MANNHEIMER
ZENTRUM FÜR
EUROPÄISCHE
SOZIALFORSCHUNG

Stereotype und Vorurteile:

Konzeptualisierung, Operationalisierung und Messung

Stephan Ganter

Arbeitsbereich III / Nr. 22
Mannheim 1997

ISSN 0948-0099

Working papers

Arbeitspapiere

Stephan Ganter

Stereotype und Vorurteile:

Konzeptualisierung, Operationalisierung und Messung

Stephan Ganter:
Stereotype und Vorurteile : Konzeptualisierung, Operationalisierung und Messung.
Mannheimer Zentrum für Europäische Sozialforschung (MZES).
Mannheim, 1997 (überarbeitete Version).
(Arbeitspapiere Arbeitsbereich III / 22)
ISSN 0948-0099

Deckblattlayout: Uwe Freund

Nicht im Buchhandel erhältlich
Schutzgebühr: DM 5,--
Bezug:
Mannheimer Zentrum für Europäische Sozialforschung (MZES), Postfach, 68131 Mannheim

Redaktionelle Notiz:

Stephan Ganter ist wissenschaftlicher Mitarbeiter am Mannheimer Zentrum für Europäische Sozialforschung (MZES). Im Rahmen des Forschungsschwerpunktes „Migration, Integration, ethnische Konflikte“ des Arbeitsbereichs III des MZES bearbeitet er ein Projekt zum Thema „Determinanten und Prozesse ethnischer Grenzziehung“.

Zusammenfassung
Bereits seit mehreren Jahrzehnten werden in den Sozialwissenschaften vielfältige und umfangreiche Anstrengungen unternommen, die Inhalte, die Verbreitung und die sozialen Konsequenzen von Stereotypisierungen und Vorurteilen gegenüber Angehörigen ethnisch differenzierter Gruppen zu erfassen und zu erklären, da davon ausgegangen wird, daß diesen Formen sozialer Distanzierung eine zentrale Bedeutung bei der Strukturierung interethnischer Beziehungen zukommt. In auffälligem Kontrast zur weitverbreiteten Einschätzung der Bedeutsamkeit von Stereotypen und Vorurteilen stehen jedoch zahlreiche Schwierigkeiten und Unstimmigkeiten in der Konzeptualisierung, Operationalisierung und Messung dieser Konstrukte. Ziel des Arbeitspapiers ist es, in einer Übersicht über den aktuellen Forschungsstand Möglichkeiten und Probleme einer angemessenen Konzeptualisierung, Operationalisierung und Messung zu klären. Da sowohl der Stereotypen- als auch der Vorurteilsbegriff in der sozialwissenschaftlichen Forschung häufig nur sehr diffus und unheitlich spezifiziert ist, werden zunächst, ausgehend von einer knappen Zusammenfassung der wichtigsten Konzeptualisierungsvarianten, Vorschläge zu einer methodologisch angemessenen Begriffsbestimmung dargestellt und die zentralen Dimensionen, auf die sich die Begriffe beziehen, herausgearbeitet. Im Anschluß daran werden die wichtigsten Methoden der Operationalisierung und Messung von Stereotypen und Vorurteilen dargestellt und einer kritischen Einschätzung unterzogen. Den Anfang bilden jeweils die "traditionellen" Methoden, also diejenigen Operationalisierungs- und Messverfahren, die bereits seit langem auch die Bevölkerungsumfragen eingesetzt werden und gewissermaßen den "Standard" darstellen. Dann folgt eine Bestandsaufnahme der wichtigsten Vorschläge zur Ergänzung oder Modifikation dieser Verfahren, die vor allem in Auseinandersetzung mit den Problemen der Anfälligkeit direkter Befragungsmethoden für die sogenannten "social-desirability-effects" entwickelt wurden. Daraus abzuleitende Konsequenzen werden abschließend kurz zusammengefaßt.

Inhalt

1.Einleitung ***1***

2.Stereotype und Vorurteile: Konzepte, Operationalisierungen und Messverfahren ***2***

2.1 Stereotype **2**

2.1.1 Der Begriff des Stereotyps 2

2.1.2 Operationalisierungen und Messverfahren 6

2.1.2.1 „Traditionelle“ Methoden 6

- (a) Die Katz/Braly-Methode 6
- (b) Die Brigham-Methode und die „diagnostic ratio“ 8
- (c) Die Methodik des „semantischen Differentials“ 11
- (d) Zwischenfazit 13

2.1.1.1 Bogus-Pipeline und neuere Paradigmen 14

- (a) Das Bogus-Pipeline-Paradigma 14
- (b) Neuere Paradigmen 17

2.2 Vorurteile **21**

2.2.1 Der Begriff des Vorurteils 21

2.2.2 Operationalisierungen und Messverfahren 24

2.2.2.1 Modifikationen der Einstellungsskalen und Ratingverfahren 25

- (a) Die „Modern Racism“-Skala 25
- (b) Die „Subtle Prejudice“-Skala: 28
- (c) Einschätzung der modifizierten Einstellungsskalen 29

2.2.2.2 „Unauffällige“ und indirekte Messverfahren 33

- (a) Die Kombination von Priming und Reaktionszeitenmessung 33
- (b) Verhaltensindikatoren 37

Zusammenfassung und Schluß ***39***

Literaturverzeichnis ***42***

1.Einleitung

Nicht erst seit der in jüngster Zeit vermehrt zu beobachtenden konflikthaften Aufladung interethnischer Beziehungen und den auch in Westeuropa immer wieder festzustellenden fremdenfeindlichen Tendenzen stehen Stereotypisierungen, Vorurteile und andere Formen sozialer Distanzierung gegenüber Angehörigen ethnisch differenzierter Gruppen im Zentrum sozialwissenschaftlichen Interesses. Vor allem in der Soziologie und in der Sozialpsychologie werden vielfältige und umfangreiche Anstrengungen unternommen, die Inhalte, die Verbreitung und die sozialen Konsequenzen von Stereotypisierungen und Vorurteilen zu erfassen und zu erklären, da davon ausgegangen wird, daß diese Formen sozialer Distanzierung zentrale Bestimmungsgründe auch des Verhaltens gegenüber den jeweiligen Personengruppen sind. In auffälligem Kontrast zur weitverbreiteten Einschätzung der Bedeutsamkeit von Stereotypen und Vorurteilen stehen jedoch zahlreiche Schwierigkeiten und Unstimmigkeiten in der Konzeptualisierung, Operationalisierung und Messung dieser Konstrukte[1].

Vor diesem Hintergrund gibt dieses Arbeitspapier eine kritische Bestandsaufnahme der Möglichkeiten und Probleme einer angemessenen Konzeptualisierung, Operationalisierung und Messung von Stereotypen und Vorurteilen. Da sowohl der Stereotypen- als auch der Vorurteilsbegriff in der sozialwissenschaftlichen Forschung häufig nur sehr diffus und uneinheitlich spezifiziert ist, werden zunächst, ausgehend von einer knappen Zusammenfassung der wichtigsten Konzeptualisierungsvarianten, Vorschläge zu einer methodologisch angemessenen Begriffsbestimmung dargestellt und die zentralen Dimensionen, auf die sich die Begriffe beziehen, herausgearbeitet. Nur so kann schließlich geklärt werden, mit Hilfe welcher Indikatoren die latenten Konstrukte mit beobachtbaren Daten in Beziehung zu setzen sind. Im Anschluß daran werden die wichtigsten Methoden der Operationalisierung und Messung von Stereotypen und Vorurteilen dargestellt und einer kritischen Einschätzung unterzogen. Den Anfang bilden jeweils die „traditionellen“ Methoden, also diejenigen Operationalisierungs- und Messverfahren, die bereits seit langem und am häufigsten in der empirischen Forschung eingesetzt werden und gewissermaßen den „Standard“ darstellen. Dann folgt eine kritische Bestandsaufnahme der wichtigsten Vorschläge zur Ergänzung oder Modifikation dieser Verfahren, die größtenteils in sozialpsychologischen Studien erarbeitet wurden.

Wie sich zeigen wird, zielen diese Ergänzungs- und Modifikationsvorschläge in erster Linie auf ein grundlegendes Problem der empirischen Erfassung von Stereotypen und Vorurteilen: die offensichtliche soziale Brisanz der angesprochenen Meinungen und Bewertungen und die damit zusammenhängende Anfälligkeit „gewöhnlicher“ Erhebungsmethoden insbesondere - aber keineswegs ausschließlich - für

[1] Der Arbeitsbericht entstand im Zuge der Vorbereitung eines Forschungsprojekts zum Thema „Determinanten und Prozesse ethnischer Grenzziehung“, das im Rahmen des Forschungsschwerpunktes „Migration, Integration, ethnische Konflikte“ (MIC) unter der Leitung von Prof. Dr. Hartmut Esser am Arbeitsbereich III des MZES bearbeitet wird. Während sich der vorliegende Bericht in erster Linie auf eher „technische“ Aspekte beschränkt, wird in einem zweiten Arbeitspapier zum Thema „Determinanten ethnischer Grenzziehung“ eine ausführliche Darstellung der zentralen mikroanalytischen Ansätze zur Erklärung solcher sozialen Distanzierungsprozesse gegeben (MZES Working Paper AB III/No.21). Dort findet sich auch eine eingehendere Einordnung des Stellenwerts von Stereotypisierungen und Vorurteilen im Kontext der Analyse interethnischer Beziehungen und ethnischer Konflikte.

Verzerrungen durch die Tendenz zur Abgabe sozial erwünschter Antworten. Diese Problematik ist zwar zweifellos seit langem bekannt, wird jedoch in den meisten Erhebungen im Rahmen der Stereotypen- und Vorurteilsforschung ausgeblendet. Nicht zuletzt deshalb werden in dieser Übersicht relativ ausführlich einige Studien vorgestellt, in denen - mit interessanten Resultaten - der Versuch unternommen wurde, das Ausmaß der Anfälligkeit direkter Befragungsmethoden für die sogenannten „social-desirability-effects“ genauer zu bestimmen. Darüber hinaus können aus diesen Studien wichtige Hinweise auf mögliche Verbesserungen der Meßinstrumentarien gewonnen werden. Daraus abzuleitende Konsequenzen werden im Schlußkapitel knapp zusammengefaßt.

Der Bericht beschränkt sich explizit auf die Konzeptualisierung, Operationalisierung und Messung von Stereotypen und Vorurteilen. Die für die Untersuchung ethnischer Grenzziehungen bzw. sozialer Distanzierungen gegenüber ethnisch differenzierten Gruppen ebenfalls relevanten Konzepte wie „Ethnozentrismus“ oder „Fremdenfeindlichkeit“ werden hier bewußt ausgeklammert. Diese Konzepte sind keineswegs identisch mit Stereotypen oder Vorurteilen, doch ihre Operationalisierung und Messung - sowie die damit verbundenen Probleme - sind in den „technischen“ Details weitgehend ähnlich; die wesentlichen Unterschiede liegen in der inhaltlichen Spezifikation der Konstrukte. Die im folgenden dargestellten Verfahren und die Einschätzungen ihrer Vorzüge und Nachteile lassen sich also im großen und ganzen auch auf die Operationalisierung und Messung etwa von „Ethnozentrismus“ oder „Fremdenfeindlichkeit“ übertragen. Dasselbe sollte für die Hinweise auf mögliche Verbesserungen der Messverfahren gelten.

2.Stereotype und Vorurteile: Konzepte, Operationalisierungen und Messverfahren

2.1 Stereotype

2.1.1 Der Begriff des Stereotyps

Der Begriff des Stereotyps gehört sicherlich zu den besonders problematischen und umstrittenen Begriffen der Sozialwissenschaften. Seit Walter Lippmann (1922) den Begriff von einem terminus technicus der Druckersprache in ein Konzept für die Analyse sozialer Phänomene umgemünzt hat, ist er gekennzeichnet durch mangelnde Präzision und starke evaluative Konnotationen (Gardner 1994: 1-5). Schon in Lippmanns „Public Opinion“ war der Stereotypenbegriff als „pictures in our heads“, die als abstrakte Repräsentationen einer überkomplexen Umwelt Teil einer Art subjektiven „pseudo-environments“ seien, eher metaphorisch und vage gefaßt. Und dort fand sich bereits ein Verständnis von Stereotypen vorgezeichnet, demzufolge diese zwar für die kognitive Orientierung in der „world outside“ als unumgänglich, zugleich aber auch als sehr mangelhaft anzusehen seien, denn: „it precedes the use of reason; is a form of perception, imposes a certain character on the data of our senses before the data reach the intelligence“ (Lippmann 1922: 98).

Die Bestimmung des Stereotypenkonzepts unter Bezug auf mangelhafte Prozesse der Meinungs- und Urteilsbildung hat zahlreiche konzeptionelle Ansätze in der Stereotypenforschung geprägt. Stereotype waren demnach definiert durch „inkorrekte Generalisierungen“, „Generalisierungen mit nicht angebbarer Validität“, „Generalisierungen mit Rigiditätscharakter“ oder auch als „Produkte eines fehlerhaften Denkprozesses“ (Brigham 1971: 17-19; Ashmore/Del Boca 1981: 12-16). Begriffsbestimmungen, die sich

auf solche Charakterisierungen stützen, sind mit der alltagssprachlichen Verwendung des Stereotypenbegriffs leicht in Einklang zu bringen und reflektieren darüber hinaus offensichtlich Überzeugungen oder Vermutungen über als unerwünscht betrachtete Auswirkungen von Stereotypisierungen auf soziales Verhalten. Für Zwecke sozialwissenschaftlicher Beschreibung und Erklärung sind diese Ansätze der Konzeptspezifikation indes ungeeignet.

Die Bezugnahme auf Merkmale wie „sachlich unkorrekte und starre Übergeneralisierungen“ oder „Ergebnisse eines fehlerhaften Denkprozesses“ als Bestandteile des Definiens erweist sich allein schon deshalb als unzweckmäßig, weil die damit verbundenen Werturteile („unkorrekt“, „starr“, „fehlerhaft“, „falsch“ usw.) keine klare und einfache Präzisierung des Konzeptes erlauben. Die Frage, inwiefern Stereotype und Stereotypisierungen sinnvoll mit solchen Eigenschaften bezeichnet werden können, wäre auf der Grundlage einer brauchbaren Operationalisierung empirisch zu beantworten und kann nicht einfach definitorisch als geklärt vorausgesetzt werden (Simpson/Yinger 1985: 97). Ob eine solche empirische Klärung allerdings überhaupt möglich ist, wird häufig grundsätzlich bezweifelt, da allgemein akzeptierte Vergleichs- und Beurteilungsstandards für den zumeist implizit behaupteten „wahren“ Zusammenhang zwischen Stereotypen und Realität oder auch für „korrekte“ Denkprozesse kaum zu formulieren seien (Ashmore/Del Boca 1981: 16f; Gardner 1994: 12; Rehm 1986: 18-27; Stroebe 1985: 3f; vgl. Brigham 1971: 31f). Außerdem tragen Vorschläge zur Definition von Stereotypen, die in erster Linie Merkmale wie „Übergeneralisierung“ oder „Fehlbeurteilungen“ in den Mittelpunkt stellen, wenig zur Klärung bei, da solche Charakteristika vermutlich für alle Urteilsprozesse zutreffen können und somit keine differentia specifica von Stereotypen darstellen (Tajfel 1969; Rehm 1986: 20f).

In neueren Beiträgen zur Stereotypenforschung werden diese Unzulänglichkeiten der Konzeptspezifikation durch einfachere und zugleich abstraktere Definitionen umgangen. Als eine Art „standard viewpoint“ kann gelten: „stereotypes are beliefs about the characteristics, attributes, and behaviors of members of certain groups“ (Hilton/von Hippel 1996: 240). Die Bezeichnung von Stereotypen als eine spezielle Variante der Klasse der „Meinungen“ (beliefs) ist insbesondere in der sozialpsychologischen Forschung mittlerweile weitgehend unumstritten (vgl. Ashmore/Del Boca 1981: 16; Gardner 1994: 3; Leyens/Yzerbyt/Schadron 1994: 11; Stroebe/Insko 1989: 5). Ein Stereotyp ist demnach ein (subjektives) „Wahrscheinlichkeitsurteil über das Bestehen einer Verbindung zwischen einem Objekt und einem Attribut“ (Stroebe 1985: 8), das sich von anderen „beliefs“ insbesondere dadurch unterscheiden läßt, daß es sich auf soziale Gruppen bzw. auf Personen aufgrund ihrer Zuordnung zu solchen Gruppen bezieht.

Als ein weiterer Bestandteil des Definiens kann das damit bereits angedeutete Element der „Generalisierung“ aufgefaßt werden, also die Verallgemeinerung der Wahrscheinlichkeitsurteile über spezifische Zusammenhänge zwischen Personen und Attributen auf alle Personen, die einer bestimmten Kategorie zugeordnet. Demgemäß wäre die Aussage „Schwaben sind sparsam“ ein Stereotyp im Sinne eines Wahrscheinlichkeitsurteils über die Verbindung zwischen dem Objekt („Schwaben“) und dem Attribut („sparsam“), das sich dadurch auszeichnet, daß das Urteil für Personen, die der entsprechenden Kategorie zugeordnet werden, aufgrund dieser Zuordnung verallgemeinert bzw. generalisiert wird. Das heißt: „stereotypes are generalizations based on the membership to a category, i.e. beliefs that derive from the inference that all members of a given category share the same properties and are, therefore, interchangeable“

(Leyens/Yzerbyt/Schadron 1994: 17; vgl. McCauley/Stitt/Segal 1980: 196). Objekte einer solchen Aussage können sowohl Angehörige einer irgendwie bestimmten „Eigengruppe“ (Autostereotypen) als auch Angehörige anderer Gruppen (Heterostereotypen) sein.

Vor allem in Beiträgen aus dem Umfeld der „social cognition“-Forschung wird der Stereotypenbegriff auch definiert als „a cognitive structure that contains the perceiver’s knowledge, beliefs, and expectations about a human group“ (Hamilton/Trolier 1986: 133; Hamilton/Sherman 1994: 2f; vgl. Stangor 1996: 628). Diese Konzeptualisierungen knüpfen explizit an das Schema-Konzept der Kognitionsforschung an, das sich wiederum allgemein als „a cognitive structure that represents knowledge about a concept or type of stimulus, including its attributes and the relations among those attributes“ (Fiske/Taylor 1991: 98) definieren läßt. Auch das Schema-Konzept wird, ähnlich wie die gelegentlich in ähnlicher Bedeutung verwendeten Begriffe „frame“ oder „Prototyp“, nicht immer einheitlich und eindeutig definiert, doch es besteht zumindest darüber weitgehender Konsens, daß damit allgemeine Wissensstrukturen gemeint sind, die die wichtigsten Merkmale eines bestimmten Gegenstandsbereichs sowie die Beziehungen zwischen diesen Merkmalen wiedergeben (Schwarz 1985: 273). Der Ausdruck „allgemeine Wissensstrukturen“ bezeichnet dabei nicht notwendig ein „objektiv richtiges“ Wissen, sondern wird im allgemeinen eher im Sinne subjektiver Hypothesen und Erwartungen der Akteure über Merkmale und Eigenschaften von Personen und Gruppen von Personen oder als „implizite Persönlichkeitstheorien“ verstanden (Hilton/von Hippel 1996: 240; Stangor/ Lange 1994: 357f). Die kognitiven Prozesse der Herausbildung, die Struktur der kognitiven Repräsentation von Stereotypen und deren Einfluß auf die Verarbeitung von Informationen und Eindrücken über Personen stehen im Mittelpunkt der kognitionstheoretischen Stereotypenforschung und werden im Rahmen dieser Ansätze zumeist, der Konzeptspezifikation entsprechend, als Spezialfälle der Schemata-Forschung behandelt.

Diese Spezifikationen des Stereotypenbegriffs unter Bezug auf das Schemakonzept lassen sich mit den bereits erwähnten Definienselementen („Meinungen“ bzw. „Wahrscheinlichkeitsurteile“ über die Attribute einer sozialen Kategorie zugeordneter Personen und deren „Generalisierung“) relativ problemlos in Einklang bringen, da auch damit weitgehend identische Begriffe zur Darstellung des Inhalts des Definiendums vorgeschlagen werden. Ob diese dann dem Schemabegriff als einer übergeordneten Kategorie zugeordnet werden, mag für die Zuordnung der Stereotypen zu einer allgemeineren Klasse kognitiver und sozialer Phänomene ebenso wie für das Verständnis kognitiver Ansätze hilfreich sein; für die Definition des Stereotypenbegriffs ist dies jedoch nicht entscheidend. Dafür ist letztlich nur wichtig, ob die intensionale Bedeutung dieses Begriffs durch die angeführten Merkmale ausreichend geklärt ist und ob diese Merkmale sich wiederum mit Hilfe von Korrespondenzregeln in Indikatoren für beobachtbare Sachverhalte transformieren lassen.
Der Anwendungsbereich des Stereotypenbegriffs, also dessen extensionale Bedeutung, ist mit den angeführten Merkmalen sehr weit gefaßt. Jede Meinung, Überzeugung oder Erwartung über die Merkmale und Eigenschaften etwa von Frauen, Alten, Türken oder Linkshändern, die derart generalisiert ist, daß sie sich auf alle Personen bezieht, die diesen Kategorien zugeordnet werden, oder auf eine Person *aufgrund* der Zuordnung zu einer solchen Kategorie, wäre demnach als ein Stereotyp zu bezeichnen. Dieser große Begriffsumfang ist indes für die Zwecke einer allgemeinen Konzeptspezifikation unproblematisch. Er läßt sich vor allem durch eine genauere Bestimmung der jeweils als relevant erachteten sozialen Kategorie, auf

die sich die „beliefs“ beziehen, präzisieren. „Ethnische Stereotype“ können demgemäß leicht dadurch definiert werden, daß damit alle generalisierten Meinungen über eine nach ethnischen Kriterien bestimmte Gruppe als Ganzes bzw. über einzelne Personen aufgrund der Zuschreibung zu einer solchen Gruppe (Kategorisierung) bezeichnet werden - vorausgesetzt, das Attribut „ethnisch“ ist ebenfalls ausreichend spezifiziert.

Mit dieser sehr allgemeinen und verhältnismäßig abstrakten Begriffsbestimmung ist keineswegs gesagt, daß die Charakterisierungen von Stereotypen etwa als stark vereinfachte, negativ konnotierte Meinungen „falsch“ oder „unwahr“ seien. Vielmehr hat sich in zahlreichen Studien gezeigt, daß Stereotype über Angehörige einer „out-group“ in den allermeisten Fällen weitaus stärkere negative Konnotationen aufweisen, als Stereotype über Angehörige der „in-group“ derjenigen Personen, die die entsprechende Meinung äußern. Auch die folgende Aussage kann - zumindest vor einem entsprechenden normativen Hintergrund - durchaus als „richtig“ angesehen werden: „Stereotypes are usually simple, overgeneralized assertions about what ‘they’ are like, ‘they’ being the members of social categories who are robbed of their individuality by having applied to them a set of beliefs that ascribe to them, one and all, a set of shared attributes of character and prospensities of behavior“ (Snyder/Miene 1994: 34). Problematisch ist an diesem Satz nur, daß er für eine Konzeptspezifikation mit dem Ziel, empirische Forschung zu betreiben, ebensowenig geeignet ist, wie die Aussage: „stereotypes are based not on an inductive collection of data, but on hearsay, rumor and anecdotes - in short, on evidence which is insufficient to justify the generalization“ (Klineberg 1951: 505, zit. in Brigham 1971: 18). Während die eine Aussage eine wahrscheinlich häufig zutreffende Vermutung über die Bildung von Stereotypen beinhaltet, konkretisiert die andere eine häufig festgestellte Regelmäßigkeit des Stereotypisierungsprozesses (und deren Bewertung). Beide Sätze stellen jedoch keine sinnvollen Definitionen dar, sondern empirische Aussagen oder zumindest Tatsachenvermutungen. Diese Einschätzung scheint sich vor allem in der sozialpsychologischen Stereotypenforschung mittlerweile weitgehend durchgesetzt zu haben und findet ihren Ausdruck in entsprechend abstrakteren Konzeptspezifikationen jenseits der sogenannten „stereotype-as-bad“-Kontroverse, während die Problematik etwa der negativen Konnotationen von Stereotypen oder der Folgen der Stereotypisierung für soziale Interaktionen im Zusammenhang empirischer Untersuchungen thematisiert wird.

Eine weitere, auch in neueren Beiträgen zur Präzisierung des Stereotypenbegriffs immer wieder behandelte konzeptionelle Unklarheit betrifft die Frage, ob Stereotype als Meinungen und Überzeugungen einzelner Individuen über Angehörige einer sozialen Kategorie bzw. Gruppe (Ashmore/Del Boca 1981: 19) oder als „societally shared beliefs“ (Stangor 1996: 628) zu definieren seien. Vor allem Ashmore und Del Boca (1981) haben dafür plädiert, definitorisch zwischen Stereotypen einzelner Individuen („stereotypes“) und konsensualen oder kulturell geteilten Stereotypen („cultural stereotypes“) zu unterscheiden, da dies wesentlich zur Präzisierung des Begriffs beitrage und darüber hinaus auf wichtige Fragestellungen verweise, insbesondere auf die Frage, inwieweit tatsächlich ein Konsens über eine im Stereotyp ausgedrückte (vermutete) Objekt-Attribut-Assoziation besteht. Dieses Argument ist, wie auch Gardner deutlich macht, keineswegs trivial: „it is evident that processes underlying judgments that individuals make on traits that are consensually attributed to an ethnic group may be very different from those underlying relatively idiosyncratic beliefs“ (Gardner 1994: 29; vgl. Brewer 1994: 322-325; Simpson/Yinger 1985: 97). Auch diese Überlegungen verweisen aber letztlich auf empirisch zu beantwortende Fragestellungen und folglich auf die

Notwendigkeit der Entwicklung angemessener Operationalisierungen und Messverfahren für die Erhebung individueller „beliefs“ ebenso wie für die Einschätzung des Ausmaßes des Konsenses über den spezifischen Inhalt dieser Meinungen.

Auf eine ähnlich pragmatische Weise läßt sich offensichtlich auch die ebenfalls immer wieder diskutierte Frage klären, ob Stereotype auf solche Merkmale oder Attribute bezogen sind, die als allgemeine Charakteristika einer spezifischen Gruppe gelten, oder auf solche, die diese Gruppe von anderen unterscheiden. So haben etwa McCauley, Stitt und Segal (1980) die Auffassung vertreten, daß in erster Linie distinktive Generalisierungen maßgebend für die Stereotypenforschung seien, da diese für die Akteure selbst den größten Informationswert hätten, während zum Beispiel Ashmore und Del Boca (1981) vor einer Beschränkung auf distinktive Charakteristika warnen, mit der Begründung, daß auch allgemeine bzw. nicht-distinktive Charakteristika Bestandteile von Stereotypen seien, so daß eine solche Beschränkung nur eine sehr eingeschränkte Beschreibung der „pictures in the head“ erlaube. Diese Diskussion ist für die Beantwortung mancher konkreten Fragestellungen gewiß von Bedeutung. Doch auch sie verweist, wie im folgenden noch deutlich werden wird, auf Probleme der Operationalisierung und Messung; für die Spezifikation des Stereotypenkonzepts ist sie nicht entscheidend.

Zusammenfassend läßt sich also festhalten, daß zur Spezifikation des Stereotypenkonzept die genannten Merkmale als ausreichend angesehen werden können, d.h., es wird hier vorgeschlagen, Stereotypen zu definieren als Meinungen bzw. Wahrscheinlichkeitsurteile über die Merkmale, Eigenschaften oder Attribute von Personen, die bestimmten Kategorien (oder Gruppen) zugeordnet werden, und zwar *aufgrund* dieser Zuordnung bzw. Kategorisierung. Die in der Stereotypenforschung üblichen Verfahren zur Operationalisierung und Messung dieser Merkmale sind Gegenstand des nächsten Abschnitts.

2.1.2 Operationalisierungen und Messverfahren

2.1.2.1 „Traditionelle“ Methoden

- (a) Die Katz/Braly-Methode

Die sogenannte Katz/Braly-Methode, die auch als Eigenschaftslisten-Verfahren („adjective selection technique“, „checklist technique“; „typicality rating“) bezeichnet wird, ist eines der ältesten und zugleich am häufigsten verwendeten Verfahren der Operationalisierung und Messung von Stereotypen. Entsprechend dem zugrunde gelegten Verständnis von Stereotypen und den Erkenntnisinteressen von Katz und Braly zielt die Methode auf die Messung individueller Eigenschaftszuschreibungen für soziale Gruppen und soll im Ergebnis eine Erfassung (a) des spezifischen Inhalts ethnischer Stereotype und (b) des Konsenses innerhalb einer Population oder Gruppe über diesen Inhalt des Stereotyps ermöglichen.

Die Umsetzung dieses Verfahrens ist einfach: In ihrer inzwischen schon klassischen Studie baten Katz und Braly 100 Studenten der Princeton University, aus einer in einem Pretest erstellten Liste von 84 Adjektiven (z.B. „intelligent“, „pleasure-loving“, „frivolous“, „loyal to family ties“) zunächst eine beliebige Anzahl von Merkmalen auszuwählen, die sie als charakteristisch für verschiedene ethnische Gruppen einschätzten. In einem zweiten Schritt sollten die Befragten ihre Antworten erneut durchsehen und fünf Merkmale

ankreuzen, die ihnen besonders typisch erschienen. Die zwölf von allen Befragten meistgenannten „besonders typischen“ Charakteristika einer Gruppe wurden dann nach der Häufigkeit der Nennung in einer Rangliste zusammengefaßt (vgl. Katz/Braly 1933: 284, Tab. 1) und als Stereotyp dieser Gruppe interpretiert. Um den Grad der Übereinstimmung in der Zuordnung der Merkmale zu den jeweiligen Gruppen zu berechnen, wurde außerdem die kleinste Anzahl der Merkmale („least number of traits“) berechnet, die erforderlich war, um 50% aller 500 überhaupt möglichen Antworten zu erfassen. Je geringer dieser Wert für eine Gruppe ausfiel, desto größer war nach Katz und Bralys Interpretation die Prägnanz („definiteness“) des Stereotyps (vgl. Katz/Braly 1933: 287, Tab. 2). Die Resultate der Studie zeigten für einige ethnische Gruppen eine hohe Übereinstimmung in den Antworten der Befragten, die nach Einschätzung der Autoren überdies in Einklang mit gängigen Charakterisierungen dieser Gruppen etwa in Tageszeitungen und Zeitschriften standen.

Die Katz/Braly-Methode wurde in zahlreichen Studien zur Untersuchung unterschiedlicher Typen von Stereotypen angewendet (vgl. für eine Übersicht: Brigham 1971: 20-22; Stroebe 1985: 5-8; Dovidio/ Gaertner 1986: 4f) und erhielt, wie Ashmore und Del Boca (1981: 12) feststellten, in der Stereotypenforschung den Status eines „exemplars“ im Kuhnschen Sinn, da dieses Verfahren die zu lösenden „puzzles“ definierte und gewissermaßen als Substitut für explizite Regeln zur Ausrichtung der konkreten Forschung fungierte. Den vermutlich zentralen Grund für diese Attraktion des Verfahrens hat Gardner (1994: 7) auf den Punkt gebracht: „The obvious advantage of this technique is its simplicity. It is easy to administer and easy to interpret, at least at the group level“. Dennoch hat sich bereits seit längerem die Erkenntnis durchgesetzt, daß dieses Verfahren in mehrfacher Hinsicht zu einfach ist. Zu den zentralen Kritikpunkten zählen:

- die einseitig methodisch-technische Operationalisierung des Stereotypenkonzepts als simple Addition individueller Zuschreibungen von Eigenschaften zu sozialen Gruppen (Schäfer 1988: 14),
- die implizite Annahme, Stereotypen seien durch binäre Zuordnung gekennzeichnet (z.B. „faul“/“nicht faul“), mit der Folge, daß die zumeist eher probabilistische Struktur von Stereotypen nur stark vereinfacht reproduziert wird (Linville/Salovey/Fischer 1986: 204);
- die Unbestimmtheit des Prägnanzkriteriums in Bezug auf ein zu forderndes Maß der Übereinstimmung und die inhaltlich nicht spezifizierte Abhängigkeit der Übereinstimmungsmaße von der Anzahl der in der Liste enthaltenen Eigenschaftsbezeichnungen (Schäfer/Six 1978: 81f),
- die unzureichende Differenzierung zwischen der Erfassung der *Kenntnis* eines Stereotyps und der persönlichen *Akzeptanz* des Stereotyps (Brigham 1971: 29; Schäfer/Six 1978: 82; Devine/Elliot 1995: 1140),
- die Problematik der Ermittlung reiner Artefakte bei der Bestimmung der Prägnanz des Stereotyps auf der Basis einer einfachen Addition unterschiedlicher Eigenschaftskombinationen von einzelnen Befragten oder Teilgruppen von Befragten (Schäfer 1988: 17) und
- die Problematik der Ermittlung von Artefakten aufgrund der eingeschränkten Antwortmöglichkeiten und der „forced choice“-Situation, in der die Befragten explizit aufgefordert werden, Generalisierungen und Kategorisierungen vorzunehmen (Brigham 1971: 24; Gardner 1994: 7; McCauley/Stitt/ Segal 1980: 197; Rehm 1986: 24).

Vor dem Hintergrund der Ergebnisse der neueren kognitionstheoretischen Forschung erscheint ferner auch die der Katz/Braly-Methode implizit zugrundeliegende Vorstellung von der kognitiven Struktur bzw. Repräsentation von Stereotypen als zumindest einseitig: „the cognitive structure implicitly assumed by this methodology is one in which various attributes are connected by association with the subject's group concept. No other form of organization in the cognitive representation is suggested by this procedure" (Hamilton/Trolier 1986: 138). Insbesondere die inzwischen mehrfach aufgezeigte Problematik subkategorialer Differenzierungen innerhalb eines Stereotyps läßt sich mit diesem Modell nicht erfassen.

Weitere sehr grundsätzliche Einwände gegen dieses Verfahren betreffen dessen starke Anfälligkeit für Antwortverzerrungen („response errors"). Neben der bereits angesprochenen Problematik der „forced choice"-Situation zählen dazu vor allem Verzerrungen durch die Tendenz zur Abgabe sozial erwünschter Antworten („social desirability and impression management effects"), die zu Über- oder Untertreibungen in Bezug auf die subjektive Neigung zur Stereotypisierung oder den vermuteten Inhalt des Stereotyps einer bestimmten Gruppe führen kann, oder durch die Abfolge von Fragen („Positionseffekte") und Gruppennennungen (Hamilton/Trolier 1986: 140; Brigham 1971: 25)[2]. Die möglichen Verzerrungen der empirischen Befunde der Stereotypen- und Vorurteilsforschung durch solche „response errors" haben in diesem Forschungszweig zu zahlreichen Bemühungen um die Entwicklung neuer, weniger reaktiver Messverfahren geführt, von denen einige im folgenden ebenfalls kurz dargestellt werden. Zuvor soll jedoch eine kurze Übersicht über einige weitere Beiträge der Operationalisierung und Messung von Stereotypen gegeben werden, die ebenfalls von einer kritischen Einschätzung der Katz/Braly-Methode ausgehen und Modifikationen im Rahmen dieses Paradigmas beinhalten.

- (b) Die Brigham-Methode und die „diagnostic ratio"

Die sogenannte Brigham-Methode („percentage rating") stellt einen Versuch dar, vor allem zwei Mängel der Katz/Braly-Methode zu korrigieren: Im Unterschied zu dieser soll sie erstens ein Maß für die individuellen Träger von Stereotypen bereitstellen und zweitens eine Operationalisierung des Grades der in Stereotypen enthaltenen „Übergeneralisierungen" ermöglichen. Die Notwendigkeit der Suche nach einem Maß für die Übergeneralisierung ergibt sich für Brigham aus einem speziellen Verständnis von (ethnischen) Stereotypen, die er definiert als „a generalization made about an ethnic group, concerning a trait attribution, which is considered to be unjustified by an observer" (Brigham 1971: 31).

Die von ihm vorgeschlagene Methode zur Umsetzung dieser Ziele besteht im Kern darin, subjektive Erwartungen bzw. Vermutungen über die Häufigkeitsverteilung bestimmter Attribute für kategorial bestimmte Gruppen von Personen zu erheben. Im Unterschied zur Katz-Braly-Methode sollen die Befragten also nicht einfach nur „typische" Merkmale dieser Personen angeben, sondern den Prozentsatz aller dieser (ethnischen) Gruppe zugeordneten Personen, die nach Einschätzung der Befragten ein bestimmtes Merkmal (aus einer Liste von 25 bis 30 Attributen) aufweisen. Auf diese Weise läßt sich zum einen - durch

[2] Brigham verweist beispielsweise auf eine Studie von Diab (1963), in der gezeigt wurde, daß sich die Zuschreibung negativer Merkmale zu einer offenbar geringgeschätzten Gruppe (bei Diab: „die Russen") noch verstärkte, wenn deren Gruppenname unmittelbar neben dem einer positiv bewerteten Gruppe (bei Diab: „die Amerikaner") plaziert wurde. In einer weiteren Studie fand Diab zudem Hinweise darauf, daß die Konnotationen der Charakterisierung eher geringgeschätzter Gruppe positiver ausfielen, wenn bei der Befragung auf eine Einbeziehung positiv bewerteter Gruppen verzichtet wurde (vgl. Brigham 1971: 25).

Berechnung der Mittelwerte der Prozentsatzangaben für jedes einzelne der jeweiligen Gruppe zugeschriebenen Merkmal - ein Indikator für Ausrichtung und Intensität von Stereotypen auf der Ebene der Gruppe der Befragten ermitteln. Zum anderen kann damit auch für jeden einzelnen Befragten ein „stereotyping score“ berechnet werden, der Inhalt und/oder Intensität der von ihnen jeweils für wahr gehaltenen Stereotype anzeigt (Brigham 1971: 32f). Außerdem soll auf dieser Grundlage der Grad der mit einem Stereotyp verbundenen „Übergeneralisierung“ bestimmt werden können. Wann allerdings eine solche „Übergeneralisierung“ vorliegt, ist, so Brigham, abhängig von der subjektiven Einschätzung des Beobachters. Es könne demnach etwa dann von einer ungerechtfertigten oder überzogenen Generalisierung - und somit von einem Stereotyp - gesprochen werden, wenn „extreme“ Prozentwerte (z.B. mehr als 80% oder unter 20%) in der Zuordnung der Merkmale auftauchen oder wenn die Prozentangaben einzelner Befragten deutlich vom Mittelwert des gesamten Samples abweichen.

Dieses Verfahrens hat im Vergleich zur Katz-Braly-Methode einige Vorzüge (vgl. Gardner 1994: 8; Stroebe/Insko 1989: 6; Leyens/Yzerbyt/Schadron 1994: 24):

- es können Aussagen sowohl für konsensuale Stereotypen innerhalb einer Gruppe von Befragten als auch für die speziellen „beliefs“ einzelner Befragter aus den Befragungsergebnissen abgeleitet werden;
- es kann eher davon ausgegangen werden, daß die Befragten bei ihren Angaben jedes in der vorgelegten Eigenschaftsliste präsentierte Attribut berücksichtigen, so daß sich ein umfassenderes, nicht nur auf wenige besonders „typische“ Merkmale beschränktes Bild der jeweiligen Stereotypen ergibt;
- mit der Möglichkeit, eine Gruppe von Personen durch eine größere Anzahl von (positiven und negativen) Attributen zu charakterisieren, wird den Befragten ermöglicht, eine differenziertere Einschätzung dieser Gruppe anzugeben, wodurch sich zugleich die in einigen Untersuchungen festgestellte Zurückhaltung bei der Beantwortung der Fragen im Vergleich zum „typicality rating“ verringern läßt.

Inwieweit die Brigham-Methode allerdings substantiell andere Ergebnisse erbringt als das traditionelle Eigenschaftslistenverfahren ist umstritten. Während Brigham selbst auf eigene Untersuchungen hinweist, in denen die beiden Verfahren unterschiedliche Resultate erbrachten, verweisen Stroebe und Insko (1989: 6f) auf neuere Studien, denen zufolge „typicality rating“ und „percentage rating“ zu sehr ähnlichen Ergebnissen führen und hohe Korrelationen (zwischen .70 und .90) aufweisen.

Grundsätzlich ist jedoch bei einer kritischen Einschätzung der Brigham-Methode festzustellen, daß auch bei diesem Verfahren letztlich nicht klar ist, durch welche Charakteristika ein spezifisches Stereotyp gekennzeichnet ist. Die Bestimmung des Stereotyps auf der Gruppenebene durch (i) die Berechnung der Mittelwerte der vermuteten Häufigkeitsverteilungen bestimmter Merkmale für verschiedene Gruppen und (ii) die Zuordnung der mit sehr hohen oder niedrigen Prozentwerten verbundenen Attribute führt offensichtlich zu ambivalenten Ergebnissen. Dies zeigt sich etwa daran, daß Merkmale mit sehr hohen Prozentwerten verbunden sein können, die im Extremfall als allgemeine Merkmale aller Menschen angesehen werden; solche Merkmale können infolge von Brighams Operationalisierung des Stereotypenkonzepts als zentrale Elementen von Stereotypen erscheinen, obwohl sie vermutlich für die „beliefs“ der Befragten in Bezug auf die spezifischen Eigenschaften einer kategorial bestimmten Gruppe von Personen - insbesondere in Relation zu anderen Gruppen - eher irrelevant sind (Stroebe/Insko 1989: 7). Ähnliche Schwierigkeiten ergeben sich bei der Interpretation niedriger Prozentwerte für einzelne Merkmale, da im Grunde ungeklärt

bleibt, ob etwa eine Angabe, wonach 15% der Angehörigen einer bestimmten Gruppe nach Einschätzung der Befragten „fleißig“ seien, darauf hinweist, daß zum Beispiel das Attribut „faul“ als ein Element des Stereotyps aufzufassen ist, oder ob dies lediglich anzeigt, daß „fleißig“ nicht als Charakteristikum dieser Gruppe angesehen wird (Gardner 1994: 8).

Solche Unklarheiten betreffen in erster Linie die Operationalisierung der Stereotypenkonzepts im engeren Sinn und können teilweise durch andere Operationalisierungsverfahren verringert werden. So haben zum Beispiel McCauley und Stitt (1978) vorgeschlagen, Stereotypen als solche Generalisierungen aufzufassen, die eine kategorial bestimmte Personengruppe von anderen unterscheidet. Zentrales Element in dem dazu entwickelten Verfahren ist die „diagnostic ratio“, die auf der Ermittlung bedingter Wahrscheinlichkeiten beruht[3]. In einem ersten Schritt werden bei dieser Methode die Befragten - weitgehend analog zur Brigham-Methode - aufgefordert, den Prozentsatz derjenigen Angehörigen einer kategorial bestimmten (ethnischen) Gruppe (E) anzugeben, die nach ihrer Einschätzung eine Reihe vorgegebener Merkmale (T) aufweisen. Im zweiten Schritt wird dann darum gebeten, den Prozentsatz aller Menschen mit den angegebenen Eigenschaften zu schätzen. Beide Prozentsatzangaben werden dabei als Ausdruck (subjektiver) Wahrscheinlichkeitsvermutungen über das Bestehen eines Zusammenhangs zwischen „Objekten“ und Attributen interpretiert. Das Verhältnis dieser beiden geschätzten Häufigkeitsverteilungen ergibt die „diagnostic ratio“; formal ausgedrückt:

$$DR = p(T/E)/p(T).$$

Tendiert der DR-Wert für ein bestimmtes Merkmal (T) gegen 1.0, deutet dies darauf hin, daß es von den Befragten für ein allgemeines Merkmal gehalten wird, das (tendentiell) für alle Menschen - unabhängig von einer kategorial bestimmten Zuschreibung zu einer Gruppe - gilt; ergibt die Messung DR-Werte, die größer (oder kleiner) als 1.0 sind, kann dies als Indikator dafür angesehen werden, daß die jeweiligen Merkmale nach Einschätzung der Befragten für die betreffende Personengruppe besonders charakteristisch (oder für DR < 1.0: besonders uncharakteristisch oder selten) sind. Je stärker der Wert von 1.0 abweicht, desto eher handelt es sich bei dem Merkmal um ein distinktes Attribut und somit um ein Element eines Stereotyps im Sinne einer auf diese Gruppe bezogenen generalisierten Meinung oder Überzeugung.

Mit Hilfe dieses Verfahrens lassen sich einige Interpretationsprobleme, die sich bei der Brigham-Methode oder auch bei der Katz-Braly-Methode ergeben können, umgehen. Es ist generell besser geeignet, die in der Regel eher probabilistische Struktur von Stereotypen zu erfassen (McCauley/Stitt/Segal 1980: 197). Und die Konzentration auf die Analyse speziell distinkter Attribute unter Bezug auf eine Art „base rate“ dürfte für den Inhalt einiger Stereotypen generell aussagekräftigere Resultate erbringen, als die in dieser Hinsicht eher unspezifischen Verfahren des „typicality rating“ und des „percentage rating“. Allerdings ist fraglich, ob das im Vergleich zu diesen etwas aufwendigere Verfahren in jedem Fall erforderlich ist, da auch bei den Methoden von Katz und Braly oder Brigham zumindest implizit nach distinkten Merkmalen gefragt wird

[3] Das Verfahren stellt im Prinzip eine Übertragung des Bayesschen Theorems auf die Analyse von Stereotypen dar; die (subjektiven) Wahrscheinlichkeiten werden dabei direkt unter Bezug auf die angegebenen Prozentualverteilungen geschätzt. Für eine ausführliche Ableitung der „diagnostic ratio“ aus dem Bayesschen Theorem und eine Auseinandersetzung mit diversen Einwänden gegen dessen empirische Gültigkeit vgl. McCauley/Stitt (1978: 930-933) und McCauley/Stitt/Segal (1980: 197-201).

und dies von den Befragten offenbar in diesem Sinn verstanden wird (Stroebe/Insko 1989: 8). Ferner kann bezweifelt werden, daß Angaben über „die Menschen im allgemeinen“ stets eine sinnvolle Bezugsgröße für den distinkten Gehalt eines Stereotyps darstellen. Denn erstens kann davon ausgegangen werden, daß darauf bezogene Antworten eher ad hoc gegeben werden und zweitens dürfte diese „base rate“ für viele Fragestellungen zu allgemein und diffus sein, da sich der distinkte Charakter mancher Attribute vermutlich eher unter Bezug auf eine oder mehrere spezifische Gruppen (z.B. die „ingroup“ der Befragten) ergibt (Gardner 1994: 10; vgl. aber McCauley/Stitt/Segal 1980: 199).

Varianten des Rating-Verfahrens, die eine Bestimmung des Stereotyps einer Gruppe in Relation zu den Stereotypen anderer Gruppen zu erfassen suchen, beruhen auf der Erhebung sogenannter semantischer oder Eindrucksdifferentiale. Diese Verfahren sind generell von besonderem Interesse im Hinblick auf die vergleichende Inhaltsanalyse von Stereotypen auf der Basis allgemeiner Urteilsskalen und ermöglichen zum Beispiel auch eine Untersuchung der Relationen zwischen Auto- und Heterostereotypen. Die damit angedeuteten Problemstellungen spielen in der Stereotypen- und Vorurteilsforschung etwa im Zusammenhang mit Phänomen der ingroup-outgroup-Differenzierung eine zentrale Rolle. Deshalb sollen diese wesentlichen Merkmale dieser Verfahren ebenfalls kurz zusammengefaßt werden, ehe dann die bereits für die Katz-Braly-Methode festgestellten und für die zuletzt dargestellten Rating-Verfahren ebenso beobachtbaren Probleme möglicher Antwortverzerrungen („response errors“) ausführlicher behandelt werden.

- (c) Die Methodik des „semantischen Differentials“

Die von Osgood u.a. zur Identifikation linguistisch universeller Bedeutungsdimensionen entwickelte Methodik des „semantischen Differentials“ wurde im deutschen Sprachraum vor allem von Hofstätter in Form des „Polaritätenprofils“ in die Stereotypenforschung eingeführt; im englischen Sprachraum stellt die „Gardner-Methode“ ein prominentes Beispiel für die Entwicklung von „stereotype differentials“ dar. In beiden Fällen werden die Befragten - ähnlich wie bei den bereits dargestellten Verfahren - darum gebeten, Meinungen über Zusammenhänge zwischen kategorial bestimmten Personengruppen und bestimmten Attributen anzugeben. Allerdings werden dazu - im Unterschied zu den anderen Eigenschaftslistenverfahren - Listen von bipolaren Adjektivpaaren verwendet. Sowohl Hofstätter als auch Gardner geben für jedes Adjektivpaar eine 7stufige Ratingskala vor, deren Endpunkte durch die jeweils als entgegengesetzt eingeschätzten Adjektive (z.B. faul versus fleißig, schweigsam versus geschwätzig usw.) markiert sind (Hofstätter 1986:83f; Gardner 1994: 10); der Mittelwert mit der Ziffer 4 verweist auf eine neutrale Position. Für jede zu berücksichtigende Gruppe von Personen sollen die Befragten auf jeder Adjektivskala diejenigen Punkte markieren, die ihre Meinung über diese Gruppe am besten wiedergeben. Anschließend kann für jedes (für die Auswertung gleichsinnig gepolte) Gegensatzpaar und jede kategorial bestimmte Gruppe der Mittelwert über alle Befragten berechnet und dessen Abweichung vom neutralen Mittelwert der Skala (μ = 4.0) festgestellt werden. Das Ausmaß, in dem eher das eine oder das andere Adjektiv als Charakteristikum einer Gruppe anzusehen ist, läßt sich dabei auf der Grundlage eines Mittelwertdifferenzentests (t-Test) für

alle unabhängigen Stichproben derselben Grundgesamtheit (also für alle Befragten) ermitteln (Gardner/Lalonde/Nero/Young 1988: 41; vgl. Schnell/Hill/Esser 1995: 413f; Schäfer/Six 1978: 79f)[4].

Auf diese Weise ergeben sich relativ einfache operationale Bestimmungen von Stereotypen: „Operationally, the stereotype of a particular group is defined in terms of those attributes [...] for which there is the greatest polarization (i.e., agreement that the attribute defining that end of the scale characterizes the group)“, vorausgesetzt die beobachtete Mittelwertdifferenz ist signifikant; und: „An individual-difference measure of stereotyping can be achieved by summing an individual's ratings on those scales defining the stereotype“ (Gardner/Lalonde/Nero/Young: 1988: 41f; vgl. Gardner 1994: 11). Außerdem können auf der Basis einer solchen Datenerhebung und -analyse nach der Konzeption Hofstätters „Eigenschaftsprofile“ im Sinne charakteristischer Merkmalsmuster für Gruppen erstellt werden, zwischen denen sich durch die Berechnung des von ihm entwickelten Q-Maßes Ähnlichkeitskorrelationen feststellen lassen[5]. Dieses Verfahren kann zum Beispiel für den Vergleich von Autostereotypen und Heterostereotypen verwendet werden und ermöglicht zudem eine Darstellung der Ähnlichkeits- oder Distanzmaße zwischen unterschiedlichen Urteilsgegenständen in einem zwei- oder dreidimensionalen Bedeutungsraum (vgl. Schäfer 1988: 17f).

Grundsätzlich weist die Verwendung semantischer Differentiale nach Einschätzung von Gardner (1994: 11) fünf wesentliche Vorzüge auf:

- die gestellte Aufgabe ist für die Befragten relativ leicht zu verstehen;
- die Befragten zeigen im allgemeinen keine Vorbehalte gegen die gewünschte Einschätzung von Gruppen auf der Basis differenzierter Adjektivskalen;
- es können in der Regel Einschätzungen für alle Adjektivskalen erfaßt werden;
- es handelt sich um eine relativ einfache und intersubjektiv überprüfbare Erfassung der Indizes für Stereotypen sowohl auf der Individual- als auch auf der Gruppenebene;
- es ermöglicht eine direkte Messung der Beziehung zwischen Meinungen, die sich auf Gruppen als Ganzes, und solchen, die sich auf einzelne Personen beziehen, so daß das Verfahren auch für Untersuchungen über das Verhältnis zwischen Stereotypen und Personenwahrnehmung geeignet ist.

In einer der wenigen empirischen Untersuchung über die Implikationen der unterschiedlichen Verfahren zur Operationalisierung und Messung von Stereotypen haben Gardner, Lalonde, Nero und Young (1988) am Beispiel des Stereotyps der Frankokanadier analysiert, in welchem Verhältnis die mit dem „Stereotypendifferential“ ermittelten Resultate zu den erwähnten Eigenschaftslistenverfahren stehen. Ausgehend von einer Berechnung der verschiedenen Indizes zum einen für die Gruppe als Analyseeinheit und zum anderen für die einzelnen Befragten wurden die für die insgesamt 30 Skalen ermittelten Werte (scores) einer Korrelationsanalyse unterzogen. Dabei zeigte sich im Hinblick auf die für die Gruppe der

---

[4] Die Gleichung für die Berechnung des „Stereotypendifferentials“ nach Gardner lautet: $(\bar{x} - \mu) / (s / \sqrt{n})$; x steht für den Mittelwert der für ein item ermittelten Angaben, μ für den neutralen Mittelwert der Skala, s für die Standardabweichung und n für die Anzahl der Befragten (vgl. Gardner/Lalonde/Nero/Young 1988: 50).

[5] Bei der Berechnung des Q-Maßes handelt es sich im Kern um einen Spezialfall der Korrelationsrechnung, der sich dadurch auszeichnet, daß nicht Eigenschaften, sondern die Träger von Eigenschaften verglichen werden sollen. Dazu werden die Koeffizienten der Produkt-Moment-Korrelation berechnet, wobei sowohl die Eigenschaften als auch die Polaritäten (die Prozentsätze der befragten Personen, die jede einzelne der vorgegebenen Eigenschaften einer bestimmten Gruppe zuschreiben) als Stichproben aus der Gesamtheit aller möglichen Eigenschaften bzw. aller möglichen Polaritäten betrachtet werden (Hofstätter 1986: 234-236).

Befragten als Ganzes ermittelten Werte, daß das Stereotypendifferential zu einem ähnlichen Ergebnis führt wie die Mittelwertberechnungen des „percentage rating“ nach Brigham ($r = .91$, $p < .0001$) und der „diagnostic ratio“ nach McCauley und Stitt ($r = .63$, $p < .001$). Hingegen ergaben sich negative Korrelationen für den Index des Stereotypendifferentials und die Mittelwerte der „diagnostic ratios“ der einzelnen Befragten ($r = -.61$, $p < .001$) sowie für die Beziehung zwischen diesem Mittelwert des „diagnostic ratio“ und dem Mittelwert des „percentage ratings“ für die Gruppe als Ganzes ($r = -.51$, $p < 0.1$). Die Korrelationen zwischen Brighams Indikator für die Häufigkeit, mit der die Befragten extreme, „übergeneralisierte“ Zuordnungen vornahmen, und den anderen Indikatoren ergaben stets nur Werte unter .34 (Gardner/Lalonde/Nero/Young 1988: 51-53). Aufgrund dieser Ergebnisse vermuten Gardner et al. (1988: 53; vgl. Gardner 1994: 13f), daß das, was jeweils als Stereotyp einer bestimmten Gruppe erfaßt wird, maßgeblich davon abhängt, welches Verfahren angewendet wird. Folglich können auch die mit den unterschiedlichen Verfahren ermittelten Resultate nur bedingt miteinander verglichen werden.

- (d) Zwischenfazit

Das Eigenschaftslistenverfahren nach Katz und Braly und die Brigham-Methode sowie, in geringerem Maße, die Verfahren unter Verwendung der „diagnostic ratios“ und der Stereotypendifferentiale stellen zweifelsohne die vorherrschenden Methoden der Operationalisierung und Messung von Stereotypen dar; sie wurden, wie Stangor und Lange noch 1994 feststellten, „in virtually all stereotyping studies over the past 20 years“ verwendet (Stangor/Lange 1994: 374). Bei allen Unterschieden in den Details handelt es sich bei den an diesen Verfahren orientierten Untersuchungen stets um Stichprobenerhebungen von Meinungen auf der Grundlage standardisierter Items und Ratingskalen, die Rückschlüsse auf Verteilungen von Meinungen in einer bestimmten Population und Vergleiche zwischen den befragten Personen ermöglichen sollen. Sie unterscheiden sich in der konkreten Umsetzung nicht wesentlich von den Standardverfahren der Meinungsforschung und teilen mit diesen einige wesentliche Vorzüge: Sie sind relativ leicht zu handhaben, können für große Stichproben in allgemeinen Bevölkerungsumfragen eingesetzt werden usw.

Mit diesen Vorzügen sind indes auch einige Nachteile und Schwierigkeiten verbunden, die in der Stereotypen- und Vorurteilsforschung zu einer eher kritischen Einschätzung dieser Verfahren geführt haben. Vor allem zwei Aspekte stehen dabei im Mittelpunkt: (1) die Einschränkung der Antwortmöglichkeiten durch vorgegebene Eigenschaftslisten und (2) die Verzerrung der von den Befragten angegebenen Meinungen und Überzeugungen durch „response errors“, etwa durch Abgabe sozial erwünschter Antworten.

Einwände gegen die Vorlage standardisierter Items und Ratingskalen beziehen sich zum einen auf Aspekte der Itemformulierung und der Abfolge von Fragen bzw. Gruppennennungen; dabei wird in erster Linie auf die raum-zeitliche Kontextabhängigkeit der Interpretation bestimmter Attribute (vgl. Devine/Elliot 1995: 14; Mummendey/Bolten/Isermann-Gerke 1982: 302; Sigall/Page 1971: 251f) oder auch auf die bereits erwähnten Befunde zur Beeinflußung der Charakterisierung einer Gruppe durch unmittelbar zuvor genannte andere Gruppen hingewiesen. Zum anderen wird die Festlegung der Befragten auf vorgegebene Antwortmöglichkeiten vor allem in der kognitionspsychologischen Stereotypenforschung als problematisch angesehen, da auf diese Weise Existenz, Intensität und Strukturierung mentaler Assoziationen zwischen Objekten (z.B einer Gruppe bzw. einem Gruppennamen) und Attributen nicht sinnvoll erfaßt werden kann (Stangor/Lange 1994: 375). Während den erstgenannten Einwänden in der Regel durch eine sorgfältige

Auswahl und Formulierung der Items sowie durch Berücksichtigung eventueller Positionseffekte Rechnung getragen werden kann, sind die aus kognitionspsychologischer Perspektive formulierten Bedenken zumeist mit der Entwicklung alternativer Operationalisierungs- und Messverfahren verbunden.

Auch die Hinweise auf mögliche Antwortverzerrungen durch die Abgabe sozial erwünschter Meinungen und ähnlicher „response errors“ sind häufig verknüpft mit umfassenderen Überlegungen zur Ausarbeitung „subtilerer“ und „unaufdringlicher“ Operationalisierungs- und Messverfahren, die die Beeinflußung des Meßergebnisses durch nicht kontrollierte Merkmale der Erhebungssituation („Reaktivität“) verringern sollen. Die Tatsache, daß gerade mit Hilfe von Interviews erhobene Daten durch Reaktivitätseffekte verfälscht sein können und somit nur stark eingeschränkte Validität aufweisen, stellt für die empirische Sozialforschung keine Neuigkeit dar (vgl. Schnell/Hill/Esser 1995: 327-330; Diekmann 1995: 382-403). Einigen der in diesem Zusammenhang immer wieder festgestellten Antwortverzerrungen wird indes in der Stereotypen- und Vorurteilsforschung gelegentlich ein besonders hoher Stellenwert zugemessen, da beispielsweise davon ausgegangen werden kann, daß in diesem Bereich zwangsläufig viele sogenannte „unangenehme Fragen“ gestellt werden, die die Befragten dazu veranlassen können, in ihrem Antwortverhalten von ihren tatsächlichen Meinungen oder Einstellungen abzuweichen (Hamilton/Trolier 1986: 140).

Im folgenden Abschnitt werden einige der grundlegenden Verfahren präsentiert, die vor dem Hintergrund der kritischen Einschätzung der traditionellen Operationalisierungs- und Meßmethoden entwickelt wurden. Den Anfang bildet eine kurze Darstellung des sogenannten „Bogus-Pipeline-Paradigmas“, mit Hilfe dessen Sigall und Page bereits 1971 versuchten, die Auswirkungen unterschiedlicher Meßmethoden und den Einfluß sozialer Erwünschtheit auf die Angaben der Befragten in einer Untersuchung über Stereotypen zu demonstrieren. Mit dieser Darstellung soll auf der Grundlage eines experimentellen Untersuchungsdesigns vor allem verdeutlicht werden, welche Bedeutung den möglichen Antwortverzerrungen durch „response errors“ zuzumessen ist. Anschließend werden neuere Operationalisierungs- und Messverfahren vorgestellt, die vor allem im Umfeld der „social cognition“-Forschung ausgearbeitet wurden und in diesem Forschungsbereich eine zentrale Rolle spielen. Alle diese „Paradigmen“ beziehen sich nicht ausschließlich auf die Analyse von Stereotypen im Sinne der bereits ausgeführten Konzeptspezifikation, sondern werden häufig auch für die Untersuchung von Vorurteilen verwendet. Die damit verbundenen Fragen nach dem Zusammenhang der beiden Konstrukte sollen jedoch an dieser Stelle noch nicht behandelt werden, da zu diesem Zweck zuerst eine Präzisierung des Vorurteilbegriffs notwendig ist.

2.1.1.1 Bogus-Pipeline und neuere Paradigmen

- (a) Das Bogus-Pipeline-Paradigma

Den Ausgangspunkt für die Entwicklung des experimentalpsychologischen Ansatzes des sogenannten „Bogus-Pipeline-Paradigmas“ und dessen Anwendung in der Stereotypenforschung bildete, den Angaben von Sigall und Page zufolge, der in zwei Replikationsstudien zur „Princeton-Studie“ von Katz und Braly (1933) ermittelte Befund, demzufolge sich eine im Zeitablauf zunehmende Abschwächung der negativen Konnotationen der Stereotypen von Schwarzen in den USA erkennen läßt (vgl. Gilbert 1951 und

Karlins/Coffman/Walters 1969). Sigall und Page schlossen zwar nicht grundsätzlich aus, daß diese auf der Basis des Eigenschaftslistenverfahrens berechneten Daten tatsächlich einen Wandel in den Meinungen weißer Amerikaner über „Negroes“ anzeigten, vermuteten aber, die Befunde könnten aufgrund von „social desirability or other demand-characteristic-related variables“ stark verzerrt sein (Sigall/Page 1971: 247). Diese Vermutung - und die damit verbundene Hypothese über die Wechselwirkung zwischen Meßmethode und Antwortverhalten der Befragten - sollte mit der von Sigall in Zusammenarbeit mit Jones entwickelten „Bogus-Pipeline“ untersucht werden.

Spezifikum des Bogus-Pipeline-Paradigmas ist der simulierte Elektromyograph (EMG), eine kompliziert ausstaffierte Versuchsapparatur, die den Befragten im Experiment als eine Art Lügendetektor vorgestellt wird. In einer ersten Phase des Experiments werden die der EMG-Bedingung zugeordneten Personen mit Hilfe manipulierter Tests davon „überzeugt“, daß der Elektromyograph tatsächlich in der Lage sei, wahre von verfälschten Antworten zu unterscheiden (vgl. Sigall/Page 1971: 248f; Mummendey/ Bolten/Isermann-Gerke 1982: 304-306; Stahlberg/Frey 1996: 228). Damit soll sichergestellt werden, daß die Befragten in der EMG-Bedingung in geringerem Maße sozial erwünschte Antworten geben, da sie davon ausgehen müssen, daß ihre tatsächliche Meinung oder Einstellung in Bezug auf die zu charakterisierende Personengruppe dennoch erkennbar wird. Auf diese Annahme stützten Sigall und Page (1971: 248) letztlich ihre zentrale Hypothese: „We expected that using this technique would lead subjects to present less socially desirable stereotypes than would result in a more typical rating-scale situation, where subjects are relatively free to distort“.

Die Ergebnisse ihrer Studie scheinen diese Hypothese weitgehend zu bestätigen. Neben einer Reihe signifikanter „ethnic group’s main effects“ ergaben sich für sechs der 22 vorgebenen Adjektive signifikante Interaktionseffekte zwischen Meßmethode und Beurteilungsobjekt („Americans“ bzw. „Negroes“), die darauf schließen ließen, daß Schwarze unter EMG-Bedingungen im Unterschied zur gewöhnlichen Rating-Bedingung anders charakterisiert wurden als Weiße. Eine signifikante Wechselwirkung ergab sich ferner für eine 2x2-Varianzanalyse des Bewertungsscores („favorability score“), mit dem summarisch der Grad der positiven Bewertung der beiden Gruppen erfaßt werden sollte. Es zeigte sich, daß der mittlere Grad der Bewertung beider Gruppen unter der Rating-Bedingung nahezu identische Werte aufwies (.53 für „Americans“ und .49 für „Negroes“), während die entsprechenden Werte unter der EMG-Bedingung stark unterschiedlich ausfielen (.84 für „Americans“ und -.03 für „Negroes“) (Sigall/Page 1971: 252f).

Ähnliche Resultate ergaben sich unter anderem bei einer experimentellen Überprüfung des - teilweise modifizierten - Bogus-Pipeline-Paradigmas in einer Untersuchung über Einstellungen gegenüber Türken, Deutschen und Holländern, die von Mummendey, Bolten und Isermann-Gerke (1982) in Deutschland durchgeführt wurde. Auch in diesem Fall zeigte sich eine signifikante Wechselwirkung zwischen der Meßmethode („Bogus-Pipeline-Messung“ vs. „Papier-Bleistift-Messung“) und den Angaben der Befragten. Besonders deutlich wurde dieser Interaktionseffekt hinsichtlich der Meinungen bzw. Einstellungen gegenüber Türken, die unter EMG-Bedingungen wesentlich ungünstigere Einschätzungen erhielten als unter der Paper-Pencil-Bedingung. Eine Reihe weiterer Studien erbrachten ähnliche Resultate (vgl. Wagner/Zick 1995: 46-50; Brown 1995: 211).

Diese experimentalpsychologischen Befunde veranschaulichen die angeführten Probleme möglicher Antwortverzerrungen durch Reaktivitätseffekte und deuten auf die häufig vermutete Beeinflußung des Antwortverhaltens der Befragten durch Abgabe sozial erwünschter Antworten hin. Inwieweit jedoch die Divergenzen in den angegebenen Meinungen hinsichtlich der jeweiligen Gruppen in den unterschiedlichen Experimentalbedingungen auf „social desirability needs“ zurückgeführt werden können, läßt sich letztlich nicht eindeutig klären. Streng genommen stützen die experimentellen Überprüfungen des Bogus-Pipeline-Paradigmas „nur“ die Hypothese einer Wechselwirkung zwischen Meßmethode und Meßergebnis. Ob aber die unterschiedlichen Untersuchungsbedingungen und Meßmethoden selbst wieder das Meßergebnis beeinflußen, kann im Rahmen dieses Paradigmas nicht ausreichend geprüft werden (vgl. Sigall/Page 1971: 254f; Mummendey/Bolten/Isermann-Gerke 1982: 310f) [6]. Somit bleibt auch ungewiß, ob die Befunde unter EMG-Bedingungen tatsächlich die „realen“ oder „wahren“ Stereotypen und Vorurteile anzeigen, und es bleibt zumeist ebenso ungewiß, ob die im Rahmen des Bogus-Pipeline-Paradigmas erfaßten Meinungen und Einstellungen sich als bessere Prädikatoren tatsächlichen Verhaltens erweisen.

Trotz dieser Einwände und Bedenken gegen eine Überbewertung der im Rahmen des Bogus-Pipeline-Paradigmas ermittelten Resultate kann aber dennoch davon ausgegangen werden, daß die sich dabei ergebenden Unterschiede zu Ergebnissen normaler Ratingverfahren zumindest wichtige, wenn auch nur vage Hinweise auf den Einfluß der Variablen „soziale Erwünschtheit“ in Untersuchungen zu Stereotypen und Vorurteilen vermitteln. Diese Einschätzung wird zum Beispiel durch Studien wie die von Mummendey und Bolten (1985) gestützt, die bei ihren Experimenten feststellten, daß Bogus-Pipeline-Effekte keineswegs immer zu beobachten sind; vielmehr seien diese vor allem dann zu erkennen, wenn die untersuchten Einstellungs- und Verhaltensbereiche „öffentlich hinreichend normiert sind, so daß die hier vorherrschenden sozialen Normen der Versuchsperson eine deutliche Orientierung für ‘sozial erwünschtes’ Verhalten liefern“ (Mummendey/Bolten 1985: 147). In Inhaltsbereichen, in denen diese Bedingung nicht oder nur eingeschränkt erfüllt ist, ergäben sich hingegen keine oder nur sehr schwache „BPL-Effekte“. Sofern eine solche soziale „Normierung“ etwa für die (öffentliche) Charakterisierung und Bewertung ethnischer Gruppen besteht, könnte also zumindest mit einiger Berechtigung vermutet werden, daß die Meßergebnisse unter EGM-Bedingungen zuverlässigere Indikatoren für die tatsächlichen Meinungen und Einstellungen gegenüber solchen Gruppen ergeben als vergleichbare Ergebnisse unter gewöhnlichen Interviewbedingungen.

Zur Kontrolle von Befunden über Stereotypen auf der Grundlage gewöhnlicher Eigenschaftslistenverfahren im Rahmen größerer Erhebungen ist das Bogus-Pipeline-Paradigma freilich ungeeignet. Außerdem ist auch die Umsetzung im Laborexperiment mit hohem Aufwand verbunden, kann nur mit Personen durchgeführt werden, die die Täuschung mit dem simulierten Elekromyographen nicht durchschauen und stößt zudem wegen der zwangsläufigen Irreführung der befragten Personen häufig auf Bedenken (vgl. Stahlberg/Frey 1996: 228f). Daraus kann zumindest geschlossen werden, daß sich dieses Verfahren kaum als

[6] Sigall und Page (1971: 254) weisen selbst darauf hin, daß die Erwartung, die Bogus-Pipeline erweise sich als „a lie detection device which facilitates truthful reporting“, eine Vermutung bleiben müsse. Es könne zum Beispiel nicht ausgeschlossen werden, daß durch den gesamten Aufbau der Bogus-Pipeline affektive Komponenten der untersuchten Konstrukte im Vergleich zum normalen Ratingverfahren in größerem Maße salient gemacht werden, so daß die Meßergebnisse der beiden Untersuchungsbedingungen allein schon deshalb nur bedingt vergleichbar seien und folglich keine adäquate Identifikation der „social-desirability-effects“ erlauben.

„Routineverfahren“ der Stereotypenforschung eignet. Aus der Perspektive der „social cognition“-Forschung wird darüber hinaus bemängelt, daß diese Methode ebenso wie die traditionellen Operationalisierungs- und Messverfahren auf individuellen „self-reports“ mit vorgegeben Adjektivskalen beruht und somit keine zuverlässigen Informationen über Stereotypen als mentale Repräsentationen erlaubt (vgl. Fazio/Jackson/Dunton/Williams 1995: 1014; Stangor/Lange 1994: 375). Für diese Zwecke wurden in diesem Forschungsbereich eine Reihe von Meßtechniken entwickelt, die im folgenden kurz vorgestellt werden sollen.

- (b) Neuere Paradigmen

Die Entwicklung spezieller Verfahren zur Operationalisierung und Messung von Stereotypen (und Vorurteilen) im breiten Spektrum der „social cognition“-Forschung geht einher mit einer besonderen Konzeptualisierung dieser Konstrukte. Sie werden demnach aufgefaßt als abstrakte kognitive Repräsentationen sozialer Gruppen, in denen die mit einem bestimmten „category label“ (z.B. „die Deutschen“) mental assoziierten Charakteristika (z.B. „fleißig“) enthalten sind. Solche auch als „Prototypen“ oder „Schemata“ bezeichneten Repräsentationen sind, so eine zentrale Annahme, im Langzeitgedächtnis abgespeichert und können von dort unter bestimmten Bedingungen wieder abgerufen werden. Die Analyse der Struktur dieser kognitiven Repräsentationen sowie der spezifischen Bedingungen und Konsequenzen ihrer Aktivierung steht - unter anderem - im Mittelpunkt der kognitionstheoretischen Stereotypen- und Vorurteilsforschung. Die angeführte Konzeptualisierung von Stereotypen ist zwar mit diesen Analysen eng verknüpft. Doch die in diesem Rahmen entwickelten Operationalisierungs- und Messverfahren lassen sich hier zunächst auch ohne eingehende Darstellung der umfangreichen Forschungsergebnisse zusammenfassen.

Ausgehend von der Annahme über die zentralen Charakteristika von Stereotypen ergibt sich eine einfache Grundüberlegung für die Operationalisierung und Messung dieser Konstrukte: Wenn Stereotype aus Meinungen („beliefs“) bestehen, die mit einer bestimmten Kategorie bzw. einer bestimmten sozialen Gruppe assoziiert werden, dann kann auch der konkrete Inhalt eines Stereotyps ermittelt werden, indem analysiert wird, welche Meinungen besonders stark mit einem bestimmten „category label“ kognitiv verbunden sind. Und ausgehend wiederum von dieser Grundüberlegung ergibt sich ein ebenso einfacher Vorschlag etwa für die Operationalisierung von „rassisch“ orientierten Stereotypen: „Attributes that are strongly associated with a particular racial concept, then, would compose the content of the stereotype about that group. In contrast, content that is less strongly associated with the concept presumably is less central to the stereotype“ (Hamilton/Trolier 1986: 141). Folglich ergibt sich aus dieser Perspektive die Notwendigkeit, Verfahren zu entwickeln und anzuwenden, die insbesondere eine Messung der Stärke der kognitiven Verknüpfung („associative strength“) zwischen Attributen und Kategorien bzw. Konzepten ermöglichen.

Ein prominentes Verfahren, das eine solche Messung ermöglicht, beruht auf der Verwendung einer Variante des sogenannten „lexical decision task“ (LDT). Die dabei übliche Vorgehensweise kann am Beispiel einer Studie von Gaertner und McLaughlin (1983) über Rassenstereotype in den USA kurz illustriert werden. Die Teilnehmer des Experiments, weiße Amerikaner mit hohen und niedrigen „prejudice-scores“, wurden mit einer langen Reihe von Ausdruckpaaren („stimulus terms“) konfrontiert, bei denen es sich jeweils zum einen um ein Rassenkonzept („Blacks“ oder „Whites“) und zum anderen um negativ konnotierte Attribute (z.B. „lazy“, „stupid“), positiv konnotierte Attribute (z.B. „smart“, „clean“) oder völlig sinnlose Ausdrücke (z.B.

„clopt“, „zumap“) handelte. Die Aufgabe der Teilnehmer bestand darin, für jedes Ausdruckspaar anzugeben, ob beide Stimuli Wörter darstellen, oder nicht. In Anlehnung an ältere Studien im Rahmen des „lexical decision task“ vermuteten Gaertner und McLaughlin, daß diejenigen Konzept-Attribut-Kombinationen schneller verarbeitet werden können, die mit bereits verfügbaren Konzept-Attribut-Assoziationen übereinstimmen. Wenn also etwa für die befragten Weißen die Charakterisierung bzw. das Stereotyp der Weißen eher mit positiven Attributen verbunden ist als die Charakterisierung der Schwarzen, dann wäre, so die Hypothese, zu erwarten, daß die Reaktionszeit bei Kombinationen von „whites“ und positiven Attributen kürzer ist als bei Kombinationen von „whites“ und negativen Attributen - und umgekehrt in Bezug auf die Charakterisierung von Schwarzen.

Die Ergebnisse der Studie von Gaertner und McLaughlin stützten diese Hypothese zumindest teilweise: Die Teilnehmer des Experiments konnten die Aufgabe wesentlich schneller lösen, wenn es sich bei dem gegebenen Ausdruckspaar um „whites“ und positive Attribute handelte, als bei Kombinationen von positiven Attributen und „blacks“. Dieser Befund ergab sich für die meisten Teilnehmer - unabhängig von den zuvor ermittelten „prejudice scores“. Hingegen zeigten sich bei negativen Attributen sowohl in Kombination mit „blacks“ als auch mit „whites“ keine Unterschiede in den Reaktionszeiten (vgl. Gaertner/Dovidio 1986: 81).

Eine zweite, ebenfalls häufig angewendete Methode für die Analyse von Stereotypen geht von dem vor allem von Rosch geprägten Priming-Paradigma aus. Im Unterschied zu der beschriebenen LDT-Variante wird bei diesem Verfahren zunächst nur die lexikalische Bezeichnung für eine bestimmte soziale Kategorie (z.B. „Türken“ oder „Deutsche“) vorgegeben. Erst dannach folgt die Präsentation eines Adjektivs. Die zentrale Annahme ist dabei, daß die Präsentation der lexikalischen Bezeichnung („cue“ oder „prime“) mit der Aktivierung einer entsprechenden Kategorie („priming“) einhergeht, die die Aufmerksamkeit für nachfolgende Stimuli und deren Wahrnehmung nach Maßgabe der spezifischen kognitiven Repräsentationen in Bezug auf diese Kategorie ausrichtet. Aus der Kenntnis der unterschiedlichen Verarbeitung der Stimulusinformationen bei gleichen Priming-Bedingungen lassen sich dann, so die Hypothese, Rückschlüsse auf die kognitiven Repräsentationen hinsichtlich der jeweiligen Kategorien bzw. sozialen Gruppen ziehen.

Dementsprechend haben zum Beispiel Dovidio, Evans und Tyler (1986) das Priming-Verfahren in einer Studie über Rassenstereotype in den USA angewendet. In dieser Untersuchung wurde den Teilnehmern des Experiments jeweils zuerst die Bezeichnung „blacks“ oder „whites“ präsentiert, um so die Aufmerksamkeit auf eine bestimmte (soziale) Kategorie zu lenken („priming“). Dann wurde jeweils ein Adjektiv vorgegeben, dessen positive oder negative Konnotation in Bezug auf die jeweiligen sozialen Kategorien aufgrund vorhergehender Untersuchungen eingeschätzt werden konnte. Die Aufgabe der Versuchspersonen bestand nun darin, möglichst rasch und genau anzugeben, ob das jeweilige Adjektiv ihrer Ansicht nach zu der zuvor aktivierten Kategorie paßt. Als Indikator der Assoziationsstärke diente ebenso wie beim „lexical decision task“ die Reaktionsgeschwindigkeit („response latency“). Die Ergebnisse der Studie zeigten einen deutlichen Interaktionseffekt zwischen „Prime Type“ und „Trait Type“: Die Reaktionszeit für die Zuordnung eines positiven Attributs war wesentlich kürzer, wenn zuvor die Kategorie „white“ präsentiert wurde (767 msec), als im Vergleichsfall, in dem die Kategorie „black“ vorgegeben wurde (908 msec). Allerdings ergaben sich auch in dieser Studie keine signifikanten Unterschiede in den Reaktionszeiten, wenn negative Attribute zugeordnet werden sollten (891 msec vs. 885 msec).

Sowohl die „lexical decision task“-Studien als auch die Untersuchungen im Rahmen des Priming-Paradigmas stellen exemplarische Versuche dar, mit Hilfe möglichst subtiler Messverfahren Stereotype ohne Verzerrungen durch sozial erwünschte Antworten und ähnliche „response errors“ zu erfassen. In beiden Varianten wird deshalb besonders Wert darauf gelegt, daß den befragten Personen keine explizit bewertende Einschätzung einer bestimmten Person oder Personengruppe abverlangt wird, da sich damit, so jedenfalls die Vermutung, die ansonsten auftretenden Reaktivitätseffekte verringern lassen (Gaertner/ Dovidio 1986: 80f). Schlußfolgerungen über Inhalte und Ausrichtung spezifischer Stereotypen sind folglich auch nur indirekt zu ziehen, indem aus den ermittelten Reaktionszeiten Aussagen über die mental verfügbaren kognitiven Verknüpfungen zwischen einem bestimmten „category label“ und dessen assoziierten Merkmalen abgeleitet werden können. Die „Daumenregel“ für eine solche Ableitung lautet etwa so: Je kürzer die erforderliche Reaktionszeit für die Verarbeitung einer gegebenen Konzept-Attribut-Kombination ausfällt, desto stärker ist das darin enthaltene Attribut im Langzeitgedächtnis mit dem Konzept verknüpft und kann folglich als ein Element des Stereotyps der entsprechenden kategorial bestimmten sozialen Gruppen angesehen werden. In ähnlicher Weise können diese Verfahren auf die Analyse von Vorurteilen angewandt werden, wobei dann allerdings die Erfassung explizit evaluativer Dimensionen der assoziativen Verknüpfungen und häufig auch deren konativer Korrelate beabsichtigt wird (vgl. dazu ausführlicher 2.2.2.2).

Aus kognitionstheoretischer Perspektive stellen diese Verfahren nicht nur wegen der erwarteten Verringerung von „response errors“ einen wichtigen methodischen Fortschritt dar. Ihre Vorzüge gegenüber den verschiedenen Varianten der Eigenschaftslistenverfahren werden darüber hinaus auch in der besseren Anwendbarkeit für die Analyse der kognitiven Strukturierung von Stereotypen, deren Konsequenzen für die Informationsverarbeitung und Personenwahrnehmung und andere eher prozessuale Aspekte der Stereotypisierung gesehen (vgl. Hamilton/Trolier 1986: 141f). Allerdings steht diese grundsätzlich positive Einschätzung auch innerhalb der eher kognitionspsychologisch orientierten Forschung noch unter Vorbehalten. So ist beispielsweise sowohl bei den LDT- als auch bei den Priming-Verfahren keineswegs auszuschließen, das die damit ermittelten Resultate nicht ebenfalls durch bestimmte, wenngleich verringerte Reaktivitätseffekte beeinflußt werden. Vor allem beim Priming-Verfahren kann die Verzögerung zwischen der Präsentation des „primes“ und des Attributs, zumal bei mehreren aufeinanderfolgenden Präsentationen, dazu führen, daß die Befragten ihre Reaktionen abwägen und sich dabei auch von Vermutungen über sozial erwünschte Antworten oder über spezifische Aufforderungscharakteristika (demand characteristics) der Erhebungssituation beeinflußen lassen (vgl. Messick/Mackie 1989: 49f; Stangor/ Lange 1994: 365; Dovidio/Fazio 1992: 224). Solche Einflüsse lassen sich jedoch zumindest teilweise durch Variationen der zeitlichen Verzögerung (stimulus onset asynchrony, SOA) gezielt kontrollieren - eine Überlegung, die sich überdies für Untersuchungen über eher automatisch oder eher kontrolliert ablaufende Prozesse der Aktivierung von Stereotypen oder Vorurteilen als nützlich erwiesen hat (vgl. 2.2.2.2).

Grundlegender sind indes einige Vorbehalte und Kritikpunkte, die zum einen die Konzeptualisierung des Stereotypenkonstrukts im Rahmen dieser Verfahren und zum anderen die Interpretation der Resultate betreffen. Vorbehalte hinsichtlich der spezifischen Konzeptualisierung von Stereotypen beziehen sich in erster Linie auf die Gültigkeit der Annahme, diese seien als mentale Assoziationen zwischen Kategorien und

Attributen zu erfassen und könnten durch das meßbare Ausmaß der Assoziationsstärke näher bestimmt werden. Das dem zugrunde liegende „Assoziationsmodell“ der Repräsentation von Stereotypen wird zwar im allgemeinen als relativ gut bewährt angesehen (Stangor/Lange 1994: 366). Doch es gibt auch einige andere empirisch gestützte „respresentational models“, deren Implikationen für die Bestimmung von Stereotypen bislang auch in der kognitionstheoretischen Forschung kaum überprüft wurden. Die damit verbundenen Aspekte der kognitiven Strukturierung mentaler Repräsentationen stellen allerdings Spezialprobleme der Kognitionspsychologie dar und müssen an dieser Stelle nicht ausführlicher behandelt werden (vgl. Hilton/von Hippel 1996: 241-244; Messick/Mackie 1989: 50f; Dovidio/Evans/ Tyler 1986: 23).

Schwierigkeiten bei der Interpretation der Resultate aus den Untersuchungen im Rahmen des LDT und des Priming-Paradigmas sind zunächst mit einem bereits aus den Eigenschaftslistenverfahren bekannten Problem verbunden: Es ist letztlich nicht eindeutig auszumachen, ob und in welchem Maß die Reaktionen (und die Reaktionszeiten) tatsächlich Rückschlüsse auf die subjektiv für wahr gehaltenen Meinungen der Befragten über die Eigenschaften bestimmter Personengruppen zulassen; die Befunde könnten ebenso darauf beruhen, daß die Befragten ein als konsensual wahrgenommenes Stereotyp wiedergeben, ohne daß dieses zwangsläufig ihren eigenen Meinungen korrespondieren muß. Diese Schwierigkeiten ergeben sich in erster Linie dann, wenn - wie in den Studien von Dovidio, Evans und Tyler (1986) oder von Gaertner und McLaughlin (1983) - sehr allgemein danach gefragt wird, ob es sich bei einer gegebenen Kategorie-Attribut-Kombination um sinnvolle Wörter bzw. um eine plausible Reihung handelt. Diese sehr offen gehaltene Aufgabenstellung mag zwar im Hinblick auf die angestrebte Verringerung von sozial erwünschten Antworten und ähnlichen Antwortverzerrungen sinnvoll sein, läßt damit aber auch den Referenzrahmen der Reaktion unbestimmt, mit der Folge, daß allein aufgrund der somit ermittelten Ergebnisse keine Differenzierung zwischen individuellen und konsensualen Stereotypen möglich ist.

Außerdem bereiten auch die Resultate der LDT- und Priming-Studien hinsichtlich der unterschiedlichen Reaktionszeiten Interpretationsprobleme. Die Annahme, kurze Reaktionszeiten seien Indikatoren eines leicht verfügbaren kognitiven Modells einer Personengruppe und damit eines Stereotyps (und umgekehrt), ist keineswegs zwingend. Deutlich wird dies vor allem im umgekehrten Fall, bei relativ langen Reaktionszeiten: Weisen diese darauf hin, daß kein Stereotyp der betreffenden Kategorie bzw. Gruppe verfügbar ist, oder daß dieses zumindest nur relativ schwach verankert ist? Oder ist die längere Reaktionszeit mancher Befragter darauf zurückzuführen, daß sie ihre Antwort - aus unterschiedlichen Gründen - länger überdenken, obwohl vielleicht ein entsprechendes Stereotyp verfügbar ist? Verlängert sich in jedem Fall die Reaktionszeit, wenn keine vorgefertigte Meinung über die typischen Merkmale einer bestimmten Personengruppe verfügbar ist?

Solche Fragen lassen sich allein auf der Grundlage der Befunde der LDT- und Priming-Studien nicht beantworten. Sie erfordern letztlich eine Überprüfung der Ergebnisse durch zusätzliche Untersuchungen, mit Hilfe derer die Kriteriumsvalidität dieser Messverfahren genauer bestimmt werden kann. Einige dieser Untersuchungen werden im folgenden Abschnitt ausführlicher dargestellt.

2.2 Vorurteile

2.2.1 Der Begriff des Vorurteils

Ähnlich wie der Begriffs des Stereotyps ist auch der Vorurteilsbegriff in den Sozialwissenschaften durch eine enorme Bedeutungsvielfalt - oder genauer: durch „a great deal of excess meaning“ (Gardner 1994: 1) - gekennzeichnet, so daß auf den ersten Blick eine allgemeine Verständigung über den Begriffsinhalt kaum möglich erscheint. Besonders auffällig ist, daß auch in Vorschläge zur Bestimmung des Vorurteilskonzepts zahlreiche für die Zwecke empirischer Forschung problematische Grundannahmen und Bewertungen eingebaut werden. Dazu zählen insbesondere Begriffsbestimmungen wie Allports Definition des „ethnischen Vorurteils“ als „an antipathy based upon faulty and inflexible generalization“ (Allport 1954: 9), aber auch zahlreiche andere Definitionen, in denen als zentrale Komponenten des Konzepts Merkmale wie „Abweichungen von der Realität“, „Abweichungen von allgemein akzeptierten Wertvorstellungen“, „Änderungsresistenz“, „Negativität“ oder „unbewiesene Pseudo-Urteile“ benannt werden, bis hin zu jenen, in denen Hypothesen über die Entstehung oder Funktionsweise von Vorurteilen in die Definitionen miteinbezogen werden (vgl. Schäfer/Six 1978: 13-18; Duckitt 1992: 9-11 und 15-17).
Die Hintergründe solcher Begriffsbestimmungen sind vermutlich diesselben wie die der entsprechenden Charakterisierungen von Stereotypen: Sie schließen unmittelbar an das alltagssprachliche Verständnis von Vorurteilen an und bringen zugleich Vermutungen oder Überzeugungen über die negativen Auswirkungen von Vorurteilen auf soziale Interaktionsprozesse zum Ausdruck. Allerdings sind auch sie im Hinblick auf die Ziele sozialwissenschaftlicher Beschreibung und Erklärung ungeeignet (Duckitt 1992: 23). Sinnvoller sind stattdessen jene Vorschläge, die den Vorurteilsbegriff, losgelöst von den normativen Konnotationen, auf bekannte Oberbegriffe wie „Urteile“, „Antipathie“ oder „Einstellungen“ zurückführen und auf dieser Grundlage präzisieren.

Insbesondere das Einstellungskonzept hat sich vor allem in der sozialpsychologischen Vorurteilsforschung als ein brauchbarer Bezugspunkt erwiesen und kennzeichnet mittlerweile die meisten Konzept-spezifikationen. Vorurteile sind demnach als eine spezielle Variante von Einstellungen aufzufassen, die im wesentlichen dadurch bestimmt ist, daß sie sich auf bestimmte Einstellungsobjekte beziehen, nämlich auf Gruppen bzw. auf die diesen Gruppen kategorisch zugeordneten Personen (vgl. Stroebe/Insko 1989: 8; Brown 1995: 8). Einstellungen wiederum können bestimmt werden als „a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor“ (Eagly/Chaiken 1993: 1) oder als „Bereitschaft zur negativen oder positiven Bewertung eines Einstellungsobjekts, die auf Gefühlen und Meinungen über den Einstellungsgegenstand beruht“ (Stroebe 1980: 10, zit. in Stroebe 1985: 9). Zwar gibt es auch für das Einstellungskonzept keine allgemein akzeptierte Spezifikation, doch es besteht zumindest darüber weitgehend Konsens, daß Einstellungen (a) Bewertungen beinhalten, (b) im Gedächtnis abgespeicherte und abrufbare Wissensstrukturen oder „Repräsentationen“ darstellen und (c) affektive, kognitive und konative Komponenten aufweisen (Olson/Zanna 1993: 119-122). Unter der affektiven Komponente wird im allgemeinen sowohl die Bewertung als auch die emotionale Orientierung auf Personen oder soziale Sachverhalte verstanden, die etwa für Likert, Guttman oder Thurstone noch den Inbegriff der „Attitüde“ darstellten; die kognitive Komponente verweist auf Wahrnehmungen, Überzeugungen und Erwartungen in Bezug auf Personen oder soziale Sachverhalte; und die konative Komponente schließlich

soll die Verhaltensrelevanz bzw. die Prädisposition zu einem bestimmten Handeln zum Ausdruck bringen (Duckitt 1992: 11-15; Ehrlich 1979: 31; Schäfer/Six 1978: 16).

Diese sogenannte „3-Komponentenkonzeption“ der Einstellungen („tripartite model“) ist vor allem von Fishbein und Ajzen (1975) nachdrücklich kritisiert worden, weil sich die drei Komponenten empirisch, wie die mehrfach nachgewiesenen Interkorrelationen zwischen den Konstrukten anzeigen, kaum voneinander unterscheiden lassen. Doch zumindest als begrifflicher Bezugsrahmen ist sie für die Spezifikation des Vorurteilbegriffs hilfreich[7]. Demgemäß kann als zentraler Bestandteil des Begriffs zunächst das Element der Bewertung von Gruppen bzw. von diesen Gruppen zugeordneten Personen angeführt werden, für das wiederum unterschiedliche „response categories“ (affektiv, kognitiv und konativ) und unterschiedliche „response modes“ (verbal und nonverbal) differenziert werden können (Manstead 1996: 47; Eagly/Chaiken 1993: 10-14). Für einige Autoren (Olson/Zanna 1993: 120-122; Eagly/Chaiken 1993: 14-16) dient diese Differenzierung darüber hinaus auch als eine Art „Heuristik“ für die gedankliche Ordnung der Voraussetzungen und Implikationen von Einstellungen, aufgrund derer sich unterschiedliche Entstehungsbedingungen (zum Beispiel: „affect-based attitudes“, „cognition-based attitudes“ usw.) und Konsequenzen der Aktivierung von Einstellungen typologisch zuordnen lassen.

Bezogen auf das Vorurteilskonzept bedeutet dies, daß darunter (a) mit Bewertungen verbundene Überzeugungen, Meinungen bzw. „beliefs“ über die Eigenschaften und Merkmale bestimmter Personengruppen (z.B. „Neger sind faul“), (b) emotionale Reaktionen in der Beziehung zu bestimmten Personengruppen (Mißtrauen, Furcht usw.) und (c) Verhaltensprädispositionen (z.B. Kontaktvermeidung) zu fassen wären (vgl. auch Harding/Proshansky/Kutner/Chein 1969: 1023-1029; Ehrlich 1979: 13-30; Heckmann 1992: 121-125; Stroebe/Insko 1989: 8). Die Konzeptspezifikation ist somit sehr allgemein, da zumindest auf der Ebene dieser begrifflichen Präzisierung zum Beispiel ungeklärt bleibt, ob eine solche Bewertung „positiv“ oder „negativ“, „richtig“ oder „falsch“ ist, ob sie mit Gefühlen wie Haß, Mißtrauen oder Furcht gegenüber dem Einstellungsobjekt verbunden ist, oder ob ein Zusammenhang zwischen Vorurteil und Verhalten besteht. Sie ist in dieser Allgemeinheit folglich auch in Einklang zu bringen etwa mit Browns Definition von Vorurteilen als „the holding of derogatory social attitudes or cognitive beliefs, the expression of negative affect, or the display of hostile or discriminatory behaviour towards members of a group on account of their membership of that group“ (Brown 1995: 8) oder mit der Begriffsbestimmung von Allport, der Vorurteile definiert als „an aversive or hostile attitude toward a person who belongs to a group, simply because he belongs to that group, and is therefore presumed to have the objectionable qualities ascribed to the group“ (Allport: 1954: 7).

Die Bestimmung von Vorurteilen durch negative Affekte oder abwertende Meinungen ist, wie auch Brown und Allport deutlich machen, nicht zwingend. Auch positiv konnotierte Einstellungen gegenüber Personen aufgrund deren Zugehörigkeit zu einer bestimmten sozialen Kategorie oder Gruppe können als Vorurteile bezeichnet werden[8]. Allerdings gibt es gerade im Hinblick auf interethnische Beziehungen offensichtlich

[7] In neueren Beiträgen aus der Einstellungsforschung sind allerdings deutliche Anzeichen für eine auch empirisch begründete Rehabilitierung des „tripartite model“ erkennbar (vgl. Manstead 1995: 48f; Stroebe/Insko 1989: 8).

[8] Gardner (1994: 2) kritisiert in diesem Zusammenhang zu Recht eine häufig übersehene logische Implikation der Bestimmung von Vorurteilen als „negative judgments“: „Note, with this definition cannot be positive, nor can one have a prejudice about an ingroup. That is, ingroup favoritism, by definition, is not a form of prejudice“.

genügend Gründe für ein besonderers Interesse an negativ ausgerichteten Vorurteilen, das sich in den entsprechenden Begriffsbestimmungen widerspiegelt.

Die Zweckmäßigkeit dieser allgemeinen, stark an neueren sozialpsychologischen Beiträgen orientierten Konzeptspezifikation muß sich in der empirischen Forschung erweisen. Dann wird sich herausstellen, ob die angeführten Bestandteile des Definiens, also „Einstellung“ bzw. „Bewertung“ von Personen auf der Grundlage einer kategorialen Zuschreibung, eine angemessene Präzisierung der Untersuchungsgegenstände erlauben und inwiefern die Konstrukte beobachtbaren Indikatoren zugeordnet werden können. Auch die Frage nach dem Sinn der angeführten Unterscheidung der „Komponenten“ oder „response categories“ von Vorurteilen - einschließlich der vielfach diskutierten Fragen nach den Zusammenhängen zwischen Stereotypen und Vorurteilen einerseits und zwischen Vorurteilen und Diskriminierung andererseits - kann letztlich erst auf der Ebene der Hypothesenformulierung und anschließender empirischer Analysen geklärt werden.

Schon auf der Ebene der begrifflichen Präzisierung ist indes ein Manko der bisher umrissenen Konzeption von Vorurteilen festzustellen: Die vor allem von der sozialpsychologischen Vorurteilsforschung geprägte Begriffsbestimmung hat offensichtlich eine „psychologistische Schlagseite“, da der Vorurteilsbegriff zumindest definitorisch nur auf kognitive, affektive oder verhaltensbezogene Aspekte von Vorurteilen im Sinne von Einstellungen einzelner Personen bezogen wurde. Damit wird aber leicht jene Dimension von Vorurteilen übersehen, die Heckmann (1992: 118 und 138-154) als die „gesamtgesellschaftliche Existenzweise von Vorurteilen als Ideologien“ bezeichnet hat. Speziell im Hinblick auf interethnische Beziehungen ist damit gemeint, daß etwa abwertende Aussagen und Urteile über ethnische Gruppen nicht einfach nur isolierte Einzelurteile und Aussagen einzelner Personen darstellen, sondern darüber hinaus auch „objektivierte Aussagensysteme“, verstanden als „gesellschaftliche Ideologien“, repräsentieren. So wäre es zum Beispiel verkürzt und irreführend, Aussagen wie „Juden sind raffgierig“ oder „Neger sind faul“ ausschließlich als Ausdruck einer individuellen Einstellung einer einzelnen Person zu interpretieren, ohne zu berücksichtigen, daß solche Aussagen in allgemeine, häufig in hohem Maß sozial geteilte „belief systems“ oder Ideenkomplexe eingebettet sind, die sich nicht einfach auf eine Aggregation singularer Vorurteile reduzieren lassen.

Entstehung, Ausgestaltung und soziale Bedeutung dieser „belief systems“, Ideenkomplexe oder Ideologien sind bekanntlich seit langem Gegenstände einer umfangreichen Forschung, die sich mit „Rassismus“, „Sexismus“, „Antisemitismus“ und verwandten „Ismen“ befaßt. Insbesondere in der neueren sozialpsycholgischen Forschung werden diese Beiträge zur „gesamtgesellschaftlichen Existenzweise von Vorurteilen“ und die darin thematisierten Fragestellungen jedoch kaum zur Kenntnis genommen; es dominiert der Blick auf die einzelnen Träger von Vorurteilen. Für die speziellen Forschungsinteressen dieser Disziplin mag dies genügen. Eine allgemeine sozialwissenschaftliche Erklärung der Entstehung, Aktivierung und Folgen von Vorurteilen kann diese Aspekte jedoch nicht ausblenden.

Allerdings folgt daraus nicht, daß die vorgeschlagene Konzeptspezifikation für die Zwecke einer auf diese Zielsetzung bezogenen empirische Forschung unzulänglich ist. Die sozial relevanten Auswirkungen der in „objektivierten Aussagensystemen“ eingebetteten Vorurteile werden letztlich erst erfaßbar in Meinungen,

Einstellungen und Verhaltensweisen konkreter Akteure - und eine der interessantesten Fragen in diesem Zusammenhang ist, unter welchen Bedingungen diese Auswirkungen beobachtbar sind. Die begrifflichen Grundlagen einer solchen Analyse sollten mit der angeführten Begriffsbestimmung ausreichend präzisiert sein.

2.2.2 Operationalisierungen und Messverfahren

Die am häufigsten verwendeten Verfahren der Informationsgewinnung über Inhalte, Intensität und Verbreitung von Vorurteilen beruhen auf unterschiedlichen Varianten direkter mündlicher und schriftlicher Befragungen, insbesondere auf der Grundlage von Einstellungsskalen und Ratingverfahren. Ähnliches gilt für die Erfassung von Konstrukten wie Fremden- bzw. Ausländerfeindlichkeit oder Ethnozentrismus. Die dabei üblichen Techniken entsprechen logischerweise weitestgehend denen, die aus der allgemeinen Einstellungsforschung bekannt sind. Das heißt: Vorurteile, verstanden als eine spezielle Form von Einstellungen, werden im wesentlichen dadurch operationalisiert, daß den zu befragenden Personen zunächst eine Reihe von Fragen bzw. Aussagen („Statements“) vorgelegt wird, denen diese zustimmen oder nicht zustimmen können. Die so erfaßten Aussagen bzw. verbalen Verhaltensweisen auf die einzelnen Items sollen Rückschlüsse auf die Ausprägung der latenten Variablen bei den Befragten ermöglichen, zum Beispiel auf die Ausprägung des Vorurteils gegenüber einer bestimmten sozialen Gruppe oder der nationalen Identifikation. Einzelne Items, deren Indikatorfunktion sich in der Konstruktionsphase des Instruments bewährt haben, können dann zu Indizes bzw. zu Skalen zusammengefaßt werden, wobei nach Maßgabe unterschiedlicher Skalierungsverfahren spezifische Anforderungen erfüllt sein müssen, damit ein Item in die Skala aufgenommen werden kann. Je nach bevorzugter Spezifikation der Dimensionen der latenten Variablen und Skalierungsverfahren ergeben sich dann die bekannten Skalenvarianten: Thurstone-Skalen, Likert-Skalen, Guttman-Skalen, Bogardus’ Skala der sozialen Distanz usw.

Die methodologischen Einzelheiten der Konstruktion dieser Skalen, deren Unterschiede hinsichtlich des meßtheoretischen Status und der Itemcharakteristika oder auch die Details der Auswertung der damit ermittelten Einstellungs-Meßwerte müssen hier nicht dargestellt werden (vgl. dazu allgemein Himmelfarb 1993: 30-57; Schnell/Hill/Esser 1995: 128-199; speziell zur Vorurteilsforschung: Petermann 1988: 133-139; Schäfer/Six 1978: 66-80; Simpson/Yinger 1985: 94-97). Die Konstruktion eines Meßinstrumentarium zur Erfassung etwa von Vorurteilen gegenüber ethnischen Minderheiten oder von Fremdenfeindlichkeit im allgemeinen unterscheidet sich in den technischen Details nicht von der Konstruktion eines Instrumentariums zur Messung von Einstellungen zu irgendwelchen anderen „issues“. Spezifische Probleme der Operationalisierung und Messung von Vorurteilen und ähnlichen Konstrukten ergeben sich allerdings aus dem besonderen inhaltlichen Gegenstandsbereich, vor allem aus der starken Anfälligkeit darauf bezogener Umfragen für „response errors“. Die bereits im Zusammenhang mit der Erforschung von Stereotypen angeführten Tendenzen der Antwortverzerrungen durch Abgabe sozial erwünschter Antworten oder durch Reaktionen auf spezielle Itemformulierungen erschweren auch in diesen Fällen die Erfassung der „wahren“ Einstellungen und Verhaltensweisen. Es kann sogar eine weitere Verschärfung dieser Probleme bei der Untersuchung von Vorurteilen oder Fremdenfeindlichkeit vermutet werden, sofern dabei nicht nur Meinungen sondern explizite Bewertungen erhoben werden sollen, deren Offenlegung für die Befragten eventuell besonders unangenehm sein kann.

Diese besonderen Schwierigkeiten sind seit längerem Ausgangspunkt und Gegenstand zahlreicher Vorschläge zur Entwicklung von Meßinstrumentarien, die eine höhere Validität und Reliabilität gewährleisten sollen. Im großen und ganzen lassen sich dabei zwei Hauptrichtungen unterscheiden: einerseits Vorschläge zur Verbesserung der Itemformulierungen und andererseits Ansätze zur Ausarbeitung neuer Messverfahren[9]. Einige wichtige Vorschläge werden im folgenden ausführlicher dargestellt.

2.2.2.1 Modifikationen der Einstellungsskalen und Ratingverfahren

Bei allen Unterschieden hinsichtlich der inhaltlichen Schwerpunkte, des methodischen Zugangs oder der bevorzugten Grundgesamtheit der Befragten zeigt sich in den meisten Analysen zu Vorurteilen, Fremdenfeindlichkeit und ähnlichen Themen als grundlegende Gemeinsamkeit nicht nur die ungebrochene Dominanz der auf Einstellung- und Ratingskalen beruhenden Umfrageforschung. Auch die den Befragten üblicherweise vorgelegten Fragebögen weisen deutliche Ähnlichkeiten auf. Sie enthalten in den meisten Fällen Fragen bzw. Aussagen hinsichtlich der allgemeinen Wahrnehmung und Bewertung von „Fremdgruppen“, der sozialen Kontakte mit Angehörigen dieser Gruppen, der bevorzugten Politik ihnen gegenüber usw. So wurden zum Beispiel die Befragten in den ALLBUS-Umfragen seit 1980 gebeten, folgende jeweils gleichlautende Einstellungsitems auf einer Skala von „Stimme überhaupt nicht zu“ bis „Stimme voll und ganz zu“ zu beantworten: „Gastarbeiter sollten ihren Lebensstil ein bißchen besser an den der Deutschen anpassen“, oder „Man sollte Gastarbeitern jede politische Betätigung in Deutschland untersagen“. Der deutsche Fragebogen des ISSP 1995 („National Identity“) enthielt unter anderem folgende Statements: „Zuwanderer erhöhen die Kriminalitätsrate“ und „Ausländern sollte es nicht erlaubt sein, in Deutschland Grund und Boden zu erwerben“. Die Liste solcher Itemformulierungen ließe sich beliebig verlängern. Sie finden sich in zahllosen repräsentativen nationalen Surveys und diversen Spezialuntersuchungen.

- (a) Die „Modern Racism“-Skala

Insbesondere in der US-amerikanischen „Racism“-Forschung ist in den vergangenen Jahren verstärkt Kritik an solchen Itemformulierungen, wie sie zum Beispiel auch im verbreiteten „Inventory of Verbal Racial Attitudes“ (Woodmansee/Cook 1967) enthalten sind, geäußert worden. In empirischer Hinsicht gründete diese Kritik in dem Verdacht, daß die in den NORC-Surveys, Harris-Polls und anderen Umfragen ausgewiesenen Veränderungen der Einstellungen und Verhaltensweisen gegenüber Schwarzen in Richtung auf eine zunehmende Abschwächung traditioneller Vorurteile und geringere Diskriminierungsneigungen

[9] Einige weitere Vorschläge zielen auf eine Modifikation der Skalenbildung. Hintergrund dieser Überlegungen ist die Annahme, daß die Befragten vor allem bei negativ konnotierten Statements in Bezug auf andere Personengruppen zur Abgabe sozial erwünschter Antworten neigen. Dovidio, Mann und Gaertner (1989) haben deshalb in einer Untersuchung positive und negative Items in separaten Skalen plaziert („good“: 1 = „not at all“ bis 7 = „extremely“, „bad“: 1 = „not at all“ bis 7 = „extremely“). Aufgrund ähnlicher Überlegungen ersetzten Gaertner und McLaughlin (1983) die konventionellen positiven und negativen Skalen durch moderat positive (negative) bis sehr positive (negative) Skalen. In beiden Fällen zeigte sich, daß bei den Ratings auf den negativen Skalen keine wesentlichen Unterschiede in der Bewertung von Schwarzen und Weißen auftraten, während in Bezug auf die positiven Skalen deutliche Bewertungsunterschiede zugunsten der „ingroup“ erkennbar waren.

nicht sonderlich glaubwürdig seien. Zu offensichtlich erschien ihr Widerspruch zu Erkenntnissen über faktische Verhaltensweisen in Bezug auf alltägliche interethnische Beziehungen oder auf politische Präferenzen (vgl. Dovidio/Gaertner 1986: 3-12). Dieser Befund untermauerte die Kritik an der bereits zuvor mehrfach experimentell nachgewiesenen Reaktivität des Meßinstruments, die sich nicht zuletzt auch in den Umfragen in hohen Verweigerungsraten niederschlug (vgl. McConahay 1986: 93f).

Als zentrale Ursachen der zunehmend verringerten Validität des traditionellen Meßinstrumentariums wurde jedoch nicht allein die vermutlich generell höhere Störanfälligkeit der üblichen Verfahren der Umfrageforschung angesehen. Vielmehr wurde darüber hinaus davon ausgegangen, daß die bislang verwendeten Einstellungs- und Ratingskalen nicht mehr in der Lage seien, einem tatsächlichen Einstellungswandel Rechnung zu tragen, der allerdings nicht zwangsläufig die Vorurteile unmittelbar betrifft, sondern die Erwartungen über die als sozial erwünschten bzw. „angemessenen“ Antworten: „what may have changed is what people regard as a socially desirable position rather than their racial attitudes per se“ (Dovidio/Gaertner 1986: 9). Der Wandel der Erwartungen über „opportune“ Meinungen und Einstellungen reflektiert dieser Einschätzung zufolge durchaus die faktischen sozialen und politisch-institutionellen Veränderungen in den interethnischen Beziehungen in den USA. Er kann aber nicht einfach mit einer Verringerung rassischer Vorurteile oder Diskriminierungen ineins gesetzt werden.

Vor diesem Hintergrund haben einige Autoren wie McConahay, Sears und Kinder versucht eine Reihe von neuen Items zu formulieren, die dem vermuteten Einstellungswandel besser gerecht werden sollten als die bis dato verwendeten Formulierungen. Das wohl bekannteste Ergebnis dieser Versuche ist die mittlerweile in zahlreichen Studien verwendete „Modern Racism Scale“ (MRS). Grundlage der Skala ist die Vermutung, daß neben den Rassismus alter Prägung („Old Fashioned Racism“) ein neues „set of beliefs“ getreten sei, das auf wesentlich subtilere Vorurteile gegenüber Schwarzen verweist. Als Kennzeichen des „Old Fashioned Racism“ gelten unter anderem: „beliefs about black intelligence, ambition, honesty, and other stereotyped characteristics, as well as support for segregation and support for open discrimination“ (McConahay 1986: 93).

Der „moderne Rassismus“ hingegen zeichne sich gerade dadurch aus, daß er auf einer expliziten Ablehnung dieser überlieferten Überzeugungen und Einstellungen beruhe. An deren Stelle sei nunmehr ein neues Einstellungssyndrom getreten, das sich im Kern auf die Verteidigung des status quo - einschließlich der damit verbundenen Ungleichheiten der Lebenschancen - richte. Dessen konkrete Merkmale seien vor allem: die Verneinung des Fortbestehens faktischer Benachteiligungen und Diskriminierungen; die Auffassung, die Schwarzen drängten zu hartnäckig und zu schnell in anspruchsvolle Tätigkeitsfelder, in denen sie häufig überfordert seien; deren Ansprüche und Forderungen seien zumeist überzogen und ungerechtfertigt; und ihre sozialen Errungenschaften seien im großen und ganzen unverdient, da sie häufig lediglich auf zu starken Bevorzugungen und nicht auf tatsächlichen Leistungen beruhen. Solche Aussagen werden, so ein weiteres Charakteristikum des „Modern Racism“, nicht mehr wie früher rassenideologisch begründet, sondern als empirische Feststellungen ausgegeben, denen allerdings implizit die säkularisierten Werte der traditionellen Protestantischen Ethik zugrunde gelegt sind (McConahay 1986: 92f). Kurz: Der moderne Rassismus erweist sich dieser Diagnose zufolge als „the expression in terms of abstract ideological symbols and symbolic beha-

viors of the feeling that blacks are violating cherished values and making illegitimate demands for changes in the racial *status quo*“ (McConahay/Hough 1976: 38, zit. in McConahay 1986: 95).

Diese neue Form des Rassismus, die in zunehmendem Maß neben den „altmodischen Rassismus“ tritt, soll mit der eigens dafür entwickelten „Modern Racism Scale“ erfaßt werden können. Ihre wesentlichen Unterschiede zu den traditionellen Einstellungs- bzw. Vorurteilsskalen lassen sich am besten an einigen Beispielen verdeutlichen (vgl. McConahay 1986: 106; Tab. 3):

Items der „Old Fashioned Racism Scale“:

- „Generally, do you feel blacks are smarter, not as smart or about as smart as whites?“
- „How strongly would you object if a member of your family had friendship with a black - strongly, somewhat, slightly or not at all?“

Items der „Modern Racism Scale“

- „Over the past few years, the governments and news media have shown more respect for blacks than they deserve.“
- „Blacks are getting too demanding in their push for equal rights.“

Solche Modifikationen der Item-Formulierungen sollen eine valide und zugleich nicht-reaktive Messung der neuen Varianten der Vorurteile gewährleisten. Der Erwartung geringerer Antwortverzerrungen stützt sich dabei vor allem auf eine zentrale Annahme: „We expected them to be less reactive in part because they tap into current issues about which there is no clear consensus on the prejudiced and nonprejudiced position and in part because for each item, there is a plausible, nonprejudiced explanation for endorsing the position scored as prejudiced on the scale“ (McConahay 1986: 97). Das heißt, die Item-Formulierungen zielen explizit auf Aussagen, für die der Ort sozial erwünschter Antworten nicht leicht erkennbar ist und die den Befragten die Möglichkeit offen läßt, ihre Äußerungen subjektiv in einem (scheinbar) vorurteilsfreien Bezugsrahmen zu begründen. Damit soll dem Einstellungswandel hinsichtlich der allgemeinen Vorstellungen über die Regulierung der interethnischen Beziehungen einerseits und der als „opportun“ erscheinenden Meinungen und Einstellungen gegenüber Schwarzen andererseits Rechnung getragen werden. Gleichzeitig soll aber die Skala als Ganzes sicher stellen, daß die wesentlichen Charakteristika des „modernen Rassismus“ - vor allem nach wie vor bestehende, eher unterschwellige „negative Affekte“ und die Orientierung auf eine Erhaltung des status quo - identifiziert werden können.

Die ausschließlich auf die US-amerikanischen Verhältnisse bezogenen Überlegungen zur Notwendigkeit einer Modifikation der üblichen Vorurteils- bzw. Rassismusskalen sind inzwischen auch in der auf Europa bezogenen Forschung auf Resonanz gestoßen (vgl. Brown 1995: 221-225). Hier treten sie vor allem unter Bezeichnungen wie „neuer Rassismus“, „alltäglicher Rassismus“, „latente Vorurteile“ und ähnlichen Titulierungen zum Vorschein. Von besonderem Interesse ist in diesem Zusammenhang der relativ sorgfältig ausgearbeitete Vorschlag von Pettigrew und Meertens zur Entwicklung und Anwendung eines Instrumentariums, das auf die Erfassung „subtiler Vorurteile“ im europäischen Kontext zugeschnitten ist.

Dieses Meßinstrument, dessen Items bereits 1988 in den Eurobarometer-Survey No. 30 aufgenommen wurden, wird im folgenden kurz dargestellt, ehe im Anschluß daran generelle Vorzüge und Nachteile solcher Modifikationen der traditionellen Einstellungs- bzw. Vorurteilsskalen, einschließlich der bislang vorliegenden Erkenntnisse über deren Reliabilität und Validität, beleuchtet werden.

- (b) Die „Subtle Prejudice"-Skala:

Ebenso wie der Entwicklung der „Modern Racism Scale" liegt auch der Erarbeitung der „Subtle Prejudice Scale" die Überlegung zugrunde, daß die bisher üblichen Erhebungsinstrumentarien zur Analyse von Vorurteilen, Fremdenfeindlichkeit und ähnlichen Phänomenen mittlerweile unzulänglich geworden sind. Diese erfassen nur die hinlänglich bekannten Formen der „blatant prejudices", die nach Pettigrew und Meertens (1995: 58) insbesondere durch zwei Komponenten gekennzeichnet sind: erstens durch eine generelle affektive Abneigung oder gar Abscheu gegen „Fremde" und Bedrohungswahrnehmungen, die sich unter Umständen mit dem Glauben an grundlegende, genetisch bedingte „Wesensunterschiede" vermischen; und zweitens durch Widerstände gegen allzu häufige Begegnungen und enge Kontakte mit den Angehörigen der „out-group", sei es im Wohnumfeld, am Arbeitsplatz oder erst recht in intimeren persönlichen Lebensbereichen. Die darin zum Ausdruck kommenden Vorurteile sind, so die Diagnose, keineswegs verschwunden. Allerdings unterliegen sie zunehmend öffentlicher Kritik und erscheinen in weiten Bevölkerungskreisen als inakzeptabel und sozial unerwünscht. Diese Entwicklung reflektiert einen faktischen Einstellungswandel, der sich in Anlehnung an Dovido und Gaertner zumindest als eine Verschiebung der „socially desirable position" hinsichtlich irgendwelcher Äußerungen über „Ausländer", ethnische Minoritäten und andere „Fremdgruppen" interpretieren läßt. Auch dieser Wandel kann jedoch nicht einfach als Indiz für einen generellen Bedeutungsverlust von Vorurteilen und fremdenfeindlichen Dispositionen gewertet werden. Vielmehr treten sie nun teilweise in eher verdeckten Formen „subtiler Vorurteile" in Erscheinung.

„Subtile Vorurteile" sind nach Einschätzung von Pettigrew und Meertens in erster Linie gekennzeichnet durch „three more covert components, each of which is expressed in ways deemed normative and acceptable in western societies" (Pettigrew/Meertens 1995: 58). Diese Komponenten sind: (a) „the defence of traditional values": Die Verhaltensweisen der „outgroup"-Angehörigen werden als inakzeptabel und für deren Fortkommen als hinderlich eingeschätzt, wobei als normativer Bezugsrahmen der vorherrschende Wertekanon der „ingroup" fungiert; (b) „the exaggeration of cultural difference": faktische Deprivilegierungen und soziale Probleme ethnischer Minderheiten werden nun nicht mehr als „natürlicher" Ausdruck genetischer Unterschiede aufgefaßt, sondern als Resultat tatsächlich bestehender, aber stark überbewerteter kultureller Divergenzen; (c) „the denial of positive emotions": an die Stelle genereller affektiver Abneigungen gegen die „Fremden" treten nun subtilere Formen affektueller Reaktionen, die sich vorrangig in der Negation positiver Eigenheiten der „outgroups" und der relativen Überbewertung der Charakteristika der eigenen Gruppe zeigen (Pettigrew/Meertens 1995: 58-60).

Um diese subtilen Formen von Vorurteilen ermitteln zu können, haben Pettigrew, Meertens und die anderen Mitarbeiter eines „joint research project", eine 10-Item-Likert-Skala entwickelt, deren einzelne Items faktorenanalytisch drei Gruppen zugeordnet werden können, die den drei angeführten Komponenten entsprechen. Sie umfaßt demnach vier „traditional values factor items", vier „cultural differences factor

items“ und zwei „positive emotions factor items“ (vgl. Pettigrew/Meertens 1995: 62, Tab. 1). Diese wurden zusammen mit einer weiteren 10-Item-Skala zur Messung des Konstrukts „blatant prejudice“ und neun unabhängigen Variablen im Eurobarometer-Survey 30 erhoben und ausgewertet.

- (c) Einschätzung der modifizierten Einstellungsskalen

Sowohl für die „Modern Racism Scale“ (MRS) als auch für die „Subtle Prejudice Scale“ (SPS) liegen inzwischen einige empirische Untersuchungen vor, die zumindest eine ungefähre Einschätzung der Güte dieser modifizierten Meßinstrumente ermöglichen. Da die MRS in den USA bereits seit längerem verwendet wird, ist in diesem Fall die Bewertungsgrundlage weitaus günstiger als im Fall der SPS, die noch relativ neuen Datums ist. Wenngleich die MRS ausschließlich auf die Beziehungen zwischen weißen und schwarzen Amerikanern zugeschnitten ist, sollen die darauf bezogenen empirischen Tests und kritischen Einschätzungen knapp zusammengefaßt werden, da sie infolge der ähnlichen Grundanlage einige vorsichtige Rückschlüsse auf die Validität der SPS und auf andere Meßinstrumente dieses Zuschnitts zulassen.

Die für die MRS vorliegenden Analysen ergeben nach Angaben von McConahay (1986) folgendes Bild:

1. Die Faktorenanalyse der MRS und der „Old Fashioned Racism Scale“ (OFRS) zeigt, daß die beiden Item-Cluster auf unterschiedlichen Faktoren laden. Allerdings sind die beiden Faktoren nicht vollständig voneinander getrennt. Die standardisierten Regressions- oder Pfadkoeffizienten für die Items der MRS ergaben in zwei Erhebungen (N= 879 und N=709) für den ersten Faktor („Modern Racism Factor“) Werte zwischen .300 und .800 und für den zweiten Faktor („Old Fashioned Racism Factor“) Werte deutlich unter .300. Umgekehrt luden die Items der OFRS mit Werten unter .180 auf den ersten Faktor und mit Werten über .300 auf den zweiten Faktor. Die Faktorenkorrelation lag zwischen .68 und .70. In einer weiteren Erhebung (N=167) mit leicht modifizierten Items wurden diese Ergebnisse weitgehend bestätigt; allerdings wiesen die Faktorladungen in diesem Fall ein weniger einheitliches Muster auf; der Koeffizient der Faktorenkorrelation sank auf .59. „Hence, it appears that we have two distinct but correlated empirical cognitive dimensions: modern racism beliefs and old-fashioned racism beliefs“ (McConahay 1986: 110).

2. Die Reaktivität der MRS ist im Vergleich zur OFRS deutlich geringer. Dies zeigen Experimente mit weißen und schwarzen Interviewern, bei denen die OFRS-Scores im einen Fall (schwarze Interviewer) deutlich geringer ausfielen als im anderen (weiße Interviewer), während die MRS-Scores in beiden Experimentalbedingungen keine signifikanten Unterschiede aufwiesen. Weitere Experimente deuten darauf hin, daß die Items der OFRS von Befragten eher als „rassistisch“ eingeschätzt werden als die Items der MRS. Allerdings: „the racial implications of the Modern Racism Scale items were not completely lost on the subjects“ (McConahay 1986: 114).

3. Die MRS ist ein valides Instrument zur Messung von Vorurteilen bzw. „racism“. Untersuchungen zur Kriteriumsvalidität im Rahmen einer Analyse des Wahlverhaltens zeigten, daß Personen mit hohen MRS-Scores ihre Stimme eher einem weißen Kandidaten geben als einem konkurrierenden schwarzen Bewerber. Die Korrelationen in einer Reihe solcher Untersuchungen ergaben Werte zwischen .300 und .400. Weiterhin zeigten sich deutliche Zusammenhänge zwischen hohen MRS-Werte und der Stärke des Widerstands gegen

„Busing“-Programme (.511 und .391). Und außerdem fanden sich in einem Experiment mit simulierten Entscheidungen über die Einstellung von weißen oder schwarzen Stellenbewerbern zusätzliche Evidenzen für die Validität der MRS (McConahay 1986: 115-120; vgl. Brown 1995: 221).

Im Vergleich zu diesen Befunden, die auf zahlreichen Studien beruhen (vgl. auch Brown 1995: 221f; Dovidio/Fazio 1992: 218), kann sich die Einschätzung der SPS bislang nur auf einige wenige empirische Untersuchungen stützen. Die wichtigste ist die Studie von Pettigrew und Meertens auf der Grundlage der Daten des Eurobarometers 30, die in Frankreich, den Niederlanden, Großbritannien und Westdeutschland (N(ges) = 3788) erhoben wurden[10]. Deren zentralen Ergebnisse sind:

1. Die Faktorenanalyse deutet darauf hin, daß die beiden konzeptionell unterschiedenen Varianten des Vorurteils („blatant prejudice“ und „subtle prejudice“) auch empirisch sinnvoll getrennt und gemessen werden können, gleichzeitig aber moderate Interkorrelationen aufweisen. Die Ergebnisse der konfirmatorischen Faktorenanalyse unter Verwendung von EQS für alle sieben unabhängigen Samples zeigen, daß die Korrelationsmatrix für alle 20 Items am besten durch ein korreliertes Zwei-Faktoren-Modell oder durch ein „second-order hierarchical model“ repräsentiert wird. Dies kann als Indiz dafür gewertet werden, daß SPS und BPS entweder zwei grundsätzlich unterscheidbare, jedoch miteinander verbundene Konstrukte messen, oder aber zwei latente Faktoren erster Ordnung, die gleichermaßen auf einen latenten Faktor zweiter Ordnung (Vorurteil) laden. Cronbachs Alpha-Koeffizient für die Korrelation zwischen den beiden Skalen reicht von .70 (Frankreich) bis .48 (Niederlande) (Pettigrew/Meertens 1995: 64-66).

2. Als abhängige Variable teilen BP und SP ähnliche Korrelate, die allerdings von unterschiedlicher Bedeutung sind. Als stärkster Prädiktor für beide Skalen und alle Samples erweist sich eine von diesen unabhängig erhobene Drei-Item-Ethnozentrismus-Skala. Vergleichsweise stark erscheint auch der Einfuß der Akzeptanz rassistischer Bewegungen („racist movement approval“) auf die BPS-Scores. Für eine Reihe weiterer unabhängiger Variablen sind die Zusammenhänge mit den beiden Skalen allerdings nur sehr schwach, weisen aber zumindest die gemeinhin erwarteten Vorzeichen auf. Positive Korrelationen mit beiden Skalen zeigen sich für folgende Konstrukte: „politischer Konservatismus“, „fraternale Deprivation“, „Nationalstolz“ und „Alter“. Negative Korrelationen ergeben sich für „interethnische Kontakte“, „politisches Interesse“ und „Bildungsniveau“ (Pettigrew/Meertens 1995: 66-68; vgl. Wagner/Zick 1995: 42-46).

3. Signifikante und konsistente Zusammenhänge ergeben sich zwischen den BP- bzw. SP-Scores und den ebenfalls erhobenen Statements zur rechtlichen Stellung von Immigranten, zur bevorzugten Immigrationspolitik und zu den präferierten Strategien einer Verbesserung der interethnischen Beziehungen. Diejenigen Befragten, die sowohl auf der BPS als auch auf der SPS hohe Werte erreichten, befürworteten generell einen restriktiven Umgang mit Immigranten, während diejenigen, die auf beiden Skalen niedrige Werte aufwiesen, eine Erweiterung der Recht für Immigranten und ähnliche politische Maßnahmen befürworteten. Für die „subtles“, mit niedrigen Werten auf der BPS und relativ hohen Werten auf der SPS,

[10] Mit Ausnahme von Westdeutschland, wo sich die Eurobarometer-Erhebung auf Einstellungen gegenüber Türken (N=989) beschränkte, wurden in den anderen Ländern jeweils zwei getrennte Stichproben gezogen: in Frankreich für Nordafrikaner (N=455) und Asiaten (N=475), in Großbritannien für Asiaten (N=482) und „West Indians“ (N=471) und in den Niederlanden für Türken (N=476) und Surinamesen (N=462) (vgl. Pettigrew/Meertens 1995: 60f; Wagner/Zick 1995: 42f).

ergab sich indessen folgendes Bild: „They adopt intermediate positions that reject minorities in socially acceptable ways. Thus, they would neither restrict further nor enlarge immigrants' rights. They would not send all immigrants 'home', but would send 'home' those for whom there is an ostensibly non-prejudicial reason to do so" (Pettigrew/Meertens 1995: 72).

Insgesamt lassen sich diese Ergebnisse als Indizien dafür ansehen, daß die Unterscheidung zwischen „offenkundigen" und „subtilen" Varianten des Vorurteils fruchtbar sein kann und daß die beiden Konstrukte mit dem verwendeten Instrumentarium einigermaßen valide und reliabel gemessen werden können. Diese Einschätzung ist allerdings zwangsläufig tentativ. Anlaß zur Vorsicht geben unter anderem die in der Studie von Pettigrew und Meertens ausgewiesenen statistischen Kennziffern, die vielfach hinter den üblichen Standards zurückbleiben. Und vor allem mangelt es an weiteren Validierungsstudien, in denen die vorliegenden Ergebnisse einer kritischen Analyse unterzogen werden.

Die Überlegungen, die sowohl zur Konstruktion der „Modern Racism Scale" als auch der „Subtle Prejudice Scale" geführt haben, gehen über die ansonsten gelegentlich vorgenommenen Modifikationen solcher Meßinstrumente hinaus. Im Unterschied etwa zu den kürzlich vorgenommenen Reformulierungen der sogenannten „Gastarbeiterfragen" im ALLBUS, die sich auf den Austausch des Begriffs „Gastarbeiter" durch den Begriff „in Deutschland lebende Ausländer" beschränkten[11], zeichnen sich diese Bemühungen vor allem durch zwei Besonderheiten aus: (1) Sie beruhen auf der Diagnose, daß sich infolge gesellschaftlicher Veränderungen in den USA wie in Westeuropa die normativen Bezugspunkte sozial erwünschter Einstellungen in Bezug auf ethnische Minderheiten, „Fremde" und andere Gruppen derart verschoben haben, daß die „altmodischen", „offenkundigen" Vorurteile und „belief systems" in zunehmendem Maße der Kritik ausgesetzt sind. Vorurteile und diskriminierende Einstellungsmuster sind damit aber, der Diagnose zufolge, nicht verschwunden, sondern treten nun teilweise in „modernisierten", „subtileren" Formen in Erscheinung. (2) Aus dieser Einschätzung wird die Annahme abgeleitet, zur Erfassung dieser subtilen Vorurteile bedürfe es eines neuen, ebenso subtilen Meßinstruments, da die bisher üblichen Instrumente dazu nicht in der Lage seien.

Sollten diese Diagnosen und Annahmen - über die gesellschaftlichen Veränderungen ebenso wie über die Tauglichkeit der Meßinstrumente - zutreffen, muß wohl davon ausgegangen werden, daß in einer großen Zahl von Untersuchungen die Verbreitung und Stärke genereller diskriminierender Einstellungen systematisch unterschätzt wird. Inwieweit dies der Fall ist, kann an dieser Stelle nicht genauer überprüft werden. Die bisher vorliegenden Erkenntnisse sprechen jedoch dafür, diese Überlegungen weiterzuverfolgen und zu überprüfen, ob die Entwicklung von Einstellungsskalen vom Typ der „Modern Racism Scale" und der „Subtle Prejudice Scale" tatsächlich reliable und valide Messungen einer neuen Variante des Vorurteils erlauben, die mit den „traditionellen" Einstellungsskalen nicht möglich sind.

[11] Die Reformulierung der Gastarbeiter-Items im ALLBUS wurde vor allem aus zwei wichtigen Gründen vorgenommen: Zum ersten erschien die mit den Begriffskomponenten „Gast" und „Arbeiter" verknüpfte Begriffsverengung auf bestimmte Statuspositionen und Personen mit nur vorübergehender Aufenthaltsdauer zunehmend als problematisch, zumal der Ausdruck „Gastarbeiter" auch in der Alltagssprache und in den Medien immer seltener benutzt wird. Zum zweiten wurde erkannt, daß der Stimulus „Gastarbeiter" nicht mehr die prototypischen Zielgruppen fremdenfeindlicher Einstellungen erfaßt, da neue Gruppen wie Asylbewerber und Flüchtlinge hinzugekommen sind, die vermehrt zum hauptsächlichen Zielobjekt der Fremdenfeindlichkeit wurden (Blank/Wasmer 1996: 45f).

Bei solchen Reformulierungen der Einstellungsitems müßten in jedem Fall eine Reihe von Schwierigkeiten berücksichtigt werden, die bereits aus der Diskussion um die „Modern Racism Scale“ und ähnliche Skalen in den USA bekannt sind. Dazu zählen die bislang noch nicht hinreichend spezifizierten empirischen Zusammenhänge zwischen den konzeptionell differenzierten Konstrukten und deren Erklärung (vgl. Brown 1995: 219f und 225; Sniderman/Tetlock 1986). Ferner hat sich gezeigt, daß die explizit „subtilen“ Item-Formulierungen, mit denen die subtilen Vorurteile enthüllt werden sollen, gravierende Interpretationsprobleme aufwerfen können. Gerade weil bewußt solche „Statements“ präsentiert werden, für die es eine plausible, nicht unmittelbar vorurteilsbezogene Erklärungs- oder Einordnungsoption gibt, kann häufig nicht eindeutig geklärt werden, inwiefern darauf bezogene Antworten der Befragten zuverlässige Rückschlüsse auf „Vorurteile“ zulassen. So muß zum Beispiel auch McConahay konstatieren, daß ein Item wie „Blacks shouldn’t push themselves where they’re not wanted“ nicht zwangsläufig „rassistische“ Einstellungen indiziert. Ähnliches läßt sich für einige Items der „Subtle Prejudice Scale“ feststellen. McConahay hat dieser Kritik zwar entgegengehalten: „the same nonracial, non-prejudiced explanation cannot be used for the prejudiced response across all the items“ (McConahay 1986: 97); entscheidend sei letztlich die Indikatorqualität der gesamten Skala, nicht die der einzelnen Items. Doch die inzwischen mehrfach aufgezeigten Interkorrelationen der MRS mit Indikatoren des politischen Konservatismus lassen sich mit diesem Argument ebensowenig entkräften wie einige andere Unklarheiten (Brown 1995: 225f; Fazio/Jackson/Dunton/ Williams 1995: 1020-1025; Sniderman/Piazza/Tetlock/ Kendrick 1991).

Wenn sich solche Schwierigkeiten bei künftigen Reformulierungen des Items und Skalen besser in Griff bekommen lassen, könnten auf diese Weise Meßinstrumente konstruiert werden, die eine Verringerung systematischer Antwortverzerrungen ermöglichten, „subtile“ Formen von Vorurteile erfassten und insofern „an important departure from traditional bipolar scales“ (Dovidio/Fazio 1992: 218) darstellten. Selbst dann wird es sich aber weiterhin um verbale „self-reports“ handeln, die mit den bekannten Vor- und Nachteilen verbunden sind: Sie sind relativ leicht in standardisierter Form für große Populationen zu erheben und können mit ausgefeilten Auswertungsverfahren analysiert werden. Doch auch „subtile“ Vorurteilsskalen bleiben mit dem, wenngleich möglichst minimierten Problem behaftet, daß der ihnen zugrunde liegende Erhebungszweck leicht durchschaut werden kann und unter Umständen Reaktionen in Richtung auf sozial erwünschte Antworten hervorruft. Und sie erlauben streng genommen nur Aussagen über „elicited verbal attitudes“, nicht aber über andere Einstellungsreaktionen oder über die Stärke einer Einstellung bzw. eines Vorurteils. Im folgenden sollen deshalb einige Möglichkeiten dargestellt werden, mit denen diese Beschränkungen umgangen werden können.

2.2.2.2 „Unauffällige“ und indirekte Messverfahren

Zu den in der allgemeinen Einstellungsforschung angewendeten „unauffälligen“ und indirekten Messverfahren zählen neben physiologischen Messverfahren eine Reihe von „verfälschungsimmunisierten“ Methoden, die im Kern darauf abzielen, durch Kontrolle der Randbedingungen des Entscheidungsverhaltens der Befragten die Möglichkeiten einer bewußten Verzerrungen der zu erhebenden Reaktionen zu minimieren (vgl. Petermann 1988: 129-133). Einige dieser Verfahren sind auch in Untersuchungen zu Vorurteilen und ähnlichen Konstrukten wiederholt eingesetzt worden. Mit physiologischen Messverfahren („galvanic skin response“, „pupillary response“, „facial electromyographic activity“ u.a.) wurde dabei in erster Linie versucht, emotionale Reaktionen gegenüber einem Einstellungsobjekt zu erfassen, um somit Erkenntnisse über affektive Korrelate von Vorurteilen zu gewinnen (als Beispiel: Vanman/Miller 1993)[12]. Diese Methoden sind allerdings sehr aufwendig und kaum außerhalb des Experimentallabors einzusetzen. Außerdem kann auf diese Weise häufig nur die Intensität emotionaler Reaktionen, nicht aber deren Richtung gemessen werden; es ist nicht auszuschließen, daß die gemessenen Reaktionen auch auf die Meßapparatur zurückzuführen sind usw. Diese Verfahren werden deshalb im folgenden ausgeklammert. Ebenfalls ausgeklammert wird die Bogus-Pipeline-Methode, die sowohl für die Untersuchung von Stereotypen als auch für die Analyse von Vorurteilen verwendet wird, hier aber nicht erneut dargestellt werden muß.

- (a) Die Kombination von Priming und Reaktionszeitenmessung

Wie bereits im Abschnitt über neuere Messverfahren der Stereotypenforschung aufgezeigt wurde, beruhen Priming-Verfahren auf der Grundannahme, daß abstrakte mentale Repräsentationen von Objekt-Attribut-Verknüpfungen erfaßt und gemessen werden können, indem Reaktionszeiten für unterschiedliche Objekt- und Attribut-Stimuli ermittelt werden. Entsprechende Verfahren sind inzwischen mehrfach in der Einstellungsforschung für unterschiedliche Objekte mit dem Ziel eingesetzt worden, das Vorliegen und die Bedingungen der Aktivierung von Einstellungen, verstanden als „an association in memory between a given object and one’s evaluation of that object“ (Fazio 1990: 81), zu analysieren. Ihre theoretische Grundlage ist im Grunde stets die Gleiche. Es wird davon ausgegangen, daß die Reaktionszeit, die benötigt wird, um eine Beurteilung von Zusammenhängen zwischen gegebenen Objekten und Attributen abzugeben, ein zuverlässiger Indikator für die Verfügbarkeit und Stärke einer entsprechenden Einstellung ist (Fazio et al. 1995: 1013). Als Hypothese formuliert: Je kürzer die Reaktionszeit, desto größer ist die Verfügbarkeit und Stärke der betreffenden Einstellung.

Diese auf zahlreiche kognitionspsychologischen Studien gestützten Grundüberlegungen sind mittlerweile in mehreren Untersuchungen auf die Analyse von Vorurteilen übertragen worden (zum Beispiel: Devine 1989; Gilbert/Hixon 1991; Dovidio/Gaertner 1991; Perdue et al. 1990). Zur Demonstration der dabei üblichen Vorgehensweise eignet sich vor allem ein Beitrag von Fazio, Jackson, Dunton und Williams (1995), in dem Priming-Verfahren und Reaktionszeitenmessungen über das gängige Prozedere hinausgehend verwendet wurden, um Indikatoren für die Einstellungen einzelner Personen zu bilden und diese dann mit externen Daten zu überprüfen.

[12] Für eine knappe Übersicht über diese physiologischen Messverfahren vgl. Himmelfarb (1993: 60-63).

In den Untersuchungen von Fazio und seinen Mitarbeitern wurden den Teilnehmern des Experiments, das angekündigt wurde als eine Studie über „word meaning as an automatic skill“, zunächst verschiedene (positive und negativ konnotierte) Adjektive auf einem Bildschirm präsentiert, die sie möglichst schnell und wahrheitsgemäß durch Drücken einer Taste („gut“ oder „schlecht“) beurteilen sollten; aus den dabei ermittelten Reaktionszeiten wurden später die „baseline latencies“ berechnet. In einer zweiten Phase sollten sie sich verschiedene Gesichter von Schwarzen, Weißen und Asiaten einprägen, die nacheinander auf dem Bildschirm gezeigt wurden. Die dritte Phase diente der Überprüfung des Erinnerungsvermögens in Bezug auf die zuvor präsentierten Gesichter, die zusammen mit einigen zuvor nicht gezeigten Gesichtern vorgelegt wurden. Erst die vierte Phase enthielt schließlich die eigentliche Priming-Prozedur. Hier bestand die angekündigte Aufgabe darin, erneut die bereits präsentierten Adjektive zu bewerten und sich gleichzeitig die in der zweiten Phase gezeigten Gesichter zu merken, da diese anschließend ein weiteres Mal identifiziert werden sollten. Vor jedem Adjektiv, das auf dem Bildschirm gezeigt wurde, wurde jedoch für 315 Millisekunden das Gesicht eines Schwarzen oder Weißen eingeblendet, ehe nach einem weiteren 135 Millisekunden-Intervall tatsächlich das Adjektiv erschien[13]. Dabei wurde sichergestellt, daß jedes Photo jeweils zusammen mit zwei positiven und zwei negativen Adjektiven kombiniert wurde. Die fünfte Phase umfaßte dann die angekündigte Überprüfung des Erinnerungsvermögens und in der letzten Phase wurde ein explizites Rating der zuvor präsentierten Gesichter auf den Photos erbeten (Fazio et al. 1995: 1015f).

Die Auswertung der so erhobenen Daten ergab - unter Berücksichtigung der individuellen „baseline latencies“ - eine deutliche und signifikante „Race of Student x Race of Photo x Valence of Adjective interaction“ ($F(1, 51) = 25.88$, $p < .001$). Die Reaktionszeiten der weißen Studenten waren deutlich kürzer, wenn vor positiv bewerteten Adjektiven unbemerkbar „White primes“ eingeblendet wurden und fielen länger aus, wenn die Primes Gesichter von Schwarzen repräsentierten; hingegen wurde die Reaktionszeit bei der Beurteilung negativer Adjektive verlängert, wenn zuvor die Gesichter von Weißen eingeblendet wurden, und verkürzt, wenn es sich um „Black primes“ handelte. Für die schwarzen Studenten ergaben sich genau umgekehrte Resultate. Diese unterschiedlichen „patterns of facilitation“ konnten also als deutliche Hinweise darauf interpretiert werden, daß Weiße die bildlich präsentierten Einstellungsobjekte „Weiße“ mit positiv bewerteten Adjektiven assoziierten, während die Einstellungsobjekte „Schwarze“ mit negativ konnotierten Charakteristiken verbunden wurden - und umgekehrt für die Schwarzen. Stimmten die im Gedächtnis abgespeicherten Bewertungen der als Prime präsentierten Einstellungsobjekte mit der Bewertungsqualität der Adjektive überein, konnten diese als Stimulusobjekt vorgelegten Adjektive aufgrund der erhöhten Zugänglichkeit schneller bewertet werden; Bewertungsdiskrepanzen bei Objekt-Adjektiv-Kombinationen (z.B. positiv bewertete Primes und negativ bewertete Negative) hingegen verlängerten die zur Informationsverarbeitung benötigte Zeit.

Auf der Grundlage dieser Daten berechneten Fazio und seine Mitarbeiter für jeden Befragten einen Index zur Einschätzung der individuellen Einstellungen gegenüber Schwarzen (für Details vgl. Fazio et al. 1995: 1017). Die Häufigkeitsverteilungen dieser individuellen Einstellungsindizes zeigten abermals, daß die Bewertungen von Schwarzen durch Weiße zumeist negativ (Mittelwert des Scores: -.26) und durch

[13] Die gewählten Zeitintervalle sollten sicherstellen, daß die Primes nur unbewußt bzw. „subliminal“ wahrgenommen werden konnten (vgl. Fazio et al. 1995: 1016, Anm. 1).

Schwarze durchgehend positiv (.33) waren, wobei es in der Gruppe der weißen Befragten deutliche individuelle Unterschiede gab. Weitaus interessanter als diese Resultate sind indes die Ergebnisse einer externen Validierung der individuellen Einstellungsindizes mit möglichst unauffällig erhobenen Daten zum Verhalten der Befragten gegenüber einer schwarzen Versuchsleiterin sowie zur Bewertung des Gerichtsurteils im Rodney King-Prozeß und der anschließenden Unruhen.

Diese externe Validierung zeigt, daß die auf der Basis der Reaktionszeitenmessung geschätzten individuellen Einstellungen durchaus zuverlässige Prädikatoren des Verhaltens gegenüber einer schwarzen Versuchsleiterin waren. Weiße mit negativen Einstellungswerten verhielten sich unfreundlicher und desinteressierter als solche mit einem Bewertungsscore, der eher positive Einstellungen gegenüber Schwarzen indiziert. Vergleichbare Zusammenhänge ergaben sich außerdem zwischen den Einstellungsindizes und der Bewertung des Rodney King-Prozesses sowie der Frage nach der Verantwortung für die damaligen „racial riots“. Die ebenfalls ermittelten Einstellungsindikatoren auf der Grundlage der „Modern Racism Scale“ erwiesen sich im Vergleich dazu als wenig überzeugend: Schon bei der Reaktionszeitenmessung im Priming-Verfahren waren keine signifikanten Unterschiede zwischen Weißen mit unterschiedlichen MRS-Scores zu erkennen und auch in Bezug auf das Verhalten gegenüber der schwarzen Versuchsleiterin erwiesen sich diese als relativ untaugliche Prädikatoren (vgl. Fazio et al. 1995: 1020-1025).

Insgesamt lassen diese Befunde den Schluß zu „that the priming methodology may yield a valid, unobtrusive measure of racial attitudes“ (Fazio et al. 1995: 1020). Im Unterschied zu den üblichen Messverfahren wird darauf verzichtet, Einstellungen bzw. Vorurteile gegenüber Personengruppen durch explizite Angaben von Statements zu erheben; und ferner wird im Unterschied zu anderen Varianten des Priming-Verfahrens versucht, die Möglichkeit einer reflektierten Reaktion durch subliminale Präsentation des Einstellungsobjekts zusätzlich zu verringern. Damit kann sichergestellt werden, daß die Problematik einer Verzerrung durch Abgabe sozial erwünschter Antworten weitestgehend ausgeschaltet wird.

Außerdem ermöglicht das Verfahren eine Einschätzung des Ausmaßes, in dem die mit einem Einstellungsobjekt assozierte Bewertung bereits unmittelbar durch dessen Präsentation aktiviert wird. Das heißt, auf der Grundlage des Priming-Verfahrens und der Reaktionszeitenmessung kann ermittelt werden, ob und in welchem Maß bereits die - unter Umständen kaum oder gar nicht bewußt registrierte - Wahrnehmung etwa eines Schwarzen bestimmte, mit dieser sozialen Kategorie kognitiv verknüpfte (negative) Bewertungen automatisch zugänglich macht. Und es kann außerdem festgestellt werden, inwieweit sich die befragten Personen von der automatisch aktivierten Einstellung leiten lassen. Dies wiederum ist von besonderem Interesse im Hinblick auf die im MODE-Modell von Fazio (1990) unterschiedenen Modi der Informationsverarbeitung, insbesondere für den „spontaneous processing“-Modus, der in erster Linie für jene Situationen relevant ist, in denen lediglich „low cost“-Entscheidungen zu bewältigen sind, geringe Anreize für sorgfältige Kalkulationen bestehen oder keine Möglichkeiten zur aufwendigen Informationsverarbeitung gegeben sind.

Zusammenfassend lassen sich also - nach den bisher verfügbaren Erkenntnissen - einige beachtenswerte Vorzüge der mit dem Priming-Verfahren kombinierten Reaktionszeitenmessung zur Untersuchung von Einstellungen bzw. Vorurteilen feststellen:

- das Messverfahren ist „unauffällig“ und ermöglicht eine Minimierung möglicher Antwortverzerrungen durch „social desirability effects“[14];
- das Meßinstrument ist valide; Meßergebnisse, die mit dem Instrument erzielt werden, reflektieren die zwischen den „known groups“ (hier: Schwarze und Weiße) erwartbaren Einstellungsunterschiede und stehen in einer plausiblen und theoretisch begründbaren Beziehung zu separat gemessenen Außenkriterien (hier: Verhaltensbeobachtungen)[15];
- die prädikative Qualität des Meßinstruments (bezogen auf Verhaltensbeobachtungen) ist besser als die direkter Erhebungsverfahren;
- auf der Grundlage der Meßergebnisse können Aussagen abgeleitet werden über (a) das Vorliegen einer Einstellung gegenüber einem Objekt bzw. einer Gruppe von Personen, (b) über die Stärke bzw. Verfügbarkeit dieser Einstellung und (c) über die Bedingungen der Aktivierung der Einstellung.

Diese Verfahren eignen sich generell in erster Linie für die Erfassung jener Einstellungen die auf einem „attitude-to-nonattitude“-Kontinuum an den „attitude“-Extrempunkten zu verorten wären. Gemeint sind damit genau jene Einstellungen, die durch derart starke Objekt-Bewertungs-Assoziationen charakterisiert sind, daß bereits die bloße Wahrnehmung des Einstellungsobjekts die Einstellung aktiviert, welche wiederum - unter entsprechenden Bedingungen - den Orientierungsrahmen für „spontane“, unreflektierte (verbale oder nonverbale) Verhaltensweisen gegenüber dem Objekt bzw. der Person abgibt (Dovidio/Fazio 1992: 215f; Fazio et al. 1995: 1025f). Diese Typen von Einstellungen bzw. Vorurteilen, die möglicherweise nicht einmal den Personen selbst bewußt sind, können mit anderen Messverfahren nicht angemessen identifiziert werden.

Allerdings sind diese Verfahren auch mit einigen unübersehbaren Problemen verbunden. Dazu zählt, erstens, daß sie bislang ausschließlich auf relativ klar bestimmbare Einstellungsobjekte (v.a. Schwarze oder Weiße) und Bewertungsdimensionen (in der Regel: positiv oder negativ bewertete Attribute) angewendet wurden. Inwieweit sie allerdings auch zur Analyse komplexerer Urteile oder zur inhaltlichen Präzisierung bestimmter Vorurteile geeignet sind, ist noch unklar. Zweitens ergeben sich einige Interpretationsschwierigkeiten, wenn nach den möglichen Gründen relativ langer Reaktionszeiten gefragt wird. Deuten sie darauf hin, daß die befragte Person über keine Einstellung gegenüber dem jeweiligen Objekt verfügt (nonattitude) und erst die anzugebende Reaktion reflektiert? Oder zeigen sie an, daß die befragten

[14] Fazio und seine Mitarbeiter (1990) haben in einer Reihe von Untersuchungen mit „socially sensitive“ und „nonsensitive issues“ festgestellt, daß die unter gewöhnlichen Bedingungen („self-reports“) abgegeben Statements zu unproblematischen Themen mit den entsprechenden Ergebnissen der Reaktionszeitenmessung weitgehend übereinstimmten. Sehr starke Unterschiede ergaben sich jedoch für sensible Themen (z.B. Pornographie, Abtreibung, Sex); hier änderten sich die angegebenen Bewertungen unter den unterschiedlichen Erhebungsbedingungen teilweise drastisch (vgl. Dovidio/Fazio 1992: 220f).

[15] Ein weiteres Argument für die Validität des Meßinstruments ergibt sich aus Ergebnissen verschiedener Experimente, in denen der Zusammenhang zwischen kürzeren Reaktionszeiten und der Stärke der Einstellung kontrolliert wurde. Das Ergebnis: „Regardless of whether strong associations were identified by measurement or created experimentally, such attitude objects yielded more facilitation on evaluatively congruent trials than did objects involving relatively weaker associations“ (Dovidio/Fazio 1992: 219; vgl. Fazio et al. 1986: 219).

Personen - eventuell trotz verfügbarer, automatisch aktivierter Einstellung - von einem spontanen Entscheidungsmodus zu einem eher reflektierten Modus übergegangen sind? Solche Fragen lassen sich allein auf der Grundlage der Meßergebnisse nicht eindeutig klären. Die damit verbundenen Probleme stellen sich jedoch noch verstärkt bei reinen Reaktionszeitenmessungen (ohne Priming), da in diesen Fällen mit einer vergleichsweise größeren Anfälligkeit des Antwortverhaltens für SD-Effekte und andere situationale Einflüße - bis hin zu inneren Konflikten bei widersprechenden Einstellungen (vgl. Bassili 1995) - zu rechnen ist. Je größer aber die Zahl potentiell relevanter situativer Einflüsse wird, desto komplizierter wird auch die Identifikation der jeweiligen Bestimmungsgründe verzögerter Reaktionen.

In unmittelbarem Zusammenhang damit stellt sich, drittens, die Frage, welche Aussagekraft den mit dem Meßinstrument erfaßten Einstellungen bzw. Vorurteile zuzumessen ist, wenn zum Beispiel Prognosen über verbale oder nonverbale Verhaltensweisen unter „gewöhnlichen“ Bedingungen abgeleitet werden sollen. Wie Dovidio und Fazio (1992: 226) selbst betonen, sollten die mit dem Priming-Verfahren kombinierten Reaktionszeitenmessungen generell gute Vorhersagen für eine spezielle Klasse von Verhaltensweisen erlauben: „the behavior of interest is spontaneous in nature and is not constrained by social norms“. Diese Varianten des Verhaltens sind zwar sicherlich nicht bedeutungslos, aber sie umfassen allenfalls einen kleinen Teilbereich der sozialen Phänomene, die im Mittelpunkt des Interesses sozialwissenschaftlicher Analyse und Erklärung stehen.

Trotz dieser Einwände gegen die dargestellten Verfahren sprechen einige der bereits angeführten Argumente dafür, diese Meßmethoden in Untersuchungen über Stereotype, Vorurteile oder ethnozentristische Bewertungen stärker zu berücksichtigen[16]. Wie zahlreiche Untersuchungen in unterschiedlichen Forschungsgebieten gezeigt haben, könnten sie vor allem eingesetzt werden, um Ausmaß und Richtung von SD-Effekten zu bestimmen und könnten somit auch bei der Überprüfung der dadurch zu erwartenden Antwortverzerrungen bei „traditionellen“ Erhebungsverfahren nützlich sein. Außerdem ermöglichen sie Rückschlüsse über die Verfügbarkeit und Intensität von Vorurteilen und ähnlichen Bewertungstendenzen und erschließen so interessante Untersuchungsgegenstände (z.B.: Wie lassen sich interpersonale oder gruppenspezifische Unterschiede in der Verfügbarkeit und Intensität von Vorurteilen erklären?).

- (b) Verhaltensindikatoren

Die überwiegende Zahl der empirischen Untersuchungen über Stereotype, Vorurteile und diesen verwandte Konstrukte stützt sich nach wie vor auf Messverfahren, mit denen auf mehr oder weniger offensichtliche Weise Meinungen über Charakteristika kategorial bestimmter Personengruppen und/oder deren Bewertungen bei den Befragten direkt erhoben werden. Auch die meisten methodischen Vorschläge zur Modifikation dieser Verfahren unterscheiden sich darin nicht grundlegend. Es wird zwar versucht, durch unauffällig gestaltete Meßinstrumente und -verfahren bessere, möglichst zuverlässige und „unverzerrte“

[16] Reaktionszeitenmessungen haben sich auch in „normalen“ Erhebungssituationen bewährt. Für Telefonsurveys eignet sich vor allem das von Bassili und anderen entwickelte Verfahren des „computer-assisted telephone interviewing“ (CATI), bei dem die Reaktionszeiten der Befragten mit Hilfe eines „voice key“ und einer Computeruhr relativ genau und zuverlässig erfaßt werden können (vgl. Bassili 1995). In face-to-face-Interviews kann auf das Verfahren des „computer-assisted personal interviewing“ (CAPI) und ähnliche Erhebungsmethoden zurückgegriffen werden (vgl. Dovidio/Fazio 1992: 227f und 231f).

Indikatoren der latenten Konstrukte zu gewinnen; doch sie beruhen letztlich ebenso wie die „traditionellen“ Erhebungsverfahren auf einer nur im Ausmaß der Direktheit und Offensichtlichkeit divergierenden Abfrage der interessierenden Informationen. Damit zwangsläufig verbunden ist eine einseitige Fixierung auf eher kognitive Aspekte, insofern - je nach Verfahren - lediglich generelle Meinungen und Einstellungen zu bestimmten Personengruppen, allgemeine kognitive Repräsentationen, retrospektiv berichtetes Verhalten oder Absichtserklärungen zu künftigen Verhaltensweisen oder Ähnliches erfaßt werden. Unmittelbar verhaltensbezogene Aspekte von Vorurteilen werden im Vergleich dazu sehr selten systematisch untersucht - ein Befund, der noch zusätzliches Gewicht erhält, wenn die oft nur beschränkt generalisierbaren Studien im Rahmen reiner Laborexperimente ausgeklammert werden.

Die relativ kleine Anzahl der Studien, in denen versucht wurde, Einstellungsindikatoren auf der Grundlage von direkten und strukturierten Verhaltensbeobachtungen unter kontrollierten „natürlichen“ Bedingungen zu ermitteln[17], umfaßt ein relativ heterogenes Spektrum. Ein vor allem in den USA lange Zeit beliebtes Beobachtungsfeld waren die sogenannten „helping behavior studies“, in denen untersucht wurde, ob und inwieweit die beobachteten Personen (Schwarze und Weiße) willens waren, einem Angehörigen der ethnisch definierten „ingroup“ bzw. „outgroup“ in einer (simulierten) Notsituation zu helfen. Diese Studien wurden mit sehr unterschiedlichen Spezifikationen („wrong number“ und „lost letter technique“, „lady in distress paradigm“ usw.) und in verschiedenen Situationen (u.a. „face-to-face“- und „remote“-Bedingungen) durchgeführt. Ihre Resultate sind, wie Crosby, Bromley und Saxe (1980: 549) feststellten „more generalizable than is usually the case with social psychological studies“. Es zeigte sich, daß sowohl Schwarze als auch Weiße - unter gleichen Bedingungen - eher Angehörigen der eigenen „Rasse“ zu helfen bereit waren als Angehörigen der „Fremdgruppe“; diese Unterschiede waren aber nicht immer zu beobachten, sondern traten vor allem dann auf, wenn kein direkter Kontakt zur (vermeintlich) hilfsbedürftigen Person bestand und somit auch keine möglicherweise unangenehmen Nebenfolgen des Verhaltens zu befürchten waren.

Einige Studien untersuchten die Bereitschaft von Weißen, sich mit Schwarzen photographieren zu lassen oder eine Petition zugunsten von Schwarzen zu unterzeichnen. Andere analysierten das Verhalten von Weißen gegenüber Schwarzen in Restaurants, im Supermarkt, bei der Vergabe von Wohnungen und Arbeitsplätzen oder auch nonverbale Verhaltensweisen gegenüber Angehörigen anderer ethnischer Gruppen. Die Liste ließe sich noch um einige weitere Untersuchungen verlängern (vgl. Crosby/Bromley/Saxe 1980; Benninghaus 1976: 70-131; Schäfer/Six 1978: 85-89; Brown 1995: 211-214). Auf weitere Beispiele und Details kann an dieser Stelle verzichtet werden. Festzustellen ist indes, daß die meisten dieser Studien bereits vor 1980 entstanden sind. Aus der Zeit danach liegen fast nur noch Ergebnisse über kontrollierte Verhaltensbeobachtungen im Rahmen sozialpsychologischer Laborexperimente vor; Verhaltensbeobachtungen unter „normalen“ Bedingungen beschränken sich entweder auf sehr spezielle Verhaltensformen (v.a. Wahlverhalten), auf besondere Subpopulationen (z.B. als fremdenfeindlich bekannte Jugendgruppen) oder beruhen auf unsystematischen „Alltagsbeobachtungen“.

[17] Zur Unterscheidung von strukturierten und unstrukturierten, direkten und indirekten Beobachtungsverfahren vgl. Schnell/Hill/Esser (1995: 356-358)

Dieser Mangel an kontrollierten, systematischen und generalisierbaren Verhaltensbeobachtungen ist im Hinblick auf die Messung von Vorurteilen vor allem deshalb zu monieren, weil damit häufig sinnvolle zusätzliche Indikatoren für die Zuverlässigkeit und Gültigkeit der Erhebungsinstrumente fehlen. Verhaltensbeobachtungen können zwar an sich nie valide Indikatoren für Vorurteile und ähnliche Konstrukte sein, weil Handlungen stets erst aus einer Kombination unterschiedlicher Determinanten entstehen, unter denen Einstellungen bzw. Vorurteile allenfalls einen Faktor darstellen. Doch sie können im Sinne des Konzepts multipler Indikatoren wichtige Hinweise auf eventuelle systematische Meßfehler geben. Diese Feststellung läßt sich weitgehend losgelöst von der Frage treffen, welche spezifischen Zusammenhänge zwischen Einstellungen (Vorurteilen) und Verhalten (Diskriminierung) unter unterschiedlichen Bedingungen zu erwarten sind. Sie setzt lediglich voraus, daß irgendein Zusammenhang etwa derart angenommen werden kann, daß positive Einstellungen gegenüber einer Person tendenziell eher mit positiven als mit negativen Verhaltensweisen einher gehen. Solange plausible Gründe für diese Annahme bestehen, sollten die verwendeten Verfahren zur Erhebung von Vorurteilen auch durch solche Verhaltensindikatoren ergänzt und kontrolliert werden.

Zusammenfassung und Schluß

Die wichtigsten Ergebnisse dieser Übersicht über Möglichkeiten und Probleme der Konzeptualisierung, Operationalisierung und Messung von Stereotypen und Vorurteilen lassen sich analog zur Darstellung am sinnvollsten in zwei Gruppen unterteilen: zum einen die Befunde in Bezug auf die zwar eher mühseligen, aber gleichwohl für die empirische Forschung unumgängliche Spezifikation der Konzepte und deren Dimensionen; zum anderen die Resultate im Hinblick auf die Operationalisierung und Messung der latenten Konstrukte. Die Ergebnisse der Konzeptspezifikation können relativ kurz abgehandelt werden; die methodischen Aspekte bedürfen einer ausführlicheren Schlußbetrachtung.

Im Hinblick auf die begriffliche Präzisierung der Stereotypen- und Vorurteilskonzepte kann vor allem in der neueren sozialpsychologischen Literatur eine deutliche Tendenz zu abstrakteren und weniger normativ konnotierten Konzeptspezifikationen festgestellt werden. Diese Tendenz ist in erster Linie die Konsequenz der methodologischen Kritik an den häufig favorisierten Begriffsbestimmungen unter Bezug auf Merkmale wie „inkorrekte Generalisierungen“, „Produkte fehlerhafter Denkprozesse“ oder „Abweichungen von allgemein akzeptierten Wertvorstellungen“, die schon aufgrund der damit verbundenen begrifflichen Unklarheiten und/oder Werturteile unzweckmäßig sind. Weitaus sinnvoller und theoretisch fruchtbarer dürfte es sein, Stereotypen allgemein als *Meinungen* bzw. Wahrscheinlichkeitsurteile über die Charakteristika und Attribute einer kategorial bestimmten Personengruppe zu definieren und Vorurteile als positive oder negative *Bewertungen* einer solchen Personengruppe. Entscheidend ist in beiden Fällen, daß damit Generalisierungen gemeint sind, also Meinungen bzw. Bewertungen, die sich auf Personen aufgrund ihrer Zugehörigkeit oder Zuschreibung zu einer bestimmten sozialen Kategorie oder Gruppe beziehen; die Unterscheidung von Stereotypen und Vorurteilen soll lediglich ermöglichen, relativ wertneutrale „typisierte“ Meinungen von explizit bewertenden und eventuell auch affektiven Reaktionen gegenüber einer Personengruppe begrifflich unterscheiden zu können. Welche Beziehungen zwischen diesen Meinungen und

Bewertungen bestehen, ist dann ebenso eine letztlich theoretisch und empirisch zu klärende Frage wie etwa auch die Frage nach den Ursachen oder Konsequenzen von Stereotypisierungen und Vorurteilen.

Die Vorschläge zur Übersetzung der begrifflich abgegrenzten Konstrukte in empirisch beobachtbare Sachverhalte sind, wie die Übersicht gezeigt hat, zahlreich und unterscheiden sich teilweise nur in Details. Eine erneute Auflistung und Einschätzung der einzelnen Verfahren erübrigt sich. Lohnender ist indes ein kurzer Rückblick auf die Befunde zur Reaktivität der üblicherweise verwendeten Verfahren zur Messung von Stereotypen und Vorurteilen. Diese Befunde sich vor allem in zweierlei Hinsicht interessant: erstens im Hinblick auf die Präzisierung der zumeist nur vermuteten „response errors“ und zweitens in Bezug auf die in diesem Zusammenhang aufgezeigten Möglichkeiten der Entwicklung von weniger „störanfälligen“ Messverfahren.

Die Erkenntnis, daß das Antwortverhalten von Personen in direkten Befragungen - zumal in „sensitiven“ Bereichen - durch „response errors“ und insbesondere durch „social desirability effects“ beeinflußt und verzerrt werden kann, wird kaum überraschen. Angesichts der Ergebnisse der Studien im Rahmen des Bogus-Pipeline-Paradigmas oder der Priming-Verfahren gewinnt sie jedoch stark an Gewicht. Die unter Bogus-Pipeline- oder Priming-Bedingungen ermittelten Angaben, die als weitgehend zuverlässige und valide Indikatoren der „wahren“ Werte angesehen werden können, unterschieden sich in einem so großen Ausmaß von den mit „traditionellen“, direkten Meßmethoden erfaßten Werten, daß die Validität dieser Verfahren grundsätzlich in Frage zu stellen ist. Dafür sprechen nicht zuletzt auch die bislang vorliegenden Ergebnisse der vergleichenden Untersuchungen zur prädikativen Validität der Meßinstrumente, die gezeigt haben, daß die aus „traditionellen“ Meßmethoden ableitbaren Vorhersagen für tatsächliche Verhaltensweisen sich weitaus weniger bewähren als die aus indirekten Messungen zu gewinnenden Prognosen. Allerdings scheinen diese Unterschiede in der prädikativen Validität in jenen Situationen am deutlichsten zu sein, in denen das beobachtete Verhalten weitgehend unabhängig von bestimmten Konsequenzerwartungen spontan und unreflektiert abläuft.

Grundsätzlich legen die angeführten Befunde die Schlußfolgerung nahe, daß die „traditionellen“ direkten Meßmethoden aufgrund ihrer Anfälligkeit für sozial erwünschte Antworten insbesondere dann als unzulänglich einzuschätzen sind, wenn im Hinblick auf eine spezifische Fragestellung eine möglichst vollständige Eliminierung der dadurch zu erwartenden Verzerrungen unumgänglich ist. Dies wiederum dürfte in erster Linie dann der Fall sein, wenn automatisch aktivierbare und spontan verfügbare - also von überlegten und kalkulierten Prozessen noch „unberührte“ - Meinungen oder Bewertungen im Mittelpunkt des Interesses stehen. Richtet sich das Interesse hingegen auf „normale“ Situationen in „natürlichen“ Handlungsumgebungen, in denen sich das „äußere“ und „innere“ Tun der Akteure kalkulierend und reflektiert an der Logik der Situation orientiert, ist die Schlußfolgerung zu relativieren. Sofern die in solchen Situationen getroffenen Entscheidungen etwa über die „Angemessenheit“ einer spezifischen Einstellung und/oder overtes Verhalten gegenüber anderen Personen (Einstellungsobjekten) auch auf situational relevante normative Erwartungen Bezug nehmen, kann beispielsweise über die prädikative Qualität direkter und indirekter Messverfahren a priori kein eindeutiges Urteil gegeben werden. Es könnte ja durchaus der Fall sein, daß Personen, die in direkten Befragungen zur Abgabe sozial erwünschter Antworten tendieren, auch in „gewöhnlichen“ Entscheidungssituationen dazu neigen, normative Vorgaben relativ stark zu

gewichten, so daß die direkte Meßmethode eventuell eine höhere prädikative Validität aufweist als die indirekten Erhebungsverfahren.

Diese Einschätzungen sind angesichts des derzeitigen Forschungsstands zwangsläufig tentativ. Das liegt nicht zuletzt daran, daß die verfügbaren, zumeist unter Experimentalbedingungen durchgeführten Validierungsstudien nur selten mit indirekten Validierungen unter „natürlichen“ Bedingungen verbunden werden. Außerdem zeigt sich ein unübersehbarer Mangel in Bezug auf eine allgemeine theoretisch orientierte Erklärung der dargestellten Befunde; in der Regel wird lediglich darauf hingewiesen, *daß* Antwortverzerrungen durch „social-desirability-effects“ zu erwarten sind, um dann zur Entwicklung von Techniken zur Eliminierung dieser Effekte überzugehen. Doch die eigentlich naheliegende Frage, *warum* und *unter welchen speziellen Bedingungen* sie zu erwarten sind, wird entweder vollständig ausgeklammert oder lediglich am Rande behandelt. Die Möglichkeiten einer Erklärung der ermittelten Ergebnisse im Rahmen einer allgemeinen Theorie des Befragtenverhaltens werden offensichtlich nicht gesehen; eine umfassende Studie zur Validierung der diversen Messverfahren im Bereich der Stereotypen- und Vorurteilsforschung auf der Grundlage einer solchen Theorie - und zwar unter „natürlichen“ Umständen - fehlt gänzlich.

In eher pragmatischer Hinsicht ergibt sich aus den vorliegenden Befunden über Möglichkeiten und Probleme der Messung von Stereotypen und Vorurteilen die Erkenntnis, daß die in der empirischen Forschung üblicherweise eingesetzten direkten Erhebungsmethoden bereits dadurch erheblich verbessert werden könnten, indem sie durch Reaktionszeitenmessungen ergänzt werden. Wie Bassili, Fazio und andere Autoren gezeigt haben, können mit solchen Messungen auf unauffällige Weise wichtige Einsichten in Bezug auf die Stärke, Verfügbarkeit und Abrufbarkeit von Einstellungen gewonnen werden. Insbesondere für die Identifikation von stark verankerten, leicht zugänglichen und unter Umständen automatisch aktivierbaren Vorurteilen gegenüber „Fremdgruppen“ sind solche Reaktionszeitenmessungen von großem Interesse und haben zudem die angenehme Eigenschaft, daß sie relativ problemlos in Telephonsurveys oder in Computergestützten face-to-face-Interviews eingesetzt werden können. Mit Hilfe derart ergänzter direkter Messverfahren sollte es zum Beispiel in Untersuchungen über fremdenfeindliche Tendenzen möglich sein, Personen mit starken bzw. schwachen Vorurteilen gegenüber „Ausländern“ zu identifizieren, um diese Teilpopulationen anschließend eingehender analysieren können, ohne (ausschließlich) auf „known groups“ zurückzugreifen.

Literaturverzeichnis

Allport, Gordon W., 1954: The Nature of Prejudice. Boston: Beacon Press.

Ashmore, Richard D. und Frances K. Del Boca, 1981: Conceptual Approaches to Stereotypes and Stereotyping. S.1-35 in: *David L. Hamilton* (Hg.): Cognitive Processes in Stereotyping and Intergroup Behavior. Hillsdale, NJ: Lawrence Erlbaum.

Bassili, John N., 1995: Response Latency and the Accessibility of Voting Intentions: What Contributes to Accessibility and How It Affects Vote Choice, Personality and Social Psychology Bulletin 21: 686-695.

Benninghaus, Hans, 1976: Ergebnisse und Perspektiven der Einstellungs-Verhaltens-Forschung (Kölner Beiträge zur Sozialforschung und angewandten Soziologie; 20). Meisenheim am Glan: Hain.

Blank, Thomas und Martina Wasmer, 1996: Gastarbeiter oder Ausländer? Ergebnisse des Splits mit den reformulierten Gastarbeiterfragen im ALLBUS 1994, ZUMA-Nachrichten 20: 45-69.

Brewer, Marilynn B., 1994: The Social Psychology of Prejudice: Getting it All Together. S.315-329 in: *Mark P. Zanna und James M. Olson* (Hg.): The Psychology of Prejudice (Ontario Symposium on Personality and Social Psychology; 7). Hillsdale, NJ: Erlbaum.

Brigham, John C., 1971: Ethnic Stereotypes, Psychological Bulletin 76: 15-38.

Brown, Rupert, 1995: Prejudice. Its Social Psychology. Oxford: Blackwell.

Crosby, Faye, Stephanie Bromley und Leonard Saxe, 1980: Recent Unobtrusive Studies of Black and White Discrimination and Prejudice: A Literature Review, Psychological Bulletin 87: 546-563.

Devine, Patricia G., 1989: Stereotype and Prejudice: Their Automatic and Controlled Components, Journal of Personality and Social Psychology 56: 5-18.

Devine, Patricia G. und Andrew J. Elliot, 1995: Are Racial Stereotypes Really Fading? The Princeton Trilogy Revisited, Personality and Social Psychology Bulletin 21: 1139-1150.

Diekmann, Andreas, 1995: Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt.

Dovidio, John F., Nancy Evans und Richard B. Tyler, 1986: Racial Stereotypes: The Contents of Their Cognitive Representations, Journal of Experimental Social Psychology 22: 22-37.

Dovidio, John F. und Russell H. Fazio, 1992: New Technologies for the Direct and Indirect Assessment of Attitudes. S.204-237 in: *Judith M. Tanur* (Hg.): Questions About Questions: Inquiries into the Cognitive Bases of Surveys. New York: Russell Sage Foundation.

Dovidio, John F. und Samuel L. Gaertner, 1986: Prejudice, Discrimination, and Racism: Historical Trends and Contemporary Approaches. S.1-34 in: *diess.:* Prejudice, Discrimination, and Racism. San Diego et al.: Academic Press.

Dovidio, John F. und Samuel L. Gaertner, 1991: Changes in the Expression and Assessment of Racial Prejudice. S.119-148 in: *Harry J. Knopke, Robert J. Norrell und Ronald W. Rogers* (Hg.): Opening Doors: An Appraisal of Race Relations in Contemporary America. Tuscaloosa, AL/London: University of Alabama Press.

Duckitt, John, 1992: The Social Psychology of Prejudice. New York/Westport/London: Praeger.

Eagly, Alice und Shelly Chaiken, 1993: The Psychology of Attitudes. Fort Worth, TX: Harcourt Brace Jovanovich.

Ehrlich, Howard J., 1979: Vorurteil. Eine sozialpsychologische Bestandsaufnahme der Lehrmeinungen amerikanischer Vorurteilsforschung. München/Basel: Reinhardt.

Fazio, Russell H., 1990: Multiple Processes by which Attitudes Guide Behavior: The MODE Model as an Integrative Framework. *Mark P. Zanna* (Hg.): Advances in Experimental Social Psychology. Vol. 23. San Diego et al.: Academic Press.

Fazio, Russell H., Joni R. Jackson, Bridget C. Dunton und Carol J. Williams, 1995: Variability in Automatic Activation as an Unobstrusive Measure of Racial Attitudes: A Bona Fide Pipeline?, Journal of Personality and Social Psychology 69: 1013-1027.

Fiske, Susan T. und Shelley E. Taylor, 1991: Social Cognition. 2. Aufl. New York: McGraw-Hill.

Gaertner, Samuel L. und John F. Dovidio, 1986: The Aversive Form of Racism. S.61-89 in: *John F. Dovidio und Samuel L. Gaertner* (Hg.): Prejudice, Discrimination, and Racism. San Diego et al.: Academic Press.

Gaertner, Samuel L. und John P. McLaughlin, 1983: Changing Not Fading: Racial Stereotypes Revealed by a Non-Reactive, Reaction Time Measure, Social Psychological Quarterly 46: 23-30.

Gardner, R. C., 1994: Stereotypes as Consensual Beliefs. S.1-31 in: *Mark P. Zanna und James M. Olson* (Hg.): The Psychology of Prejudice (The Ontario Symposium on Personality and Social Psychology; 7). Hillsdale, NJ: Lawrence Erlbaum.

Gardner, R. C., R. N. Lalonde, A. M. Nero und M. Y. Young, 1988: Ethnic Stereotypes: Implications of Measurement Strategy, Social Cognition 6: 40-60.

Gilbert, G. M., 1951: Stereotype Persistence and Change Among College Students, The Journal of Abnormal and Social Psychology 46: 245-254.

Hamilton, David L. und J. W. Sherman, 1994: Stereotypes. S.1-68 in: *Robert S. Wyer und Thomas K. Srull* (Hg.): Handbook of Social Cognition. Vol. 2: Applications, 2. Aufl. Hillsdale/NJ: Erlbaum.

Hamilton, David L. und Tina K. Trolier, 1986: Stereotypes and Stereotyping: An Overview of the Cognitive Approach. S.127-163 in: *John F. Dovidio und Samuel L. Gaertner* (Hg.): Prejudice, Discrimination, and Racism. San Diego et al.: Academic Press.

Harding, John, Harold Proshansky, Bernard Kutner und Isidor Chein, 1969: Prejudice and Ethnic Relations. S.1021-1061 in: *Gardner Lindzey und Elliot Aronson* (Hg.): The Handbook of Social Psychology. Vol. 5, 2. Aufl. Reading, Mass.: Addison-Wesley.

Heckmann, Friedrich, 1992: Ethnische Minderheiten, Volk und Nation. Soziologie inter-ethnischer Beziehungen. Stuttgart: Enke.

Hilton, James L. und William von Hippel, 1996: Stereotypes, Annual Review of Psychology 47: 237-271.

Himmelfarb, Samuel, 1993: The Measurement of Attitudes. S.23-87 in: *Alice H. Eagly und Shelly Chaiken*The Psychology of Attitudes. Fort Worth, TX: Harcourt Brace Jovanovich.

Hofstätter, Peter R., 1986: Gruppendynamik. Kritik der Massenpsychologie. 3., rev. Aufl. Reinbek: Rowohlt.

Karlins, Marvin, Thomas L. Coffman und Gary Walters, 1969: On the Fading of Social Stereotypes: Studies in Three Generations of College Students, Journal of Personality and Social Psychology 13: 1-16.

Katz, Daniel und Kenneth W. Braly, 1933: Racial Stereotypes in One Hundred College Students, Journal of Abnormal and Social Psychology 28: 280-290.

Leyens, Jacques-Philippe, Vincent Yzerbyt und Georges Schadron, 1994: Stereotypes and Social Cognition. London et al.: Sage.

Linville, Patricia W., Peter Salovey und Gregory W. Fischer, 1986: Stereotyping and Perceived Distributions of Social Characteristics: An Application to Ingroup-Outgroup-Perception. S.165-208 in: *John F. Dovidio und Samuel L. Gaertner* (Hg.): Prejudice, Discrimination, and Racism. San Diego: Academic Press.

Lippmann, Walter, 1922: Public Opinion. New York: Harcourt Brace.

Manstead, Antony S. R., 1996: Attitude Theory and Research. S.47-52 in: *Antony S. R. Manstead und Miles Hewstone* (Hg.): The Blackwell Encyclopedia of Social Psychology. Oxford: Blackwell.

McCauley, Clark und Christopher L. Stitt, 1978: An Individual and Quantitative Measure of Stereotypes, Journal of Personality and Social Psychology 36: 929-940.

McCauley, Clark, Christopher L. Stitt und Mary Segal, 1980: Stereotyping: From Prejudice to Prediction, Psychological Bulletin 87: 195-208.

McConahay, John B., 1986: Modern Racism, Ambivalence, and the Modern Racism Scale. S.91-125 in: *John F. Dovidio und Samuel L. Gaertner* (Hg.): Prejudice, Discrimination, and Racism. San Diego et al.: Academic Press.

Messick, David M. und Diane M. Mackie, 1989: Intergroup Relations, Annual Review of Psychology 40: 45-81.

Mummendey, Hans D. und Heinz-Gerd Bolten, 1985: Zur Überprüfung des Bogus-Pipeline-Paradigmas: Verhaltens-Bericht und Verhaltens-Bewertung in vier Bereichen sozialen Verhaltens, Zeitschrift für Sozialpsychologie 16: 139-147.

Mummendey, Hans D., Heinz-Gerd Bolten und Margret Isermann-Gerke, 1982: Experimentelle Überprüfung des Bogus-Pipeline-Paradigmas: Einstellungen gegenüber Türken, Deutschen und Holländern, Zeitschrift für Sozialpsychologie 13: 300-311.

Olson, James M. und Mark P. Zanna, 1993: Attitudes and Attitude Change, Annual Review of Psychology 44: 117-154.

Perdue, Charles W., John F. Dovidio, Michael B. Gurtman und Richard B. Tyler, 1990: Us and Them: Social Categorization and the Process of Intergroup Bias, Journal of Personality and Social Psychology 59: 475-486.

Petermann, Franz, 1988: Erfassung von Einstellungen und Vorurteilen. S.127-151 in: *Bernd Schäfer und Franz Petermann* (Hg.): Vorurteile und Einstellungen. Sozialpsychologische Beiträge zum Problem sozialer Orientierung (Festschrift für Reinhold Bergler). Köln: Deutscher Instituts-Verlag.

Pettigrew, Thomas F. und Roel W. Meertens, 1995: Subtle and Blatant Prejudice in Western Europe, European Journal of Social Psychology 25: 57-75.

Rehm, Jürgen, 1986: Theoretische und methodologische Probleme bei der Erforschung von Vorurteilen: Vorurteil und Realität - Ist das traditionelle Forschungsprogramm der Vorurteilsforschung gescheitert?, Zeitschrift für Sozialpsychologie 17: 18-30.

Schäfer, Bernd, 1988: Entwicklungslinien der Stereotypen- und Vorurteilsforschung. S.11-65 in: *Bernd Schäfer und Franz Petermann* (Hg.): Vorurteile und Einstellungen. Sozialpsychologische Beiträge zum Problem sozialer Orientierung (Festschrift für Reinhold Bergler). Köln: Deutscher Instituts-Verlag.

Schäfer, Bernd und Bernd Six, 1978: Sozialpsychologie des Vorurteils. Stuttgart et al.: Kohlhammer.

Schnell, Rainer, Paul B. Hill und Elke Esser, 1995: Methoden der empirischen Sozialforschung. 5., völlig überarb. u. erw. Aufl. München: Oldenbourg.

Schwarz, Norbert, 1985: Theorien konzeptgesteuerter Informationsverarbeitung. S.269-291 in: *Dieter Frey und Martin Irle* (Hg.): Theorien der Sozialpsychologie. Bd. 3: Motivations- und Informationsverarbeitungstheorien. Bern/Göttingen/Toronto/Seattle: Huber.

Sigall, Harold und Richard Page, 1971: Current Stereotypes: A Little Fading, a Little Faking, Journal of Personality and Social Psychology 18: 247-255.

Simpson, George E. und J. M. Yinger, 1985: Racial and Cultural Minorities: An Analysis of Prejudice and Discrimination. 5. Aufl. New York/London: Plenum Press.

Sniderman, Paul M., Thomas Piazza, Philip E. Tetlock und Ann Kendrick, 1991: The New Racism, American Journal of Political Science 35: 423-447.

Sniderman, Paul M. und Philip E. Tetlock, 1986: Reflections on American Racism, Journal of Social Issues 42: 129-150.

Snyder, Mark und Peter Miene, 1994: On the Functions of Stereotypes and Prejudice. S.33-54 in: *Mark P. Zanna und James M. Olson* (Hg.): The Psychology of Prejudice (The Ontario Symposium on Personality and Social Psychology; 7). Hillsdale, NJ: Lawrence Erlbaum.

Stahlberg, Dagmar und Dieter Frey, 1996: Einstellungen: Struktur, Messung und Funktion. S.219-252 in: *Wolfgang Stroebe, Miles Hewstone und Geoffrey M. Stephenson* (Hg.): Sozialpsychologie. Eine Einführung, 3., erw. u. überarb. Aufl. Berlin et al.: Springer.

Stangor, Charles G., 1995: Stereotyping. S.628-633 in: *Antony S. R. Manstead und Miles Hewstone* (Hg.): The Blackwell Encyclopedia of Social Psychology. Oxford: Blackwell.

Stangor, Charles G. und James E. Lange, 1993: Cognitive Representations of Social Groups: Advances in Conceptualizing Stereotypes and Stereotyping. S.357-416 in: *Mark P. Zanna* (Hg.): Advances in Experimental Social Psychology. Vol. 26. San Diego, CA: Academic Press.

Stroebe, Wolfgang, 1985: Stereotyp, Vorurteil und Diskriminierung. Tübingen: Psychologisches Institut der Universität Tübingen.

Stroebe, Wolfgang und Chester A. Insko, 1989: Stereotype, Prejudice, and Discrimination: Changing Conceptions in Theory and Research. S.3-34 in: *Daniel Bar-Tal, Carl F. Graumann, Arie W. Kruglanski und Wolfgang Stroebe* (Hg.): Stereotyping and Prejudice. Changing Conceptions (Springer Series in Social Psychology). New York et al.: Springer-Verlag.

Tajfel, Henri, 1969: Cognitive Aspects of Prejudice, Journal of Social Issues 25: 79-97.

Vanman, Eric J. und Norman Miller, 1993: Applications of Emotion Theory and Research to Stereotyping and Intergroup Relations. S.213-238 in: *Diane M. Mackie und David L. Hamilton* (Hg.): Affect, Cognition, and Stereotyping. Interactive Processes in Group Perception. San Diego, CA: Academic

Press.

Wagner, Ulrich und Andreas Zick, 1995: The Relation of Formal Education to Ethnic Prejudice: Its Reliability, Validity and Explanation, European Journal of Social Psychology 25: 41-56.

Woodmansee, John J. und Stuart W. Cook, 1967: Dimensions of Verbal Racial Attitudes: Their Identification and Measurement, Journal of Personality and Social Psychology 7: 240-250.