Erhebungszeitraum	2011
Grundgesamtheit	Repräsentative Bevölkerungsumfrage
Fallzahl	966
Erhebungsmodus	CAWI Interview
Fragebogeninhalt	- Vorurteile gegenüber verschiedenen Minderheiten - Religiosität - religiöser Fundamentalismus

Verwendete Skalen

Skala	Item
SES-17	Ich habe schon einmal geliehene Sachen behalten. Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte. Manchmal werfe ich Müll einfach in die Landschaft oder auf die Straße.
BIDR - Selbsttäuschung	Mein erster Eindruck von Menschen stellt sich gewöhnlich als richtig heraus. Ich bin mir oft unsicher in meinem Urteil. Ich weiß immer genau, wieso ich etwas mag.
BIDR - Fremdtäuschung	Ich habe schon mal zu viel Wechselgeld zurückbekommen und es behalten. Ich bin immer ehrlich zu anderen. Ich habe gelegentlich mal jemanden ausgenutzt.
Antisemitismus	Juden haben in Deutschland zuviel Einfluss. Durch ihr Verhalten sind die Juden an ihren Verfolgungen mitschuldig.
Islamophobie	Muslimen sollte die Zuwanderung nach Deutschland untersagt werden. Muslimen sollte die Zuwanderung nach Deutschland untersagt werden.

Tabellen Anhang

Für die Interpretation der Werte werden im Folgenden nochmals zusammenfassend die Skalierungen der einzelnen Variablen sowie ihre Bezeichnungen dargestellt.

Variablen- bezeichnung	Konstrukt	Skalierung
FF	Vorurteile gegenüber Ausländern (Fremdenfeindlichkeit)	0 = keine Zustimmung zu Vorurteilitems bis 1 = große Zustimmung zu Vorurteilitems
AS	Vorurteile gegenüber Juden (Antisemitismus)	
IP	Vorurteile gegenüber Muslimen (Islamophobie)	
LZA	Vorurteile gegenüber Langzeitarbeitslosen	
MC-Skala	Messung des Anerkennungsbedürfnisses nach Marlow und Crowne Skala	1 = niedriges Anerkennungsbedürfnis bis 4 = hohes Anerkennungsbedürfnis
FT-Skala	Freumdtäuschungsskala	
ST-Skala	Selbsttäuschungsskala	
SES17	Messung des Anerkennungsbedürfnisses nach Stöber	
ErWahrn	Erwünschtheitswahrnehmung	-1 = Vorurteile öffentlich erwünscht bis +1 = Vorurteile öffentlich unerwünscht
Anonymität	Wahrgenommene Anonymität	Modell nach Stocké: 0 = nicht anonym und 1 = anonym Modell Kombination: 0 = nicht anonym bis 1 = anonym
ErWahrn*Ano	Interkationseffekt aus Erwünschtheitswahrnehmung und Anonymität	
SES*ErWahrn	Interkationseffekt aus Erwünschtheitswahrnehmung und Anerkennungsbedürfnis	
SES*Ano	Interkationseffekt aus Anerkennungsbedürfnis und Anonymität	
SES*ErWahrn*Ano = Anreiz SD-Bias	Interkationseffekt aus Erwünschtheitswahrnehmung, Anonymität und Anerkennungsbedürfnis	
Umfrageeinstellung		1 = negative Umfrageeinstellung bis 4 = positive Umfrageeinstellung
Umfrageerfahrung		0 = an weniger als vier Umfragen teilgenommen und 1 = an vier oder mehr Umfragen teilgenommen

Anreiz*Einstellung	Interaktionseffekt aus SD-Anreiz und Umfrageeinstellung	
Anreiz*Erfahrung	Interaktionseffekt aus SD-Anreiz und Umfrageerfahrung	
Erfahrung*Einstellung	Interaktionseffekt aus Umfrageerfahrung und Einstellung	
Anreiz*Erf*Einst	Interaktionseffekt aus SD-Anreiz, Umfrageerfahrung und Einstellung	
Geschlecht		0 = männlich und 1 = weiblich
Alter		metrisch
Bildungsgrad niedrig		Haupt- und Volkshochschulabschluss, Abschluss an der polytechnischen Oberschule nach der 8. Klasse und Befragte ohne Schulabschluss.
Bildungsgrad mittel		Realschulabschluss bzw. einen Abschluss an der polytechnischen Oberschule nach der 10. Klasse.
Bildungsgrad hoch		Abitur, Fachhochschulabschluss und Universitätsabschluss

Tabelle A1: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (Allbus 2006)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,118	0,121		0,000	3,374	0,157		0,000
MC-Skala	0,048	0,016	0,053	0,003	-0,009	0,016	-0,010	0,589
Geschlecht					-0,375	0,058	-0,109	0,000
Alter					0,014	0,002	0,135	0,000
Bildung niedrig					0,472	0,071	0,134	0,000
Bildung mittel								
Bildung hoch					-0,535	0,075	-0,137	0,000
R2	0,002			0,003	0,099			0,000
N		3158				3158		

Tabelle A2: Lineare Regression der Fremdenfeindlichkeit auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (Allbus 2006)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	4,139	0,099		0,000	3,949	0,127		0,000
MC-Skala	0,057	0,013	0,076	0,000	0,015	0,013	0,020	0,274
Geschlecht					0,127	0,049	0,045	0,009
Alter					0,108	0,023	0,090	0,000
Bildung niedrig					0,322	0,059	0,111	0,000
Bildung mittel								
Bildung hoch					-0,526	0,063	-0,162	0,000
R2	0,006			0,000	0,080			0,000
N		3146				3146		

Tabelle A3: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der FT-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,282	0,175		0,000	1,698	0,233		0,000
FT-Skala	0,091	0,055	0,082	0,097	0,057	0,058	0,051	0,328
Geschlecht					-0,208	0,075	-0,136	0,006
Alter					0,002	0,002	0,037	0,484
Bildung niedrig					0,112	0,114	0,058	0,325
Bildung mittel								
Bildung hoch					-0,167	0,086	-0,111	0,052
R2	0,004			0,097	0,038			0,001
N		405				405		

Tabelle A4: Lineare Regression der Fremdenfeindlichkeit auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der FT-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,525	0,206		0,000	1,571	0,264		0,000
FT-Skala	0,185	0,064	0,138	0,004	0,086	0,065	0,064	0,190
Geschlecht					0,132	0,085	0,072	0,121
Alter					0,006	0,003	0,099	0,045
Bildung niedrig					0,107	0,127	0,046	0,398
Bildung mittel								
Bildung hoch					-0,494	0,096	-0,275	0,000
R2	0,017			0,004	0,116			0,000
N		424				424		

Tabelle A5: Lineare Regression der Islamophobie auf die Einflussfaktoren im Rahmen des RC-Modus von Marlow und Crowne mit der FT-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,372	0,201		0,000	1,876	0,266		0,000
FT-Skala	0,151	0,063	0,116	0,017	0,108	0,066	0,083	0,101
Geschlecht					-0,126	0,086	-0,071	0,141
Alter					0,000	0,003	0,004	0,931
Bildung niedrig					0,104	0,128	0,046	0,421
Bildung mittel								
Bildung hoch					-0,373	0,097	-0,213	0,000
R2	0,011			0,017	0,062			0,000
N		421				421		

Tabelle A6: Lineare Regression der Vorurteile gegenüber Langzeitarbeitslosen auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der FT-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,723	0,222		0,000	3,297	0,293		0,000
FT-Skala	-0,052	0,069	-0,036	0,455	-0,010	0,072	-0,007	0,895
Geschlecht					-0,046	0,094	-0,024	0,622
Alter					-0,010	0,003	-0,171	0,001
Bildung niedrig					0,185	0,141	0,075	0,190
Bildung mittel								
Bildung hoch					-0,233	0,107	-0,122	0,030
R2	-0,001			0,455	0,039			0,001
N		423				423		

Tabelle A7: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der SES17-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,301	0,236		0,000	1,740	0,268		0,000
SES-17	0,076	0,066	0,057	0,254	0,035	0,073	0,026	0,636
Geschlecht					-0,209	0,076	-0,137	0,006
Alter					0,002	0,003	0,042	0,445
Bildung niedrig					0,117	0,114	0,060	0,304
Bildung mittel								
Bildung hoch					-0,168	0,086	-0,112	0,051
R2	0,001			0,254	0,036			0,001
N		405				405		

Tabelle A8: Lineare Regression der Fremdenfeindlichkeit auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der SES17-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,524	0,279		0,000	1,756	0,305		0,000
SES-17	0,165	0,078	0,102	0,036	0,006	0,082	0,004	0,944
Geschlecht					0,137	0,085	0,075	0,110
Alter					0,007	0,003	0,118	0,023
Bildung niedrig					0,117	0,127	0,050	0,357
Bildung mittel								
Bildung hoch					-0,500	0,097	-0,278	0,000
R2	0,008			0,036	0,112			0,000
N		424				424		

Tabelle A9: Lineare Regression der Islamophobie auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der SES17-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,635	0,275		0,000	2,166	0,309		0,000
SES-17	0,059	0,077	0,037	0,447	-0,016	0,083	-0,010	0,850
Geschlecht					-0,116	0,086	-0,065	0,178
Alter					0,002	0,003	0,035	0,517
Bildung niedrig					0,115	0,129	0,051	0,371
Bildung mittel								
Bildung hoch					-0,384	0,098	-0,219	0,000
R2	-0,001			0,447	0,056			0,000
N		421				421		

Tabelle A10: Lineare Regression der Vorurteile gegenüber Langzeitarbeitslosen auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der SES17-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,708	0,298		0,000	3,141	0,338		0,000
SES-17	-0,042	0,084	-0,024	0,617	0,051	0,091	0,030	0,579
Geschlecht					-0,053	0,094	-0,027	0,578
Alter					-0,011	0,003	-0,185	0,001
Bildung niedrig					0,183	0,140	0,074	0,194
Bildung mittel								
Bildung hoch					-0,227	0,107	-0,119	0,034
R2	-0,002			0,617	0,039			0,001
N		423				423		

Tabelle A11: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der ST-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,509	0,178		0,000	1,818	0,250		0,000
ST-Skala	0,020	0,060	0,017	0,740	0,004	0,059	0,004	0,940
Geschlecht					-0,205	0,075	-0,134	0,007
Alter					0,002	0,002	0,053	0,294
Bildung niedrig					0,118	0,114	0,060	0,303
Bildung mittel								
Bildung hoch					-0,171	0,086	-0,114	0,047
R2	-0,002			0,740	0,036			0,002
N		405				405		

Tabelle A12: Lineare Regression der Fremdenfeindlichkeit auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der ST-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,758	0,205		0,000	1,578	0,278		0,000
ST-Skala	0,118	0,069	0,084	0,086	0,071	0,065	0,050	0,274
Geschlecht					0,135	0,085	0,074	0,112
Alter					0,006	0,003	0,114	0,016
Bildung niedrig					0,118	0,127	0,051	0,351
Bildung mittel								
Bildung hoch					-0,495	0,096	-0,275	0,000
R2	0,005			0,086	0,115			0,000
N		424				424		

Tabelle A13: Lineare Regression der Islamophobie auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der ST-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,334	0,199		0,000	1,721	0,280		0,000
ST-Skala	0,174	0,067	0,127	0,009	0,151	0,065	0,110	0,021
Geschlecht					-0,125	0,085	-0,070	0,143
Alter					0,001	0,003	0,020	0,689
Bildung niedrig					0,118	0,128	0,052	0,358
Bildung mittel								
Bildung hoch					-0,370	0,097	-0,211	0,000
R2	0,014			0,009	0,068			0,000
N		421				421		

Tabelle A14: Lineare Regression der Vorurteile gegenüber Langzeitarbeitslosen auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der ST-Skala (GMF Graduierten Survey)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,044	0,217		0,000	2,763	0,305		0,000
ST-Skala	0,177	0,073	0,118	0,015	0,189	0,072	0,126	0,009
Geschlecht					-0,053	0,093	-0,027	0,567
Alter					-0,011	0,003	-0,185	0,000
Bildung niedrig					0,187	0,139	0,076	0,180
Bildung mittel								
Bildung hoch					-0,217	0,106	-0,113	0,041
R2	0,012			0,015	0,054			0,000
N		423				423		

Tabelle A15: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der SES17-Skala (GMF Graduierten Pretest)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,947	0,429		0,000	2,628	0,451		0,000
SES-17	-0,119	0,128	-0,069	0,356	-0,240	0,136	-0,140	0,079
Geschlecht					-0,227	0,105	-0,150	0,033
Alter					0,003	0,003	0,084	0,293
Bildung niedrig					0,397	0,152	0,217	0,010
Bildung mittel								
Bildung hoch					-0,281	0,122	-0,193	0,022
R2	-0,001			0,356	0,137			0,000
N		182				182		

Tabelle A16: Lineare Regression der Fremdenfeindlichkeit auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der SES17-Skala (GMF Graduierten Pretest)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,146	0,499		0,023	1,620	0,526		0,002
SES-17	0,233	0,148	0,113	0,118	0,101	0,158	0,049	0,525
Geschlecht					-0,102	0,122	-0,056	0,404
Alter					0,002	0,004	0,049	0,522
Bildung niedrig					0,675	0,171	0,312	0,000
Bildung mittel								
Bildung hoch					-0,237	0,138	-0,136	0,087
R2	0,008			0,118	0,158			0,000
N		194				194		

Tabelle A17: Lineare Regression der Islamophobie auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der SES17-Skala (GMF Graduierten Pretest)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,764	0,487		0,000	2,391	0,523		0,000
SES-17	0,009	0,145	0,004	0,951	-0,138	0,157	-0,069	0,379
Geschlecht					-0,207	0,121	-0,117	0,090
Alter					0,004	0,004	0,074	0,348
Bildung niedrig					0,586	0,170	0,281	0,001
Bildung mittel								
Bildung hoch					-0,187	0,138	-0,111	0,177
R2	-0,005			0,951	0,121			0,000
N		191				191		

Tabelle A18: Lineare Regression der Vorurteile gegenüber Langzeitarbeitslosen auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der SES17-Skala (GMF Graduierten Pretest)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,823	0,529		0,000	2,885	0,571		0,000
SES-17	-0,076	0,158	-0,035	0,631	0,061	0,171	0,028	0,724
Geschlecht					0,018	0,132	0,009	0,894
Alter					-0,010	0,004	-0,185	0,020
Bildung niedrig					0,528	0,185	0,231	0,005
Bildung mittel								
Bildung hoch					-0,262	0,149	-0,144	0,080
R2	-0,004			0,631	0,103			0,000
N		193				193		

Tabelle A19: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der FT-Skala (Websample)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,449	0,147		0,000	1,808	0,182		0,000
FT-Skala	0,058	0,050	0,044	0,245	0,087	0,051	0,066	0,091
Geschlecht					-0,206	0,057	-0,138	0,000
Alter					-0,002	0,002	-0,037	0,350
Bildung niedrig					0,098	0,126	0,031	0,434
Bildung mittel								
Bildung hoch					-0,114	0,059	-0,076	0,055
R2	0,001			0,245	0,021			0,002
N		686				686		

Tabelle A20: Lineare Regression der Islamophobie auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der FT-Skala (Websample)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,323	0,178		0,000	2,842	0,217		0,000
FT-Skala	-0,071	0,061	-0,044	0,245	-0,019	0,062	-0,012	0,762
Geschlecht					-0,081	0,068	-0,045	0,231
Alter					-0,009	0,002	-0,142	0,000
Bildung niedrig					0,175	0,148	0,046	0,239
Bildung mittel								
Bildung hoch	2,323			0,000	-0,324	0,070	-0,178	0,000
R2	0,001			0,245	0,043			0,000
N		706				706		

Tabelle A21: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der SES17-Skala (Websample)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,135	0,158		0,000	2,411	0,184		0,000
SES-17	-0,155	0,047	-0,126	0,001	-0,156	0,050	-0,127	0,002
Geschlecht					-0,180	0,056	-0,121	0,001
Alter					0,001	0,002	0,025	0,540
Bildung niedrig					0,009	0,125	0,003	0,945
Bildung mittel								
Bildung hoch					-0,113	0,059	-0,076	0,056
R2	0,014			0,001	0,031			0,000
N		686				686		

Tabelle A22: Lineare Regression der Islamophobie auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der SES17-Skala (Websample)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,444	0,193		0,000	2,843	0,221		0,000
SES-17	-0,097	0,057	-0,065	0,087	-0,018	0,060	-0,012	0,767
Geschlecht					-0,084	0,068	-0,046	0,215
Alter					-0,008	0,002	-0,138	0,001
Bildung niedrig					0,174	0,148	0,046	0,241
Bildung mittel								
Bildung hoch					-0,325	0,070	-0,178	0,000
R2	0,003			0,087	0,043			0,000
N		705				705		

Tabelle A23: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der ST-Skala (Websample)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,536	0,172		0,000	1,960	0,215		0,000
ST-Skala	0,028	0,058	0,018	0,633	0,019	0,058	0,013	0,740
Geschlecht					-0,197	0,057	-0,132	0,001
Alter					-0,001	0,002	-0,023	0,547
Bildung niedrig					0,069	0,125	0,022	0,582
Bildung mittel								
Bildung hoch					-0,113	0,060	-0,076	0,058
R2	-0,001			0,633	0,017			0,005
N		686				686		

Tabelle A24: Lineare Regression der Islamophobie auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne mit der ST-Skala (Websample)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,401	0,207		0,000	2,131	0,254		0,000
ST-Skala	0,245	0,070	0,131	0,000	0,233	0,069	0,125	0,001
Geschlecht					-0,091	0,067	-0,050	0,174
Alter					-0,009	0,002	-0,152	0,000
Bildung niedrig					0,208	0,146	0,055	0,153
Bildung mittel								
Bildung hoch					-0,293	0,070	-0,161	0,000
R2	0,016			0,000	0,058			0,000
N		706				706		

Tabelle A25: Lineare Regression des Antisemitismus bei negativer Erwünschtheitswahrnehmung¹ auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	1,744	0,401		0,000	2,294	0,435		0,000
SES-17	-0,103	0,119	-0,078	0,392	-0,125	0,132	-0,095	0,346
Geschlecht					-0,215	0,101	-0,187	0,036
Alter					-0,001	0,003	-0,020	0,839
Bildung niedrig					0,208	0,151	0,149	0,172
Bildung mittel								
Bildung hoch					-0,220	0,123	-0,195	0,075
R2	-0,002			0,392	0,082			0,010
N	123				123			

Tabelle A26: Lineare Regression des Antisemitismus bei positiver Erwünschtheitswahrnehmung² auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,319	1,009		0,026	3,054	1,007		0,004
SES-17	-0,146	0,303	-0,070	0,633	-0,337	0,311	-0,162	0,284
Geschlecht					-0,299	0,249	-0,160	0,237
Alter					0,008	0,008	0,146	0,352
Bildung niedrig					0,821	0,373	0,346	0,033
Bildung mittel								
Bildung hoch					-0,233	0,272	-0,131	0,397
R2	-0,016			0,633	0,169			0,022
N	49				49			

Tabelle A27: Lineare Regression der Fremdenfeindlichkeit bei negativer Erwünschtheitswahrnehmung auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,319	1,009		0,026	3,054	1,007		0,004
SES-17	-0,146	0,303	-0,070	0,633	-0,337	0,311	-0,162	0,284
Geschlecht					-0,299	0,249	-0,160	0,237
Alter					0,008	0,008	0,146	0,352
Bildung niedrig					0,821	0,373	0,346	0,033
Bildung mittel								
Bildung hoch					-0,233	0,272	-0,131	0,397
R2	-0,016			0,633	0,169			0,022
N	49				49			

1 Die negative Erwünschtheitswahrnehmung umfasst Befragte die Antisemitismus als unerwünscht in der Gesellschaft wahrnehmen.

2 Die positive Erwünschtheitswahrnehmung umfasst Befragte die Antisemitismus als erwünscht in der Gesellschaft wahrnehmen.

Tabelle A28: Lineare Regression der Fremdenfeindlichkeit bei positiver Erwünschtheitswahrnehmung auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	0,471	0,639		0,462	1,053	0,666		0,116
SES-17	0,459	0,192	0,208	0,018	0,261	0,204	0,118	0,204
Geschlecht					-0,087	0,156	-0,046	0,577
Alter					0,003	0,005	0,055	0,555
Bildung niedrig					0,702	0,210	0,323	0,001
Bildung mittel								
Bildung hoch					-0,147	0,172	-0,083	0,396
R2	0,036			0,018	0,153			0,000
N		129				129		

Tabelle A29: Lineare Regression der Islamophobie bei negativer Erwünschtheitswahrnehmung auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,241	0,671		0,001	3,123	0,755		0,000
SES-17	-0,182	0,197	-0,105	0,359	-0,232	0,220	-0,134	0,295
Geschlecht					-0,261	0,173	-0,171	0,135
Alter					-0,246	0,253	-0,130	0,333
Bildung niedrig					0,000	0,006	-0,004	0,974
Bildung mittel								
Bildung hoch	2,241			0,001	-0,431	0,196	-0,289	0,031
R2	-0,002			0,359	0,038			0,167
N		79				79		

Tabelle A30: Lineare Regression der Islamophobie bei positiver Erwünschtheitswahrnehmung auf die Einflussfaktoren im Rahmen des RC-Modells von Marlow und Crowne (GMF Graduierten Pretest)

	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,241	0,671		0,001	3,123	0,755		0,000
SES-17	-0,182	0,197	-0,105	0,359	-0,232	0,220	-0,134	0,295
Geschlecht					-0,261	0,173	-0,171	0,135
Alter					-0,246	0,253	-0,130	0,333
Bildung niedrig					0,000	0,006	-0,004	0,974
Bildung mittel								
Bildung hoch	2,241			0,001	-0,431	0,196	-0,289	0,031
R2	-0,002			0,359	0,038			0,167
N		79				79		

Tabellen A31: Lineare Regression Fremdenfeindlichkeit auf die Einflussfaktoren im Rahmen des RC-Modells über die SES17-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	B	St.Err.	Beta	Sign.
Konstante	3,354	0,284		0,000
SES-17	0,398	0,233	0,074	0,088
Alter	0,005	0,010	0,022	0,623
Geschlecht	-0,104	0,054	-0,084	0,057
adj. R ² (F-Test)	0,007			0,077
N	538			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,399	0,283		0,000	3,464	0,283		0,000
SES-17	0,318	0,233	0,059	0,173	0,232	0,236	0,043	0,327
ErWahrn	0,121	0,047	0,112	0,010	0,103	0,047	0,095	0,029
Anonymität	-0,093	0,085	-0,047	0,277	-0,201	0,095	-0,102	0,034
ErWahrn*Ano					-0,037	0,029	-0,060	0,195
SES*ErWahrn					-0,026	0,027	-0,042	0,336
SES*Ano					-0,084	0,031	-0,130	0,006
SES*ErWahrn*Ano					-0,024	0,028	-0,040	0,385
Alter	0,005	0,010	0,020	0,650	0,004	0,010	0,018	0,685
Geschlecht	-0,101	0,054	-0,081	0,063	-0,096	0,054	-0,078	0,075
adj. R ² (F-Test)	0,019			0,009	0,027			0,005
N	538				538			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,377	0,282		0,000	3,381	0,282		0,000
SES-17	0,336	0,233	0,062	0,150	0,330	0,233	0,061	0,158
ErWahrn	0,126	0,047	0,116	0,007	0,126	0,047	0,116	0,007
SES*ErWahrn					-0,014	0,027	-0,023	0,594
Alter	0,005	0,010	0,020	0,646	0,005	0,010	0,021	0,636
Geschlecht	-0,101	0,054	-0,082	0,062	-0,102	0,054	-0,083	0,060
adj. R ² (F-Test)	0,019			0,007	0,017			0,013
N	538				538			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,433	0,284		0,000	3,509	0,284		0,000
SES-17	0,325	0,233	0,060	0,163	0,225	0,234	0,042	0,337
ErWahrn	0,122	0,047	0,113	0,009	0,104	0,046	0,096	0,026
Anonymität	-0,154	0,102	-0,065	0,132	-0,209	0,104	-0,087	0,045
ErWahrn*Ano					-0,019	0,027	-0,030	0,485
SES*ErWahrn					-0,025	0,027	-0,040	0,351
SES*Ano					-0,095	0,027	-0,150	0,001
SES*ErWahrn*Ano					-0,028	0,027	-0,044	0,299
Alter	0,005	0,010	0,020	0,644	0,005	0,010	0,020	0,639
Geschlecht	-0,103	0,054	-0,083	0,057	-0,103	0,054	-0,083	0,057
adj. R ² (F-Test)	0,021			0,006	0,038			0,000
N	538				538			

Tabellen A32: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells über die SES17-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	B	St.Err.	Beta	Sign.
Konstante	3,889	0,166		0,000
SES-17	-0,020	0,133	-0,007	0,881
Alter	-0,003	0,006	-0,026	0,563
Geschlecht	0,053	0,031	0,075	0,093
adj. R ² (F-Test)	0,001			0,304
N	523			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,822	0,168		0,000	3,822	0,168		0,000
SES-17	-0,041	0,134	-0,013	0,761	-0,051	0,134	-0,017	0,704
ErWahrn	0,078	0,035	0,098	0,026	0,075	0,036	0,094	0,035
Anonymität	0,012	0,050	0,010	0,815	-0,020	0,052	-0,018	0,695
ErWahrn*Ano					-0,018	0,016	-0,048	0,284
SES*ErWahrn					0,000	0,016	0,000	0,998
SES*Ano					-0,018	0,017	-0,049	0,283
SES*ErWahrn*Ano					-0,053	0,021	-0,115	0,013
Alter	-0,002	0,006	-0,016	0,721	-0,002	0,006	-0,018	0,687
Geschlecht	0,052	0,031	0,074	0,096	0,059	0,031	0,085	0,058
adj. R ² (F-Test)	0,007			0,123	0,016			0,045
N	523				523			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,824	0,168		0,000	3,823	0,168		0,000
SES-17	-0,043	0,133	-0,014	0,744	-0,048	0,133	-0,016	0,721
ErWahrn	0,078	0,035	0,098	0,026	0,081	0,036	0,102	0,023
SES*ErWahrn					0,008	0,016	0,022	0,623
Alter	-0,002	0,006	-0,016	0,721	-0,002	0,006	-0,016	0,715
Geschlecht	0,052	0,031	0,074	0,095	0,053	0,031	0,076	0,089
adj. R ² (F-Test)	0,009			0,071	0,007			0,115
N	523				523			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,822	0,170		0,000	3,866	0,169		0,000
SES-17	-0,043	0,133	-0,014	0,747	-0,095	0,133	-0,031	0,478
ErWahrn	0,079	0,035	0,099	0,026	0,074	0,035	0,093	0,036
Anonymität	0,005	0,059	0,004	0,935	-0,005	0,059	-0,004	0,936
ErWahrn*Ano					-0,025	0,015	-0,071	0,110
SES*ErWahrn					0,002	0,016	0,004	0,926
SES*Ano					-0,033	0,016	-0,092	0,038
SES*ErWahrn*Ano					-0,036	0,016	-0,101	0,024
Alter	-0,002	0,006	-0,016	0,722	-0,003	0,006	-0,019	0,667
Geschlecht	0,052	0,031	0,074	0,095	0,051	0,031	0,073	0,101
adj. R ² (F-Test)	0,007			0,125	0,022			0,014
N	523				523			

Tabellen A33: Lineare Regression der Islamophobie auf die Einflussfaktoren im Rahmen des RC-Modells über die SES17-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	B	St.Err.	Beta	Sign.
Konstante	3,962	0,193		0,000
SES-17	0,171	0,158	0,047	0,279
Alter	-0,009	0,007	-0,059	0,179
Geschlecht	-0,027	0,037	-0,033	0,460
adj. R ² (F-Test)	0,000			0,374
N	538			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,966	0,194		0,000	3,975	0,195		0,000
SES-17	0,154	0,159	0,042	0,331	0,161	0,160	0,044	0,315
ErWahrn	0,021	0,033	0,027	0,529	0,013	0,034	0,018	0,691
Anonymität	-0,039	0,059	-0,029	0,506	-0,054	0,063	-0,040	0,388
ErWahrn*Ano					-0,003	0,020	-0,007	0,892
SES*ErWahrn					0,011	0,018	0,026	0,553
SES*Ano					-0,002	0,020	-0,005	0,908
SES*ErWahrn*Ano					-0,031	0,021	-0,071	0,142
Alter	-0,009	0,007	-0,059	0,184	-0,010	0,007	-0,063	0,156
Geschlecht	-0,027	0,037	-0,032	0,467	-0,026	0,037	-0,031	0,485
adj. R ² (F-Test)	-0,002			0,554	-0,004			0,647
N	538				538			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,956	0,194		0,000	3,961	0,194		0,000
SES-17	0,163	0,158	0,045	0,303	0,161	0,158	0,044	0,310
ErWahrn	0,021	0,033	0,028	0,520	0,023	0,033	0,030	0,496
SES*ErWahrn					0,014	0,018	0,033	0,444
Alter	-0,009	0,007	-0,058	0,186	-0,009	0,007	-0,060	0,173
Geschlecht	-0,027	0,037	-0,032	0,463	-0,026	0,037	-0,031	0,483
adj. R ² (F-Test)	-0,001			0,474	-0,002			0,533
N	538				538			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,983	0,196		0,000	3,998	0,196		0,000
SES-17	0,157	0,158	0,043	0,321	0,138	0,160	0,038	0,389
ErWahrn	0,020	0,033	0,026	0,543	0,015	0,034	0,019	0,659
Anonymität	-0,069	0,070	-0,042	0,328	-0,071	0,071	-0,044	0,318
ErWahrn*Ano					-0,004	0,019	-0,009	0,837
SES*ErWahrn					0,011	0,018	0,027	0,536
SES*Ano					-0,023	0,019	-0,054	0,224
SES*ErWahrn*Ano					-0,015	0,019	-0,035	0,424
Alter	-0,009	0,007	-0,058	0,185	-0,009	0,007	-0,060	0,174
Geschlecht	-0,028	0,037	-0,033	0,448	-0,027	0,037	-0,033	0,463
adj. R ² (F-Test)	-0,001			0,482	-0,003			0,607
N	538				538			

Tabellen A34: Lineare Regression der negativen Einstellungen gegenüber Langzeitarbeitslosen auf die Einflussfaktoren im Rahmen des RC-Modells über die SES17-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	B	St.Err.	Beta	Sign.
Konstante	2,591	0,354		0,000
SES-17	-0,454	0,288	-0,069	0,116
Alter	0,024	0,012	0,084	0,057
Geschlecht	-0,095	0,067	-0,063	0,154
adj. R ² (F-Test)	0,013			0,021
N	524			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,568	0,356		0,000	2,627	0,356		0,000
SES-17	-0,424	0,290	-0,064	0,144	-0,507	0,292	-0,077	0,084
ErWahrn	-0,054	0,058	-0,041	0,348	-0,079	0,059	-0,060	0,180
Anonymität	0,055	0,108	0,022	0,615	-0,040	0,118	-0,016	0,735
ErWahrn*Ano					-0,072	0,040	-0,087	0,074
SES*ErWahrn					0,002	0,035	0,002	0,966
SES*Ano					-0,052	0,038	-0,064	0,175
SES*ErWahrn*Ano					-0,072	0,042	-0,083	0,083
Alter	0,023	0,013	0,083	0,062	0,023	0,013	0,081	0,069
Geschlecht	-0,096	0,067	-0,063	0,154	-0,094	0,067	-0,062	0,159
adj. R ² (F-Test)	0,011			0,052	0,014			0,058
N	524				524			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,582	0,354		0,000	2,580	0,355		0,000
SES-17	-0,434	0,289	-0,066	0,133	-0,428	0,289	-0,065	0,140
ErWahrn	-0,056	0,058	-0,043	0,329	-0,055	0,058	-0,041	0,343
SES*ErWahrn					0,016	0,034	0,021	0,639
Alter	0,023	0,013	0,083	0,063	0,023	0,013	0,082	0,065
Geschlecht	-0,095	0,067	-0,063	0,154	-0,095	0,067	-0,063	0,157
adj. R ² (F-Test)	0,013			0,030	0,011			0,053
N	524				524			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,589	0,357		0,000	2,669	0,359		0,000
SES-17	-0,435	0,289	-0,066	0,133	-0,542	0,293	-0,082	0,065
ErWahrn	-0,057	0,058	-0,043	0,325	-0,065	0,058	-0,049	0,266
Anonymität	-0,020	0,128	-0,007	0,875	-0,055	0,130	-0,019	0,673
ErWahrn*Ano					-0,040	0,035	-0,050	0,257
SES*ErWahrn					0,005	0,035	0,006	0,888
SES*Ano					-0,054	0,035	-0,068	0,123
SES*ErWahrn*Ano					-0,053	0,036	-0,065	0,140
Alter	0,023	0,013	0,083	0,063	0,023	0,013	0,081	0,070
Geschlecht	-0,095	0,067	-0,063	0,154	-0,093	0,067	-0,062	0,163
adj. R ² (F-Test)	0,011			0,057	0,014			0,059
N	524				524			

Tabellen A35: Lineare Regression der Fremdenfeindlichkeit auf die Einflussfaktoren im Rahmen des RC-Modells über die Kurzversion der SES17-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	B	St.Err.	Beta	Sign.
Konstante	3,147	0,266		0,000
SES-17_k	0,735	0,170	0,188	0,000
Alter	0,002	0,010	0,010	0,812
Geschlecht	-0,049	0,055	-0,040	0,369
adj. R ² (F-Test)	0,036			0,000
N	538			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,166	0,265		0,000	3,210	0,266		0,000
SES-17_k	0,698	0,170	0,179	0,000	0,628	0,175	0,161	0,000
ErWahrn	0,118	0,046	0,109	0,010	0,099	0,047	0,092	0,034
Anonymität	-0,075	0,084	-0,038	0,374	-0,175	0,098	-0,088	0,075
ErWahrn*Ano					-0,033	0,029	-0,054	0,253
SES*ErWahrn					-0,012	0,028	-0,019	0,662
SES*Ano					-0,089	0,034	-0,134	0,009
SES*ErWahrn*Ano					-0,019	0,032	-0,028	0,562
Alter	0,002	0,010	0,009	0,836	0,002	0,010	0,009	0,840
Geschlecht	-0,050	0,054	-0,040	0,359	-0,049	0,054	-0,039	0,370
adj. R ² (F-Test)	0,046			0,000	0,052			0,000
N	538				538			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,149	0,264		0,000	3,149	0,265		0,000
SES-17_k	0,710	0,169	0,182	0,000	0,712	0,170	0,182	0,000
ErWahrn	0,122	0,046	0,112	0,008	0,121	0,046	0,112	0,009
SES*ErWahrn					0,003	0,027	0,005	0,903
Alter	0,002	0,010	0,009	0,835	0,002	0,010	0,009	0,838
Geschlecht	-0,049	0,054	-0,040	0,363	-0,049	0,054	-0,040	0,366
adj. R ² (F-Test)	0,046			0,000	0,045			0,000
N	538				538			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,200	0,267		0,000	3,241	0,270		0,000
SES-17_k	0,699	0,169	0,179	0,000	0,650	0,174	0,167	0,000
ErWahrn	0,119	0,046	0,110	0,010	0,107	0,046	0,099	0,021
Anonymität	-0,140	0,101	-0,058	0,167	-0,157	0,104	-0,066	0,132
ErWahrn*Ano					-0,007	0,027	-0,012	0,784
SES*ErWahrn					-0,006	0,028	-0,009	0,840
SES*Ano					-0,048	0,028	-0,076	0,085
SES*ErWahrn*Ano					-0,012	0,028	-0,020	0,655
Alter	0,002	0,010	0,009	0,831	0,002	0,010	0,009	0,837
Geschlecht	-0,052	0,054	-0,042	0,341	-0,056	0,055	-0,045	0,305
adj. R ² (F-Test)	0,048			0,000	0,046			0,000
N	538				538			

Tabellen A36: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells über die Kurzversion der SES17-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	B	St.Err.	Beta	Sign.
Konstante	3,790	0,158		0,000
SES-17_k	0,138	0,099	0,062	0,167
Alter	-0,004	0,006	-0,028	0,524
Geschlecht	0,061	0,032	0,088	0,054
adj. R ² (F-Test)	0,005			0,137
N	523			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,718	0,161		0,000	3,729	0,161		0,000
SES-17_k	0,129	0,100	0,058	0,196	0,108	0,101	0,048	0,288
ErWahrn	0,075	0,035	0,094	0,033	0,072	0,035	0,090	0,041
Anonymität	0,017	0,050	0,015	0,727	-0,003	0,051	-0,003	0,950
ErWahrn*Ano					-0,010	0,016	-0,028	0,534
SES*ErWahrn					-0,002	0,017	-0,005	0,902
SES*Ano					-0,020	0,018	-0,052	0,255
SES*ErWahrn*Ano					-0,025	0,016	-0,070	0,120
Alter	-0,002	0,006	-0,019	0,675	-0,003	0,006	-0,020	0,662
Geschlecht	0,060	0,032	0,085	0,060	0,063	0,032	0,091	0,047
adj. R ² (F-Test)	0,010			0,068	0,009			0,129
N	523				523			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,721	0,160		0,000	3,720	0,161		0,000
SES-17_k	0,127	0,099	0,057	0,203	0,128	0,100	0,058	0,201
ErWahrn	0,075	0,035	0,094	0,032	0,075	0,035	0,094	0,033
SES*ErWahrn	-0,002	0,006	-0,019	0,675	-0,002	0,017	-0,005	0,901
Alter	0,060	0,032	0,085	0,060	-0,002	0,006	-0,018	0,681
Geschlecht	3,721			0,000	0,060	0,032	0,085	0,060
adj. R ² (F-Test)	0,012			0,038	0,010			0,071
N	523				523			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,718	0,162		0,000	3,741	0,163		0,000
SES-17_k	0,127	0,099	0,057	0,202	0,096	0,102	0,043	0,346
ErWahrn	0,075	0,035	0,095	0,032	0,075	0,035	0,094	0,035
Anonymität	0,008	0,059	0,006	0,891	0,002	0,060	0,001	0,973
ErWahrn*Ano					-0,009	0,016	-0,024	0,594
SES*ErWahrn					-0,006	0,017	-0,015	0,731
SES*Ano					-0,019	0,016	-0,052	0,245
SES*ErWahrn*Ano					-0,024	0,016	-0,067	0,131
Alter	-0,002	0,006	-0,019	0,677	-0,003	0,006	-0,020	0,658
Geschlecht	0,060	0,032	0,085	0,060	0,059	0,032	0,085	0,062
adj. R ² (F-Test)	0,010			0,071	0,010			0,122
N	523				523			

Tabellen A37: Lineare Regression Islamophobie auf die Einflussfaktoren im Rahmen des RC-Modells über die Kurzversion der SES17-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	B	St.Err.	Beta	Sign.
Konstante	3,944	0,183		0,000
SES-17_k	0,198	0,118	0,075	0,092
Alter	-0,010	0,007	-0,064	0,149
Geschlecht	-0,012	0,038	-0,014	0,758
adj. R ² (F-Test)	0,003			0,188
N	538			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,944	0,184		0,000	3,971	0,184		0,000
SES-17_k	0,189	0,118	0,071	0,110	0,176	0,119	0,066	0,139
ErWahrn	0,020	0,033	0,027	0,537	0,001	0,034	0,001	0,977
Anonymität	-0,037	0,059	-0,028	0,525	-0,071	0,061	-0,052	0,246
ErWahrn*Ano					-0,013	0,020	-0,032	0,515
SES*ErWahrn					0,006	0,018	0,015	0,743
SES*Ano					-0,026	0,021	-0,055	0,223
SES*ErWahrn*Ano					-0,046	0,020	-0,114	0,021
Alter	-0,010	0,007	-0,063	0,155	-0,011	0,007	-0,068	0,122
Geschlecht	-0,012	0,038	-0,014	0,748	-0,009	0,038	-0,010	0,816
adj. R ² (F-Test)	0,001				0,349			
N	538				538			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,936	0,184		0,000	3,938	0,184		0,000
SES-17_k	0,194	0,118	0,073	0,100	0,196	0,118	0,074	0,097
ErWahrn	0,021	0,033	0,027	0,528	0,021	0,033	0,027	0,536
SES*ErWahrn					0,009	0,017	0,024	0,586
Alter	-0,010	0,007	-0,063	0,156	-0,010	0,007	-0,064	0,150
Geschlecht	-0,012	0,038	-0,014	0,754	-0,012	0,038	-0,014	0,757
adj. R ² (F-Test)	0,002				0,296			
N	538				538			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,962	0,186		0,000	3,980	0,186		0,000
SES-17_k	0,189	0,118	0,071	0,109	0,172	0,120	0,065	0,152
ErWahrn	0,020	0,033	0,026	0,550	0,014	0,034	0,018	0,684
Anonymität	-0,067	0,070	-0,041	0,344	-0,064	0,071	-0,040	0,364
ErWahrn*Ano					-0,003	0,019	-0,007	0,865
SES*ErWahrn					0,007	0,017	0,019	0,670
SES*Ano					-0,020	0,019	-0,046	0,294
SES*ErWahrn*Ano					-0,023	0,018	-0,056	0,199
Alter	-0,010	0,007	-0,063	0,155	-0,010	0,007	-0,065	0,145
Geschlecht	-0,013	0,038	-0,016	0,728	-0,016	0,038	-0,019	0,679
adj. R ² (F-Test)	0,002				0,299			
N	538				538			

Tabellen A38: Lineare Regression der negativen Einstellungen gegenüber Langzeitarbeitslosen auf die Einflussfaktoren im Rahmen des RC-Modells über die Kurzversion der SES17-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	B	St.Err.	Beta	Sign.
Konstante	2,056	0,335		0,000
SES-17_k	0,422	0,212	0,089	0,047
Alter	0,022	0,012	0,079	0,073
Geschlecht	-0,075	0,068	-0,050	0,270
adj. R ² (F-Test)	0,016			0,011
N	524			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,038	0,336		0,000	2,106	0,334		0,000
SES-17_k	0,440	0,213	0,092	0,039	0,311	0,215	0,065	0,149
ErWahrn	-0,063	0,058	-0,048	0,275	-0,082	0,058	-0,062	0,154
Anonymität	0,080	0,108	0,032	0,461	-0,025	0,113	-0,010	0,823
ErWahrn*Ano					-0,057	0,037	-0,068	0,129
SES*ErWahrn					-0,050	0,035	-0,064	0,147
SES*Ano					-0,110	0,040	-0,130	0,006
SES*ErWahrn*Ano					-0,111	0,044	-0,119	0,012
Alter	0,022	0,013	0,078	0,080	0,022	0,012	0,079	0,073
Geschlecht	-0,074	0,068	-0,049	0,280	-0,075	0,067	-0,050	0,268
adj. R ² (F-Test)	0,015			0,023	0,031			0,003
N	524				524			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,054	0,335		0,000	2,040	0,335		0,000
SES-17_k	0,430	0,212	0,090	0,043	0,425	0,212	0,089	0,046
ErWahrn	-0,066	0,058	-0,050	0,250	-0,072	0,058	-0,054	0,214
SES*ErWahrn					-0,043	0,036	-0,052	0,236
Alter	0,022	0,013	0,077	0,081	0,023	0,013	0,080	0,072
Geschlecht	-0,074	0,068	-0,049	0,276	-0,074	0,068	-0,049	0,276
adj. R ² (F-Test)	0,016			0,014	0,017			0,016
N	524				524			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,059	0,338		0,000	2,158	0,339		0,000
SES-17_k	0,430	0,213	0,090	0,044	0,325	0,215	0,068	0,131
ErWahrn	-0,067	0,058	-0,050	0,249	-0,068	0,058	-0,051	0,239
Anonymität	-0,013	0,128	-0,004	0,921	-0,028	0,128	-0,009	0,829
ErWahrn*Ano					-0,029	0,035	-0,036	0,414
SES*ErWahrn					-0,044	0,035	-0,056	0,206
SES*Ano					-0,073	0,035	-0,094	0,036
SES*ErWahrn*Ano					-0,073	0,037	-0,090	0,048
Alter	0,022	0,013	0,077	0,081	0,021	0,012	0,074	0,094
Geschlecht	-0,074	0,068	-0,049	0,276	-0,085	0,068	-0,056	0,214
adj. R ² (F-Test)	0,014			0,028	0,023			0,013
N	524				524			

Tabellen A39: Lineare Regression der Fremdenfeindlichkeit auf die Einflussfaktoren im Rahmen des RC-Modells über die FT-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	B	St.Err.	Beta	Sign.
Konstante	3,402	0,258		0,000
FT	0,345	0,131	0,114	0,008
Alter	0,006	0,010	0,027	0,534
Geschlecht	-0,115	0,054	-0,093	0,036
adj. R ² (F-Test)	0,015			0,012
N	538			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,411	0,257		0,000	3,397	0,258		0,000
FT	0,324	0,130	0,107	0,013	0,326	0,131	0,108	0,013
ErWahrn	0,124	0,046	0,115	0,008	0,124	0,046	0,114	0,008
Anonymität	-0,086	0,085	-0,043	0,312	-0,119	0,089	-0,060	0,179
ErWahrn*Ano					-0,018	0,027	-0,030	0,505
SES*ErWahrn					-0,021	0,027	-0,033	0,440
SES*Ano					-0,036	0,028	-0,058	0,190
SES*ErWahrn*Ano					0,015	0,026	0,025	0,575
Alter	0,006	0,010	0,025	0,568	0,006	0,010	0,025	0,562
Geschlecht	-0,112	0,054	-0,091	0,039	-0,105	0,054	-0,085	0,053
adj. R ² (F-Test)	0,027			0,002	0,026			0,007
N	538				538			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,395	0,256		0,000	3,395	0,256		0,000
FT	0,333	0,130	0,110	0,011	0,331	0,130	0,110	0,011
ErWahrn	0,128	0,046	0,118	0,006	0,128	0,046	0,118	0,006
SES*ErWahrn					-0,016	0,027	-0,026	0,544
Alter	0,006	0,010	0,025	0,562	0,006	0,010	0,026	0,557
Geschlecht	-0,113	0,054	-0,091	0,038	-0,113	0,054	-0,091	0,038
adj. R ² (F-Test)	0,027			0,001	0,026			0,002
N	538				538			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,445	0,258		0,000	3,464	0,258		0,000
FT	0,335	0,130	0,111	0,010	0,332	0,130	0,110	0,011
ErWahrn	0,125	0,046	0,115	0,007	0,121	0,046	0,112	0,009
Anonymität					-0,196	0,104	-0,082	0,059
ErWahrn*Ano					-0,020	0,027	-0,031	0,476
SES*ErWahrn					-0,020	0,027	-0,031	0,466
SES*Ano					-0,060	0,027	-0,095	0,029
SES*ErWahrn*Ano	-0,161	0,102	-0,067	0,114	0,005	0,027	0,007	0,867
Alter	0,006	0,010	0,025	0,560	0,005	0,010	0,022	0,613
Geschlecht	-0,115	0,054	-0,093	0,034	-0,108	0,054	-0,088	0,046
adj. R ² (F-Test)	0,030			0,001	0,033			0,001
N	538				538			

Tabellen A40: Lineare Regression des Antisemitismus auf die Einflussfaktoren im Rahmen des RC-Modells über die FT-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	Beta		Sign.	
Konstante	3,932	0,149		0,000
FT	-0,068	0,076	-0,039	0,372
Alter	-0,004	0,006	-0,032	0,468
Geschlecht	0,049	0,032	0,069	0,119
adj. R ² (F-Test)	0,002		0,275	
N	545			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	Beta		Sign.		Beta		Sign.	
Konstante	3,864	0,152	0,000		3,860	0,152	0,000	
FT	-0,075	0,076	-0,043	0,323	-0,067	0,078	-0,038	0,390
ErWahn	0,074	0,035	0,090	0,037	0,076	0,036	0,092	0,039
Anonymität	0,002	0,049	0,002	0,967	-0,010	0,052	-0,008	0,854
ErWahn*Ano					-0,007	0,019	-0,019	0,716
SES*ErWahn					0,015	0,018	0,038	0,399
SES*Ano					-0,042	0,019	-0,115	0,024
SES*ErWahn*Ano					-0,017	0,024	-0,041	0,463
Alter	-0,003	0,006	-0,024	0,591	-0,004	0,006	-0,027	0,536
Geschlecht	0,049	0,031	0,068	0,122	0,053	0,031	0,074	0,092
adj. R ² (F-Test)	0,006		0,143		0,018		0,028	
N	545				545			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	Beta		Sign.		Beta		Sign.	
Konstante	3,864	0,152	0,000		3,862	0,152	0,000	
FT	-0,076	0,076	-0,043	0,320	-0,079	0,076	-0,045	0,301
ErWahn	0,074	0,035	0,090	0,036	0,077	0,036	0,094	0,031
SES*ErWahn					0,011	0,017	0,029	0,509
Alter	-0,003	0,006	-0,024	0,590	-0,003	0,006	-0,023	0,596
Geschlecht	0,049	0,031	0,068	0,121	0,049	0,031	0,069	0,117
adj. R ² (F-Test)	0,008		0,082		0,007		0,122	
N	545				545			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	Beta		Sign.		Beta		Sign.	
Konstante	3,861	0,154	0,000		3,871	0,152	0,000	
FT	-0,076	0,076	-0,043	0,319	-0,071	0,075	-0,040	0,350
ErWahn	0,074	0,035	0,090	0,036	0,079	0,035	0,096	0,026
Anonymität	0,009	0,059	0,006	0,885	0,019	0,059	0,014	0,750
ErWahn*Ano					-0,034	0,016	-0,093	0,038
SES*ErWahn					0,011	0,017	0,027	0,532
SES*Ano					-0,026	0,016	-0,070	0,111
SES*ErWahn*Ano					-0,051	0,017	-0,133	0,004
Alter	-0,003	0,006	-0,023	0,593	-0,004	0,006	-0,027	0,528
Geschlecht	0,049	0,031	0,069	0,120	0,042	0,031	0,059	0,178
adj. R ² (F-Test)	0,006		0,142		0,028		0,004	
N	545				545			

Tabellen A41: Lineare Regression der Islamophobie auf die Einflussfaktoren im Rahmen des RC-Modells über die FT-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	Beta		Sign.	
Konstante	4,032	0,170		0,000
FT	0,051	0,087	0,025	0,561
Alter	-0,008	0,007	-0,055	0,206
Geschlecht	-0,034	0,036	-0,040	0,355
adj. R ² (F-Test)	-0,001		0,498	
N	560			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	Beta		Sign.		Beta		Sign.	
Konstante	4,032	0,171	0,000		4,041	0,172	0,000	
FT	0,045	0,087	0,022	0,608	0,039	0,088	0,019	0,657
ErWahrn	0,020	0,032	0,026	0,543	0,015	0,033	0,019	0,653
Anonymität	-0,046	0,057	-0,035	0,417	-0,048	0,057	-0,036	0,407
ErWahrn*Ano					-0,002	0,018	-0,004	0,933
SES*ErWahrn					0,025	0,019	0,055	0,200
SES*Ano					-0,005	0,018	-0,010	0,807
SES*ErWahrn*Ano					-0,017	0,020	-0,039	0,385
Alter	-0,008	0,007	-0,055	0,208	-0,008	0,007	-0,056	0,196
Geschlecht	-0,033	0,036	-0,040	0,363	-0,032	0,036	-0,039	0,374
adj. R ² (F-Test)	-0,003		0,636		-0,005		0,743	
N	560				560			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	Beta		Sign.		Beta		Sign.	
Konstante	4,022	0,171	0,000		4,023	0,171	0,000	
FT	0,050	0,087	0,025	0,567	0,045	0,087	0,022	0,608
ErWahrn					0,020	0,032	0,026	0,544
SES*ErWahrn	0,020	0,032	0,026	0,533	0,025	0,019	0,055	0,197
Alter	-0,008	0,007	-0,054	0,215	-0,008	0,007	-0,054	0,216
Geschlecht	-0,034	0,036	-0,040	0,354	-0,033	0,036	-0,039	0,366
adj. R ² (F-Test)	-0,002		0,599		-0,001		0,489	
N	560				560			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	Beta		Sign.		Beta		Sign.	
Konstante	4,046	0,173	0,000		4,023	0,175	0,000	
FT	0,051	0,087	0,025	0,557	0,054	0,088	0,026	0,540
ErWahrn	0,019	0,032	0,025	0,556	0,023	0,033	0,030	0,482
Anonymität					-0,053	0,070	-0,033	0,451
ErWahrn*Ano					0,020	0,018	0,049	0,270
SES*ErWahrn					0,028	0,019	0,062	0,149
SES*Ano					-0,002	0,019	-0,005	0,913
SES*ErWahrn*Ano	-0,066	0,068	-0,041	0,334	0,021	0,018	0,054	0,225
Alter	-0,008	0,007	-0,055	0,208	-0,007	0,007	-0,049	0,260
Geschlecht	-0,035	0,036	-0,042	0,339	-0,036	0,037	-0,043	0,333
adj. R ² (F-Test)	-0,002		0,594		-0,005		0,732	
N	560				560			

Tabellen A42: Lineare Regression der negativen Einstellungen gegenüber Langzeitarbeitslosen auf die Einflussfaktoren im Rahmen des RC-Modells über die FT-Skala

RC-Modell nach Marlow und Crowne

RC-Variablen	Modell 1			
	Beta		Sign.	
Konstante	2,224	0,319		0,000
FT	0,170	0,160	0,046	0,289
Alter	0,026	0,012	0,091	0,037
Geschlecht	-0,134	0,066	-0,089	0,044
adj. R ² (F-Test)	0,014		0,014	
N	540			

RC-Modell nach Stocké

RC-Variablen	Modell 1				Modell 2			
	Beta		Sign.		Beta		Sign.	
Konstante	2,208	0,320		0,000	2,176	0,322		0,000
FT	0,181	0,160	0,049	0,258	0,186	0,162	0,050	0,252
ErWahrn	-0,070	0,057	-0,053	0,221	-0,069	0,058	-0,052	0,240
Anonymität	0,085	0,106	0,035	0,422	0,077	0,114	0,031	0,501
ErWahrn*Ano					-0,038	0,039	-0,047	0,322
SES*ErWahrn					0,025	0,034	0,033	0,454
SES*Ano					-0,003	0,036	-0,004	0,925
SES*ErWahrn*Ano					0,027	0,037	0,035	0,468
Alter	0,026	0,012	0,090	0,040	0,027	0,012	0,093	0,033
Geschlecht	-0,134	0,066	-0,089	0,044	-0,129	0,066	-0,086	0,052
adj. R ² (F-Test)	0,015		0,025		0,013		0,069	
N	540				540			

RC-Modell nach Skarbek-Kozietulska et al.

RC-Variablen	Modell 1				Modell 2			
	Beta		Sign.		Beta		Sign.	
Konstante	2,227	0,319		0,000	2,222	0,319		0,000
FT	0,171	0,160	0,046	0,283	0,177	0,160	0,048	0,268
ErWahrn	-0,073	0,057	-0,055	0,202	-0,071	0,057	-0,054	0,211
SES*ErWahrn					0,028	0,033	0,036	0,400
Alter	0,025	0,012	0,089	0,042	0,025	0,012	0,089	0,042
Geschlecht	-0,133	0,066	-0,088	0,046	-0,132	0,066	-0,088	0,046
adj. R ² (F-Test)	0,015		0,016		0,015		0,024	
N	540				540			

RC-Modell nach Lischewski

RC-Variablen	Modell 1				Modell 2			
	Beta		Sign.		Beta		Sign.	
Konstante	2,229	0,322		0,000	2,194	0,323		0,000
FT	0,171	0,160	0,046	0,283	0,187	0,160	0,051	0,244
ErWahrn	-0,073	0,057	-0,055	0,202	-0,072	0,057	-0,054	0,212
Anonymität	-0,007	0,126	-0,003	0,953	0,004	0,127	0,001	0,977
ErWahrn*Ano					-0,024	0,035	-0,030	0,500
SES*ErWahrn					0,028	0,033	0,037	0,391
SES*Ano					0,046	0,034	0,059	0,176
SES*ErWahrn*Ano					0,073	0,035	0,093	0,037
Alter	0,025	0,012	0,089	0,043	0,026	0,012	0,092	0,035
Geschlecht	-0,133	0,066	-0,088	0,046	-0,131	0,066	-0,087	0,048
adj. R ² (F-Test)	0,013		0,032		0,020		0,021	
N	540				540			

Tabellen A43: Lineare Regression zu den Einflussfaktoren im Rahmen des Frame-Selektion-Modells auf die Fremdenfeindlichkeit

RC-Variablen	Modell nach Stocké				Modell nach Lischewski			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,778	0,285		0,000	3,752	0,288		0,000
Anreiz SD-Bias	-0,026	0,032	-0,040	0,417	0,000	0,029	0,000	0,992
Umfrageeinstellung	-0,064	0,055	-0,051	0,245	-0,054	0,055	-0,043	0,327
Umfrageerfahrung	0,000	0,006	0,003	0,950	0,000	0,006	-0,001	0,981
Anreiz*Einstellung	-0,093	0,030	-0,162	0,002	-0,052	0,027	-0,088	0,058
Anreiz*Erfahrung	0,017	0,031	0,026	0,587	0,004	0,032	0,006	0,900
Erfahrung*Einstellung	-0,058	0,029	-0,090	0,044	-0,063	0,029	-0,098	0,028
Anreiz*Erf*Einst	0,015	0,025	0,030	0,550	-0,032	0,029	-0,055	0,268
Alter	0,005	0,010	0,023	0,600	0,005	0,010	0,023	0,597
Geschlecht	-0,098	0,054	-0,080	0,071	-0,101	0,055	-0,082	0,064
adj. R ² (F-Test)	0,019			0,026	0,009			0,135
N	538				538			

Tabellen A44: Lineare Regression zu den Einflussfaktoren im Rahmen des Frame-Selektion-Modells auf den Antisemitismus

RC-Variablen	Modell nach Stocké				Modell nach Lischewski			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	3,970	0,168		0,000	4,010	0,167		0,000
Anreiz SD-Bias	-0,028	0,017	-0,076	0,095	-0,030	0,016	-0,083	0,065
Umfrageeinstellung	-0,032	0,032	-0,045	0,318	-0,041	0,032	-0,058	0,198
Umfrageerfahrung	0,001	0,036	0,002	0,974	0,002	0,036	0,002	0,965
Anreiz*Einstellung	-0,011	0,018	-0,029	0,529	-0,030	0,016	-0,086	0,057
Anreiz*Erfahrung	0,010	0,017	0,028	0,564	-0,022	0,018	-0,056	0,219
Erfahrung*Einstellung	-0,009	0,017	-0,024	0,602	-0,013	0,017	-0,035	0,442
Anreiz*Erf*Einst	0,001	0,017	0,002	0,965	0,000	0,018	-0,001	0,980
Alter	-0,004	0,006	-0,030	0,508	-0,004	0,006	-0,032	0,475
Geschlecht	0,057	0,031	0,081	0,071	0,051	0,031	0,074	0,102
adj. R ² (F-Test)	-0,002			0,514	0,007			0,187
N	523				523			

Tabellen A45: Lineare Regression zu den Einflussfaktoren im Rahmen des Frame-Selektion-Modells auf die Islamophobie

RC-Variablen	Modell nach Stocké				Modell nach Lischewski			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	4,014	0,195		0,000	4,021	0,195		0,000
Anreiz SD-Bias	-0,038	0,018	-0,095	0,031	-0,019	0,018	-0,047	0,286
Umfrageeinstellung	0,023	0,037	0,027	0,537	0,033	0,037	0,039	0,371
Umfrageerfahrung	0,018	0,042	0,019	0,666	-0,011	0,004	-0,115	0,010
Anreiz*Einstellung	-0,014	0,015	-0,039	0,370	-0,028	0,017	-0,072	0,111
Anreiz*Erfahrung	0,008	0,019	0,018	0,682	-0,010	0,020	-0,024	0,608
Erfahrung*Einstellung	-0,029	0,019	-0,067	0,135	-0,015	0,019	-0,034	0,440
Anreiz*Erf*Einst	-0,019	0,016	-0,053	0,229	-0,014	0,020	-0,033	0,481
Alter	-0,010	0,007	-0,064	0,149	-0,009	0,007	-0,056	0,205
Geschlecht	-0,027	0,037	-0,032	0,466	-0,036	0,037	-0,043	0,331
adj. R ² (F-Test)	0,005			0,239	0,014			0,062

Tabellen A46: Lineare Regression zu den Einflussfaktoren im Rahmen des Frame-Selektion-Modells auf die negativen Einstellungen gegenüber Langzeitarbeitslosen

RC-Variablen	Modell nach Stocké				Modell nach Lischewski			
	B	St.Err.	Beta	Sign.	B	St.Err.	Beta	Sign.
Konstante	2,533	0,358		0,000	2,517	0,357		0,000
Anreiz SD-Bias	-0,058	0,047	-0,062	0,216	-0,060	0,039	-0,074	0,123
Umfrageeinstellung	-0,076	0,068	-0,050	0,259	-0,072	0,067	-0,047	0,286
Umfrageerfahrung	0,027	0,077	0,016	0,723	0,005	0,010	0,023	0,603
Anreiz*Einstellung	-0,002	0,045	-0,003	0,958	-0,040	0,033	-0,055	0,234
Anreiz*Erfahrung	-0,009	0,050	-0,010	0,858	-0,001	0,046	-0,001	0,985
Erfahrung*Einstellung	-0,033	0,036	-0,042	0,354	-0,032	0,035	-0,041	0,365
Anreiz*Erf*Einst	-0,013	0,035	-0,026	0,708	0,000	0,034	0,001	0,991
Alter	0,024	0,013	0,084	0,060	0,023	0,013	0,083	0,062
Geschlecht	-0,101	0,068	-0,067	0,135	-0,104	0,067	-0,069	0,124
adj. R ² (F-Test)	0,008			0,112	0,008			0,150
N		524				524		

Tabellen A47: Darstellung der unterschiedlichen Operationalisierungen der zentralen RC-Variablen

Variable	Ursprüngliche Codierung	Umcodierung nach Stocké
Anerkennungsbedürfnis	Wertebereich von 1="niedriges Anerkennungsbedürfnis" bis 4="hohes Anerkennungsbedürfnis"	Festlegung des neuen Wertebereichs: 1=0 2=0.33 3=0.66 4=1 → 0="niedriges Anerkennungsbedürfnis" bis 1="hohes Anerkennungsbedürfnis" → Umcodierung aller Items einzeln und anschließende Indexbildung aus dem Mittelwert
Erwünschtheitswahrnehmung	Wertebereich von 1="Vorurteile öffentlich erwünscht" bis 5="Vorurteile öffentlich unerwünscht"	Festlegung des neuen Wertebereichs: 1=-1 1.5=-0.75 2=-0.5 2.5=-0.25 3=0 3.5=0.25 4=0.5 4.5=0.75 5=1 → -1="Vorurteil öffentlich erwünscht" bis +1="Vorurteile öffentlich unerwünscht"
Anonymität	Wertebereich von 1="anonym" bis 5="nicht anonym"	Festlegung des neuen Wertebereichs: 1=0 2=0 3=0 4=1 5=1 → 0="anonym" 1="nicht anonym"

Tabellen A47: Darstellung der unterschiedlichen Operationalisierungen der zentralen RC-Variablen

Variablen	Erklärungsmodelle			
	Marlow und Crowne	Stocké	Skarbek-Kozietulska et al.	Lischewski
Anerkennungsbedürfnis	Marlow-Crowne-Skala	Kurzversion der MC-Skala	Kurzversion der MC-Skala + 1 Item der BIDR-Skala	SES17-Skala SES17_K-Skala FT-Skala
Erwünschtheitswahrnehmung	-	Persönliche Bewertung	Extrinsische Bewertung	Extrinsische Bewertung
Anonymität	-	Experimentell	-	Wahrgenommene Anonymität

Eidesstattliche Erklärung

„Ich versichere, dass ich die eingereichte Dissertation „Soziale Erwünschtheit im Licht des RC-Ansatzes“ selbstständig und ohne unerlaubte Hilfsmittel verfasst habe. Anderer als der von mir angegebenen Hilfsmittel und Schriften habe ich mich nicht bedient. Alle wörtlich oder sinngemäß den Schriften anderer Autoren entnommenen Stellen habe ich kenntlich gemacht.“