### Skript zur Vorlesung

# Methoden der empirischen Wirtschafts- und Sozialforschung

Prof. Dr. Peter Kappelhoff

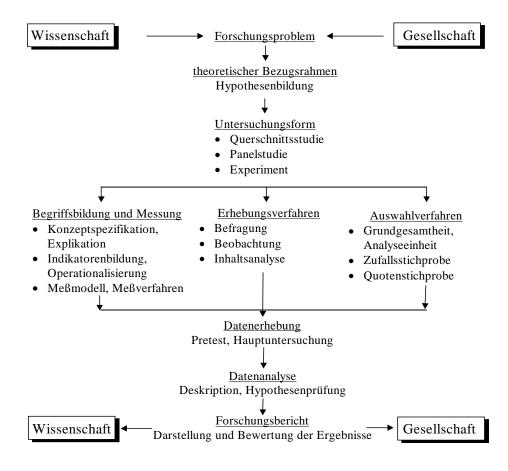
### 1. Der Forschungsprozeß

Allgemeines Ziel der Wissenschaft im Sinne der Aufklärung über soziale Zusammenhänge ist

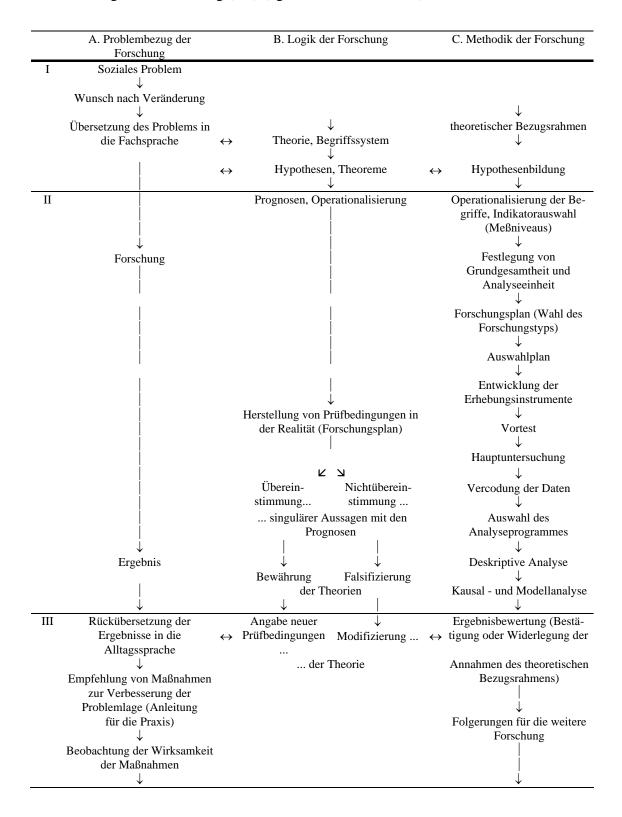
- 1. in theoretischer Hinsicht: die Konstruktion und Überprüfung von allgemeinen Theorien
- 2. in praktischer Hinsicht: ein Beitrag zur Lösung sozialer Probleme durch Anwendung empirisch bewährter Theorien.

Ziel der empirischen Wirtschafts- und Sozialforschung ist die empirische Überprüfung von Theorien im Rahmen eines erfahrungswissenschaftlichen Erkenntnisprogrammes. Empirische Sozialforschung ist daher immer theoriegeleitet - die methodische Vorgehensweise ist vom Forschungsablauf her der inhaltlich forschungsleitenden Fragestellung nachgeordnet. In diesem Rahmen leistet die empirische Sozialforschung einen unverzichtbaren Beitrag zum Wissenschaftsfortschritt, indem sie dazu zwingt, Theorien, deren Aussagen sich empirisch nicht bewähren, zu modifizieren und weiterzuentwickeln, oder sogar ganz aufzugeben (vgl. Kapitel 2).

Trotz der nahezu grenzenlosen Vielfalt der Formen empirischer Untersuchungen (vgl. Kapitel 3) können einige allgemeine Entscheidungsprobleme herausgestellt werden, die sich aus dem folgenden schematischen Ablaufdiagramm des Forschungsprozesses ergeben:



Spezifiziert auf den Problembezug, die Logik (vgl. Kapitel 2.) und die Methodik (vgl. Kapitel 3. bis 8.) der Forschung, läßt sich der Ablauf des Forschungsprozesses wie folgt in der Abfolge von Entdeckungszusammenhang (I), Begründungszusammenhang (II) und Verwendungszusammenhang (III.) (vgl. von Alemann 1977) darstellen:



### Anhang zu Kap. 1:

## <u>Einige exemplarische empirische Untersuchungen</u> (die Daten sind für Sekundäranalysen verfügbar):

#### ALLBUS-Studien (werden seit 1980 alle 2 Jahre durchgeführt):

Die <u>ALL</u>gemeine <u>B</u>evölkerungs<u>U</u>mfrage der <u>S</u>ozialwissenschaften (ALLBUS) ist eine Studie, die seit 1980 alle zwei Jahre durchgeführt wird. Bei jeder dieser Umfragen wird ein repräsentativer Querschnitt der wahlberechtigten Bevölkerung in der Bundesrepublik Deutschland und Westberlin befragt. Mit der Basisumfrage 1991 wurde der ALLBUS auch auf die neuen Bundesländer ausgeweitet. Die Stichprobe (Zufallsauswahl) umfaßt jeweils etwa 3.000 Personen. Im Rahmen des International Social Survey Program (ISSP) geht darüber hinaus seit 1985 im Rahmen einer Zusammenarbeit sozialwissenschaftlicher Forschungsinstitute aus der Bundesrepublik Deutschland, Großbritanien, den USA, Italien, Australien, Österreich, den Niederlanden und anderen Ländern jährlich eine in allen diesen Ländern vergleichbare schriftliche Erhebung ins Feld. Es ist das erklärte Ziel des ALLBUS- und des ISSP-Programms, Fragen in kleineren oder größeren Abständen zu wiederholen. Mit den so erhobenen Daten werden die Voraussetzungen geschaffen, um sowohl systematisch Gesellschaften zu vergleichen, als auch in einzelnen Ländern und ländervergleichend sozialen Wandel abzubilden (Trendstudien). Mit jeder Erhebung nehmen die Informationen, die als Zeitreihe verfügbar sind, zu.

Die im ALLBUS und ISSP erhobenen Fragen decken ein weites Feld von Problembereichen der Sozialwissenschaften ab, für die eine kontinuierliche Erhebung wünschenswert ist und für die Daten nicht aus anderen Erhebungsprogrammen, beispielsweise der amtlichen Statistik, zur Verfügung stehen. Neben einer ausführlichen, bei jeder Erhebung wiederholten Standarddemographie und Einzelfragen zu vielfältigen Einstellungen und Verhaltensweisen, wie z.B. zum Postmaterialismus, zur Ausländerfeindlichkeit, zur sozialen Sicherung, zur Familie usw., hat jede Erhebung ein inhaltliches Schwerpunktthema. Bisherige Schwerpunkte von ALLBUS und ISSP waren:

1980: ALLBUS: Politik; Freundschaftsbeziehungen 1982: ALLBUS: Religion und Weltanschauung

1984: ALLBUS: Soziale Ungleichheit und Wohlfahrtsstaat 1985: ISSP: Einstellungen zu Staat und Regierung

1986: ALLBUS: Bildung und Kulturfertigkeiten

ISSP: Soziale Netzwerke und Unterstützungsbeziehungen

1987: ISSP: Soziale Ungleichheit

1988: ALLBUS: Einstellungen zum politischen System und politische Partizipation

ISSP: Familie und sich ändernde Geschlechtsrollen

1989: ISSP: Arbeitsorientierungen

1990: ALLBUS: Politik, Freundschaftsbeziehungen II

ISSP: Einstellungen zu Staat und Regierung II 1991: ALLBUS: (Basisumfrage): Probleme der Vereinigung

ISSP: Religion

1992 ALLBUS: Religion und Weltanschauung II

ISSP: Soziale Ungleichheit II

1993 ISSP: Umwelt

1994 ALLBUS: Soziale Ungleichheit und Wohlfahrtsstaat II ISSP: Familie und sich ändernde Geschlechtsrollen II

→ Müller u.a. (Hrsg.): Blickpunkt Gesellschaft

→ Mohler/Bandilla (Hrsg.): Blickpunkt Gesellschaft 2

→ Braun/Mohler (Hrsg.): Blickpunkt Gesellschaft 3

### Sozioökonomisches Panel [SOEP](seit 1984 jährlich):

In dieser größten nichtkommerziellen Paneluntersuchung in der BRD werden jährlich alle über 16jährigen Haushaltsangehörigen von ca. 6000 Haushalten (mehr als 12000 Personen) befragt. Seit 1990 werden auch die neuen Bundesländer durch das SOEP erfaßt. Die Themenbereiche des SOEP betreffen die Veränderungen der Haushaltszusammensetzung, die Erwerbsbeteiligung und die berufliche Mobilität, Einkommensverläufe sowie die Wohnsituation und die regionale Mobilität. Darüberhinaus werden neben objektiven Lebensbedingungen auch subjektive Einstellungen, wie z.B. Werthaltungen und Zufriedenheiten, erhoben. Jede Welle enthält zusätzlich einen variablen Erhebungsteil, so z.B.:

- 1. Welle (1984): Erwerbsbiographie seit dem 15. Lebensjahr
- 2. Welle (1985): Ehe- und Familienbiographie
- 8. Welle (1991): Familie und soziale Unterstützung (wird in der 13. Welle [1996] repliziert)
- → Hanefeld: Das sozioökonomische Panel

#### Schleswig-Holstein-Survey 1989:

Die Erhebung bezieht sich auf eine repräsentative, disproportional geschichtete Zufallsstichprobe von 1000 Frauen und 200 Männern (als Kontrollgruppe). Die Studie untersucht die soziale Situation von Frauen im Spannungsfeld von Familie und Beruf. Sie enthält Angaben zur Aufgabenverteilung im Haushalt, zum Machtungleichgewicht in Ehe und Partnerschaft, zur Zusammensetzung der persönlichen Netzwerke und eine Vielzahl von Einstellungsskalen wie z.B. zur Geschlechtsrollen-, Berufsrollen- und Karriereorientierung. Vervollständigt wird die Studie durch eine Erhebung der Bildungs- und Erwerbsbiographie sowie der Familienbiographie.

→ Kappelhoff/Schulz: Determinaten der Berufsverläufe von Frauen

### Konsumverhalten und postmaterielle Werthaltungen (Konsumsucht-Studie):

Die Studie ist Teil eines Forschungsprojekts, das von G. Scherhorn am Institut für Haushaltsund Konsumökonomik der Universität Stuttgart-Hohenheim durchgeführt wurde. Die Daten wurden als Teil einer Mehrthemenbefragung des EMNID-Instituts im Oktober 1991 erhoben. Befragt wurde eine repräsentative Zufallsstichprobe, die die alten (N = 1001) und die neuen (N = 526) Bundesländer umfaßt. Neben den üblichen soziodemographischen Variablen enthält die Studie Skalen zur Kaufsucht, zur Positionalität, zur materiellen Gütergebundenheit, zum Selbstwertgefühl und zur Selbstakzeptanz, zur Naturverträglichkeit, zur Sozialverträglichkeit, zu Lebenszielen, usw.. Von besonderem Interesse ist der Versuch, unterschiedliche Kausalitätsorientierungen (autonom, kontrollorientiert, impersonal) zu erheben, um so die Bedeutung psychischer Dispositionen für die Entstehung von Kaufsucht untersuchen zu können.

Die Daten der ALLBUS-Studien, des Schleswig-Holstein-Survey und der Konsumsucht-Studie sind (wie viele andere Studien auch; vgl. den <u>Datenbestandskatalog des ZA</u>) über das Zentralarchiv für empirische Sozialforschung in Köln (ZA) für sekundäranalytische Zwecke erhältlich (das SOEP unterliegt strengeren Datenschutzbestimmungen).

### 2. Wissenschaftstheorie

Nach Kant bemüht sich die Philosophie um Antworten auf die folgenden drei <u>Grundfragen</u> der menschlichen Existenz:

Was kann ich wissen?

Was soll ich tun?

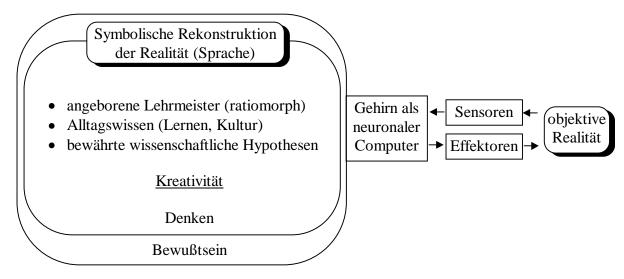
Was darf ich hoffen?

Die erste dieser drei Grundfragen wird im Rahmen der Erkenntnisphilosophie und Wissenschaftstheorie (englisch: philosophy of science) behandelt. Jede mögliche Antwort kann nur vor dem Hintergrund ihrer jeweiligen philosophischen Leitideen verstanden werden. Diese Leitideen stellen grundlegende philosophische Orientierungen dar, über die zwar, auch vor dem Hintergrund des gegenwärtig akzeptierten erfahrungswissenschaftlichen Wissens (vgl. das Buch von Albert 1987 über die "Kritik der reinen Erkenntnislehre"), kritisch diskutiert werden kann, die sich aber einer Begründung im Sinne erfahrungswissenschaftlicher Überprüfung letztlich entziehen. Im folgenden stellen wir eine Antwort dar, nämlich die des von Karl Popper entwickelten Kritischen Rationalismus (KR).

### 2.1 Kritischer Rationalismus

#### 2.1.1 Erkenntnis als Re-Konstruktion von Realität

Die Grenzen unserer Erkenntnisfähigkeit sind zunächst einmal die Grenzen unseres Erkenntnisapparates. Unumstritten ist, daß Erkenntnis rekonstruktiven Charakter hat, also kein "naturgetreues" Abbild der Welt darstellt. Unser Erkenntnisapparat kann als selbstreferentiell geschlossenes System verstanden werden (das Gehirn als neuronaler Computer mit einem eigenen neuronalen Code und einer eigenen (sprachlichen) Symbolik, das über Sensoren und Effektoren an die Außenwelt ("Realität") gekoppelt ist):



Der menschliche Weltbildapparat enthält genetisch fundierte Erkenntnisprogramme (sog. angeborene Lehrmeister mit ratiomorphem Charakter), die als arterhaltende Anpassung an Umweltbedingungen (Mesokosmos) zu verstehen sind (Leben als erkenntnisgewinnender Prozeß). Dies ist die Grundthese der von Konrad Lorenz entwickelten sog. evolutionären Erkenntnistheorie, die in dieser Form heutzutage allseits akzeptiert wird.

Ein darüber hinausgehender entscheidender Anteil unseres Wissens ist gelernt und wird als kulturelles Wissen tradiert (Alltagswissen: kultureller Überbau des Weltbildapparates). In diesem Sinne ist der These von Berger und Luckmann (1980) zuzustimmen, daß unsere Wirklichkeit gesellschaftlich konstruiert ist. Dieses Alltagswissen wird zunächst unreflektiert und unkritisch übernommen. Es stellt eine notwendige Voraussetzung für die Orientierung des Menschen in der Welt dar. Der Mensch ist im Alltagshandeln grundsätzlich immer auf der Suche nach Regelmäßigkeiten, die ihm eine Orientierung in einer ansonsten als chaotisch empfundenen Welt ermöglichen. Wie die Experimente mit nichtkontingenten Belohnungen (vgl. Watzlawick 1989) zeigen, konstruiert der Mensch sogar dort eine Ordnung, wo "in Wirklichkeit" keine vorhanden ist. Ein weiteres instruktives Beispiel in diesem Zusammenhang ist die Konstruktion von Sternbildern, die den Völkern unterschiedlichsten Kulturstufen zur "Ordnung des Himmels" dienten. Weltbilder sind also notwendige Orientierungshilfen und haben als solche einen inhärent ideologischen Charakter. Mit Topitsch (1988) können wir sie als plurifunktionale Führungssysteme verstehen, die gleichzeitig zur Informationsvermittlung (das Wahre), zur Verhaltenssteuerung (das Gute) und zur Steuerung der emotionalen Reaktionen (das Schöne) dienen. Weltbilder konstituieren und legitimieren also die gesellschaftliche Ordnung. Daraus erwächst die immer gegenwärtige Gefahr der Schließung von Weltbildern und ihrer Immunisierung gegen kritische Einwände (sofern diese überhaupt möglich sind und gesellschaftlich toleriert werden). Ein extremes Beispiel für solche Immunisierungsstrategien ist die Untersuchung einer den Weltuntergang erwartenden Sekte durch Festinger, Riecken und Schachter (1956). Die Studie zeigt, daß die Sekte an ihrem Weltbild festhielt, obwohl die Prophezeiung des Sektenführers nicht eintraf. Die Reaktion auf die nicht eingetroffene Prophezeiung war sogar eine noch rigorosere Schließung des Weltbildes.

Um solche, Weltbildern inhärenten, Immunisierungsstrategien zu unterlaufen, ist eine kritische Einstellung erforderlich. Historisch gesehen bildet sich eine solche kritische Einstellung im Zuge der Differenzierung der oben angesprochenen plurifunktionalen Führungssysteme heraus, konkret durch die gesellschaftliche Ausdifferenzierung eines (relativ!) selbständigen Wissenschaftssystems, verbunden mit der grundsätzlichen Unterscheidung zwischen Sein und Sollen, also zwischen Tatsachenaussagen und Werturteilen. Alleinige Instanz zur Überprüfung von Seinsaussagen ist nach diesem neuen Verständnis die Konfrontation mit der Realität. Wissenschaftsfortschritt ist danach möglich durch das Ausscheiden empirisch nicht bewährter Theorien und das ständige Erfinden von neuen Hypothesen (kühnen Vermutungen) mit einem größeren Informationsgehalt. Die

kritische Einstellung ist grundlegend für den kritischen Rationalismus (KR). Mit der Metapher vom "Mythos des Rahmens" (vgl. auch 2.2.1) verdeutlicht Popper (1974) seine Position, daß das wesentliche Element einer wissenschaftlichen Einstellung in der Kritik vorgegebener Weltbilder liegt, also in dem Vermögen der menschlichen Vernunft, den Rahmen eines Weltbildes als Mythos zu durchschauen und zu durchbrechen.

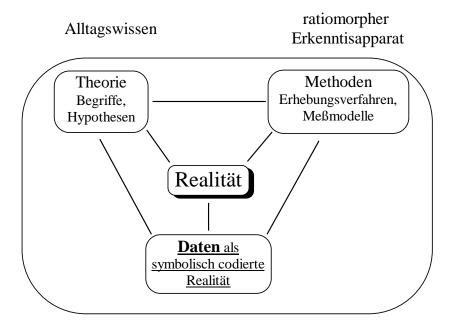
#### **Hypothetischer Realismus:**

Der kritische Rationalismus vertritt die Position eines hypothetischen Realismus. Danach gibt es eine <u>strukturierte</u> Realität, die unabhängig vom menschlichen Bewußtsein existiert. Diese strukturierte Realität ist zumindest <u>teilweise erkennbar</u>. Allerdings haben alle Aussagen über die Realität nur hypothetischen Charakter.

#### Wahrheit:

Aufbauend auf dem hypothetischen Realismus vertritt der KR die sog. Korrespondenztheorie der Wahrheit. Danach ist eine Aussage genau dann wahr, wenn sie mit der Realität übereinstimmt. (Andere, vom KR abgelehnte, Wahrheitsdefinitionen machen Wahrheit an dem gesellschaftlichen Konsens über eine Aussage (Konsenstheorie der Wahrheit) fest. Dies geschieht z.B. in der von Habermas sprachphilosophisch gewendeten kritischen Theorie.) Die Diskussion des KR wird zeigen, daß die Wahrheit einer Aussage nie endgültig bewiesen werden kann. Es kann also keinen sicheren Grund geben, aus dem heraus die Wahrheit einer Aussage abgeleitet werden kann. In dieser Position drückt sich das letztliche Scheitern eines Begründungsdenkens aus, das die philosophische Tradition seit Aristoteles beherrscht hat. Eine lesenswerte philosophische Reflexion dieses Tatbestandes stellt das Buch von Kolakowski (1977) über "Die Suche nach der verlorenen Gewißheit" dar. Auch, wenn es also kein absolutes Wahrheitskriterium gibt, kann der Wahrheitsbegriff dennoch als regulative Idee im Sinne der Annäherung an die Wahrheit sinnvoll gebraucht werden.

Als Konsequenz dieser grundsätzlichen Überlegungen folgt, daß auch in der empirischen Sozialforschung "Daten" immer nur als selektive Rekonstruktion einer überkomplexen Realität vor dem Hintergrund unseres ratiomorphen Erkenntnisapparates, unseres Alltagswissens und der zugrundeliegenden wissenschaftlichen Theorien verstanden werden können. Insbesondere ist der konstitutive Beitrag der verwendeten Methoden (z.B von Erhebungsverfahren und Meßmodellen) hervorzuheben. Auch wissenschaftliche Erfahrung ist also in einem allgemeinen und in einem spezifischen Sinne immer theoriegeleitet, schematisch:



### 2.1.2 Erklärung sozialen Handelns in den Verhaltenswissenschaften

Im Gegensatz zum Methodendualismus vertreten wir den <u>Methodenmonismus</u>, der eine methodologische Sonderstellung der Verhaltenswissenschaften (hier: Wirtschafts- und Sozialwissenschaften) ablehnt. Wir akzeptieren daher nicht das methodologische Schisma zwischen Geistes- und Naturwissenschaften, nach dem die kausale Erklärung von Tatbeständen Aufgabe der Naturwissenschaften ist, während es in den Geisteswissenschaften nur darum gehen kann, menschliches Verhalten seinem Sinn nach zu verstehen. Stattdessen vertreten wir mit Max Weber (1921) eine Position, die das Verstehen des subjektiv gemeinten Sinns einer Handlung gerade als Voraussetzung ihrer kausalen Erklärung ansieht.

Dabei orientieren wir uns an der von Weber gegebenen Definition der Soziologie:

<u>Soziologie</u> soll heißen: Eine Wissenschaft, welche soziales Handeln <u>deutend verstehen und</u> dadurch in seinem Ablauf und seinen Wirkungen ursächlich erklären will.

<u>Handeln</u> soll dabei ein menschliches Verhalten (einerlei ob äußeres oder innerliches Tun, Unterlassen oder Dulden) heißen, wenn und insofern als der oder die Handelnden mit ihm einen <u>subjektiven Sinn</u> verbinden.

<u>Soziales</u> Handeln soll ein solches Handeln heißen, welches seinem von dem oder den Handelnden gemeinten Sinn nach auf das <u>Verhalten anderer</u> bezogen wird und daran in seinem Ablauf orientiert ist.

Die Erklärung einer Handlung beginnt also mit dem deutenden <u>Verstehen</u>. Verstehen bedeutet nach diesem Verständnis die Rekonstruktion des subjektiv gemeinten Handlungssinns (Situationsinterpretation, Motivation und Ziele des Akteurs). Weber spricht in diesem Zusammenhang von der Sinnadäquanz einer Erklärung. Auf dieser Grundlage erfolgt die kausale Er-

klärung einer Handlung als Deduktion aus einem allgemeinen Gesetz (z. B. Theorien rationalen Handelns, Verhaltenstheorien). In diesem Sinne spricht Weber von der Kausaladäquanz einer Erklärung (vgl. auch die Darstellung einer deduktiv-nomologischen Erklärung in 2.4.1.2).

Auch die <u>qualitative Sozialforschung</u> hat daher keine methodologische Sonderstellung. Qualitative Sozialforschung (d.h. eine nicht standardisierte, eher verstehend beschreibende und heuristisch hypothesensuchende Vorgehensweise) wird also nicht als Konkurrenz zur quantitativen Sozialforschung betrachtet, sondern als deren Ergänzung. Innerhalb einer <u>einheitswissenschaftlichen Methodologie</u> (vgl. oben: Methodenmonismus) erhalten die qualitativen Methoden als Formen der Datenerhebung und als Arten einer systematischen Beschreibung der sozialen Wirklichkeit ihren legitimen Stellenwert. Empirische Untersuchungen können nämlich entweder <u>explorativ</u> (der Erkundung eines speziellen Forschungsfeldes bei ungenügendem Vorwissen dienend) oder <u>hypothesentestend</u> (Ex-postfacto-Anordnungen oder experimentelle Untersuchungsformen zur Überprüfung einer kausalen Hypothese) sein (vgl. Kapitel 3).

### 2.1.3 Gescheiterte Versuche einer sicheren Begründung (klassischer Empirismus, klassischer Rationalismus)

Ausgangspunkt der abendländischen Erkenntnistheorie ist die Postion, daß Wahrheit und Gewißheit menschlicher Erkenntnis unmittelbar miteinander verbunden sind. Dies kommt deutlich in der klassischen Definition des Wissens zum Ausdruck, die von Aristoteles stammt. In ihr ist davon die Rede, daß man etwas weiß, wenn man den Grund kennt, warum es so ist, und damit die Gewißheit hat, daß es nicht anders sein kann. Daraus folgt als strategisches Ziel der Wahrheitssuche die Suche nach sicheren Gründen im Sinne eines archimedischen Punktes der Erkenntnis. Im folgenden soll gezeigt werden, daß dieses Erkenntnismodell, also die Suche nach sicheren Gründen, in eine Sackgasse führt.

Englischer Empirismus: Die Erfahrung gilt als sichere Erkenntisquelle. Unser Wissen wird induktiv aus der (von Trugbildern bereinigten (Bacon)) Erfahrung gewonnen. Der menschliche Verstand ist bei der Geburt eine tabula-rasa, also ein "weißes Papier ohne alle Schriftzeichen, frei von allen Ideen" (Locke). An diesem Zitat wird die Ablehnung eines Ideen-Modells der Erkenntnis deutlich: Es gibt im Verstand nichts, das nicht vorher in den Sinnen gewesen wäre. Die Gegenargumente liegen auf der Hand: Unsere Erfahrung kann täuschen, der ratiomorphe Weltbildapparat wirkt als Filter für Erfahrung, die Erfahrung ist grundsätzlich theoriegeleitet (siehe genauer in 2.1.4). Darüberhinaus ist eine rein induktive Begründung allgemeiner Gesetze schon aus rein logischen Gründen nicht möglich (siehe unten).

Kontinentaler Rationalismus: Ausgangspunkt bei <u>Descartes</u> ist der methodische Zweifel an allem, was man zu wissen glaubt. Es wird nur das anerkannt, was diesem methodischen

Zweifel standhält, was also <u>klar und deutlich erkannt</u> wird. Dies ist zunächst einmal das denkende Subjekt selbst. <u>Sichere Erkenntnis</u> ist also auf der Grundlage unmittelbarer Intuition und <u>rationaler Einsicht möglich</u> (wie z.B. für Descartes die Erkenntnis der Existenz Gottes). Auch hier liegen die Gegenargumente auf der Hand: Was uns klar einleuchtet, sind oft nur kulturelle Selbstverständlichkeiten oder, anders ausgedrückt, die Vorurteile unseres Weltbildes (wie z.B. der Schluß von der Idee Gottes auf die Existenz Gottes bei Descartes).

Transzendentaler Idealismus (Versuch der Versöhnung von Rationalismus und Empirismus) bei Kant: Hintergrund der Überlegungen von Kant ist der Erfolg der klassischen Mechanik (Newton), die von Kant als sicheres Wissen anerkannt wird. Grundlage seiner Erkenntnistheorie ist die folgende These: Der Verstand schöpft seine Gesetze nicht aus der Natur, sondern schreibt sie dieser vor. Die Welt als Gegenstand unserer Erfahrung wird nämlich durch die apriorischen Kategorien unseres Verstandes, wie z. B. Raum, Zeit und Kausalität, geordnet. Die Gegenstände unserer Erfahrung müssen sich also nach unseren angeborenen Anschauungsformen und Vernunftbegriffen richten. Auf diese Weise kann sicheres Wissen über die Erfahrungswelt, wie z.B. in der Newtonschen Mechanik, erlangt werden. Überspitzt formuliert könnte man sagen, daß der menschliche Verstand der Natur ihre Gesetze vorschreibt. Über die auf der Grundlage dieser Kategorien mögliche Erfahrung hinaus, also über die "Dinge an sich", können wir nichts wissen.

Auch hier ist der folgende Einwand offensichtlich: Die Passung unserer apriorischen Verstandeskategorien kann nicht weiter begründet werden (wie z.B. in der evolutionären Erkenntnistheorie als Anpassung an die Umwelt, also als <u>evolutionär a posteriori</u>; vgl. Vollmer 1975). Solche Anpassungen ermöglichen zwar das Überleben der Art, können aber durchaus fehlerbehaftet sein, wie z.B. die Überwindung der klassischen Raum-Zeit-Vorstellung in der Einsteinschen Relativitätstheorie zeigt. An diesem Beispiel zeigt sich allerdings auch die Fähigkeit des menschlichen Verstandes, die Restriktionen des eigenen Erkenntnisapparates kritisch zu reflektieren und auch dort, wo sich die eigenen Anschauungsformen als unzulänglich erweisen, neue Erkenntnisse in Form von empirisch überprüfbaren mathematischen Aussagen über die Wirklichkeit zu erlangen.

Erkenntnis ohne (Letzt-)Begründung: Die fehlgeschlagenen Versuche, zu einer sicheren Begründung von Erfahrungserkenntnis zu gelangen, lassen nur eine Schlußfolgerung zu: Die für das abendländische Denken konstitutive Kopplung von Wahrheit und Gewißheit in der Erkenntnistheorie muß aufgegeben werden. Dies ist die Position des KR, wie sie von Popper entwickelt und von Albert (1968) weitergeführt wurde. Die Suche nach einer zureichenden Begründung führt danach entweder zu einem infiniten Regreß, oder zu einem logischen Zirkel, oder man ist gezwungen, das Verfahren an irgendeinem Punkt dogmatisch abzubrechen. Jeder Versuch, an der Möglichkeit einer sicheren Begründung von Erkenntnis festzuhalten, führt also unausweichlich in das von Albert sog. Münchhausen-Trilemma. Insbesondere der dogmatische Abbruch durch Berufung auf eine sichere Erkenntnisquelle (Intuition, Erfahrung, unmittelbare Evidenz, lichtvolle Gewißheit, Schöpfungsgeschichte in

der Bibel, Arbeitswertlehre bei Marx, usw.) ist beliebt und wird gewöhnlich zusätzlich durch eine Ideologietheorie abgesichert, die erklärt, warum einige Ignoranten die (angeblich) sichere Begründung des Wissens nicht anerkennen können oder wollen (Verblendung, falsches Bewußtsein usw.).

Schlußfolgerung: Wahrheit im Sinne der Übereinstimmung einer Aussage mit der Realität ist weiterhin ein sinnvolles Konzept. Wahre Erkenntnis im Sinne einer sicheren Begründung von Erkenntnis ist aber unmöglich; es gibt kein absolut sicheres Wahrheitskriterium. Hält man in dieser Situation am Ideal der sicheren Begründung fest, bleibt nur die Alternative zwischen der "Flucht ins Engagement" (vgl. Bartley 1987) und Resignation. Entweder hält man an der Idee der sicheren Begründung fest und immunisiert sein Wissen durch dogmatischen Abbruch gegen kritische Einwände, oder man resigniert und nimmt eine relativistische Position ein, nach der jedes Wissenssystem seine eigene Wahrheit hat. Zwischen diesen beiden Ausflüchten, also zwischen Dogmatismus und Relativismus, bleibt als dritte Möglichkeit die Position des KR, der das Prinzip der sicheren Begründung zwar aufgibt, aber an einem realistisch interpretierten Verständnis von Erkentnis festhält und Wahrheit weiterhin als regulative Idee versteht. Ziel der Wissenschaft ist demnach die Suche vorläufig bewährtem Wissen und Annäherung die Wahrheit durch nach an Wissenschaftsfortschritt.

Ein weiterer Ausweg wäre ein rein <u>instrumentalistisches Verständnis</u> von Erkenntnis. Dabei wird die Annahme des hypothetischen Realismus ganz aufgegeben. Wissen bedeutet danach nicht mehr die hypothetische Erkenntnis einer unabhängig vom Menschen existierenden Wirklichkeit. Stattdessen stellt Wissen lediglich ein nützliches Instrument zur Lösung von Problemen dar. Die Frage nach der Wahrheit wissenschaftlicher Aussagen kann dann überhaupt nicht mehr gestellt werden; wissenschaftliche Aussagen werden allein im Hinblick auf ihre <u>Nützlichkeit unter Anwendungsgesichtspunkten</u> beurteilt. Mit der Aufgabe des hypothetischen Realismus und einer regulativen Idee der Wahrheit im Instrumentalismus ist also eine unnötige <u>Aushöhlung des Erkenntnisbegriffs</u> verbunden. Letztlich bleibt es im Rahmen des Instrumentalismus ein Rätsel, wie die empirische Bewährung von wissenschaftlichen Aussagen zu verstehen ist, und wie die Möglichkeit einer praktischen Nutzung wissenschaftlicher Erkenntnisse überhaupt möglich ist - es sei denn, man hält doch implizit an der realistischen Überzeugung einer grundsätzlichen Erkennbarkeit der Welt fest (vgl. auch die Kritik des radikalen Konstruktivismus im Anhang von Kapitel 2).

Induktionsproblem: Im Zusammenhang mit der Position des Empirismus ist auch das Induktionsproblem von Bedeutung. Dabei geht es um die Begründung von allgemeinen Gesetzen mit einem unendlichen (raum-zeitlich-unbegrenzten) Geltungsbereich durch Erfahrung, genauer durch einzelne (endlich viele) Beobachtungen. Formal ist der folgende logische Schluß gemeint:

$$A(1) \land A(2) \land .... \land A(n) \rightarrow \forall x A(x)$$

Wenn die Aussage A für endlich viele Objekte (hier 1, 2, ..., n) gilt, dann gilt die Aussage A für alle Objekte x.

Der Induktionsschluß ist also der Schluß vom Besonderen auf das Allgemeine. Wie leicht einsehbar ist, ist der Induktionsschluß <u>logisch nicht zulässig</u>. Wählen wir als Beispiel für eine allgemeine Aussage den Satz "Alle Schwäne sind weiß", so könnte z.B. eine neue Beobachtung plötzlich einen schwarzen Schwan zeigen ( $\neg A(n+1)$ ). (In der Natur gibt es tatsächlich schwarze Schwäne.) Selbst wenn auch die Beobachtung A(n+1) einen weißen Schwan zeigen würde, könnte man nie sicher sein, daß nicht weitere Beobachtungen n+2, … ein Gegenbeispiel (einen schwarzen Schwan) erbringen würden. Durch noch so viele Einzelbeobachtungen kann also die Wahrheit einer <u>allgemeinen</u> Aussage nicht bewiesen werden (vgl. Popper 1973, Kap. 1).

Asymmetrie von Verifikation und Falsifikation: Zwar kann eine allgemeine Aussage durch noch so viele Einzelbeobachtungen nicht bewiesen werden, umgekehrt kann aber ein allgemeines Gesetz durch ein einziges Gegenbeispiel widerlegt werden, d.h. eine einzige Beobachtung eines schwarzen Schwanes genügt, um die allgemeine Aussage "Alle Schwäne sind weiß" zu widerlegen.

Formal:

$$\neg A(n+m) \rightarrow \neg \forall x A(x)$$

Diese Möglichkeit der Falsifikation von Allaussagen durch ein einziges Gegenbeispiel bildet die Grundlage der <u>Falsifikationslogik</u>.

### 2.1.4 Grundzüge des Kritischen Rationalismus (KR)

Grundlagen des Kritischen Rationalismus (vgl. Popper 1934):

- **Hypothetischer Realismus** (siehe 2.1.2)
- **Konsequenter Fallibilismus:** Unser theoretisches Wissen ist grundsätzlich fehlbar, eine sichere Begründung ist nicht möglich (siehe 2.1.3). Popper: "Wir wissen nicht, sondern wir raten." Ziel der Wissenschaftsentwicklung kann es daher nur sein, unsere Theorien schrittweise zu verbessern, indem wir <u>Fehler eliminieren</u> und uns dadurch der <u>Wahrheit</u> annähern. Dies führt zur Falsifikationslogik und zum Falsifikationismus.
- Methodologischer Falsifikationismus: Theorien werden als kühne Vermutungen in einem kreativen Prozeß generiert (Entdeckungszusammenhang: Wie kommt der Forscher zu seinen Theorien?). Entscheidend für die Akzeptierung einer Theorie ist aber nicht die Art ihrer Begründung im Entdeckungszusammenhang, sondern allein ihre empirische Bewährung im Begründungszusammenhang (empirische Überprüfung von Theorien). Theorien enthalten allgemeine Sätze, die durch ein Gegenbeispiel widerlegt werden

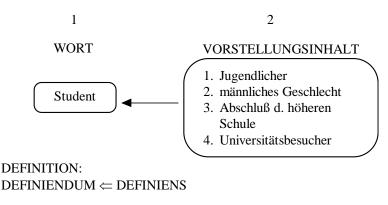
können (vgl. die in 2.1.3 abgeleitete Asymmetrie von Verifikation und Falsifikation). Der Prozeß der Elimination falscher Theorien durch Konfrontation mit der Erfahrung heißt Falsifikation. Die Falsifikation ist an <u>strenge methodische Regeln</u> gebunden; insbesondere ist die <u>Theorieabhängigkeit von Beobachtungen</u> zu berücksichtigen. Die Falsifikation gilt als Motor allen wissenschaftlichen Erkenntnisfortschritts. Popper: "Wir lernen aus unseren Fehlern".

Der Grundgedanke des methodologischen Falsifikationismus kann in der Biographie von <u>Karl Popper</u> (1902-1994) an einem Schlüsselerlebnis festgemacht werden. Einstein entwickelte 1916 die Allgemeine Relativitätstheorie (kühne Vermutung), die im Widerspruch zur klassischen Mechanik Newtons steht. Zur Überprüfung seiner Theorie schlug er vor, eine besonders riskante Vorhersage der Theorie bei der nächsten Sonnenfinsternis zu überprüfen: die Krümmung des Raumes durch Gravitationsfelder und die dadurch hervorgerufene Abweichung von Lichtstrahlen von der "geraden" Bahn. Die Messungen Eddingtons 1919 bestätigten die Einsteinsche Relativitätstheorie und falsifizierten zugleich die klassische Mechanik Newtons. An diesem Beispiel wird deutlich, wie wissenschaftlicher Erkenntnisfortschritt in einem Wechselspiel von kühnen Vermutungen und empirischer Überprüfung erreicht werden kann. Diese Einsichten hat Popper in Abgrenzung von dem logischen Positivismus des Wiener Kreises zum ersten Mal in seinem bahnbrechenden Werk "Logik der Forschung" (1934) veröffentlicht, das seine volle Wirkung aber erst nach dem zweiten Weltkrieg entfaltet hat.

### 2.1.4.1 Begriff, Aussage, Theorie

Realität kann nie unmittelbar erfaßt werden, sondern immer nur sprachlich vermittelt; wir sprechen in diesem Zusammenhang daher auch von einer symbolischen Rekonstruktion von Wirklichkeit (siehe 2.1.1). Begriffe, die in sozialwissenschaftlichen Untersuchungen verwendet werden, müssen zunächst präzise definiert werden (vgl. Prim und Tilmann 1983):

**Nominaldefinition:** Das Verfahren, mit dem Worte und Vorstellungsinhalte als "Begriffe" festgelegt werden, nennt man Definition. Hierbei wird eine Entscheidung darüber getroffen, daß ein bestimmtes sprachliches Zeichen nur noch in einer bestimmten Weise verwendet werden soll. Nominaldefinitionen sind daher nichts weiter als <u>sprachliche Konventionen</u>.

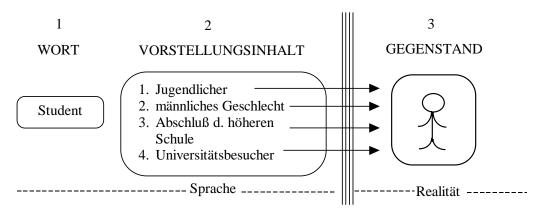


Nominaldefinitionen können nicht wahr oder falsch sein, sondern sind <u>zweckmäßige</u> <u>Setzungen</u> (Ordnungsfunktion, Kommunikationsfunktion).

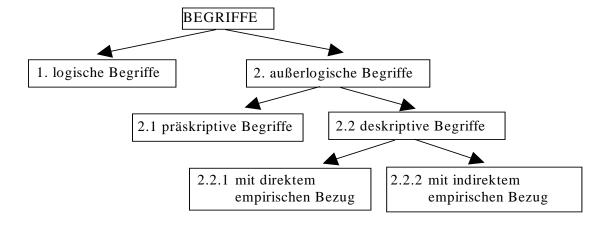
Begriffe sollten präzise definiert sein, <u>konsistent</u> verwendet werden und <u>theoretisch fruchtbar</u> sein (siehe auch 4.1).

Realdefinition: Wir lehnen die Vorstellung einer Realdefinition ab, die für sich in Anspruch nimmt, das "Wesen" einer Sache zu erfassen (vgl. auch Popper 1973, Kap. 2 und 4). Die Realdefinition setzt ein grundsätzlich nicht vorhandenes Wissen über eine tieferliegende Dimension der Wirklichkeit voraus (Ideenmodell der Erkenntnis). So sieht Hegel das "Wesen" des Staates anders als die moderne pluralistische Demokratietheorie, und über das "Wesen" des Menschen existieren die unterschiedlichsten Meinungen (homo faber, homo oeconomicus, homo ludens usw.).

Die Gleichsetzung eines qua Nominaldefinition konventionell definierten Vorstellungsgehaltes eines Begriffes mit dem Gegenstand selbst wird als <u>Reifizierung (Verdinglichung)</u> bezeichnet. Solche Reifizierungen verfehlen den konventionellen Charakter der Nominaldefinition und haben meist ideologischen Charakter, da sie ein vermeindliches "Wesen" einer Sache qua Definiton der Realität unterschieben.



Das folgende Schema unterscheidet verschiedene Begriffsarten:



Von besonderer Bedeutung ist hier der Unterschied von Begriffen mit direktem und indirektem empirischen Bezug. Letztere können sich nur indirekt, über sog. Indikatoren (direkt

beobachtbare Ersatzgrößen) auf etwas Beobachtbares beziehen. Beispiele sind Begriffe wie normative Geschlechstrollenorientierung, Kaufsucht, Intelligenz, Markentreue usw. Wie man an den Beispielen sieht, treten in den sozialwissenschaftlich relevanten Aussagen fast ausschließlich Begriffe ohne direkten empirischen Bezug auf. Das bedeutet aber, daß die empirische Überprüfung solcher Aussagen nur indirekt unter Zuhilfenahme von Indikatoren möglich ist (siehe auch Zweisprachentheorie, Basissatzproblem).

Mit Hilfe präzise definierter Begriffe können Sätze gebildet werden. Wir unterscheiden folgende Formen von Aussagen:

Einzelmerkmerkmale* Satzarten	Gültigkeitsanspruch	Realitätsbezug	Gültigkeitsbereich	Überprüfbarkeit	Informationsgehalt
Analytische Sätze	apriorisch	ohne oder mit	raum-zeitlich unbegrenzt	logisch	ohne
Präskriptive Sätze	apriorisch	ohne (selten) oder mit	raum-zeitlich begrenzt oder unbegrenzt	logisch	ohne
Deskriptive Sätze	aposteriorisch	mit	raum-zeitlich begrenzt	logisch und empirisch	mit (graduell)
Hypothetische Sätze	aposteriorisch	mit	raum-zeitlich begrenzt oder unbegrenzt	logisch und empirisch	mit (graduell)
Definitionen	apriorisch	ohne oder mit	raum-zeitlich begrenzt, oder unbegrenzt	logisch	ohne

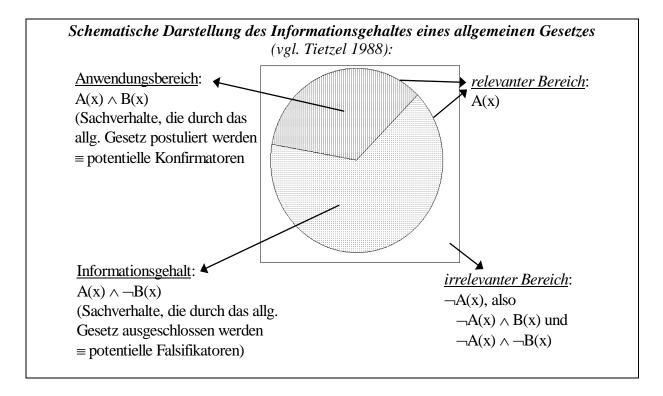
#### \* Einzelmerkmale:

1. Gültigkeitsanspruch	2. Realitätsbezug	3. Gültigkeitsbereich	4. Überprüfbarkeit	5. Informations- gehalt
a) apriorisch	a) ohne	a) raum-zeitlich begrenzt	a) logisch	a) ohne
b) aposteriorisch	b) mit - wertend - nicht wertend	b) raum-zeitlich unbegrenzt	b) logisch und empirisch	b) mit

In Hinblick auf unser Ziel, wissenschaftliche Aussagen über die Realität zu machen, sind insbesondere deskriptive und hypothetische Sätze von Interesse. Realitätsbezug können zwar auch Definitionen, analytische und präskriptive Sätze haben, nämlich immer dann, wenn sie Begriffe mit empirischem Bezug enthalten. Entscheidend ist aber der Informationsgehalt einer Aussage. Negativ formuliert kann man sagen, daß der Informationsgehalt (empirische Gehalt) eines Satzes um so höher ist, je mehr Zustände der Realität durch ihn ausgeschlossen werden (entspricht der Klasse der möglichen Falsifikatoren (siehe 2.1.4.2)). Deskriptive und hypothetische Sätze haben einen aposteriorischen Gültigkeitsanspruch und die Überprüfbarkeit ist logisch und empirisch. Der Unterschied zwischen beiden Satzarten besteht in dem Gültigkeitsbereich, also in dem Grad der Allgemeinheit. Deskriptive Sätze beziehen sich auf singuläre Tatbestände, während Hypothesen für eine größere Klasse von Tatbeständen gelten. Im Idealfall eines allgemeinen Gesetzes ist der Gültigkeitsbereich raum-zeitlich unbegrenzt. Im einfachsten Fall haben allgemeine Gesetze (G) die folgende logische Struktur:

$$\forall x \big( \mathbf{A}(x) \to B(x) \big)$$

Für alle x gilt: wenn für x die Eigenschaft A zutrifft, dann trifft auch die Eigenschaft B zu.



Der Informationsgehalt eines Gesetzes ist dabei um so größer, je allgemeiner A und je spezieller B formuliert ist. Wie die schematische Darstellung zeigt, wird nämlich durch eine Verallgemeinerung von A der relevante Bereich erweitert und durch eine Spezifikation von B der Anwendungsbereich relativ zum Informationsgehalt verringert, d.h. die Menge der potentiellen Konfirmatoren verkleinert sich zugunsten der Menge der potentiellen Falsifikatoren.

Unter einer <u>Theorie</u> versteht man im Kritischen Rationalismus eine thematisch und logisch systematisierte Menge von Wenn-Dann-Aussagen, also von allgemeinen Gesetzen, die untereinander widerspruchsfrei sind, einen möglichst hohen Informationsgehalt haben und möglichst gut empirisch bewährt sind.

### 2.1.4.2 Naiver und methodologischer Falsifikationismus

Grundlage der Falsifikationslogik ist ein <u>deduktives Argument</u>. Durch logische Deduktion werden aus einem allgemeinen Gesetz (Hypothese) und einer Randbedingung Aussagen über singuläre Tatbestände (deskriptive Sätze) abgeleitet. Diese Form der Ableitung wird als <u>deduktiv-nomologische-Erklärung</u> bezeichnet (oder auch H-O-Schema nach Hempel und Oppenheim, die das Schema ausgearbeitet haben).

**Deduktiv-nomologische Erklärung** (Erklärung als logische Deduktion aus Hypothesen und Randbedingungen):

Erklärungen folgen dem bereits bekannten Prinzip der Deduktion, nach dem das zu Erklärende, das Explanandum, aus dem Erklärenden, dem Explanans, logisch abgeleitet wird. Das

Explanandum ist in Form eines singulären deskriptiven Satzes oder mehrerer solcher Sätze repräsentiert. Das Explanans enthält 2 Aussagearten:

- a) Mindestens eine bewährte nomologische Hypothese.
- b) Mindestens eine singuläre deskriptive Aussage, die beschreibt, ob die in der Wenn-Komponente der Hypothese geforderten Bedingungen im Falle des zu erklärenden Ereignisses tatsächlich gegeben sind: die sogenannten Anfangs- oder <u>Randbedingungen</u>.

Schematisch stellt sich die Erklärungsstruktur wie folgt dar:

1. Gesetz  $\forall x (A(x) \rightarrow B(x))$ 2. Randbedingung A(1)Aus 1. und 2. zu erklärender : Explanandum B(1)Satz (Sätze)

Beispiel:

Hypothese: Alle entwickelten Industriegesellschaften verfügen über ein Schichtungs-

system.

Randbedingung: Die Bundesrepublik Deutschland ist eine entwickelte

Industriegesellschaft.

*Explanandum:* Die Bundesrepublik Deutschland verfügt über ein Schichtungssystem.

#### Falsifikationsprinzip:

Wie in 2.1.3 ausführlich erläutert, kann ein allgemeines Gesetz <u>widerlegt (falsifiziert)</u> werden, wenn es gelingt, einen einzelnen (singulären) Tatbestand zu finden, der im Widerspruch zu der Ableitung aus dem allgemeinen Gesetz steht. Formal:

$$A(1) \land \neg B(1) \rightarrow \neg \forall x (A(x) \rightarrow B(x))$$

Ein <u>allgemeines</u> Gesetz kann also durch das Auffinden eines <u>einzigen Gegenbeispiels</u> widerlegt werden, in unserem Beispiel also durch eine Industriegesellschaft, die nicht über ein Schichtungssystem verfügt.

Entscheidende Voraussetzung für die Falsifikation eines allgemeinen Gesetzes ist es, das empirisch eindeutig über das Vorliegen eines singulären Tatbestandes (gilt B(1) oder gilt  $\neg B(1)$ ) entschieden werden kann (dabei wurde der Einfachheit halber bereits vorausgesetzt, daß die Geltung von A(1) als unproblematisch unterstellt werden kann, also die Randbedingung des Gesetzes erfüllt ist). Gerade darin aber besteht das <u>Basissatzproblem</u>: Kann über die Wahrheit oder Falschheit einer singulären Aussage (als Basis der Falsifikation) eindeutig und endgültig entschieden werden?

Hier unterscheiden wir zwischen dem <u>naiven Falsifikationismus</u> (eine nur zum Zwecke der Argumentation als Kontrast konstruierte Position, die zumindest von Popper <u>nie</u> vertreten wurde) und dem <u>methodologischen Falsifikationismus</u> (die Position Poppers und des KR). Der naive Falsifikationismus geht davon aus, daß durch sorgfältiges methodisches Vorgehen Beobachtungsfehler vermieden werden können und daher über die Gültigkeit von Basissätzen wie "Dies ist ein schwarzer Schwan" oder "Die Bundesrepublik ist eine egalitäre Gesellschaft" endgültig entschieden werden kann. Dagegen erkennt der methodologische Falsifikationismus an, daß in unseren Beobachtungen Fehler nie ganz ausgeschlossen werden können und darüber hinaus jede Messung eine Theorie voraussetzt, die sog. Meßtheorie, die selbst fehlerhaft sein kann.

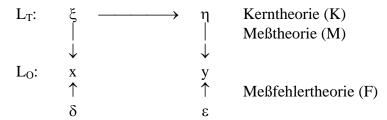
Grundlegend für das folgende Argument ist die Unterscheidung zwischen Begriffen mit direktem und solchen mit indirektem empirischen Bezug und die damit in Zusammenhang stehende Zweisprachentheorie. Nach der Zweisprachentheorie wird zwischen der theoretischen Sprache  $L_T$  (theoretische Konstrukte, d.h. Begriffe ohne direkten empirischen Bezug, die durch sog. latente Variabeln  $\xi$  und  $\eta$  dargestellt werden) und der Beobachtungssprache  $L_O$  (Indikatoren, d.h. Begriffe mit direktem empirischen Bezug, die durch die zugehörigen manifesten Variablen x und y erfaßt werden) unterschieden. Die Verbindung zwischen theoretischer Sprache und Beobachtungssprache wird durch sog. Korrespondenzregeln hergestellt (Zuordnung von Indikatoren zu theoretischen Konstrukten).

 $L_T$ : theoretische Sprache ( $\xi$ ,  $\eta$ , ...)

: (Korrespondenzregeln)

L<sub>O</sub>: Beobachtungssprache (x, y, ...)

Das Korrespondenzproblem wird im Zusammenhang mit der Frage der Operationalisierung eines theoretischen Konstrukts in Kapitel 4 wieder aufgegriffen. Hier genügt eine kurze Skizze der Lösung des Korrespondenzproblems durch den <u>kausalanalytischen Ansatz</u>, welche die für die wissenschaftstheoretische Diskussion wichtigen Elemente hervorhebt:



Auf der theoretischen Ebene wird zunächst ein Kausalgesetz ( $\xi \to \eta$ ) formuliert, das als Kerntheorie bezeichnet wird. (Je größer die soziale Distanz ( $\xi$ ) zu Ausländern, desto größer

Variablen der theoretischen Sprache (latente Variablen) werden i.d.R. durch griechische Buchstaben dargestellt ( $\xi =$  "ksi",  $\eta =$  "eta",  $\zeta =$  "zeta",  $\delta =$  "delta",  $\epsilon =$  "epsilon"), Variablen der Beobachtungssprache werden dagegen mit lateinischen Buchstaben bezeichnet: x, y.

die Ausländerfeindlichkeit ( $\eta$ ); siehe 4.3.) Ein solches Gesetz kann nicht direkt empirisch überprüft, also auch nicht direkt falsifiziert werden, da die Begriffe keinen direkten empirischen Bezug haben, also nicht direkt beobachtbar sind. Erforderlich ist die Zuordnung von direkt beobachtbaren Indikatoren, hier den Variablen x und y. ("Anzahl von Ausländern im Freundeskreis" (x) als (negativer) Indikator für soziale Distanz ( $\xi$ ) und Zustimmung zu der Aussage "Gastarbeiter sollten sich ihre Ehepartner unter den eigenen Landsleuten auswählen" (y) als (positiver) Indikator für Ausländerfeindlichkeit ( $\eta$ ); siehe 4.3.) Entscheidend ist, daß die Beziehung von  $\xi$  zu x als Kausalbeziehung konzipiert wird, d.h. x als <u>kausale Folge</u> von  $\xi$  betrachtet wird (analog ist y die kausale Folge von  $\eta$ ). (Der Grad der Zustimmung zu der Aussage "Gastarbeiter …." (y) wird kausal bestimmt durch den Grad der Ausländerfeindlichkeit ( $\eta$ ) des Befragten). Damit wird eine weitere <u>Theorie</u> formuliert, die sog. <u>Meßtheorie</u>, ohne die eine empirische Überprüfung der Kerntheorie nicht möglich wäre.

Das Schema enthält aber noch eine weitere Klasse von latenten Variablen, nämlich  $\delta$  und  $\epsilon$ , die als weitere (in diesem Zusammenhang unerwünschte) Ursachen die Messung von x und y beeinflussen (Meßfehlertheorie);  $\delta$  und  $\epsilon$  stellen also in unserem Zusammenhang Einflußfaktoren dar, die Meßfehler verursachen (x mißt nicht nur die Wirkung von  $\xi$  (wie erwünscht), sondern zusätzlich noch den (unerwünschten) Einfluß von  $\delta$ ). Im Rahmen des kausalanalytischen-Ansatzes können die Meßfehler  $\delta$  und  $\epsilon$  als Zufallsfehler oder als systematische Fehler spezifiziert werden. Insgesamt ergibt sich ein komplexes Kausalmodell, das die Kerntheorie, die Meßtheorie und die Meßfehlertheorie umfaßt. (Sind die Korrelationen zwischen den Indikatoren aus einer empirischen Untersuchung bekannt, kann ein solches Kausalmodell gelöst werden, wenn bestimmte Voraussetzungen erfüllt sind; vgl. 4.3).

Eigentlich müßte also das deduktiv-nomologische Schema neben dem allgemeinen Gesetz (Kerntheorie) auch noch die Meßtheorie und Meßfehlertheorie enthalten, denn nur so ist die Ableitung eines Basissatzes im Sinne eines deskriptiven Satzes mit direkt beobachtbaren Begriffen und damit eine empirische Überprüfung möglich. Setzt man, um das Argument nicht unnötig zu komplizieren, einmal das Vorliegen der Randbedingung A(1) als empirisch unproblematisch voraus (hier gelten im Grunde die gleichen Argumente wie für die Überprüfung des Basissatzes) so gilt:

$$(K \wedge M \wedge F) \rightarrow B(1)$$

D.h. der Basissatz B(1) kann nur bei der gleichseitigen Geltung von drei Theorien (Kerntheorie, Meßtheorie und Meßfehlertheorie) abgeleitet werden. Im Umkehrschluß (Falsifikationslogik) gilt dann:

$$\neg B(1) \rightarrow \neg (K \land M \land F)$$

Wird also der Basissatz empirisch widerlegt, so ist <u>mindestens eine</u> der drei Theorien falsch. Es kann z.B. durchaus sein, daß Kerntheorie und Meßtheorie korrekt sind, die Meß-

fehlertheorie aber falsch ist. (Es wurden z. B. Zufallsfehler angenommen, obwohl ein systematischer Meßfehler vorliegt, da die Befragten aufgrund einer Tendenz zur sozialen Erwünschtheit ihre wahre Ausländerfeindlichkeit nicht offen zugeben wollen; vgl. 6.1.) Es könnte aber auch sein, daß die Meßtheorie (allein oder in Verbindung mit der Meßfehlertheorie) falsch war (z. B. weil wir die falschen Indikatoren ausgewählt haben; zum Problem der Operationalisierung siehe Kapitel 4). In beiden Fällen würden unsere Beobachtungen ( $\neg$  B(1)) der Kerntheorie widersprechen, obwohl die Kerntheorie richtig ist (falsche Meßtheorie und/oder Meßfehlertheorie). Natürlich ist es aber auch denkbar, daß wir korrekt gemessen haben (korrekte Meßtheorie und Meßfehlertheorie) und die Kerntheorie falsch ist, wir also tatsächlich berechtigt sind, die Kerntheorie als falsifiziert anzusehen.

# (Vorläufige) Entscheidung über die Anerkennung eines Basissatzes durch ein "Gerichtsverfahren":

Grundlage der Entscheidung über die Anerkennung eines Basissatzes sind die folgenden Bedingungen:

- Es muß sich um wiederholbare, intersubjektiv überprüfbare Beobachtungen handeln.
- Die Beobachtungen dürfen <u>nicht im Widerspruch zu anderen Basissätzen</u> stehen.
- Weiter muß überprüft werden, ob die empirische Untersuchung unter Berücksichtigung des methodischen Fachwissens der wissenschaftlichen Disziplin korrekt durchgeführt wurde. Insbesondere muß überprüft werden, ob die zugrundegelegte Meßtheorie und Meßfehlertheorie (vorläufig!) akzeptiert werden kann.

Dieses <u>Prüfverfahren</u> wird nicht durch den einzelnen Wissenschaftler, sondern durch die <u>wissenschaftliche Gemeinschaft</u> in einem <u>kritischen Diskurs</u> durchgeführt (Metapher des Gerichtsverfahrens). Übersteht ein Basissatz als potentieller Falsifikator einer Theorie diese Prozedur, wird er vorläufig anerkannt und die zugrunde liegende Kerntheorie gilt als methodologisch falsifiziert.

Um noch einmal die Problematik der (im naiven Falsifikationismus angeblich sicheren) empirischen Basis zu betonen, zitieren wir Poppers' "Sumpflandmetapher" (Logik der Forschung, 1967, S. 75f.): "So ist die empirische Basis der objektiven Wissenschaft nichts «Absolutes»; die Wissenschaft baut nicht auf Felsengrund. Es ist eher ein Sumpfland, über dem sich die kühne Konstruktion ihrer Theorien erhebt; sie ist ein Pfeilerbau, dessen Pfeiler sich von oben her in den Sumpf senken - aber nicht bis auf einen natürlich «gegebenen» Grund. Denn nicht deshalb hört man auf, die Pfeiler tiefer hineinzutreiben, weil man auf eine feste Schicht gestoßen ist; wenn man hofft, daß sie das Gebäude eines Tages tragen werden, beschließt man, sich vorläufig (Unterstreichung nicht im Original) mit der Festigkeit der Pfeiler zu begnügen." Merke: Popper ist alles andere als ein Positivist, sondern ein kritischer

<u>Rationalist</u>. Der Vorwurf des Positivismus gegen Popper und den KR ist eine Unterstellung, die im sog. "Positivismusstreit" von seinen Gegnern Adorno und Habermas eingeführt wurde.

Schlußfolgerung: Über die Anerkennung eines Basissatzes kann nie auf der Grundlage unproblematischer Beobachtungen entschieden werden, sondern immer nur im Kontext eines anerkannten theoretischen Wissens und methodischer Regeln. Die isolierte Überprüfung einer einzelnen Theorie ist also nicht möglich. Damit akzeptiert der KR die schwache Form der Duhem-Quine-These, die besagt, daß eine Theorie immer nur vor dem Horizont anderer Theorien überprüft werden kann, deren Gültigkeit in den jeweiligen Überprüfungskontext als unproblematisch vorausgesetzt werden muß. Auch diese Hintergrundtheorien sind natürlich grundsätzlich fallibel und müssen im Verlauf der Wissenschaftsentwicklung schrittweise verbessert werden. Die Wissenschaftsentwicklung ist also durch das parallele Arbeiten an verschiedenen Theorien. die zu einem gegebenen Zeitpunkt jeweils Hintergrundwissen darstellen, charakterisiert. Das bedeutet, daß die Theorieentwicklung in einem bestimmten Bereich den Wissenschaftsfortschritt dadurch fördern kann, daß sich das für die Überprüfung anderer Theorien zur Verfügung stehende Hintergrundwissen erweitert. So machte sich Galilei die Entwicklungen in der optischen Theorie zunutze, indem er ein Fernrohr konstruierte, um so durch Beobachtungen die kopernikanische Theorie zu bestätigen (Entdeckung der Jupitermonde, usw.). Die Anerkennung der schwachen Form der Duhem-Quine-These durch den KR bedeutet aber keineswegs auch eine Zustimmung zu der starken Form der Duhem-Quine-These mit der pessimistischen Schlußfolgerung, daß sich Theorien grundsätzlich nicht widerlegen lassen, die Theorieabhängigkeit von Beobachtungen letztlich also zu Immunisierung von Theorien gegen Falsifikation führt. (In diesem Zusammenhang ist auch der von Lakatos geschilderte "Fall planetarischer Unart" von Interesse, ebenso wie die von Lakatos an diesem Beispiel illustrierte Unterscheidung zwischen progressiven und degenerativen Forschungsprogrammen; vgl. 2.2.2.)

Zusammenfassung: Der KR <u>kennt keine absolut sichere Erkenntnisquelle</u>. Wissenschaftliches Denken ist "systematisches Raten" mit dem Ziel einer möglichst guten <u>Annäherung an die Wahrheit</u>. Die <u>Objektivität</u> des wissenschaftlichen Prozesses ist eine <u>soziale Angelegenheit der gegenseitigen Kritik</u>. Ziel der Kritik ist die <u>Eliminierung falscher Problemlösungen (Falsifikationslogik)</u>. Je häufiger und strenger die Prüfversuche sind und je ungehinderter sich die Kritik entfalten kann, um so mehr wird der Wissenschaftsfortschritt gefördert. Die <u>wissenschaftliche Kritik</u> findet auf der Grundlage von <u>Spielregeln (methodische Regeln)</u> statt, die sich ebenfalls weiter entwickeln. Daraus folgt die Notwendigkeit der ständigen Kontrolle und Verbesserung methodischer Regeln. Genau das ist Gegenstand des Faches: Methoden der empirischen Wirtschafts- und Sozialforschung.

Auch die methodologische Falsifikation einer Theorie kann also grundsätzlich nur vorläufigen Charakter haben, da sie in Relation zu dem dabei vorausgesetzten Hintergrundwissen zu sehen ist (schwache Form der Duhem-Quine-These). In dieser Position kommt noch einmal der konsequente Fallibilismus des KR und die Ablehnung jedes

Begründungsdenkens zum Ausdruck. Der KR wendet sich daher auch nachdrücklich gegen wissenschaftstheoretische Systeme, die implizit oder explizit eine Tabula-rasa-Strategie verfolgen, die also versuchen, durch das Ausscheiden von falschen Vorstellungen die Vernunft "quasi zu reinigen", um dann auf dieser bereinigten Grundlage schrittweise zu sicheren Erkenntnissen zu kommen (wie dies in der Tradition von Descartes immer wieder versucht wurde - z.B. in der Phänomenologie Husserls). Auch die Ideologiekritik und die sich daraus entwickelnde Wissenschaftssoziologie verfolgt diese Strategie, nach der es zunächst darum geht, das hinter einer Erkenntnis stehende "Interesse" aufzudecken. So unterscheidet Habermas zwischen einem technischen, einem praktischen und einem emanzipatorischen Erkenntnisinteresse. Erst wenn sich das Letztere in einem herrschaftsfreien Diskurs artikulieren kann, ist eine wahrhaft kritische Theorie möglich. Allen diesen Positionen ist das (zumindest implizite) Festhalten an einem Begründungsdenken inhärent, daß der KR schon aus logischen Gründen (Münchhausen-Trilemma; vgl. 2.1.3) grundsätzlich ablehnt. Pointiert formuliert postuliert der KR die "Nicht-Existenz einer reinen Vernunft" (vgl. Albert 1968). Da unser Wissen grundsätzlich fallibel ist, ist eine völlige "Reinigung" der Vernunft von falschen Vorstellungen ("Ideologiekritik") und ein Neuaufbau auf sicherer Grundlage unmöglich. Wir müssen daher mit unserem unvollkommenen und teilweise falschen Wissen leben. Der einzige Ausweg ist eine kritische Haltung in Permanenz, also die schrittweise Verbesserung grundsätzlich fallibler Theorien.

Gesellschaftspolitisch ist Popper ein Verfechter der "offenen Gesellschaft" und damit ein Anhänger des liberalen Gesellschaftsmodells, da sich nur in einer freiheitlich verfaßten Gesellschaft der kritische Diskurs ungehindert entfalten kann. Für die Sozialwissenschaften vertritt Popper einen methodologischen Individualismus, der mit einem theoretischen Institutionalismus (Ablehnung des psychologischen Reduktionismus) gekoppelt ist. Seine sozialphilosophischen Überlegungen hat Popper (zuerst 1945) programmatisch in seiner Kampfschrift "Die offene Gesellschaft und ihre Feinde" dargestellt. Darin warnt Popper vor dem historischen Hintergrund der nationalsozialistischen und kommunistischen Diktaturen und dem 2. Weltkrieg vor totalitären Tendenzen von rechts (Band I: Der Zauber Platos) und links (Band II: Falsche Propheten: Hegel, Marx und die Folgen). Im Gegensatz dazu bekennt sich Popper zu der kritischen Tradition der Aufklärung, und damit zu der Forderung nach individueller Autonomie und einem kritischen Gebrauch der Vernunft, wie sie in dem Wahlspruch der Aufklärung in der Formulierung von Kant zum Ausdruck kommt: Sapere aude! (Habe den Mut, Dich Deines eigenen Verstandes zu bedienen!). Folgerichtig hat Popper sein Buch "Die offene Gesellschaft und ihre Feinde" dem "Andenken des Philosophen der Freiheit und Menschlichkeit Immanuel Kant" gewidmet.

### 2.2 Kritik des Kritischen Rationalismus

Schwachstelle in der Argumentation des KR und damit möglicher Ausgangspunkt für Kritik ist die Tatsache, daß <u>auch im KR Theorien nie endgültig falsifiziert</u> werden können. Die sich daraus entwickelnde kritische Diskussion konzentriert sich auf folgende Punkte:

- Die Anerkennung von Basissätzen durch Beschluß der wissenschaftlichen Gemeinschaft auf der Grundlage methodologischen Fachwissens enthält eine konventionalistische Komponente. Die Frage ist, wann und wie genau das Prüfverfahren als sozialer Prozeß zu einem Abschluß kommt.
- Im Gegensatz zum KR gibt Kuhn den Falsifikationsgedanken ganz auf und läßt im Rahmen der Normalwissenschaft Anomalien zu (wo kein Kläger ist, ist auch kein Richter) und Lakatos kritisiert die utopische Idee einer sofort wirkenden Rationalität und setzt stattdessen auf einen langfristigen Prozeß der Theorienkonkurrenz.
- Der <u>Theoriebegriff</u> (Kerntheorie) ist wenig elaboriert; Meßtheorie und Meßfehlertheorie, methodologische Regeln und begrifflicher Rahmen strukturieren das Umfeld der Kerntheorie, werden aber im Theoriebegriff des KR nicht explizit berücksichtigt. In der kritischen Auseinandersetzung mit dem KR wird dieser enge Theoriebegriff ersetzt durch den Paradigmabegriff bei Kuhn bzw. das Konzept wissenschaftlicher Forschungsprogramme bei Lakatos.

Die Wissenschaftstheorie des KR argumentiert normativ und weitgehend ahistorisch. Das provoziert die Frage, ob sich der tatsächliche Ablauf des Wissenschaftsprozesses vor dem Hintergrund des KR hinreichend gut verstehen läßt. Die Gegenargumente gegen den Falsifikationsgedanken bei Kuhn und Lakatos werden daher wissenschaftsgeschichtlich begründet (wissenschaftshistorische Herausforderung des KR). Beide argumentieren, daß die Wissenschaftsgeschichte ein komplexeres Bild zeigt, als es bei einem linearen Fortschritt nach der Falsifikationslogik zu erwarten wäre, ziehen daraus aber unterschiedliche Konsequenzen. Während Kuhn die Falsifikationslogik ganz aufgibt und an ihre Stelle ein weitgehend irrationales Wechselspiel zwischen Normalwissenschaft und wissenschaftlichen Revolutionen setzt, erweitert und verändert Lakatos den Falsifikationsgedanken in seinem Konzept des raffinierten Falsifikationismus.

### 2.2.1 Normalwissenschaft und wissenschaftliche Revolutionen

<u>Vorbemerkung</u>: Die folgende Darstellung stützt sich in erster Linie auf das Frühwerk von <u>T. Kuhn</u>: <u>Die Struktur wissenschaftlicher Revolutionen</u> (zuerst 1962). Die Kritik wird aus didaktischen Gründen überspitzt dargestellt, da sie die in späteren Schriften von Kuhn selbst vorgenommenen Relativierungen vernachlässigt (vgl. z.B. Hoyningen-Huene 1987).

Kuhns Untersuchung ist wissenschaftshistorisch angelegt und stellt die Frage, ob sich der "normale Wissenschaftsbetrieb", gerade in den exakten Naturwissenschaften, mit dem Falsifikationsgedanken Poppers vereinbaren läßt. Kuhn kommt zu einer ganz anderen Sicht der Theoriedynamik: Lange Phasen normaler Wissenschaft, in denen im Rahmen eines Paradigmas Rätsel gelöst werden, werden von krisenhaften Entwicklungen, sog. wissenschaftichen Revolutionen, abgelöst, in denen nicht einfach Theorien verworfen werden, sondern sich die gesamte Betrachtungweise der Realität nach Art eines Gestaltwandels verändert. Im Gegensatz zu Popper behauptet Kuhn also:

- längere Phasen relativ unkritischer <u>Normalwissenschaft</u> (Forschen im Rahmen eines Paradigmas ohne Falsifikation),
- gefolgt von relativ kurzen Phasen <u>revolutionären Wandels</u>, so daß die Wissenschaftsgeschichte insgesamt eine Abfolge inkommensurabler Paradigmen darstellt (kein Wissenschaftsfortschritt).

Grundlegend für die Kuhnsche Vorstellung von Wissenschaft ist ein <u>Paradigmabegriff</u>, der umfassender ist als der zugeordnete Theoriebegriff im KR. Ein Paradigma enthält:

- 1. eine Art der Betrachtung der Welt (Weltbild)
- 2. theoretische Aussagen
- 3. methodische Regeln
- 4. paradigmatische (beispielhafte) Anwendungen

Gerade die Punkte 1 und 4 machen den Unterschied zum traditionellen Theoriebegiff aus. So enthält z.B. das ökonomische Paradigma ein Weltbild, daß die Realität unter einem bestimmten Blickwinkel (z.B. rationales Wirtschaften mit knappen Ressourcen bei gegebenen Handlungszielen) betrachtet. Diese Art der Weltsicht kann nicht bewiesen, sondern nur durch lehrbuchartige Beispiele anschaulich gemacht werden. Auf diese Weise wird der angehende Wissenschaftler in ein wissenschaftliches Weltbild hinein sozialisiert. "Die Studierenden akzeptieren aber Theorien wegen der Autorität des Lehrers und des Lehrbuches, nicht aufgrund von Beweisen. [...] Die in den Lehrbüchern geschilderten Anwendungen stehen dort nicht als Beweis, sondern weil ihr Erlernen ein Teil des Erlernens des der derzeitigen Praxis zugrundeliegenden Paradigmas ist. Würden die Anwendungen als Beweis hingestellt, so würde das bloße Versäumnis der Lehrbücher, Alternativinterpretationen zu geben oder Probleme zu diskutieren, für welche die Wissenschaftler keine Paradigmalösungen finden konnten, von den extremen Vorurteilen ihrer Autoren zeugen. Für eine solche Anklage ist aber nicht der geringste Grund vorhanden." (S. 93f) Ein Paradigma erscheint also als ein Rahmen, in dem wissenschaftliche Probleme in spezifischer Weise definiert werden. Normalwissenschaft besteht in der Lösung solcher wohldefinierter Probleme, oder, mit den Worten von Kuhn, in der Lösung von Rätseln (puzzles). Ziel der Normalwissenschaft ist also die Integration neuer Tatsachen in ein bestehendes Weltbild.

Gelingt eine solche Integration nicht, spricht man von Anomalien (die Form der Widerständigkeit der Realität in der Kuhnschen Wissenschaftstheorie). Solche <u>Anomalien</u> werden aber zunächst als eine normale Begleiterscheinung der Forschung angesehen. Alle Theorien sehen sich zu jeder Zeit Gegenbeispielen ausgesetzt. "Die Unfähigkeit, eine Lösung zu finden, diskreditiert nur den Wissenschaftler und nicht die Theorie. Hier trifft sogar noch besser als oben das Sprichwort zu: Das ist ein schlechter Zimmermann, der seinem Werkzeug die Schuld gibt. [...] Kein bisher durch das historische Studium der wissenschaftlichen Entwicklung aufgedeckter Prozeß hat irgendeine Ähnlichkeit mit der methodologischen Schablone der Falsifikation durch unmittelbaren Vergleich mit der Natur." (S. 93/90)

Erst bei einer Häufung solcher Anomalien kommt es zu einer krisenhaften Entwicklung innerhalb des Paradigmas und die Grundlagen des wissenschaftlichen Weltbildes werden in Zweifel gezogen. Ein solcher Umsturz des Weltbildes ist meistens mit existenziellen Krisen in der Biographie von Wissenschaftlern verbunden (Einstein: "Es war, wie wenn einem der Boden unter den Füßen weggezogen worden wäre, ohne daß sich irgendwo fester Grund zeigte, auf den man hätte bauen können."). "Der Übergang von einem krisenhaften Paradigma zu einem neuen, aus dem eine neue Tradition der normalen Wissenschaft hervorgehen kann, ist weit von einem kumulativen Prozeß entfernt, wie ihn eine Artikulation oder eine Erweiterung des alten Paradigmas darstellen würde. Es ist vielmehr der Neuaufbau des Gebietes auf neuen Grundlagen, ein Neuaufbau, der einige der elementarsten theoretischen Verallgemeinerungen des Gebiets wie auch viele seiner Paradigmamethoden und -anwendungen verändert." (S. 97f) Ein Paradigmawechsel gleicht also einem Gestaltwandel. Die Verständigung über eine solche revolutionäre Trennungslinie hinweg ist daher zwangsläufig nur teilweise möglich. "Nehmen wir als Beispiel die Männer, die Kopernikus für verrückt hielten, weil er verkündete, daß die Erde sich bewege. Sie waren nicht ganz und gar im Unrecht. Was sie unter «Erde» verstanden, war u.a. feste Position. Zumindest ihre Erde konnte sich nicht bewegen. Dementsprechend bestand Kopernikus' Neuerung nicht einfach darin, die Erde in Bewegung zu setzen. Es war vielmehr eine vollkommen neue Art und Weise, die Probleme der Physik und Astronomie zu betrachten. Eine Art, die notwendigerweise die Bedeutung sowohl von «Erde» wie auch von «Bewegung» veränderte." (S. 160f)

Ein besonderer Aspekt des Kuhnschen Paradigmabegriffs ist die Verankerung eines Paradigmas in einer wissenschaftlichen Gemeinschaft. Wissenschaftliche Revolutionen sind also immer auch soziale Revolutionen. "Wie die Wahl zwischen konkurrierenden politischen Institutionen erweist sich die zwischen konkurrierenden Paradigmata als eine Wahl zwischen unvereinbaren Lebensweisen der Gemeinschaft." (S. 106)

Eine Verständigung über die Grenzen solcher wissenschaftlichen Gemeinschaften hinaus ist nicht möglich (Inkommensurabilitätsthese). "Wenn Paradigmata in eine Diskussion über die Wahl von Paradigmata eintreten - und sie müssen es ja -, dann ist ihre Rolle notwendigerweise zirkulär. Jede Gruppe verwendet ihr eigenes Paradigma zur Verteidigung eben dieses Paradigmas. Dieses im Kreis gehende Argument hat nur den Status eines Überredungsversu-

ches." (S. 106) Daher kann nach Kuhn die Frage der Paradigmawahl niemals endgültig durch Logik und Experiment allein eindeutig entschieden werden. Nur aus der Sicht des siegreichen Paradigmas kann ein Paradigmawechsel als wissenschaftlicher Fortschritt interpretiert werden. (Diese radikale These von der Irrationalität der Wissenschaftsentwicklung hat Kuhn in weiteren Schriften weitgehend abgeschwächt und konzediert, daß neue Paradigmen über eine höhere Problemlösungskapazität verfügen, allerdings nur im Sinne einer instrumentalistischen Interpretation von Wissenschaft.)

#### Zur Kritik von Kuhn:

Die Kuhnsche Wissenschaftstheorie steht und fällt mit dem Paradigmabegriff und damit mit der Behauptung, daß sich ein Wissenschaftlter stets in einem von ihm nicht hinterfragten und auch prinzipiell nicht hinterfragbarem Rahmen (Weltbild) bewegt. Popper (1974) spricht hier von dem "Mythos des Rahmens". Kuhn bringt in diesem Zusammenhang selbst das Beispiel von Experimenten zur Erkennung von Spielkarten, z.B. einer (roten) Herz-7 (♥) oder einer (schwarzen) Pik-4 (A). Was passiert, wenn Versuchspersonen nach einer sehr kurzen Betrachtungszeit Spielkarten einordnen müssen, die den Paradigmarahmen sprengen - also z.B. eine rote Pik-4 oder eine schwarze Herz-7? In Einklang mit Kuhn's Vorstellung zeigt das Experiment, daß die Befragten solche Spielkarten entsprechend des Paradigmas einordnen, also eine schwarze Herz-7 entweder als (rote) Herz-7 oder als (schwarze) Pik-7 bezeichnen, ohne den Fehler zu bemerken. Allerdings zeigt das gleiche Experiment, daß die Versuchspersonen bei längerer Betrachtungszeit durchaus in der Lage sind zu erkennen, daß es sich um außergewöhnliche (den Rahmen sprengende) Spielkarten, also z.B. um eine schwarze Herz-7, handelt. Daran, wie auch an dem Beispiel des Übergangs vom ptolemäischen zum kopernikanischen Weltbild, wird deutlich, daß der Mensch aufgrund seiner Kritikfähigkeit durchaus in der Lage ist, den Rahmen eines Weltbildes zu sprengen. Ein weiteres Beispiel in diesem Zusammenhang ist die Überwindung der dreidimensionalen Raumvorstellung durch die Einsteinsche Relativitätstheorie. Man kann also durchaus Popper Recht geben, wenn er vom Mythos des Rahmens spricht und die menschliche Fähigkeit hervorhebt, auch die Grundlagen eines Weltbildes kritisch zu reflektieren. Das schließt natürlich nicht aus, daß auch diese Kritik wieder im Rahmen eines neuen Paradigmas stattfindet, aber nun - und das ist die These des KR - eines Paradigmas oder besser gesagt, eines Theoriesystems, das realitätsadäquater ist - und dies nicht nur im Sinne einer instrumentalistischen Interpretation (wie dies vielleicht auch Kuhn einräumen würde), sondern im Sinne eines realistischen Wissenschaftsverständnisses (hypothetischer Realismus, Korrespondenztheorie der Wahrheit).

# 2.2.2 Die Methodologie wissenschaftlicher Forschungsprogramme

Die Gedanken von Lakatos - veröffentlicht in seinem zuerst 1970 erschienenen Aufsatz "Falsifikation und die Methodologie wissenschaftlicher Forschungsprogramme" - werden oft als Weiterentwicklung des KR verstanden. Dabei wird allerdings die Radikalität der Uminterpretation der Gedanken Poppers unterschätzt. Ausgangspunkt der Überlegung von Lakatos ist die Einsicht, daß Theorien nicht endgültig widerlegbar sind. Der konventionalistische Charakter der methodologischen Falsifikation im KR eröffnet einen weiten Spielraum für Strategien der Exhaurierung (vgl. das weiter unten diskutierte Beispiel "planetarischer Unart"). Lakatos kritisiert in diesem Zusammenhang die utopische Idee einer sofort wirkenden Rationalität im KR. Wie Kuhn akzeptiert er auf der Grundlage wissenschaftshistorischer Untersuchungen die Tatsache, daß viele Theorien trotz bekannter Anomalien nicht verworfen werden. Dies führt Lakatos (1974) zu der folgenden grundlegenden Einsicht: "Es gibt keine Falsifikation vor dem Auftauchen einer besseren Theorie." (S. 117) Aus der direkten Konfrontation zwischen Theorie und empirischen Tatbeständen wird ein Dreikampf zwischen zwei konkurrierenden Theorien T und T' und der empirischen Evidenz. Dabei ist die Theorie T' der Theorie T überlegen (T wird durch T' raffiniert falsifiziert), falls folgende Bedingungen erfüllt sind (S. 114):

- 1. T' erklärt auch die bewährten Tatbestände, die von T erklärt werden.
- 2. T' hat einen Gehaltsüberschuß gegenüber T, d.h. mit T' können <u>neue</u> Tatbestände erklärt werden.
- 3. Der Gehaltsüberschuß von T' ist zumindest teilweise empirisch bewährt.

Lakatos hält also grundsätzlich an der Idee der Falsifikation fest, allerdings nur in einem langfristigen Prozeß der Theorienkonkurrenz. Dabei steht mehr die Leistungsfähigkeit alternativer Theorien im Vordergrund (progressive Problemverschiebung), als die direkte Konfrontation mit der Realität. In diesem Zusammenhang ist der von Lakatos konstruierte imaginäre Fall "planetarischer Unart" als Beispiel relevant. Ein Physiker berechnet auf der Grundlage der Newtonschen Mechanik N die Bahn eines eben entdeckten Planeten p. Die Experimentalastronomen stellen aber fest, daß der Planet von der berechneten Bahn abweicht. "Glaubt unser Newtonianer, daß die Abweichung von Newtons Theorie verboten war und daß ihr Beweis die Theorie N widerlegt? - Keineswegs. Er nimmt an, daß es einen bisher unbekannten Planeten p' gibt, der die Bahn von p stört. Er berechnet Masse, Bahn etc. dieses hypothetischen Planeten und ersucht dann einen Experimentalastronomen, seine Hypothese zu überprüfen. Aber der Planet p' ist so klein, daß selbst das größte vorhandene Teleskop ihn nicht beobachten kann: Der Experimentalastronom beantragt einen Forschungszuschuß, um ein noch größeres Teleskop zu bauen. In drei Jahren ist das neue Instrument fertig. Wird der unbekannte Planet p' entdeckt, so feiert man diese Tatsache als einen neuen Sieg der

Newtonschen Wissenschaft - aber man findet ihn nicht. Gibt unser Wissenschaftler Newtons Theorie und seine Idee des störenden Planeten auf? - Nicht im mindesten! Er mutmaßt nun, daß der gesuchte Planet durch eine kosmische Staubwolke vor unseren Augen verborgen wird. Er berechnet Ort und Eigenschaften dieser Wolke und beantragt Forschungsstipendium, um einen Satelliten zur Überprüfung seiner Berechnung abzusenden. Vermögen die Instrumente des Satelliten (darunter völlig neue, die auf wenig geprüften Theorien beruhen) die Existenz der vermuteten Wolke zu registrieren, dann erblickt man in diesem Ergebnis einen glänzenden Sieg der Newtonschen Wissenschaft. Aber die Wolke wird nicht gefunden. Gibt unser Wissenschaftler Newtons Theorie, seine Idee des störenden Planeten und die Idee der Wolke, die ihn verbirgt, auf? - Nein! Er schlägt vor, daß es im betreffenden Gebiet des Universums ein magnetisches Feld gibt, daß die Instrumenten des Satelliten gestört hat. (Einschub: Meßtheorie, Meßfehlertheorie!) Ein neuer Satellit wird ausgesandt. Wird das magnetische Feld gefunden, so feiern Newtons Anhänger einen sensationellen Sieg. - Aber das Resultat ist negativ. Gilt dies als eine Widerlegung der Newtonschen Wissenschaft? - Nein. Man schlägt entweder eine neue, noch spitzfindigere Hilfshypothese vor oder .... die ganze Geschichte wird in den staubigen Wänden der wissenschaftlichen Analen begraben, vergessen und nie mehr erwähnt. Zumindest bis ein neues Forschungsprogramm das Programm Newtons überholt, das fähig ist, dieses bisher unzugängliche Phänomen zu erklären. In diesem Fall wird das Phänomen wieder ausgegraben und als ein "entscheidendes Experiment" inthronisiert." (S. 99)

An diesem Zitat wird zweierlei deutlich: Erstens der Vorgang einer degenerativen Problemverschiebung, bei der immer neue Hilfshypothesen generiert werden, ohne daß das postulierte Phänomen tatsächlich gefunden wird (die Hilfshypothesen haben einen reinen Ad-hoc-Charakter). Zweitens taucht in dem Zitat der Begriff des Forschungsprogramms (FP) auf, der bei Lakatos den Theoriebegriff weitgehend ersetzt. Abfolgen von Theorien T, T', T", ... können als Bestandteile eines wissenschaftlichen FP angesehen werden, wenn folgende Bedingungen erfüllt sind: Es gibt zentrale metaphysische Prinzipien und grundlegende theoretische Annahmen, den sog. harten Kern des FP, der in der Theoriereihe beibehalten wird und zusätzliche Hilfstheorien, wie z.B. Instrumententheorien und Annahmen über mögliche Störfaktoren (Meßfehlertheorien), die im Laufe der Theorieentwicklung von T über T' nach T'' modifiziert werden, um die theoretischen Voraussagen präziser und gleichzeitig realitätsgerechter zu machen und so zu zunehmend komplexen und informationshaltigen Modellen der Wirklichkeit zu gelangen. Damit die Entwicklung von Hilfshypothesen nicht zu einer konventionalistischen Strategie entartet (Ad-hoc-Annahmen), muß jeder neue Schritt im FP konsequent gehaltsvermehrend sein (progressive Problemverschiebung). Gerade das wird aber bei der raffinierten Falsifikation einer Theorie T durch eine neue Theorie T' gefordert (Bedingung 1 und 3). Der harte Kern eines FPs wird auch als <u>negative Heuristik</u> bezeichnet, da er eine grundlegende methodologische Regel darstellt, die angibt, welche Teile des Forschungsprogramms prinzipiell der Widerlegung entzogen werden sollen. Der harte Kern eines FP wird also durch eine Entscheidung der das FP tragenden Forscher bewußt von der Falsifikation ausgeschlossen. (Das Paradigma in der Kuhnschen Wissenschaftstheorie enthält Hintergrundannahmen, die für das Weltbild der beteiligten Forscher konstitutiv sind. Im Gegensatz zum harten Kern eines FP sind diese Hintergrundannahmen den Forschern oft so selbstverständlich, daß Alternativen überhaupt nicht denkbar sind. In diesem Sinne liegt bei Kuhn keine bewußte Entscheidung für die Paradigmaannahmen vor.) Dies bedeutet, daß Modifikationen der Kernaussagen der Theorie auf jeden Fall vermieden werden sollen und der Theorie widersprechende empirische Tatbestände auf den Schutzgürtel der Hilfstheorien umgelenkt werden. Hier gibt die positive Heuristik die Richtung an, in die die Hilfstheorien weiterentwickelt werden sollen, um die Theorie realitätsgerechter zu gestalten.

Grundsätzlich stehen nach Lakatos immer mehrere FP in Konkurrenz zueinander, die versuchen, sich unter Beibehaltung ihres jeweiligen theoretischen Kerns im Sinne einer positiven Heuristik weiterzuentwickeln. Es gibt keine endgültige Widerlegung von FP, jedoch zeigt der Vergleich von FP in Bezug auf ihre Leistungsfähigkeit (progressive bzw. degenerative Problemverschiebung), welches FP erfolgreicher ist. Progressive FP entwickeln sich und die Arbeit an degenerativen FP wird schrittweise eingestellt. (S. 144ff.) FP werden also nicht falsifiziert, sondern eliminieren sich bei Erfolglosigkeit selbst. Im Gegensatz zum KR hat Kritik also nicht nur eine negative Funktion, sondern wirkt auch positiv im Sinne der Leistungskonkurrenz von FP.

Kritik an der Methodologie wissenschaftlicher Forschungsprogramme: Im Kern wird der Konventionalismus von Basissätzen im KR (methodologische Falsifikation bei Popper) von Lakatos ersetzt durch einen Konventionalismus von Theorien (raffinierte Falsifikation bei Lakatos). Dabei ist nicht einzusehen, warum es Teile einer Theorie geben soll (der sog. harte Kern), die qua Konvention von der Kritik ausgeschlossen werden. Das Argument von Lakatos, daß ein FP durch einen harten Kern seine Identität erhält und über einen längeren Zeitraum die Chance der empirischen Bewährung (bei ausschließlicher Modifikation der Hilfstheorien) erhalten muß, überzeugt nicht ganz. Auch hier stellt sich nämlich die Frage (vgl. auch das von Lakatos selbst gegebene Beispiel "planetarischer Unart"), wie lange die Arbeit an einem degenerativen FP fortgesetzt werden soll in der Hoffnung, durch eine geschickte Modifikation in den Hilfstheorien schließlich doch noch in ein progressives FP einzumünden. Auch im KR wird ja nicht durch das Auffinden eines einzigen Gegenbeispiels sofort methodologisch falsifiziert. Insofern läuft die Kritik von Lakatos an der utopischen Idee einer sofort wirkenden Rationalität im KR ins Leere. Auch bei der methodologischen Falsifikation muß es sich um wiederholbare Beobachtungen handeln, die zunächst in Bezug auf die Einhaltung aller methodologischen Regeln, insbesondere also im Hinblick auf die verwendeten Meßtheorien und Meßfehlertheorien, überprüft werden. Erst dann wird qua Beschluß über den Basissatz (vorläufig!) entschieden. Eine solche Prozedur enthält meistens implizit auch Vorstellungen über neue Theorien, so daß bei der Falsifiktion einer bestehenden Theorie nicht automatisch ein theoretisches Vakuum ensteht, wie von Lakatos unterstellt.

#### Abschließende Bewertung der Kritik von Kuhn und Lakatos:

Ein Vorzug der Position von Lakatos gegenüber dem KR ist sicher, daß Theorien nicht einfach als Aussagensysteme verstanden werden, sondern als Forschungsprogramme im Sinne strukturierter Ganzheiten gesehen werden, die neben einigen Kernhypothesen auch Instrumentheorien und Meßfehlertheorien enthalten. In ähnlicher Weise umfaßt auch der Paradigmabegriff von Kuhn Elemente, die berücksichtigen, daß Theorien ein strukturiertes Geflecht von Hypothesen unterschiedlicher Zentralität darstellen. Allerdings ist in beiden Fällen nicht ganz klar, wie einzelne Paradigmen oder Forschungsprogramme empirisch voneinander abzugrenzen sind. Ist z.B. die neoklassische Ökonomik eine Weiterentwicklung des FP der klassischen Ökonomik? Ist die Keynes sche Revolution eine wissenschaftliche Revolution, also ein Paradigmawechsel im Sinne von Kuhn? Auf diese Fragen geben die wissenschaftshistorischen Spezialisten in der Ökonomik durchaus unterschiedliche Antworten.

Schon daran wird deutlich, daß die sog. wissenschaftshistorische Herausforderung des methodologischen Falsifikationismus durch Kuhn und Lakatos insgesamt auf schwachen Füßen steht. Grundsätzlich ist anzumerken, daß auch der methodologische Falsifikationismus anerkennt, daß Basissätze falsch sein können und daher eine Falsifikation nie endgültig und unrevidierbar ist. Insbesondere ist in diesem Zusammenhang an die Theorieabhängigkeit von Beobachtungen und an den holistischen Charakter der Überprüfung (schwache Form der Duhem-Quine-These) erinnern. Ein solcher aufgeklärter methodologischer Falsifikationismus ist durchaus mit einem längeren historischen Prozeß der Überprüfung von Theorien vereinbar, in dem über einen längeren Zeitraum durchaus unterschiedliche, oft sogar einander widersprechende Einschätzungen von Prüfsätzen bestehen können. In diesem Sinne kommt z.B. Andersson (1988) zu einer revidierten Interpretation der von Kuhn und Lakatos herangezogenen wissenschaftshistorischen Beispiele, die durchaus mit einem (auch in diesem Skript vertretenen) aufgeklärten methodologischen Falsifikationismus (wie er in den Grundzügen bereits in Poppers "Logik der Forschung" enthalten ist) vereinbar ist.

Die Kritik von Kuhn und Lakatos beruht implizit auf einer in die Irre führenden begründungsphilosophischen Deutung der Falsifikation, wie sie, dies sei noch einmal betont, mit dem aufgeklärten methodologischen Falsifikationismus nicht vereinbar ist. Umgekehrt ist der logische
Status der empirischen Gegenevidenz in den Wissenschaftsphilosophien von Kuhn und
Lakatos durchaus unklar. Kuhn spricht von Rätseln, von Anomalien und von
Gegenbeispielen, wobei er diese Begriffe eher wissenschaftspsychologisch im Sinne der
Bedrohung des das Paradigma tragenden Weltbildes interpretiert. Es werden also keine
wissenschaftstheoretisch abgeleiteten Kritierien dafür angegeben, was es bedeutet, daß sich
diese Anomalien "häufen", und wann diese Anomalien zum Ausgangspunkt einer

wissenschaftlichen Revolution werden. Auch bei Lakatos bleibt im Grunde unklar, wie eigentlich zwischen einem progressiven und einem degenerativen FP entschieden werden kann. Auf jeden Fall zeigt sich bei Lakatos in diesem Zusammenhang eine starke induktivistische Komponente, da von den Leistungen eines FPs in der Vergangenheit auf dessen zukünftige theoretische Fruchtbarkeit geschlossen wird. Darüberhinaus bedeutet das von Lakatos propagierte konventionalistische Festhalten an dem harten Kern eines FP einen Dogmatismus, der mit den Grundlagen des KR unvereinbar ist. Das Argument, das Festhalten an einem harten Kern sei für die Aufrechterhaltung der "Identität" eines FP unabdingbar, ist im Lichte der oben kurz angesprochenen wissenschaftshistorischen Evidenz nicht sehr überzeugend. Wenn nämlich die Abgrenzung von FPs wissenschaftshistorisch problematisch ist, kann die "Identität" eines FPs für das kontinuierliche Arbeiten in einer Disziplin nicht die Bedeutung haben, die Lakatos ihr zuschreibt.

In diesem Zusammenhang ist es wichtig, noch einmal zu betonen, daß der bei Popper konstatierte Konventionalismus bei der Entscheidung über Basissätze nicht dogmatisch zu verstehen ist, sondern nur relativ zu dem jeweiligen theoretischen und methodologischen Hintergrundwissen. Damit sind Kriterien gegeben, die der Prüfung von Basissätzen zugrunde liegen - dieser "Konventionalismus" ist also kritisierbar. Der Konventionalismus von Lakatos ist hingegen dogmatisch, da er den harten Kern eines FP grundsätzlich von der Kritik ausschließt. Grundgedanke der Philosophie von Popper ist aber die kritische Haltung und damit gleichzeitig die Vermeidung von dogmatischer Schließung und relativistischer Beliebigkeit. Insoweit diese Elemente. implizit oder explizit, die wissenschaftstheoretischen Positionen von Kuhn und Lakatos wieder Einzug halten, ist der aufgeklärte methodologische Falsifikationismus von Popper überlegen.

#### Exkurs: Konstruktivismus und evolutionäre Erkenntnistheorie

Konstruktivismus und evolutionäre Erkenntnistheorie stellen neuere Entwicklungen in der Erkenntnistheorie dar. Kernaussage des <u>Konstruktivismus</u> ist, daß Menschen als erkennende Systeme ihre Erkenntnis von Realität aktiv konstruieren (vgl. Watzlawick 1985). Die Frage nach der <u>Übereinstimmung</u> einer solchen Konstruktion mit der Wirklichkeit kann dabei nicht mehr gestellt werden. Entscheidend ist allein, daß sich solche <u>Konstruktionen bewähren</u>, d.h. die Anpassung eines Organismus an seine Umwelt gewährleisten. In diesem Sinne enthält der Konstruktivismus eine <u>instrumentalistische Komponente</u> (vgl. 2.1.3). Weiterhin enthält der Konstruktivismus Verbindungen zur Theorie autopoietischer Systeme und allgemein zur Theorie selbstreferentieller Systeme, da das Gehirn bzw. der menschliche Erkenntnisapparat als ein solches selbstreferentielles System verstanden wird. Entscheidend, und genau dies ist die Differenz zur evolutionären Erkenntnistheorie, ist dabei der Gedanke der <u>Autonomie</u> in der Konstruktion von Erkenntnis. Die Position des hypothetischen Realismus, die die Wirklichkeit als zumindest teilweise erkennbar ansieht, wird abgelehnt.

Auch die evolutionäre Erkenntnistheorie betont den konstruktiven Charakter unserer Erkenntnis. Leben wird generell als erkenntnisgewinnender Prozeß verstanden, wobei im Vordergrund jeweils die Passung zwischen Organismus und Umwelt steht. Entscheidend ist aber, daß unsere Erkenntnis immer eine, wenn auch partielle, Re-Konstruktion der Wirklichkeit ist. Unsere Erkenntnis (schon der ratiomorphe Weltbildapparat aber auch unsere wissenschaftlichen Hypothesen) macht also Aussagen über eine unabhängig von uns existierende Wirklichkeit (hypothetischer Realismus, Korrespondenztheorie der Wahrheit).(vgl. 2.1.1) Die biologische Evolution von Erkenntnis wird von Konrad Lorenz (1973) dargestellt. Eine Übertragung des evolutionstheoretischen Ansatzes auf die Entwicklung von Ideen und wissenschaftlichen Hypothesen findet man bei Popper (1973). In diesem Zusammenhang ist insbesondere die Drei-Welten-Theorie von Bedeutung.

Der Vorzug der evolutionären Erkenntnistheorie im Vergleich zum Konstruktivismus liegt darin, daß der hypothetische Realismus nicht aufgegeben wird. Der radikale Konstruktivismus kann meiner Ansicht nach nicht deutlich machen, was der Gedanke der Konstruktion (im Gegensatz zu einer näherungsweise adäquaten Re-Konstruktion einer gegebenen Realität) eigentlich bedeutet und in welchem Sinn in diesem Zusammenhang überhaupt noch von Erkenntnis (von was?) gesprochen werden kann. Dennoch ist vieles am Konstruktivismus bedenkenswert, insbesondere die Betonung der selbstreferentiellen Organisation und der operativen Geschlossenheit unseres Erkentnisapparates (Gehirn) (vgl. 2.1.1).

### 3. Untersuchungsformen

Wie bereits in Kapitel 2 angedeutet, kann grundsätzlich zwischen <u>explorativen</u> (erkundenden) und hypothesentestenden Untersuchungsformen unterschieden werden. Im ersten Fall geht es darum, zunächst grundlegende Informationen über den Forschungsgegenstand zu sammeln. Dabei kommen im wesentlichen <u>qualitative</u> Erhebungsverfahren zum Einsatz. Die Forschung ist zwar grundsätzlich theoriegeleitet, die un- bzw. teilstrukturierten Erhebungsverfahren bleiben aber offen für die zu erkundenden spezifischen Sichtweisen der Teilnehmer der Untersuchung. Ein Beispiel wären etwa narrative (erzählende) Interviews in der Biographieforschung, die Informationen über die Problematik der Vereinbarkeit von Familien- und Berufsrolle aus Sicht der befragten Frauen sammeln sollen. Ein anderes Beispiel sind Gruppendiskussionen, denen Testkonsumenten über ihre Beurteilung in eines neuentwickelten Produktes diskutieren. In allen Fällen ist das Ergebnis der Datensammlung ein Text (Befragungs- oder Beobachtungsprotokoll), der dann inhaltsanalytisch ausgewertet wird. In der Regel sind solche explorativen Untersuchungen Einzelfallstudien oder beziehen sich auf wenige, meistens bewußt (d.h. nicht durch einen Zufallsprozeß) ausgewählte Untersuchungseinheiten. Das Ziel einer explorativ angelegten Untersuchung ist also ein vertieftes Wissen über den Forschungsgegenstand, eine Spezifizierung des theoretischen Verständnisses des Forschungsfeldes und eine Ableitung konkreter Forschungshypothesen, die dann in einer anschließenden hypothesentestenden Untersuchung mit den Mitteln der quantitativen Sozialforschung überprüft werden können. Auf die besonderen Probleme explorativer Untersuchungen kommen wir noch einmal im Kapitel über Erhebungsverfahren im Zusammenhang mit der Befragung und speziell der Inhaltsanalyse zurück (vgl. 6.3).

Im folgenden sollen hypothesentestende Untersuchungsformen im Mittelpunkt stehen. Dabei gehen wir in allen Fällen davon aus, daß eine (bewußte oder zufällige) Auswahl von Untersuchungseinheiten (z.B. Befragte oder Versuchspersonen) getroffen wird, für die mit einem standardisierten Erhebungsinstrument Informationen über eine Vielzahl von Merkmalen (wie z.B. Bildung, Selbstwertgefühl, normative Geschlechtsrollenorientierung, Ausländerfeindlichkeit, Postmaterialismus, materielle Gütergebundenheit, Kaufsucht, usw.) gesammelt werden. Die resultierende Datenmatrix mit den Variablen x, y, z, ... ist dann die Grundlage der kausalanalytischen Untersuchung. Im folgenden unterscheiden wir Untersuchungsformen nicht nach der Art des Erhebungsverfahrens (vgl. Kapitel 6), der verwendeten Skalierungsverfahren (vgl. Kapitel 5) oder der Art des Auswahlverfahrens (vgl. Kapitel 7), sondern danach, inwieweit eine spezifische Untersuchungsform aufgrund der ihr zugrundeliegenden Struktur der Daten geeignet ist, Kausalhypothesen zu überprüfen.

Aus der Vielzahl möglicher Untersuchungsformen greifen wir für die folgende Diskussion die folgenden drei heraus:

- 1. **Querschnittsuntersuchung:** Die Untersuchung findet zu einem Zeitpunkt statt, d.h. <u>alle</u> Variablen (x, y, z, ...) werden zum gleichen Zeitpunkt erhoben.
- 2. **Panel-Untersuchung:** Die Erhebung findet zu mindestens <u>zwei Zeitpunkten</u> bei der <u>gleichen Untersuchungsgesamtheit</u> statt (x<sub>1</sub>, y<sub>1</sub>, z<sub>1</sub>, ...; x<sub>2</sub>, y<sub>2</sub>, z<sub>2</sub>, ...), so daß <u>individuelle Veränderungen</u> erfaßt werden können.
- 3. **Experiment**: Der <u>experimentelle Stimulus</u> (die unabhängige Variable x) wird vom Forscher explizit in die Untersuchungsssituation eingeführt. Dies geschieht unter kontrollierten Bedingungen (<u>Kontrolle von Störvariablen</u> z, ...). Anschließend wird die Veränderung in der experimentellen (abhängigen) Variablen y bei den Versuchspersonen in der Experimentalsituation gemessen (entweder einfache Nachher-Messung oder Vorher-Nachher-Messung).

Kausalität ist eine fundamentale Kategorie unseres Denkens. Ähnlich wie die Kategorien von Raum und Zeit existiert sie a priori (d.h. vor jeder Erfahrung; vgl. auch die Anmerkungen zur Kantschen Erkenntnistheorie in 2.1.3). Wir können zwar beobachten, daß eine Person Verbrennungen an der Hand hat, <u>nachdem</u> sie (irrtümlich) eine heiße Herdplatte berührt hat, aber der Schluß, daß sich die Person die Hand verbrannt hat, <u>weil</u> sie die Herdplatte berührt hat, ist nur auf der Grundlage einer Theorie möglich, die den Kausalitätsbegriff bereits voraussetzt.

Wir benötigen daher <u>Kausalitätskriterien</u>, um in einer empirischen Untersuchung über das Vorliegen einer Kausalitätsbeziehung entscheiden zu können. Die <u>Ursache x wirkt kausal auf die abhängige Variable y</u>, wenn folgende Bedingungen erfüllt sind:

- x geht y zeitlich voraus
- x und y sind im Rahmen einer Theorie miteinander verknüpft
- x und y sind empirisch korreliert (Möglichkeit der Schein-Nonkorrelation beachten)
- die <u>Beziehung zwischen</u> x und y verschwindet nicht bei Kontrolle anderer Einflüsse (<u>Kontrolle von Störfaktoren</u> im echten Experiment bzw. <u>Drittvariablenkontrolle</u> bei der Querschnittserhebung)
- der Einfluß von Meßfehlern wurde kontrolliert.

Im folgenden wenden wir uns den einzelnen Untersuchungsformen zu, um festzustellen, inwieweit die einzelnen Kriterien für das Vorliegen einer Kausalbeziehung überprüft werden können. Dabei klammern wir die fünfte Bedingung (Kontrolle von Meßfehlern) zunächst aus; sie wird im Zusammenhang mit dem kausalanalytischen Ansatz in 4.3 und der Klassischen Testtheorie in 5.4.1 wieder aufgegriffen.

## 3.1 Querschnitts-Design (Kausale Analyse I)

Die Querschnittsanordnung weist zwei entscheidende Nachteile auf, die die Möglichkeiten zur Überprüfung von Kausalitätsbeziehungen stark einschränken:

- 1. Da alle Variablen zum gleichen Zeitpunkt erhoben werden, kann die <u>kausale Reihenfolge</u> der Variablen nicht empirisch ermittelt, sondern muß theoretisch begründet werden.
- 2. Da eine explizite <u>Kontrolle von Störfaktoren</u>, wie sie insbesondere im Experiment möglich ist, nicht stattgefunden hat, müssen die möglichen Drittvariableneinflüsse im nachhinein kontrolliert werden (<u>Drittvariablenkontrolle</u>).

Das erste Kausalitätskriterium (zeitliche Abfolge) kann also nicht überprüft werden. Wird z.B. in einer Untersuchung zur Kaufsucht (vgl. Kapitel 1) sowohl das Selbstwertgefühl als auch die Kaufsucht der Befragten (auf einer Likert-Skala (vgl. 5.3)) gemessen, kann die kausale Reihenfolge nicht eindeutig aus der Untersuchung abgeleitet werden. (Oft enthalten Querschnittsuntersuchungen aber sog. retrospektive Fragen, wie z.B. die nach dem höchsten allgemeinbildenden Schulabschluß (Bildung). Wurde zusätzlich der gegenwärtige Berufsstatus erhoben, so ist es in diesem Fall möglich, eine eindeutige kausale Reihenfolge abzuleiten: Das Bildungsniveau bewirkt den gegenwärtigen Berufsstatus.)

Um so wichtiger ist es, die theoretisch postulierte kausale Abfolge der Variablen genauer zu begründen. In unserem Beispiel käme es also etwa darauf an zu argumentieren, welche theoretischen Gründe für eine kausale Wirkung des Selbstwertgefühls auf die Kaufsucht sprechen (Hypothese: Je geringer das Selbstwertgefühl, desto stärker ist die Kaufsucht).

Liegt eine Kausalbeziehung zwischen den Variablen x und y vor, muß in einer Querschnittsuntersuchung auch eine empirische Beziehung (Korrelation) zwischen den Variablen nachzuweisen sein. Allerdings ist auch der (sehr seltene) Fall einer Schein-Nonkorrelation denkbar, wenn nämlich beim Vorliegen einer Kausalbeziehung zwischen x und y kein entsprechender empirischer Zusammenhang festzustellen ist, weil sich der vorhandene direkte kausale Effekt und der ebenfalls vorhandene entgegengerichtete indirekte kausale Effekt gerade kompensieren (siehe unten).

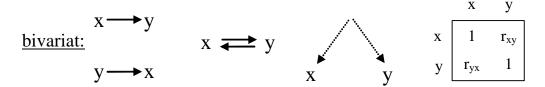
Je nach dem Meßniveau der beiden Variablen kommen zur Bestimmung der Stärke der Beziehung verschiedene Zusammenhangsmaße (Korrelationskoeffizienten) in Frage (vgl. auch die Tabelle in Kapitel 5.2). Wir betrachten im folgenden zunächst <u>intervallskalierte Variablen x, y, z, ...</u> und verwenden <u>Pearsons r</u> als Korrelationskoeffizient. Dabei wird eine <u>lineare Beziehung</u> zwischen den Variablen angenommen (vgl. die Ausführungen zur linearen Regression weiter unten), eine Annahme, die natürlich zunächst empirisch überprüft werden muß, bevor r als Korrelationskoeffizient verwendet werden darf. Zusätzlich betrachten wir dann den Fall zweier <u>dichotom gemessener Variablen X und Y</u> (vgl. die Ausführungen zur Tabellenanalyse und zur <u>Prozentsatzdifferenz</u> <u>d%(Y,X)</u> weiter unten), um auch Interaktionseffekte in einfacher Form darstellen zu können.

Das Ergebnis einer Querschnittsuntersuchung kann in einer Matrix zusammengefaßt werden, die die Stärke und Richtung der Beziehung zwischen den erhobenen Variablen in Form von Korrelationskoeffizienten enthält (Korrelationsmatrix). Generell gilt:

#### Zu jeder Korrelationsstruktur gibt es stets mehrere äquivalente Kausalstrukturen.

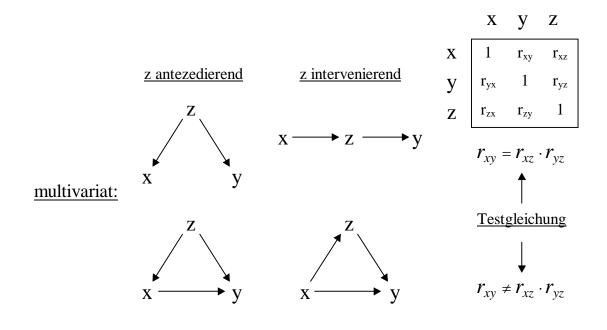
Dies kann bereits im Falle einer einzigen bivariaten Beziehung gezeigt werden:

Unterschiedliche Kausalstrukturen implizieren gleiche Korrelationsmuster



Liegt ein empirischer Zusammenhang zwischen x und y vor, kann er auf eine einseitige Kausalbeziehung von x nach y <u>oder</u> von y nach x zurückzuführen sein - aber auch eine wechselseitige Beeinflussung ist möglich.

Darüber hinaus ist aber auch denkbar, daß überhaupt keine Kausalbeziehung zwischen x und y besteht, und der empirische Zusammenhang allein auf das Wirken einer <u>Drittvariablen</u> zurückzuführen ist (gestrichelte Pfeile in der obigen Abbildung). In einem solchen Fall spricht man von einer Scheinkorrelation (siehe unten). Daran wird deutlich, daß eine bivariate Beziehung immer vor dem <u>Hintergrund weiterer kausaler Einflußfaktoren</u>, den sog. Drittvariablen gesehen werden muß. Aus der bivariaten wird dadurch eine <u>multivariate Kausalstruktur</u>. In Abhängigkeit von ihrer Stellung in ihrer Kausalstruktur wird unterschieden zwischen einer <u>antezedierenden (vorausgehenden)</u> und einer <u>intervenierenden</u> (dazwischentretenden) Drittvariablen z.



Auch über die kausale Reihenfolge in der erweiterten Kausalstruktur kann nur aufgrund theoretischer Vorüberlegungen entschieden werden. Auch hier gilt daher: Unterschiedliche Kausalstrukturen (z ist entweder antezedierend oder intervenierend) implizieren die gleichen empirischen Beziehungen zwischen denVariablen und damit eine identische Korrelationsmatrix.

Bei der Überprüfung einer Kausalhypothese (bivariate Beziehung) muß bereits bei Anlage der Untersuchung aufgrund theoretischer Überlegungen darüber entschieden werden, welche Variablen als <u>Drittvariablen</u> zusätzlich in das Kausalmodell aufgenommen werden sollen. Trivialerweise können nämlich nur die Variablen bei der Drittvariablenkontrolle berücksichtigt werden, die auch tatsächlich erhoben wurden. Bei der Planung einer Untersuchung muß daher immer gefragt werden, welche Einflüsse neben der unabhängigen Variablen zusätzlich auf die abhängige Variable wirken. Je nachdem, ob diese Variable der unabhängigen Variablen kausal vorangeht oder kausal von dieser beeinflußt wird, handelt es sich dann um eine antezedierende oder intervenierende Drittvariable. Alle theoretisch relevanten Drittvariablen müssen also bereits in der Planungsphase der Untersuchung in das zugrundeliegende Kausalmodell integriert werden, um bei der Drittvariablenkontrolle Berücksichtigung finden zu können.

Beispiele für eine intervenierende Variable z:

#### z interveniert zwischen x und y: $x \rightarrow z \rightarrow y$ , z.B.:

```
(Kirchenbindung → Einstellung zur Abtreibung → Wahl der CDU)
(Berufsstatus des Vaters → Bildung des Befragten → Berufsstatus des Befragten)
```

Beispiele für eine antezedierende Variable z:

#### z antezediert bzgl. x und y: $x \leftarrow z \rightarrow y$ , z.B.:

```
(Storchendichte ← Urbanisierung → Geburtenrate)
(Anzahl der Löschzüge ← Größe des Brandes → Brandschaden)
```

In allen Fällen kann grundsätzlich auch zusätzlich eine direkte kausale Beziehung zwischen x und y postuliert werden. Allerdings wäre eine solche Annahme (im Gegensatz zu den Beispielen für eine intervenierende Drittvariable) im Falle der beiden Beispiele für eine antezedierende Drittvariable theoretisch höchst unplausibel. Es handelt sich vielmehr in beiden Fällen um eine reine Scheinkorrelation (siehe unten).

### 3.1.1 Kausale Abhängigkeitsanalyse: Pfadanalyse

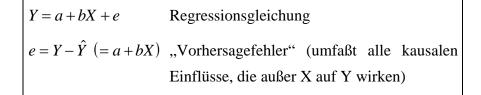
[Zur linearen Regressionsanalyse mit SPSS für Windows gibt es eine Beispieldatei (Acrobat). Diese Datei kann über das WWW bezogen werden]

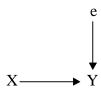
Im folgenden setzen wir <u>intervallskalierte Variablen</u> und <u>lineare Beziehungen</u> zwischen den Variablen voraus. Für den bivariaten Fall betrachten wir zunächst unstandardisierte Variablen

und gehen dann zu <u>standardisierten Variablen</u> über (bivariate <u>Pfadanalyse</u>). Für den multivariaten Fall setzen wir, wie in allen folgenden Überlegungen zur Pfadanalyse, standardisierte Variablen voraus. Die kausalen Modelle werden einmal graphisch als Pfadmodelle und zum anderen formal als Strukturgleichungsmodelle (Regressionsgleichungen) dargestellt. Wir betrachten zunächst den bivariaten Fall:

#### 3.1.1.1 Bivariate Regression

#### a) unstandardisierte Variablen (X,Y):





Methode der kleinsten Quadrate (Minimierung der Fehlervarianz):

$$\frac{1}{N} \sum_{i=1}^{N} e_i^2 = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 = \frac{1}{N} \sum_{i=1}^{N} (Y_i - (a + bX_i))^2 \implies \min!$$

$$s_{XX} = s_X^2 = \text{Varianz von } X = \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})^2$$

$$s_{XY} = \text{Kovarianz von } X \text{ und } Y = \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})$$

$$b = \frac{s_{XY}}{s_{XX}} \qquad a = \overline{Y} - b\overline{X}$$

<u>Lösung</u> des Minimierungsproblems

Für die Methode der kleinsten Quadrate gilt:

Der Mittelwert des Vorhersagefehlers ist Null ( $\overline{e} = 0$ ) und die Kovarianz von e und X und damit auch die Korrelation ist ebenfalls gleich Null ( $r_{Xe} = 0$ ). Im Fall der Pfadanalyse sprechen wir im folgenden von einem <u>Zufallsfehler</u>, wenn der Fehlerterm einen Erwartungswert von Null hat und mit den unabhängigen Variablen in der Strukturgleichung unkorreliert ist.

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \text{Korrelationskoeffizient}$$

(Produkt-Moment-Korrelationskoeffizient; Pearsons r) Varianzanalytische Interpretation von  $r^2$ :

$$\frac{1}{N} \sum_{i=1}^{N} (Y_i - \overline{Y})^2 = \frac{1}{N} \sum_{i=1}^{N} (\hat{Y}_i - \overline{Y})^2 + \frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$$

(<u>Gesamtvarianz</u> = <u>erklärte Varianz</u> + <u>Fehlervarianz</u>)

$$r_{XY}^2$$
 = Anteil der erklärten Varianz  
 $1 - r_{XY}^2$  = Anteil der Fehlervarianz

$$1 - r_{XY}^2$$
 = Anteil der Fehlervarianz

#### standardisierte Variablen (x,y):

$$X \Rightarrow x = \frac{X - \overline{X}}{s_X}$$
 Die standardisierte Variable  $x$  hat den Mittelwert 0 und die Standardabweichung 1.

Regressionsgleichung:  $y = (\alpha + )\beta x + \epsilon$ 

β ist der standardisierte Regressionskoeffizient. Nach a) gilt:

$$\beta = r_{xy}$$
  $\alpha = 0$ 

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = s_{xy} = \frac{1}{N} \sum_{i=1}^{N} x_i y_i$$
 (\*)

#### c) bivariate Pfadanalyse:

Es werden standardisierte Variablen vorausgesetzt.

$$x \xrightarrow{p_{yx}} y$$

$$y = p_{yx} x + p_{ye} \varepsilon$$
[Pfaddiagramm]
$$y = p_{yx} x + p_{ye} \varepsilon$$
[Strukturgleichung]

Mit Hilfe des Verfahrens der Multiplikation der Strukturgleichung(en) eines Pfadmodells mit der (den) jeweiligen prädeterminierten Variablen (d.h. mit der (den) unabhängigen Variablen) können Pfadmodelle auf einfache und übersichtliche Weise gelöst werden. Dazu müssen jeweils geeignete Eigenschaften für die Fehlerterme vorausgesetzt werden, üblicherweise die Unabhängigkeit der Fehlervariablen von den kausal vorangehenden Variablen (hier:  $r_{x\varepsilon} = 0$ (!)). Das Verfahren wird nun für den vorliegenden Fall einer bivariaten Strukturgleichung ausführlich erläutert.

$$x_{i} y_{i} = p_{yx} x_{i} x_{i} + p_{y\varepsilon} x_{i} \varepsilon_{i} \qquad | x_{i}$$

$$\frac{1}{N} \sum_{i=1}^{N} x_{i} y_{i} = \frac{1}{N} p_{yx} \sum_{i=1}^{N} x_{i} x_{i} + \frac{1}{N} p_{y\varepsilon} \sum_{i=1}^{N} x_{i} \varepsilon_{i} \qquad | \frac{\Sigma}{N}$$

$$r_{xy} = p_{yx} r_{xx} + p_{y\varepsilon} r_{x\varepsilon} \qquad (\text{nach *})$$

Generell ergibt sich also als Ergebnis der Multiplikation eine Gleichung, die aus der zugrundeliegenden Strukturgleichung dadurch hervorgeht, daß die neu entstandenen "gemischten" Variablen-Terme durch die entsprechende Korrelation ersetzt werden.

Weiter gilt allgemein: $r_{xx}=1$  und qua Voraussetzung zusätzlich  $\ r_{x\epsilon}=0$ . Daher erhalten wir als Lösung des bivariaten Pfadmodells:

$$p_{yx} = r_{yx} = \beta$$

 $\underline{Pfadkoeffizient} \equiv standardisierter \ Regressionskoeffizient$ 

→ K. Opp / P. Schmidt: Einführung in die Mehrvariablenanalyse, Teil 3

#### Fehlspezifikation:

Der Pfadkoeffizient (standardisierter Regressionskoeffizient=Korrelationskoeffizient) ist im bivariaten Fall nur dann ein unverzerrter Schätzer für den tatsächlichen kausalen Einfluß von x auf y, wenn die Modellvoraussetzungen erfüllt sind. Insbesondere muß es sich bei ε um ein Zufallsfehler handeln, d.h.  $\varepsilon$  ist mit der unabhängigen Variablen x unkorreliert. Nun enthält aber ε alle kausalen Einflüsse die neben x auf y wirken. Es ist daher durchaus vorstellbar, daß diese anderen Einflüsse mit x korreliert sein können. In diesem Fall ist eine entscheidende Voraussetzung des (bivariaten) Pfadmodells nicht erfüllt, das Modell ist fehlspezifiziert und der Pfadkoeffizient ist verzerrt, d.h. er spiegelt nicht den tatsächlichen kausalen Einfluß von x auf y wieder. Dazu ein Beispiel (vgl. auch 3.1.1.3). In einer empirischen Untersuchung wird eine Korrelation zwischen dem Berufsstatus des Befragten und dem seines Vaters von .48 ermittelt. Dem entspricht im bivariaten Kausalmodell ein Pfadkoeffizient von .48, der auf einen starken direkten kausalen Effekt des Berufsstatus des Vaters (sozialer Herkunftsstatus) hindeutet (Statusvererbung). Gemäß dem Voraussetzungen des Modells korreliert der Berufsstatus des Vaters des Befragten (x) nicht mit den anderen Variablen, die ebenfalls einen kausalen Einfluß auf den Berufsstatus des Befragten (y) haben, und die in dem Fehlerterm E zusammengefaßt sind  $(r_{x\varepsilon} = 0)$ . Dazu gehört u.a. auch die Bildung des Befragten (z), die sicher einen kausalen Einfluß auf den erreichten Berufsstatus des Befragten hat. Offensichtlich hängt aber das erreichte Bildungsniveau wiederum von dem Herkunftsstatus (x) ab. Die Voraussetzung  $r_{x\varepsilon} = 0$  ist also in diesem Fall nicht erfüllt, das Pfadmodell ist fehlspezifiziert. Berücksichtigt man das Bildungsniveau des Befragten als zusätzliche (intervenierende) Variable im Modell, so reduziert sich der direkte kausale Einfluß des Berufsstatus des Vaters auf den des Sohnes (pyx) von .48 auf .17 (vgl. 3.1.1.3). Die durch Fehlspezifikation hervorgerufenen Verzerrungen in den geschätzten Kausalkoeffizienten können also erheblich sein.

Um Fehlspezifikationen zu vermeiden, muß eine Drittvariable z genau dann explizit als zusätzliche Variable in das Pfadmodell aufgenommen werden, das den kausalen Einfluß von x auf y spezifiziert, wenn gilt:

- 1. z wirkt kausal auf y, d.h. insbesondere muß eine Korrelation zwischen z und y bestehen und
- 2. z ist mit x korreliert. In diesem Fall ist nämlich die Voraussetzung der bivariaten Pfadanalyse ( $r_{x\epsilon} = 0$ ) verletzt.

Je nachdem, wie die Kausalbeziehung zwischen z und x angenommen wird, muß z dann entweder als antezedierende oder als intervenierende Variable in das Pfadmodell aufgenommen werden. Offensichtlich kann dieses Argument iteriert werden, so daß in der Forschungspraxis erweiterte Kausalmodelle entstehen, die ein komplexes Geflecht von Kausalbeziehungen zwischen einer Vielzahl von Variablen enthalten. Dennoch kann man nie sicher sein, daß alle relevanten Einflußfaktoren tatsächlich erfaßt wurden. Bleiben die Pfadkoeffizienten des Modells allerdings beim Hinzufügen weiterer Variablen stabil, so kann man mit einer gewissen Sicherheit davon ausgehen, ein zumindestens annähernd realistisches Kausalmodell spezifziert zu haben. Davon unberührt bleiben natürlich Spezifikationsfehler anderer Art, wie z.B. im Hinblick auf die funktionale Form der Beziehungen (Linearität und die Richtung der Kausalität). Abschließend stellen wir daher noch einmal mit Nachdruck fest, daß eine umfassende theoretische Begründung des Kausalmodells die unabdingbare Voraussetzung einer jeden Pfadanalyse ist.

Grundlegend für die folgenden Überlegungen ist die Unterscheidung zwischen rekursiven und nichtrekursiven Kausalstrukturen. Kausalstrukturen sind <u>rekursiv</u>, wenn sie keine kausale Zyklen enthalten, also Kausalbeziehungen, die von einer Variablen ausgehen und (möglicherweise über eine längere Kette von Kausalbeziehungen) indirekt auf diese Variable zurückwirken. Das einfachste Beispiel einer <u>nichtrekursiven</u> Kausalstruktur ist die wechselseitige Kausalbeziehung zwischen zwei Variablen. Im Fall von drei oder auch mehreren Variablen können auch längere Kausalzyklen auftreten.



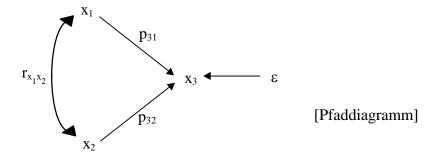
nichtrekursive Kausalstruktur



(vollständig) rekursive Kausalstruktur

Im folgenden betrachten wir nur noch rekursive Kausalstrukturen und klammern damit die schwierigen Spezifikations- und Schätzprobleme aus, die sich bei nichtrekursiven Systemen stellen können.

#### 3.1.1.2 Multivariate Regression



Der gebogene Doppelpfeil zwischen  $x_1$  und  $x_2$  soll auf eine kausal nicht weiter spezifizierte korrelative Beziehung zwischen  $x_1$  und  $x_2$  hinweisen.

$$x_3 = p_{31}x_1 + p_{32}x_2 + p_{3\varepsilon}\varepsilon$$
 [Strukturgleichung]

Multiplikation der Strukturgleichung mit den beiden prädeterminierten Variablen  $x_1$  und  $x_2$  ergibt bei entsprechenden Voraussetzungen für den Fehlerterm  $\epsilon$ :

$$r_{31} = p_{31} + p_{32}r_{21}$$
  $(r_{x_1\varepsilon} = 0 \ (!))$   
 $r_{32} = p_{31}r_{21} + p_{32}$   $(r_{x_2\varepsilon} = 0 \ (!))$ 

Daraus ergeben sich die <u>Pfadkoeffizienten</u> als <u>standardisierte partielle</u> <u>Regressionskoeffizienten</u>.

$$p_{31} = \frac{r_{31} - r_{32}r_{21}}{1 - r_{21}^2} \qquad p_{32} = \frac{r_{32} - r_{31}r_{21}}{1 - r_{21}^2}$$

Für rekursive Pfadmodelle sind die Pfadkoeffizienten (bei geeigneten Annahmen über die Fehlerterme) also stets die standardisierten Regressionskoeffizienten und können mit Hilfe der Methode der kleinsten Quadrate, also mit jedem Statistikprogramm, das über eine einfache Regressionssroutine verfügt (wie z.B. SPSS-PC+), direkt berechnet werden.

# Exkurs: erklärte Varianz und Varianzzerlegung im Falle der multivariaten Regression

Wie gesehen, ist die erklärte Varianz im bivariaten Falle gleich dem Quadrat des Korrelationskoeffizienten  $(r_{yx}^2)$ . Im multivariaten Fall ist die Situation komplexer. Die von den beiden unabhängigen Variablen  $x_1$  und  $x_2$  in y gemeinsam erklärte Varianz  $(R^2_{3.12})$  oder kurz  $R^2$ 0 kann im allgemeinen nicht addiditiv in die durch  $x_1$   $(r_{31}^2)$  und  $x_2$   $(r_{32}^2)$  erklärte Varianz zerlegt werden. Allgemein gilt:

$$R_{3.12}^2 = p_{31}^2 + p_{32}^2 + 2p_{31}p_{32}r_{21}$$

Eine eindeutige Zerlegung existiert nur für den Fall, daß die unabhängigen Variablen  $x_1$  und  $x_2$  voneinander unabhängig sind ( $r_{12}$ =0).

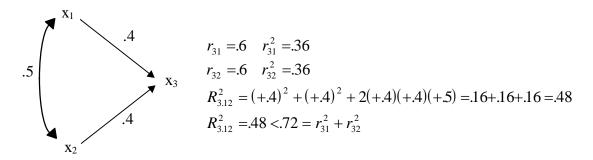
Dann gilt:

$$p_{31} = r_{31}$$
 und  $p_{32} = r_{32}$  und damit  
 $R_{3.12}^2 = r_{31}^2 + r_{32}^2$   $\left(r_{21} = 0 \ (!)\right)$ 

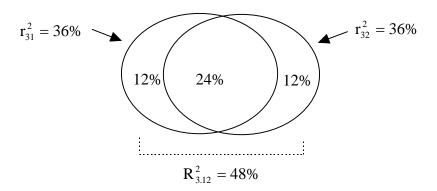
Korrelieren aber  $x_1$  und  $x_2$  positiv miteinander und sind darüberhinaus die beiden Pfadkoeffizienten positiv, so "überlappen" sich  $x_1$  und  $x_2$  bei der Erklärung der abhängigen Variaben  $x_3$ , d.h.

$$R_{312}^2 < r_{31}^2 + r_{32}^2$$

Beispiel:



Der größte Teil der von  $x_1$  und  $x_2$  erklärten Varianz ist gemeinsam erklärte Varianz, wie aus dem folgenden Mengendiagramm deutlich wird:

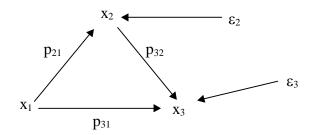


3

1

### 3.1.1.3 Effektzerlegung im vollständig rekursiven Modell

#### a) Das Modell mit drei Variablen



[Pfaddiagramm]

$$x_2 = p_{21}x_1 + p_{2\varepsilon_2}\varepsilon_2$$
  
$$x_3 = p_{31}x_1 + p_{32}x_2 + p_{3\varepsilon_3}\varepsilon_3$$

(bivariate Regression)

(multivariate Regression) [Strukturgleichungen]

Es werden Zufallsfehler vorausgesetzt ( $r_{x_1\epsilon_2}=0,\ r_{x_1\epsilon_3}=0,\ r_{x_2\epsilon_3}=0$  (!) ):

Wie bereits im Zusammenhang mit der bivariaten und der multivariaten Regression gezeigt, können die Pfadkoeffizienten als standardisierte (partielle) Regressionskoeffizienten aus der empirisch erhobenen Korrelationsmatrix berechnet werden. Aus den durch die Multiplikationsmethode gewonnenen Gleichungen ergibt sich daraus die folgende Effektzerlegung:

$$\mathbf{r}_{21} = \mathbf{p}_{21}$$

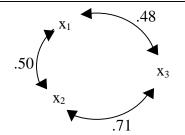
$$\mathbf{r}_{31} = \mathbf{p}_{31} + \mathbf{p}_{32}\mathbf{p}_{21}$$

$$r_{32} = p_{32} + p_{21}p_{31}$$

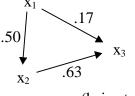
#### Beispiel: Statuszuweisungsprozeß (Müller 1972)

 $x_1 = \text{Berufstatus des Vaters} \qquad 1 \qquad 1$   $x_2 = \text{Bildung des Befragten} \qquad 2 \qquad .50 \qquad 1$   $x_3 = \text{Berufstatus des Befragten} \qquad 3 \qquad .48 \qquad .71$ 

empirische Untersuchung (Korrelationskoeffizienten)



postuliertes Kausalmodell (Pfadkoeffizienten)



(bei entsprechenden Voraussetzungen für die Fehlervariablen) Die Korrelation zwischen dem Berufstatus des Vaters und dem des Befragten kann kausal interpretiert werden:

$$r_{31} = .48 = .17 + .50 * .63 = .17 + .31$$

= direkter kausaler Effekt + indirekter kausaler Effekt über die Bildung des Befragten

Die Korrelation zwischen Bildung und Berufsstatus des Befragten kann kausal interpretiert werden:

$$r_{32} = .71 = .63 + .50 * .17 = .63 + .08$$

= direkter kausaler Effekt + scheinkorrelativer Effekt des Berufsstatus des Vaters

→ K. Opp / P. Schmidt: Einführung in die Mehrvariablenanalyse, Teil 4 und 5.

#### b) Typologie kausaler Effekte

#### 1. direkter kausaler Effekt

Der direkte kausale Effekt zwischen zwei Variablen ist der Pfadkoeffizient

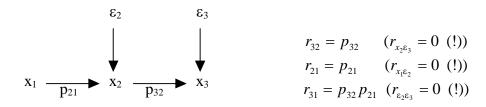
Im vollständig rekursiven Modell wird zwischen je zwei Variablen ein direkter kausaler Effekt postuliert, im unserem Fall also  $p_{21}$ ,  $p_{31}$  und  $p_{32}$ .

#### 2. indirekter kausaler Effekt

Der indirekte kausale Effekt wird durch eine intervenierende Variable erzeugt und entspricht dem Produkt der jeweiligen Pfadkoeffizienten.

In unserem Fall ist  $x_2$  die intervenierende Variable und erzeugt einen indirekten kausalen Effekt zwischen  $x_1$  und  $x_3$  von der Stärke  $p_{32}p_{21}$ 

Interpretation (reiner indirekter Effekt):  $p_{31}=0$ 



Aus dem Verschwinden des direkten kausalen Effekts ( $p_{31}$ =0) kann also eine Testgleichung abgeleitet werden:

$$\mathbf{r}_{31} = \mathbf{r}_{32}\mathbf{r}_{21}$$

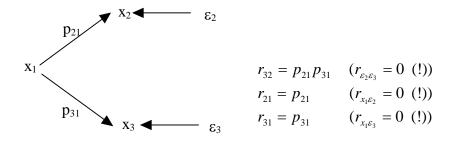
#### 3. Scheinkorrelativer Effekt

Im Gegensatz zu dem direkten und dem indirekten kausalen Effekt beschreibt der scheinkorrelative Effekt keine Kausalbeziehung zwischen den beteiligten Variablen.

Der scheinkorrelative Effekt wird durch eine antezedierende Variable hervorgerufen, die beide Variablen kausal beeinflußt. Der scheinkorrelative Effekt entspricht ebenfalls dem Produkt der zugeordneten Pfadkoeffizienten.

In unserem Fall ist  $x_1$  eine antezedierende Variable, die zwischen  $x_2$  und  $x_3$  einen scheinkorrelativen Effekt der Stärke  $p_{31}p_{21}$  erzeugt.

Erklärung (Scheinkorrelation, d.h. reiner scheinkorrelativer Effekt):  $p_{32} = 0$ 



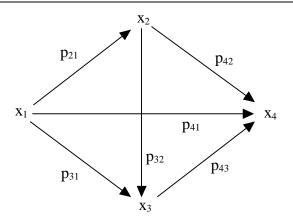
Wiederum kann aus dem Verschwinden eines direkten kausalen Effektes eine Testgleichung abgeleitet werden:

$$\mathbf{r}_{32} = \mathbf{r}_{31}\mathbf{r}_{21}$$

#### c) Das Modell mit vier Variablen

In komplexen Modellen kann eine Gesamtkorrelation sowohl direkte als auch indirekte kausale Effekte sowie darüber hinaus noch scheinkorrelative Anteile enthalten.

Gesamtkorrelation =		direkter kausaler	+	<sub>+</sub> indirekter kausaler Effekt		Schein- korrelation
		Effekt				



(<u>auf das explizite Mitführen der Fehlerterme</u> und der entsprechenden Voraussetzungen der Nicht-Korrelation der Fehlerterme mit den prädeterminierten Variablen <u>wird im folgenden verzichtet</u>)

Das Modell enthält eine zusätzliche Strukturgleichung:

$$x_4 = p_{41}x_1 + p_{42}x_2 + p_{43}x_3$$

Im folgenden soll die Effektzerlegung der Korrelation r<sub>42</sub> vorgeführt werden.

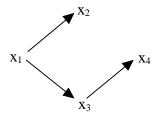
Die Multiplikation der Strukturgleichung mit x<sub>2</sub> ergibt:

$$r_{42} = p_{41}r_{21} + p_{42} + p_{43}r_{32}$$

Einsetzen der aus dem Drei-Variablen-Fall bekannten Effektzerlegungen für  $r_{21}$  und  $r_{32}$  ergibt:

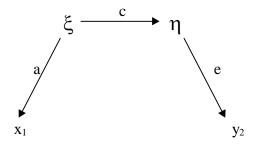
 $r_{42} = p_{42}$  direkter kausaler Effekt +  $p_{41}p_{21}$  scheinkorrelativer Effekt von  $x_1$  (antezedierende Variable) +  $p_{43}p_{32}$  indirekter kausaler Effekt über  $x_3$  (intervenierende Variable) +  $p_{43}p_{31}p_{21}$  scheinkorrelativer Effekt von  $x_1$  (antezedierende Variable)

Insbesondere der letzte scheinkorrelative Effekt ist von Interesse, da er zusätzlich einen indirekten kausalen Effekt enthält, schematisch:



Die antezedierende Variable  $x_1$  beeinflußt  $x_2$  direkt und  $x_4$  indirekt über  $x_3$ . Dadurch wird ein scheinkorrelativer Effekt zwischen  $x_2$  und  $x_4$  erzeugt.

[Im Vorgriff auf Kapitel 4.3 weisen wir darauf hin, daß dort scheinkorrelative Effekte auftreten, die die gleiche Struktur haben, z.B.:



 $r_{x_1y_2}$ =ace (scheinkorrelativer Effekt mit  $\xi$  als antezedierende Variable)]

#### 3.1.1.4 Typologie von Konstellationen bei der Drittvariablen-Kontrolle

Ausgangspunkt unserer Überlegung zur Kausalanalyse von Querschnittsdaten war eine postulierte Kausalbeziehung zwischen x und y. Eine der Bedingungen für das Vorliegen einer Kausalbeziehung ist die empirische Bestätigung einer Beziehung r<sub>xy</sub> zwischen x und y. In dem bi-

variaten Pfadmodell entspricht die Korrelation den Pfadkoeffizienten, da das bivariate Pfadmodell nur einen direkten kausalen Effekt enthält. Wie kann sich nun der direkte kausale Effekt  $p_{yx}$  bei Kontrolle einer Drittvariablen z verändern? Generell gilt:

 $r_{xy} = p_{yx} + p_{yz}p_{zx}$  : z interveniert  $r_{xy} = p_{yx} + p_{yz}p_{xz}$  : z antezediert

Es sind folgende Konstellationen denkbar (o.B.d.A. unterstellen wir  $r_{xy} > 0$ ):

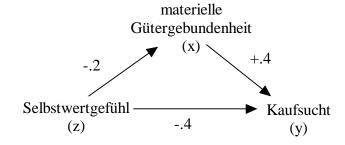
 $\underline{\text{Verst\"{a}rkung}} : \quad p_{yx} > r_{yx} \quad \text{(gegengerichteter (negativer) indirekter bzw. scheinkorrela-$ 

tiver Effekt vermindert  $r_{xy}$ )

<u>Bestätigung</u>:  $p_{yx} = r_{yx}$  (kein indirekter bzw. scheinkorrelativer Effekt)

<u>Abschwächung</u>:  $0 < p_{yx} < r_{yx}$  ( $r_{yx}$  setzt sich additiv aus einem direkten und einem gleichgerichteten (positiven) indirekten bzw. scheinkorrelativen Effekt zusammen)

Beispiel für Abschwächung (entspricht dem am häufigsten vorkommenden Fall):



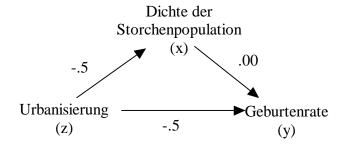
$$r_{yx} = +.4 + (-.2)(-.4) = +.48$$

$$p_{yx} = +.40$$

Der direkte Effekt ( $p_{yx}$ = +.40) ist um den (gleichgerichteten) scheinkorrelativen Effekt von +.08 gegenüber der bivariaten Korrelation von  $r_{yx}$ = +.48 abgeschwächt.

<u>Verschwinden</u>:  $p_{yx}$ =0 ( $r_{yx}$  besteht ganz aus einem (positiven) indirekten kausalen (Interpretation) oder scheinkorrelativen (Erklärung bzw. Scheinkorrelation) Effekt)

Beispiel für Scheinkorrelation:



$$p_{yx} = .00$$
 kein direkter kausaler Effekt

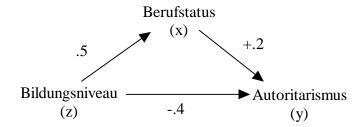
$$r_{yx} = .00 + (\text{-}.5)(\text{-}.5) = +.25 \qquad \text{die Scheinkorrelation ist allein auf den durch die Urbanisierung hervorgerufenen scheinkorrelativen Effekt zurückzuführen.}$$

<u>Umkehrung</u>:  $p_{yx} < 0 < r_{yx}$  (der (positive) indirekte kausale oder scheinkorrelative Effekt ist absolut größer als der (negative) direkte Effekt)

Von besonderem Interesse ist der Fall der Schein-Nonkorrelation:

Hier weist die bivariate Korrelation nicht auf das Bestehen einer Kausalbeziehung hin (vgl. das 3. Kausalitätskriterium). Erst bei Kontrolle einer Drittvariablen z kommt der direkte kausale Effekt zum Vorschein ( $p_{yx} \neq 0$ ).

#### Beispiel für die Schein-Nonkorrelation aus der Autoritarismusforschung:



Der Berufsstatus, d.h. eine hohe hierarchische Position, hat einen positiven direkten kausalen Effekt auf den Autoritarismus ( $p_{yx} = 0.2$ ). Andererseits sind gehobene berufliche Positionen durch Personen mit hohem Bildungsniveau besetzt ( $p_{xz} = 0.5$ ). Personen mit hohem Bildungsniveau tendieren aber zu einem geringeren Autoritarismus ( $p_{yz} = -0.4$ ), da unser Bildungssystem eher individualistisch-liberale als autoritäre Wertvorstellungen vermittelt. Daraus resultiert ein auf die antezedierende Variable Bildung zurückgehender negativer scheinkorrelativer Effekt, der den positiven direkten Effekt genau neutralisiert:

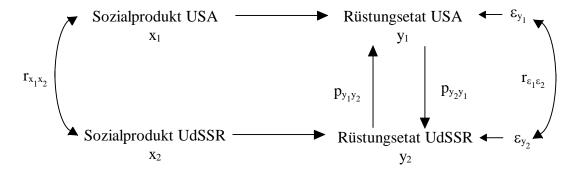
$$r_{xy} = +0.2 + (0.5)(-0.4) = 0.00$$
 (Schein-Nonkorrelation)

Formale Anregung: durch Variation von  $p_{yx}$  lassen sich leicht andere Konstellationen erzeugen, z.B.:

$$p_{yx} = +0.3 \qquad \qquad r_{xy} = +0.1 \; \text{(Verstärkung)}$$
 
$$p_{yx} = -0.2 \qquad \qquad r_{xy} = -0.4 \; \text{(Abschwächung)}$$
 
$$p_{yx} = 0.0 \qquad \qquad r_{xy} = -0.2 \; \text{(Scheinkorrelation)}$$

#### Exkurs: nichtrekursive Kausalmodelle

Das folgende Modell des Rüstungswettlaufs ist ein Beispiel für eine nichtrekursive Kausalstruktur:



Es wird unterstellt, daß die Höhe der Rüstungsausgaben der beiden Supermächte nicht nur direkt von den verfügbaren ökonomischen Ressourcen abhängen, sondern auch von den Rüstungsausgaben des jeweiligen Konkurrenten im Rüstungswettlauf. Auf diese Weise entsteht eine kausale Wechselwirkung (positive Rückkopplung) zwischen den Rüstungsausgaben, die nur von den ökonomischen Restriktionen begrenzt wird.

Wie man leicht sieht, führt die einfache Annahme unabhängiger Fehlerterme in diesem Fall zu logischen Widersprüchen. Will man z.B. mit der Methode der kleinsten Quadrate den Pfad  $p_{y_1y_2}$  schätzen, benötigt man die Annahme  $r_{y_2\epsilon_1}=0$ .  $\epsilon_1$  hat aber einen indirekten kausalen Effekt über  $y_1$  auf  $y_2$   $(p_{y_1\epsilon_1}\ p_{y_2y_1})$  und daher kann  $\epsilon_1$  nicht von  $y_2$  unabhängig sein. Es müssen daher andere Schätzverfahren verwendet werden, um die Pfadkoeffizienten des Modells zu schätzen, in diesem Fall die Methode der instrumentellen Variablen (IV) oder die zweistufige Methode der kleinsten Quadrate (2SLS) oder das Maximum-Likelihood-Verfahren (ML). (In der WINDOWS-Version von SPSS ist das LISREL-Programm von Jöreskog und Sörbon enthalten, das solche Modelle schätzen kann.) Bevor jedoch die Kausalkoeffizienten in einem nichtrekursiven Kausalmodell geschätzt werden können, muß zunächst die Frage der Identifizierbarkeit geklärt werden. In unserem Beispielmodell ist die Identifizierbarkeit der Parameter  $p_{y_1y_2}$  und  $p_{y_2y_1}$  dadurch sichergestellt, daß keine direkte kausale Wirkung von  $x_1$  auf  $y_2$  beziehungsweise von  $x_2$  auf  $y_1$  postuliert wird. Im Gegensatz zu rekursiven Kausalmodellen ist die Frage der Identifikation in nichtrekusiven Kausalmodellen oft nur mit erheblichem mathematischen Aufwand zu entscheiden. (Ein zusätzliches Problem in unserem Anwendungsbeispiel ergibt sich daraus, daß zur Schätzung der Parameter des Modells Zeitreihen über die Höhe des Sozialprodukts und der Rüstungsausgaben für die beiden Supermächte verwendet werden müssen. In diesem Fall stellt sich zusätzlich das Problem der Autokorrelation der Fehlerterme.)

## 3.1.2 Kausale Abhängigkeitsanalyse: Tabellenanalyse

[Zur Tabellenanalyse mit SPSS für Windows gibt es eine Beispieldatei (Acrobat). Diese Datei kann über das WWW bezogen werden]

Im Fall der Pfadanalyse gingen wir von linearen Beziehungen zwischen intervallskalierten Variablen aus. Nun betrachten wir den Fall von <u>dichotom</u> gemessenen Variablen X, Y, Z, ... (nominalskalierten Variablen mit genau 2 Ausprägungen; siehe 5.2). Als Maß für die

bivariate Beziehung verwenden wir die <u>Prozentsatzdifferenz</u>. Die Prozentsatzdifferenz wird nach folgender Logik berechnet:

## Gruppiere nach der unabhängigen Variablen und prozentuiere nach der abhängigen Variablen!

Dazu wird die Stichprobe zunächst nach der unabhängigen Variablen in zwei Teilgruppen aufgespalten (X=1 und X=2) und für jede Teilgruppe wird anschließend der Prozentsatz bestimmt, für den die abhängige Variable Y die Ausprägung 2 hat. Besteht eine Kausalbeziehung zwischen X und Y, so sollte eine empirische Beziehung nachweisbar sein (3. Kausalitätskriterium), d.h. der Prozentsatz in der abhängigen Variablen Y sollte in den beiden gemäß der unabhängigen Variablen X gebildeten Teilgruppen verschieden sein.

#### X,Y seien Dichotomien: $X \rightarrow Y$

	X = 1 (nein, -)	X = 2 (ja, +)	
Y = 1 (nein, -)	a	b	
Y = 2 (ja, +)	С	d	
	a+c	b+d	N

$$d\%(Y,X) = 100 \cdot \left(\frac{a}{a+c} - \frac{b}{b+d}\right)$$
$$= 100 \cdot \left(\frac{ad - bc}{(a+c)(b+d)}\right)$$

Die Prozentsatzdifferenz d%(Y,X) ist also ein <u>asymmetrisches</u> Beziehungsmaß (im Gegensatz zu Pearsons r; vgl. 3.1.1.1). Würde man umgekehrt nach Y gruppieren und nach X prozentuieren, so erhielte man ein anderes Beziehungsmaß, das zur Überprüfung der umgekehrten Kausalbeziehung  $Y \rightarrow X$  geeignet wäre:

$$d\%(X,Y) = 100 \cdot \left(\frac{ad - bc}{(a+b)(c+d)}\right)$$

#### Drittvariablenkontrolle durch Verfeinerungsanalyse (Elaboration):

Im Fall der Tabellenanalyse geschieht die Drittvariablenkontrolle durch das Aufspalten der bivariaten Tabelle in zwei <u>Partialtabellen</u> gemäß den beiden Ausprägungen der Drittvariablen Z (ebenfalls als Dichotomie vorausgesetzt). Diesen Vorgang nennt man (nach P. Lazarsfeld, dem Erfinder dieser Technik) <u>Verfeinerung</u> oder auch Elaboration. Die Ausprägung der Drittvariablen z ist also in jeder Partialtabelle konstant. Durch Konstanthalten (Kontrolle) von

Z erhalten wir daher genau zwei Partialbeziehungen, die wieder als (konditionale) Prozentsatzdifferenzen ausgedrückt werden können:

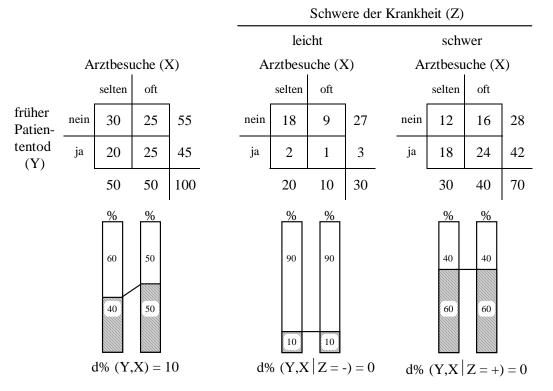
$$d\%(Y,X \mid Z = -)$$
 und  $d\%(Y,X \mid Z = +)$ .

Gemäß unserem vierten Kausalitätskriterium fordern wir, daß die beiden Partialbeziehungen nicht verschwinden dürfen. Anderenfalls würde die empirische Beziehung zwischen X und Y nur auf die Wirkung der Drittvariablen Z zurückzuführen sein, d.h. es gäbe keine direkte kausale Beziehung zwischen X und Y.

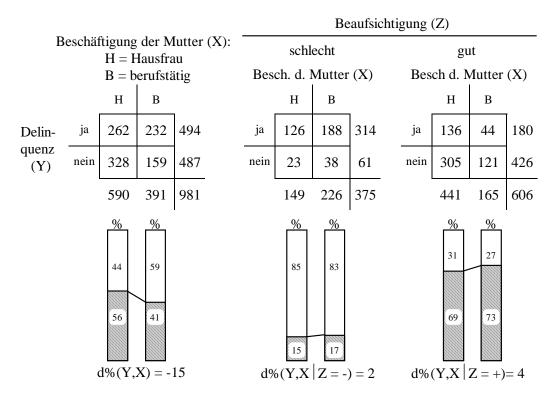
Für den Fall, daß beide Partialbeziehungen identisch sind (d%(Y,X,|Z=-)=d%(Y,X|Z=+)) entspricht die so definierte Partialbeziehung zwischen X und Y bei Kontrolle von Z logisch genau dem Pfadkoeffizienten  $p_{yx}$  in dem entsprechenden Dreivariablen-Modell der Pfadanalyse, das den direkten kausalen Effekt von X nach Y bei Kontrolle von Z beschreibt. In diesem Fall können wir die oben entwickelte Typologie von möglichen Konstellationen bei der Drittvariablenkontrolle Punkt für Punkt übertragen, wobei die bivariate Beziehung nun durch d%(Y,X) gemessen wird. Wie bei der Pfadanalyse ist also auch bei der Verfeinerungsanalyse eine Verstärkung der Beziehung möglich, wie auch eine Bestätigung, eine Abschwächung, ein Verschwinden oder sogar eine Umkehrung. Für den Fall des Verschwindens (beide Partialbeziehungen sind gleich 0) können wir wieder zwischen einer Erklärung (Scheinkorrelation) und einer Interpretation unterscheiden, je nachdem, ob die Drittvariable Z antezediert oder interveniert.

Formal hat die kausale Stellung der Drittvariablen Z (antezedierend oder intervenierend) keine Bedeutung für die Verfeinerungsanalyse. In beiden Fällen werden die Partialbeziehungen in gleicher Weise berechnet. Erst bei der Interpretation der Ergebnisse spielt die kausale Stellung von Z wieder eine entscheidende Rolle. Dazu zwei Beispiele (vgl. Benninghaus 1990, S. 282 und 1979 S. 268):

## **Erklärung** (Scheinkorrelation): Schwere der Krankheit (Z) ist antezedierende Variable



#### **Interpretation:** Beaufsichtigung (Z) ist intervenierende Variable

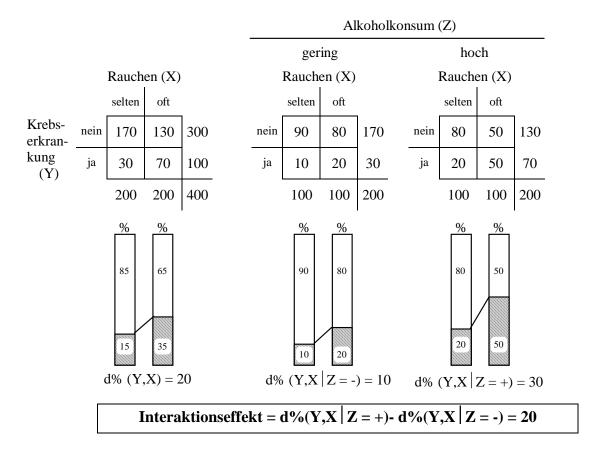


#### Interaktionseffekt:

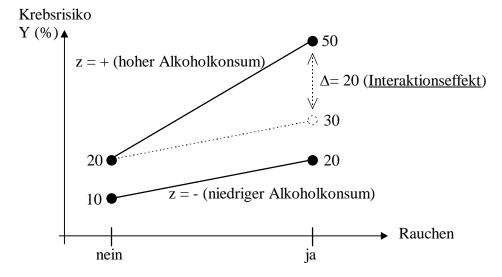
Bei der Verfeinerungsanalyse kann eine Konstellation auftreten, die in der Pfadanalyse in dieser Form nicht vorkommen kann, nämlich daß beide Partialbeziehungen (signifikant)

voneinander verschieden sind. In diesem Fall sprechen wir von einem <u>Interaktionseffekt</u>. Die Größenordnung des Interaktioneffekts entspricht der Differenz der beiden Partialbeziehungen.

Wir illustrieren das Auftreten eines Interaktionseffekts an dem folgenden (konstruierten) Beispiel, das die Höhe des Krebsrisikos (Y) in Abhängigkeit von der Häufigkeit des Rauchens (X) und des Alkoholkonsums (Z) untersucht.



Die bivariate Beziehung d%(Y,X) beträgt 20% und die beiden Partialbeziehungen 10% für geringen Alkoholkonsum und 30% für starken Alkoholkonsum. Die Erhöhung des Krebsrisikos in Folge starken Rauchens ist also abhängig davon, ob der Befragte zusätzlich einen hohen Alkoholkonsum hat oder nicht. Zur weiteren Verdeutlichung des Interaktionseffekts dient die folgende Grafik:



Kapitel 3.

Geht man von der Basisrate des Krebsrisikos von 10% aus (weder Rauchen noch Trinken), so erhöht sich das Krebsisiko für jeden Risikofaktor jeweils um 10% auf 20%. Wirken aber beide Risikofaktoren zusammen, addieren sich die beiden Einzelrisiken nicht einfach auf insgesamt 30% (gestrichelte Linie), sondern es tritt ein zusätzlicher Interaktionseffekt von 20% auf, so daß das Krebsrisiko insgesamt 50% beträgt. (Analoge Effekte können auch bei der Umweltbelastung durch Schadstoffe auftreten: Die Belastung durch das gleichzeitige Einwirken zweier Schadstoffe kann nicht einfach als additive Kombination der beiden Schädigungspotentiale berechnet werden. In diesen oft schwer oder überhaupt nicht kalkulierbaren Synergieeffekten durch das Zusammenwirken mehrerer Umweltrisiken (wobei noch zusätzlich Selbstverstärkungseffekte durch eine dynamische Wechselwirkung berücksichtigt werden müssen) liegt das eigentliche Problem bei der Bestimmung von Gefährdungspotentialen durch verschiedene Komponenten der Umweltbelastung.)

Durch das Auftreten von Interaktionseffekten verkompliziert sich unsere Typologie von Konstellationen bei der Drittvariablenkontrolle noch einmal. In allen betrachteten Fällen sind Interaktionseffekte möglich. So liegt in unserem Beispiel partiell eine Verstärkung (für Z=+) und eine Abschwächung (für Z=-) vor. Liegt ein Interaktionseffekt vor, so spricht man bei einer antezedierenden Drittvariablen Z von einer Spezifikation und bei einer intervenierenden Drittvariablen von einer Vorhersage.

#### Exkurs: Interaktionseffekte in der Pfadanalyse

Auch im Fall der Pfadanalyse können im Prinzip Interaktionseffekte modelliert werden. Dazu werden zusätzlich konstruierte Variablen in die Regressionsgleichung aufgenommen, im Falle unseres Dreivariablen-Modells etwa das Produkt der beiden Variablen  $x_1 \cdot x_2$  in die Regression von  $x_3$  auf  $x_1$  und  $x_2$  (Erweiterung der Strukturgleichung). Allerdings macht sich dabei häufig das Problem der <u>Multikolliniarität</u> unangenehm bemerkbar, da die künstlich gebildete Variable  $x_1 \cdot x_2$  gewöhnlich sowohl mit  $x_1$  wie auch mit  $x_2$  hoch korreliert ist. In diesem Fall werden die Schätzungen der Regressionskoeffizienten instabil.

Für den Fall, daß die beiden (oder auch mehrere) unabhängigen Variablen nominalskaliert sind (Faktoren) und die abhängige Variable intervallskaliert ist, bietet sich eine Varianzanalyse an, um mögliche Interaktionseffekte zwischen den beiden Faktoren (nominalskalierte unabhängige Variablen) zu berechnen. Im Fall eines Experiments ist es darüber hinaus möglich, in einem mehrfaktoriellen Design (vgl. auch 3.3) für gleiche Besetzungszahlen in den verschiedenen Experimentalgruppen zu sorgen. Ist dies der Fall, so liegen orthogonale (unkorrelierte) Faktoren vor und die Varianzzerlegung ist eindeutig möglich: Gesamtvarianz = Varianz der beiden Hauptfaktoren A, B + Varianz der Interaktion von A und B + Residualvarianz. In Ex-post-facto-Untersuchungen werden die beiden Faktoren A und B im allgemeinen korreliert sein, so daß eine eindeutige Varianzzerlegung unmöglich ist (vgl. auch den Exkurs zur Varianzzerlegung im Rahmen der Pfadanalyse).

Eine Erweiterung der Verfeinerungsanalyse stellt die <u>loglineare Tabellenanalyse</u> dar. Das loglineare Modell ermöglich die Analyse mehrdimensionaler Kontingenztabellen im Rahmen des <u>allgemeinen linearen Modells</u>. Dazu werden die logarithmierten Zellhäufigkeiten als abhängige Variable betrachtet. Unabhängige Variablen sind die nominalskalierten Merkmale, die den verschiedenen Dimensionen der Tabelle zugrunde liegen und die als <u>Dummy-Variablen</u> codiert werden. Auf der Grundlage geeigneter statistischer Modelle (Poisson-Verteilung der Residuen) können die Koeffizienten des Modells als Regressionskoeffizienten der zugeordneten Dummy-Variablen geschätzt werden. Auf diese Weise können in einem multivariaten Modell sowohl die partiellen Beziehungen zwischen den Dimensionen der Tabelle als auch deren Interaktionseffekte und Wechselwirkungen höherer Ordnungen bestimmt und auf Signifikanz geprüft werden. Zusätzlich sind Modelltests möglich, die sich an einem Maß orientieren, das als Analogon für die erklärte Varianz im Regressionsmodell verstanden werden kann.

## 3.2 Panel - Design (Kausale Analyse II)

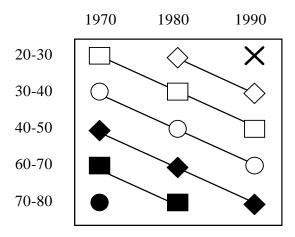
Bei einer <u>Trenduntersuchung</u> handelt es sich um eine replikative Studie (unechtes Panel), in der die gleichen Variablen mit den gleichen Operationalisierungen erhoben werden (exakt identische Fragestellung), allerdings an <u>anderen</u> Personen, die jedoch mit der gleichen Auswahlprozedur ermittelt wurden. So enthalten z.B. die ALLBUS-Studien der Jahre 1980 bis 1990 die gleichen Fragen zur Messung der Einstellung gegenüber Gastarbeitern (in allen Fällen handelt es sich um repräsentative Bevölkerungsstichproben, die nach dem gleichen dreistufigen Auswahlverfahren ermittelt wurden (vgl. 7.3)). Es zeigte sich folgender Trend (vgl. Wiegand 1991):

Ausländerfeindlichkeit in Deutschland (Prozentsatz, der den jeweiligen Items zumindest weitgehend zustimmt) in den Jahren 1980 bis 1990:

		1980	1984	1988	1990
1.	Gastarbeiter sollten ihren <u>Le-bensstil</u> ein bißchen besser an den der Deutschen anpassen.	45	42	42	34
2.	Man sollte Gastarbeitern jede politische Tätigkeit in Deutschland untersagen.	39	38	29	27
3.	Wenn <u>Arbeitsplätze</u> knapp werden, sollte man die Gastarbeiter wieder in ihre Heimat zurückschicken.	38	29	25	20
4.	Gastarbeiter sollten sich ihre Ehepartner unter den eigenen Landsleuten auswählen.	33	26	23	18

Für alle Items ist ein gleichgerichteter Trend hin zu einer abnehmenden Ausländerfeindlichkeit zu konstatieren. Solche Trends könnten mit dem vorhandenen Datenmaterial auch subgruppenspezifisch aufgebrochen werden, z.B. nach Bildung oder Geschlecht. Auf diese Weise läßt sich feststellen, ob der Trend alle Bevölkerungsgruppen in gleicher Weise erfaßt hat, oder ob er für manche Gruppen verstärkt, für andere aber nur abgeschwächt zutrifft.

Insbesondere können auch <u>Kohortenanalysen</u> durchgeführt werden (zur Einführung in die Kohortenanalyse vgl. Glenn 1977), d.h. man kann überprüfen, wie sich die Trends in den einzelnen Altersgruppen (Kohorten) entwickelt haben, schematisch:



In der Kohortenanalyse unterscheidet man zwischen:

• <u>Lebenszykluseffekten</u> (Alterseffekten): Ein reiner Lebenszykluseffekt liegt vor, wenn jede Kohorte die Einstellung der jeweils älteren Kohorte übernimmt, wenn sie die entsprechende Alterskategorie erreicht (z.B. sind die 20-30jährigen von 1970 im Jahre

1990 genauso konservativ, wie die 40-50jährigen von 1970). Die Gesamtverteilung ändert sich also nicht, die Altersabhängigkeit ist über den gesamten Erhebungszeitraum konstant (konstante Reihenprozentsätze).

- <u>Periodeneffekte</u>: Reine Periodeneffekte betreffen Veränderungen, die zu einem Erhebungszeitpunkt (Periode) für alle Kohorten in gleicher Weise zutreffen, z.B. wenn alle Kohorten im Jahre 1990 um 15% weniger ausländerfeindlich sind, als im Jahre 1980 (spaltenweise konstante Effekte).
- Kohorteneffekte: Ein reiner Kohorteneffekt liegt vor, wenn die Kohorte im Zeitverlauf ihre Einstellung beibehält (konstante Prozentsätze in der Diagonale). Sind die jüngeren Kohorten jeweils weniger ausländerfeindlich als die alten, so führt dies im Zeitverlauf zu einer Abnahme der Ausländerfeindlichkeit in der Gesamtbevölkerung, da die alten (relativ ausländerfeindlichen) Kohorten ausscheiden und die jungen (weniger ausländerfeindlichen) Kohorten nachwachsen. Es würde also, wie bei einem Periodeneffekt, ebenfalls ein Trend hin zu abnehmender Ausländerfeindlichkeit zu beobachten sein, allerdings auf der Grundlage eines anderen dynamischen Prozesses.

Trenduntersuchungen gestatten also durchaus differenzierte Analysen zeitlicher Abläufe (zur Einführung in die Zeitreihenanalyse vgl. ergänzend Ostrom 1990). Eines ist aber bei Trenduntersuchungen prinzipiell unmöglich, nämlich die Untersuchung <u>individueller Veränderungen</u>. Dazu sind echte <u>Paneluntersuchungen</u> notwendig, bei denen, im Gegensatz zu Trendstudien, zu verschiedenen Zeitpunkten in gleichen Variablen mit den gleichen Operationalisierungen an den <u>gleichen Personen</u> gemessen werden. Damit ist das erste Kausalitätskriterium (zeitliche Reihenfolge der Variablen) erfüllt, d.h. die Veränderungen in einer abhängigen Variablen y ( $\Delta y = y_2 - y_1$ ) können auf eine zeitlich vorangehende Ursache x (den zum Zeitpunkt 1 gemessenen Wert  $x_1$ ) zurückgeführt werden. Im Unterschied zum Querschnittsdesign sind damit auch Aussagen über die kausale (zeitliche) Reihenfolge von Variablen empirisch überprüfbar.

Grundlegend für die Panelanalyse ist die Möglichkeit, in einer <u>Paneltabelle</u> neben den Nettoveränderungen (Veränderungen in den Randverteilungen) auch individuelle Veränderungen (interne Fluktuationen, d.h. Veränderungen innerhalb der Tabelle) zu erfassen. Dazu das folgende hypothetische Beispiel, das einen Markenwechsel untersucht.

	$\mathbf{t_1}$				
		Marke 1	Marke 2		
$t_2$	Marke 1	30	10	40	
	Marke 2	20	40	60	
	•	50	50	-	

Nettoveränderung 10% (zugunsten von Marke 2):

Stabile: 70%

Wechsler: 30%

Aus den Randverteilungen der Paneltabelle (Nettoveränderungen, die auch durch eine Trendstudie hätten ermittelt werden können) kann nicht auf die interne Fluktuation in der Tabelle geschlossen werden, wie die folgenden konstruierten Tabellen mit den gleichen Randverteilungen (Nettoveränderungen) zeigen:

40	-	40
10	50	60
50	50	

- 40 40 50 10 60 50 50

Stabile: 90%

Stabile: 10%

Wechsler: 10%

Wechsler: 90%

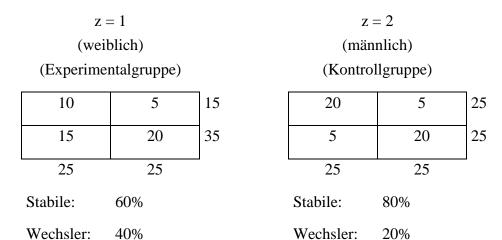
Bei der Panelanalyse unterscheiden wir drei verschiedene Arten von Variablen:

- <u>Konstante</u> Variablen: Variablen, die über die Befragungszeitpunkte (Wellen) hinweg gleichbleiben, wie z.B. Geschlecht usw.
- <u>Intermittierende</u> Variablen: Variablen, die zwischen den Befragungszeitpunkten auftreten, und denen nur ein Teil der Befragen ausgesetzt ist (z.B. ein unkontrolliertes Ereignis oder ein experimenteller Stimulus; vgl. auch 3.3).
- <u>kovariierende</u> Variablen: Mindestens zwei Variablen x und y, die sich zwischen den Befragungswellen ändern und sich dabei kausal beeinflussen, wobei die Art der kausalen Beeinflussung zunächst unspezifiziert bleibt.

#### Die Einbeziehung von Drittvariablen in der Panelanalyse:

Handelt es sich bei der Drittvariablen um eine konstante oder intermittierende Variable, so ist eine Drittvariablenkontrolle durch Elaboration leicht durchzuführen (vgl. 3.1.2). So könnte die obige Paneltabelle der Markenwechsler nach der konstanten Variable Geschlecht bzw. nach einer experimentellen Variable (Experimentalgruppe mit Werbemaßnahme für Marke 1, Kontrollgruppe ohne Werbemaßnahme) in die folgenden Partialtabellen aufgespalten werden:

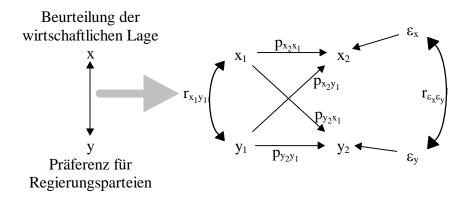
Drittvariable z (Geschlecht bzw. experimentelle Variable)



Im Falle der konstanten Drittvariablen Geschlecht zeigt die Analyse, daß der Wechsel zu Marke 2 allein auf die weiblichen Konsumenten zurückzuführen ist. Im Falle eines Experiments kann gefolgert werden, daß der Wechsel zu Marke 2 allein auf die Werbemaßnahme in der Experimantalgruppe zurückzuführen ist.

Am Beispiel der Tabelle für z=2 ist zu erkennen, daß auch bei stabilen Randverteilungen (keine Nettoveränderung) durchaus interne Fluktuationen innerhalb der Tabelle auftreten können. (Im Falle einer <u>Test-Retest-Situation</u> im Rahmen der Klassischen Testtheorie kann eine solche Tabelle zur Bestimmung der <u>Reliabilität</u> verwendet werden - perfekte Reliabilität (= keine Zufallsfehler) = keine Wechsler; vgl. 5.4.1.)

Komplexer ist die Situation im Falle kovariierender Variablen x und y. Für die klassische Analyse mit Hilfe der 16-Felder-Tabelle bei dichotomen Variablen  $(x_1 * y_1 * x_2 * y_2)$  verweisen wir auf Arminger (1976, S. 153ff). Im folgenden behandeln wir den pfadanalytischen Ansatz, wie er von Heise (1970) entwickelt wurde. Dabei wird die zunächst nicht genauer definierte Wechselwirkung zweier kovariierender Variablen x und y als zeitverzögerte Kausalbeziehung von  $x_1$  auf  $y_2$  und von  $y_1$  auf  $x_2$  spezifiziert:



Zwischen  $x_1$  und  $y_1$  besteht eine (kausal nicht genauer interpretierte) Korrelation von  $r_{x_1y_1}$ .  $p_{x_2x_1}$  und  $p_{y_2y_1}$  messen die zeitliche Stabilität von x und y, während  $p_{x_2y_1}$  und  $p_{y_2x_1}$  die

zeitverzögerte Wechselwirkung zwischen den beiden Variablen messen. Ist nur einer der beiden Koeffizienten (signifikant) von Null verschieden, liegt eine einseitige Kausalbeziehung vor, andernfalls eine zeitverzögerte kausale Wechselwirkung. Die Fehlerterme  $\varepsilon_x$  und  $\varepsilon_y$  sind mit den prädeterminierten Variablen  $x_1$  und  $y_1$  unkorreliert, dagegen wird eine Korrelation zwischen  $\varepsilon_x$  und  $\varepsilon_y$  ( $r_{\varepsilon_x \varepsilon_y}$ ) zugelassen. Über die üblichen Voraussetzungen der Pfadanalyse hinaus sind bei dem Modell der zeitverzögerten Wechselwirkungen zwei Annahmen besonders problematisch:

- Die Zeitverzögerungen in der kausalen Wirkung von  $x_1$  auf  $y_2$  und von  $y_1$  auf  $x_2$  sind gleich und entsprechen genau dem Zeitintervall  $\Delta t = t_2 t_1$  zwischen den beiden Panelwellen.
- Die Annahme, daß die Fehlerterme ε<sub>x</sub> und ε<sub>y</sub> mit den prädeterminierten Variablen unkorreliert sind, ist im Falle zeitverzögerter Variablen besonders schwer zu erfüllen. Insbesondere muß jede Variable, die mit dem Kausalsystem in Wechselwirkung steht und jede stabile Variable, die auf das System wirkt, explizit modelliert werden, um Spezifikationsfehler zu vermeiden.

Spezielle Probleme des Panel-Designs sind:

- Anzahl und Zeitpunkt der Panelwellen: Definitionsgemäß werden bei Panelstudien mindestens zwei Befragungswellen durchgeführt. Werden soziale Prozesse über einen längeren Zeitraum begleitet, sind auch 3- bis 7-wellige Panel (z.B. Wahlpanel) oder sogar wöchentliche Panel über einen längeren Zeitraum (insbesondere in der Konsumforschung) üblich. Um zwischen mangelnder Reliabilität (Zufallsfehler der Messung im Sinne der Klassischen Testtheorie) und tatsächlichen Veränderungen unterscheiden zu können, sind mindestens drei Wellen erforderlich. Wie oben in Bezug auf das Pfadmodell mit zeitverzögerten Variablen erläutert, muß das Zeitintervall zwischen den Panelwellen im Hinblick auf die zu untersuchenden sozialen Prozesse aufgrund theoretischer Überlegungen bestimmt werden.
- Anlage der Stichprobe und <u>Panelsterblichkeit</u>: Da im Panel die gleichen Personen wiederholt befragt werden, entstehen sog. verbundene Stichproben, für die sich die Stichprobenvarianz der Differenz zweier Stichprobenmittelwerte wie folgt reduziert:

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 - 2\rho_{\bar{x}_1\bar{x}_2}\sigma_{\bar{x}_1}\sigma_{\bar{x}_2}$$

Durch die reduzierte Stichprobenvarianz sind Mittelwertdifferenzen in verbundenen Stichproben eher signifikant (t-Test für verbundene Stichproben).

Das größte Problem bei Paneluntersuchungen ist es, die Untersuchungspersonen für die wiederholte Teilnahme an den Befragungen im Verlauf des Panels zu motivieren. Die typischerweise bei Panels auftretende hohe Ausfallquote (<u>Panelmortalität</u>) hat verschiedene Ursachen: Verlust der Teilnahmemotivation, Wohnortwechsel, Krankheit, usw.. Eine antezipierte hohe Ausfallquote kann durch eine entsprechende Vergrößerung der

Stichprobe der ersten Welle kompensiert werden - allerdings nur, wenn die Ausfälle unsystematisch erfolgen (was nicht wahrscheinlich ist). Langwellige Konsumentenpanel, die üblicherweise auf großen Quotenstichproben beruhen, umgehen das Problem der Panelmortalität dadurch, daß Ausfälle durch sog. statistische Zwillinge ersetzt werden (Befragte mit den gleichen Quotenmerkmalen). Darüberhinaus wird versucht, durch verschiedene Anreize (Geldprämien, Preise, usw.) die Teilnahmemotivation aufrecht zu erhalten.

• Paneleffekte: Paneleffekte betreffen den Einfluß einer wiederholten Befragung auf das Antwortverhalten der Panelteilnehmer. Dazu gehören Effekte, wie die reaktive Ausbildung von vor der Befragung noch nicht bestehenden Einstellungen, die Verfestigung bestehender Einstellungen und die Anpassung an die Befragungssituation im Sinne sokratischer Effekte (vgl. 3.3), aber auch im Sinne von Antwortverzerrungen z.B. durch Interviewereffekte (vgl. 6.2).

Gerade in der Marktforschung spielen Konsumentenpanel eine herausragende Rolle. Oft werden sie zu sog. experimentellen Mikrotestmärkten ausgebaut, wie z.B. das GfK-Behaviorscan-Testpanel in Haßloch. Haßloch ist eine kleine, ca. 10.000 Haushalte umfassende, geographisch relativ isolierte Gemeinde, die optimal verkabelt ist. Das Panel besteht aus 3.000 repräsentativen Testhaushalten. Jeder Haushalt besitzt eine Identifikationskarte, mit deren Hilfe Warenkäufe an den Scannerkassen der lokalen Geschäfte direkt erfaßt werden können. Die Testhaushalte erhalten Gratiszeitschriften und sind dadurch gezielt für Werbebeilagen ansprechbar. Darüberhinaus kann eine Teilstichprobe von 2.000 Testhaushalten, die mit sog. GfK-Boxen ausgestattet sind, spezielle TV-Werbesendungen empfangen, die über Kabel in die üblichen Werbeblöcke der kommerziellen Sender eingespeist werden. In regelmäßigen Befragungen werden zusätzliche Informationen erhoben, die für das Konsumentenverhalten der Testhaushalte von Bedeutung sind. Das experimentelle Design des GfK-Behaviorscan-Panels wird überwiegend genutzt, um die Wirksamkeit von Werbemaßnahmen zu überprüfen und die Chancen und Risiken, die mit einer Neuprodukteinführung bzw. einem Relaunch verbunden sind, zu quantifizieren.

## 3.3 Experimentelle Designs (Kausale Analyse III)

Das <u>echte Experiment</u> ist die ideale Untersuchungsform zur Überprüfung kausaler Hypothesen. Im Gegensatz zur Ex-post-facto-Anordnung (Querschnittsdesign) erlaubt das Experiment <u>die Einführung des experimentellen Stimulus (unabhängige Variable) durch den Forscher unter kontrollierten Bedingungen.</u> Damit ist zum einen die kausale Reihenfolge eindeutig bestimmt und zum anderen ist eine <u>explizite Kontrolle von Störfaktoren</u> möglich (im Gegensatz zur nachträglichen Drittvariablenkontrolle im Querschnittsdesign).

Dennoch muß die Gültigkeit (Validität) der kausalen Schlußfolgerung auch im Experiment immer kritisch hinterfragt werden. Eine Untersuchung heißt

- intern valide (iv), wenn die kausale Schlußfolgerung keine Alternativerklärung zuläßt,
- <u>extern valide</u> (ev), wenn die kausale Schlußfolgerung über die besonderen Bedingungen der Untersuchung und über den untersuchten Personenkreis hinaus verallgemeinerbar ist.

Da eine experimentelle Kontrolle von Störfaktoren im Labor besonders gut durchführbar ist, gilt das Laborexperiment im allgemeinen als intern valide. Andererseits können Störeffekte gerade durch die experimentelle Laborsituation auftreten (z.B. der Aufforderungscharakter der Experimentalsituation; siehe die Diskussion von Forschungsartefakten im Experiment weiter unten), die die externe Validität des Laborexperiments einschränken. Umgekehrt wird dem Feldexperiment, das in einer für die Untersuchungspersonen gewohnten Umgebung stattfindet, eine tendenziell höhere externe Validität zugesprochen, da reaktive Effekte aufgrund der künstlichen Laborsituation nicht auftreten können. Sofern auch im Feldexperiment eine explizite Kontrolle von Störfaktoren möglich ist, kann auch die interne Validität weitgehend gewährleistet werden. Im Gegensatz dazu verfügen Ex-post-facto-Anordnungen wie z.B. das Querschnittsdesign generell über eine geringe interne Validität, da die auf einer solchen Untersuchung basierenden kausalen Schlußfolgerungen immer auf einer schwachen empirischen Basis beruhen (vgl. die ausführliche Diskussion in 3.1). Zusammenfassend kann die interne und externe Validität von Experiment und Ex-post-facto-Anordnung im Labor und im Feld schematisch wie folgt berurteilt werden:

	Ex-post-facto Anordnung		Exp	eriment
Labor	iv: ev:	-	iv: ev:	+
Feld	iv: ev:	- +	iv: ev:	+ +

#### 3.3.1 Kontrolle von Störfaktoren

Das <u>Quasiexperiment</u> unterscheidet sich vom Experiment dadurch, daß der experimentelle Stimulus nicht explizit unter kontrollierten Bedingungen vom Forscher in die Untersuchungssituation eingeführt wird. Der Sozialforscher begleitet lediglich die Einführung einer Innovation, wie z.B. einem neuen Lehrprogramm in einer Schule, in einer natürlichen Situation und versucht die daraus resultierenden Veränderungen systematisch zu erfassen. Je stärker dabei eine explizite Kontrolle von Störfaktoren durch den Forscher möglich ist, desto mehr nähert sich das Untersuchungsdesign einem echten Experiment (Feldexperiment). Um im Experiment die Veränderung in der abhängigen Variablen kausal auf die Wirkung des

experimentellen Stimulus (unabhängige Variable) zurückführen zu können, müssen Alternativerklärungen, die auf dem Einfluß möglicher Störfaktoren beruhen, möglichst weitgehend ausgeschlossen werden können. Mögliche Störfaktoren im Experiment sind:

- Personengebundene Störfaktoren: Jede Versuchsperson ist durch eine spezifische Kombination von Persönlichkeitsmerkmalen charakterisiert, die alle neben dem experimentellen Stimulus einen Einfluß auf die experimentelle Variable (abhängige Variable) haben können. Im Sinne der experimentellen Logik handelt es sich also um Störfaktoren, die durch geeignete Techniken, insbesondere Matching und Randomisierung, kontrolliert werden müssen.
- Zwischenzeitliches Geschehen: Erstreckt sich ein Experiment über einen längeren Zeitraum, so können äußere Ereignisse auf den Ablauf des Experiments einwirken und einen Einfluß auf die experimentelle Variable ausüben. Um die kausale Wirkung solcher unkontrollierter Ereignisse auf die experimentelle Variable beurteilen zu können, muß das experimentelle Design um eine Kontrollgruppe ergänzt werden, die bis auf den experimentellen Stimulus (Experimentalgruppe) den gleichen Einflußfaktoren ausgesetzt ist. Der Interaktionseffekt von experimentellem Stimulus und unkontrollierten Ereignissen (I<sub>EU</sub>) kann grundsätzlich nicht kontrolliert werden (siehe 3.3.2).
- Reifungsprozesse der Versuchspersonen: Veränderungen in der experimentellen Variablen können auch durch unabhängig vom experimentellen Stimulus ablaufende interne Veränderungsprozesse der Versuchspersonen verursacht werden. Erstreckt sich ein Lernexperiment z.B. über ein ganzes Schuljahr, so wird ein Teil der gemessenen Lernerfolge auf die natürliche biologisch-psychologische Reifung der Schüler zurückzuführen sein. Auch hier ist der Vergleich mit einer Kontrollgruppe, die den gleichen Reifungsprozessen unterliegt, unbedingt erforderlich.
- <u>Meßeffekte</u>: Wird in einem Experiment die experimentelle Variable zweimal gemessen, nämlich einmal vor Einführung des experimentellen Stimulus (Pretest) und zum zweiten mal danach (Posttest), so muß mit einem Einfluß der ersten Messung auf den Posttest und damit auf die experimentelle Variable gerechnet werden (vgl. auch die Ausführungen zum Paneleffekt). Auch zur Kontrolle dieses Pretesteffektes ist eine Kontrollgruppe unbedingt erforderlich.
- Versuchsleitereffekte: Der Versuchsleiter ist Teil der experimentellen Situation und beeinflußt daher das Verhalten der Versuchspersonen in oft nur schwer einschätzbarer Weise. Die daraus resultierenden Forschungsartefakte (die durch den Versuchsleiter hervorgerufenen Veränderungen in der experimentellen Variablen werden irrtümlich auf den experimentellen Stimulus zurückgeführt) können eine ernsthafte Bedrohung der internen Validität eines Experiments darstellen (vgl. die eingehende Diskussion von Forschungsartefakten im Experiment weiter unten).

- Veränderungen der experimentellen Bedingungen: Werden während des Ablaufs eines Experiments die Meßinstrumente verändert oder etwa die Versuchsleiter ausgetauscht, so wird dies im allgemeinen Auswirkungen auf die experimentelle Variable haben. Es ist daher unabdingbar, daß das Experiment unter konstanten experimentellen Bedingungen durchgeführt wird.
- <u>Ausfälle von Versuchspersonen:</u> Scheiden während des Experiments Versuchspersonen aus, kann dies ebenfalls zu verzerrten Ergebnissen führen, nämlich immer dann, wenn diese Ausfälle systematisch erfolgen. Dies wäre z.B. der Fall, wenn in einem Experiment zur Überprüfung einer neuen Lehrmethode gerade die Schüler den Test abbrechen, die mit dieser neuen Lehrmethode nicht zurecht kommen.

Die angeführten Störfaktoren stellen eine Bedrohung der internen Validität eines Experiments dar, da es sich in allen Fällen um kausale Einflußfaktoren handelt, die möglicherweise neben dem experimentellen Stimulus eine Wirkung auf die experimentelle Variable haben könnten. Weitere Störfaktoren, die insbesondere die externe Validität des Experiments bedrohen, indem sie die Verallgemeinerung des experimentellen Ergebnisses auf andere Personengruppen und andere situative Kontexte einschränken, sind:

- Reaktive Effekte des Messens (Interaktionseffekt von Pretest und experimentellem Stimulus): Durch die Durchführung eines Pretestes kann die Art und Weise, wie die Versuchspersonen auf den experimentellen Stimulus reagieren, beeinflußt werden. Die Veränderung in der experimentellen Variable ist dann zwar auf den experimentellen Stimulus zurückzuführen, aber nur unter der Voraussetzung der durch den Pretest hervorgerufenen Sensibilisierung der Versuchspersonen (Interaktionseffekt I<sub>PE</sub>; siehe 3.3.2).
- Reaktive Effekte der experimentellen Situation (Aufforderungscharakter der Experimentalsituation): Die experimentelle Situation mit ihren oft artifiziellen, von Alltagssituationen
  abweichenden Bedingungen, kann bei den Versuchspersonen zu Reaktionen führen, die die
  Wirkung des experimentellen Stimulus überlagern. So mag die Versuchsperson überlegen,
  was der Versuchsleiter von ihr erwartet und welchem Zweck das Experiment dient. Veränderungen in der experimentellen Variable sind dann eher auf die Reaktionen der Versuchspersonen auf die experimentelle Situation (Meerschweincheneffekt) zurückzuführen,
  als auf die Wirkung des experimentellen Stimulus. Neben den Versuchsleitereffekten stellt
  der Aufforderungscharakter der experimentellen Situation die zweite Hauptquelle von
  Forschungsartefakten im Experiment dar (siehe unten).

Es gibt verschiedene Techniken zur Kontrolle von Störfaktoren:

• <u>Elimination:</u> Darunter versteht man die Ausschaltung denkbarer Störgrößen. Insbesondere im Labor ist die Ausschaltung vieler Störgrößen möglich, da hier die Kontrolle der experimentellen Situation am größten ist. Gerade dadurch erhöht sich aber andererseits die Reaktivität des Experiments, so daß die Gefahr von Forschungsartefakten zunimmt.

• <u>Konstanthalten:</u> Nicht alle Störfaktoren können eliminiert werden (so z.B. das zwischenzeitliche Geschehen und Reifungsprozesse bei den Versuchspersonen). Bildet man neben der Experimentalgruppe eine Kontrollgruppe, die den gleichen Einflüssen unterworfen ist, werden die Störfaktoren "konstant gehalten", d.h. sie wirken in beiden Gruppen in gleicher Weise. Durch Vergleichen der Experimental- und Kontrollgruppe kann dann die Wirkung des experimentellen Stimulus bestimmt werden.

Die Bildung von Kontrollgruppen kann auf zweierlei Weise geschehen:

- Matching: Es wird sichergestellt, daß Experimental- und Kontrollgruppe bezüglich bestimmter Merkmale vergleichbar sind. Jedes Mitglied der Experimentalgruppe hat also in der Kontrollgruppe genau einen "statistischen Zwilling", allerdings nur in bezug auf die ausgewählten Merkmale (z.B. Alter und Bildung) der Versuchsperson (paarweises Matching). Das Matching entspricht einer Kontrolle der entsprechenden personengebundenen Störfaktoren bereits durch die experimentelle Versuchsanordnung (im Gegensatz zum Querschnittsdesign, wo nur eine nachträgliche statistische Drittvariablenkontrolle möglich ist).
- Randomisierung: Wie bei der Drittvariablenkontrolle besteht das Hauptproblem des Matching darin, von vorneherein die personengebundenen Störfaktoren zu bestimmen, die für das Experiment von Relevanz sind, d.h. die Variablen, die neben dem experimentellen Stimulus einen Einfluß auf die experimentelle Variable haben können. (Darüberhinaus wird das Matching bei einer Vielzahl solcher zu kontrollierenden Drittvariablen so komplex, daß die Durchführung nicht mehr praktikabel ist.) Einen Ausweg bietet hier die sog. Randomisierung, d.h. die zufällige Zuweisung der Versuchspersonen in die Experimental- und die Kontrollgruppe. Dadurch wird erreicht, daß Experimental- und Kontrollgruppe bis auf Zufallsschwankungen in allen Merkmalen übereinstimmen systematische Unterschiede zwischen den beiden Gruppen sind dadurch ausgeschlossen. Zur Verdeutlichung sei gesagt, daß der Prozeß der zufälligen Zuordnung nicht impliziert, daß Experimental- und Kontrollgruppe für eine wie auch immer definierte Grundgesamtheit repräsentativ sind. Es wird lediglich gefordert, daß die Menge aller Versuchspersonen (wie auch immer zusammengesetzt) durch einen Zufallsmechanismus der Experimental- und Kontrollgruppe zugeordnet werden.

# 3.3.2 Experimentelle Designs und Effekte auf die experimentelle Variable

Die beiden klassischen experimentellen Designs sind:

• Die Vorher-Nachher-Messung mit Kontrollgruppe und Randomisierung (R), schematisch:

		Pretest (P)	exp. Stimulus (E)	Posttest (T)
(R)	E:	$Y_1^E$	X	$Y_2^E$
(R)	K:	$Y_1^K$		$Y_2^K$

• Die Nachher-Messung mit Kontrollgruppe und Randomisierung:

		exp. Stimulus (E)	Posttest (T)
(R)	E:	X	$Y_2^E$
(R)	K:		$Y_2^K$

Um die Wirkung der experimentellen Bedingungen auf die experimentelle (abhängige) Variable systematisch darstellen zu können, unterscheiden wir:

p: = Wirkung des Pretests (P) auf die experimentelle Variable

e: = Wirkung des experimentellen Stimulus (E) auf die experimentelle Variable (nur in der Experimentalgruppe).

u: = Wirkung der unkontrollierten Ereignisse (U) auf die experimentelle Variable (kann nie ausgeschlossen werden)

Die Veränderung d=T - P (Nachher- minus Vorher-Messung) kann daher systematisch auf die Wirkung der oben spezifizierten Effekte zurückgeführt werden. Das Bild kompliziert sich aber dadurch, daß auch die Interaktionseffekte  $I_{PE}$ ,  $I_{PU}$ ,  $I_{EU}$ ,  $I_{PEU}$  der experimentellen Faktoren berücksichtigt werden müssen. Schematisch:

Zur Erläuterung der Wirkungsweise möglicher Interaktionseffekte skizzieren wir im folgenden kurz ein Experiment zur Bestimmung der Langzeitwirkung zweier Tageszeitungen auf die politischen Einstellungen der Leser (vgl. Boden u.a. 1975). Vom Design her handelt es sich dabei um eine Vorher-Nachher-Messung mit zwei Experimentalgruppen E<sub>1</sub> und E<sub>2</sub> und einer Kontrollgruppe K, die durch Randomisierung gebildet wurden. Das Experiment

wurde in den Jahren 1971/72 durchgeführt und erstreckte sich über den Zeitraum eines Jahres. Der experimentelle Stimulus war die einjährige, nicht selbst veranlaßte Lieferung entweder der konservativen Tageszeitung "Die Welt" (E1) oder der linksliberalen Tageszeitung "Frankfurter Rundschau" (E2); den Versuchspersonen in der Kontrollgruppe wurde keine der beiden Tageszeitungen zugestellt. (Daß die beiden Tageszeitungen tatsächlich unterschiedliche politische Inhalte transportieren, wurde durch eine Befragung der Journalisten, eine Inhaltsansalyse der Leitartikel und durch eine Befragung von Personen, die sich voll mit "ihrer" Zeitung identifizieren, abgesichert.) Die Versuchspersonen waren Studenten, die zusätzlich nach Fachrichtungen (Wirtschaftswissenschaften Ingenieurwissenschaften (I)) unterschieden wurden. Danach erhält man das folgende mehrfaktorielle Design mit den beiden Faktoren "Tageszeitung" und "Fachrichtung":

		Fachrichtung		
		$\mathbf{W}$	I	
	"Die Welt"	$E_{1W}$	$E_{1I}$	$E_1$
	Welt"			
Tages-	"Frankfurter	$E_{2W}$	$E_{2I}$	$E_2$
zeitung	Rundschau"			
	keine	$K_{\mathrm{W}}$	K <sub>I</sub>	K

Für das folgende Argument vernachlässigen wir zunächst den zweiten Faktor (Fachrichtung) und konzentrieren uns ganz auf die Wirkung der Tageszeitungen (Gruppen E<sub>1</sub>, E<sub>2</sub> und K) auf die politischen Einstellungen der Leser. Die politischen Einstellungen wurden durch die folgenden Likertskalen (vgl. 5.3) erfaßt:

- Beurteilung des marktwirtschaftlichen Systems (M)
- Beurteilung staatlicher Ordnungsmaßnahmen (S)
- Beurteilung der Gewerkschaften und ihrer Politik (G)
- Beurteilung der westlichen Wirtschafts- und Gesellschaftsordnung (W)
- Beurteilung der Ostpolitik der Bundesregierung (O).

Bei der Befragung handelte es sich um eine Panelbefragung mit zwei Messungen jeweils zu Beginn und am Ende des Experiments (Vorher-Nachher-Messung). Tatsächlich wurde in dem Experiment sogar ein dreiwelliges Panel erhoben, da sich das Experiment noch über ein weiteres Jahr erstreckte, in dem die experimentellen Stimuli über die Versuchsgruppen rotiert wurden (vgl. Boden u.a., S. 758). Wir dokumentieren hier kurz einige der aufgetretenen Paneleffekte bei der Messung der politischen Einstellung (vgl. auch 3.2).

1. interne Konsistenz der Skalen (Cronbachs  $\alpha$ ) (vgl. 5.4.1):

	M	S	G	W	О
					.73
2. Welle	.83	.81	.81	.83	.82
3. Welle	.86	.85	.82	.83	.82

Generell ist eine Zunahme der Skalenkonsistenz zu beobachten, insbesondere im Vergleich der ersten mit der zweiten Welle.

2. Interkorrelationen zwischen den Skalen in der ersten Welle (untere Dreiecksmatrix) und in der dritten Welle (obere Dreiecksmatrix):

	M	S	G	W	O
M	*	.50	52	.64	41
S	.39	*	65	.57	52
G	46	44	*	43	.56
W	.50	.44	36	*	36
O	32	39	.43	33	*

Es ist eine deutliche Zunahme der Interkorrelationen zu beobachten (z.B. erhöht sich die (negative) Korrelation der Befürwortung staatlicher Ordnungsmaßnahmen mit der Einstellung gegenüber den Gewerkschaften von -.44 auf -.65).

Beide Ergebnisse sind als Lerneffekte (sog. sokratische Effekte) zu deuten. Die Befragten lernen es, im Verlauf der wiederholten Befragungen die zu einer Skala gehörenden Items konsistenter zu beantworten und die Beziehung zwischen den Skalen besser zu verstehen. Natürlich ist auch nicht auszuschließen, daß neben diesen reinen Methodeneffekten tatsächlich eine deutlichere Artikulation des politischen Weltbildes der Befragten stattgefunden hat, etwa im Sinne eines Reifungsprozesses.

Wir kommen nun zu unserem eigentlichen Thema zurück, nämlich der Interpretation von Interaktionseffekten der experimentellen Bedingungen auf die experimentellen Variablen, hier also auf die politischen Einstellungen der Versuchspersonen. Ein Interaktionseffekt von Pretest und experimentellem Stimulus I<sub>PE</sub> würde eine Veränderung in den politischen Einstellungen der Versuchspersonen bedeuten, die durch die gemeinsame Wirkung von P und E hervorgerufen wird. So könnte etwa der Pretest die Versuchspersonen für den experimentellen Stimulus besonders empfänglich machen, d.h. die Versuchspersonen wurden durch die Abfrage ihrer politischen Einstellung im Pretest besonders für die politischen

Aussagen sensibilisiert, die in den jeweiligen Tageszeitungen vermittelt werden. Ein solcher Interaktionseffekt würde zusätzlich zu dem reinen Pretesteffekt (p) und dem reinen Effekt des experimentellen Stimulus (e) wirken - es handelt sich um einen echten Interaktionseffekt im Sinne der Tabellenanalyse (vgl. 3.1.2). Im Gegensatz zu dem reinen Pretesteffekt (p) kann der Interaktionseffekt I<sub>PE</sub> nicht durch die Kontrollgruppe kontrolliert werden. Es handelt sich um einen reaktiven Effekt des Pretests, der eine ernsthafte Gefährdung der externen Validität eines Experiments darstellen kann (siehe oben).

In gleicher Weise kann ein Interaktionseffekt IPU zwischen dem Prestest und einem unkontrolunserem Beispiel stellen der auftreten. In Abschluß Viermächteabkommens über Berlin 1971 und die vorgezogene Bundestagswahl 1972 solche unkontrollierten Ereignisse dar, die während der Laufzeit des Experiments stattfanden (71/72). Auch in diesem Zusammenhang ist eine besondere Sensibilisierung für diese Ereignisse durch den Pretest denkbar, die zu einer Veränderung der politischen Einstellung führt, die ohne den Pretest nicht eingetreten wäre.

Von besonderer Bedeutung ist der Interaktionseffekt I<sub>EU</sub> zwischen experimentellem Stimulus und unkontrollierten Ereignissen. Werden die gerade angesprochenen Ereignisse von den beiden Tageszeitungen in besonderer Weise kommentiert und bewertet, so könnte es zu Veränderungen in der politischen Einstellung aufgrund der Lektüre der Zeitungen gekommen sein, die in politisch ruhigeren Zeiten (also ohne das Auftreten der unkontrollierten Ereignisse) so nicht stattgefunden hätten. Dem Charakter der unkontrollierten Ereignisse entsprechend, muß stets mit dem Auftreten des Interaktionseffekts IEU gerechnet werden. Ebenso wie IPE kann er nicht durch die Kontrollgruppe kontrolliert werden und stellt eine Gefahr für die externe Validität des Experiments dar.

Schließlich ist noch ein Interaktionseffekt IPEU denkbar, der auf das gemeinsame Wirken aller genannten experimentellen Bedingungen zurückzuführen ist (Interaktionseffekt höherer Ordnung; vgl. auch den Exkurs zur Tabellenanalyse).

Formal ergibt der Vergleich von Experimental und Kontrollgruppe im Falle der Vorher-Nachher-Messung (PEUT-PUT-Anordnung):

$$d_1 - d_2 = e + I_{EU} + I_{PE} + I_{PEU}$$

Im Vergleich dazu ergibt sich im Fall der einfachen Nachher-Messung (EUT-UT-Anordnung):

$$d_3 - d_4 = e + I_{EU}$$

Die Wirkung des experimentellen Stimulus kann also nie von dem möglicherweise gleichzeitig auftretenden Interaktionseffekt zwischen experimentellem Stimulus und unkontrollierten Ereignissen getrennt werden (grundsätzliche Gefährdung der externen Validität des Experiments).

#### **Pretestproblematik:**

Im Unterschied zur EUT-UT-Anordnung sind bei der PEUT-PUT-Anordnung zusätzliche Interaktionseffekte zu berücksichtigen, die auf das gemeinsame Wirken von Pretest und experimentellem Stimulus zurückzuführen sind (reaktive Effekte des Pretests, die ebenfalls die externe Validität des Experiments gefährden). Von daher ist die EUT-UT-Anordnung vorzuziehen, wenn die beiden folgenden Bedingungen erfüllt sind:

- Die Wirkung des experimentellen Stimulus soll lediglich auf Gruppenebene durch Vergleich von Experimental- und Kontrollgruppe nachgewiesen werden. Sollen darüberhinaus individuelle Veränderungen untersucht werden, kann auf den Pretest nicht verzichtet werden (vgl. auch die Ausführungen zum Panel-Design in 3.2).
- Die Anzahl der Versuchspersonen in Experimental- und Kontrollgruppe ist groß genug, um (bei Randomisierung!) die unterstellte Gleichheit von P<sub>3</sub> und P<sub>4</sub> mit hinreichend großer Wahrscheinlichkeit behaupten zu können. (Da P<sub>3</sub> und P<sub>4</sub> wegen des Fehlens eines Pretests nicht gemessen werden können, muß P<sub>3</sub> ≈ P<sub>4</sub> vorausgesetzt werden, um

$$d_3 - d_4 = (T_3 - P_3) - (T_4 - P_4)$$

$$= T_3 - T_4 - (P_3 - P_4)$$

$$\approx T_3 - T_4$$

berechnen zu können.)

Um festzustellen, ob ein Effekt des experimentellen Stimulus vorliegt, genügt im einfachsten Fall ein einfacher t-Test (vgl. Übungsaufgabe I,2.), der die Veränderungen in der experimentellen Variable in Experimental- und Kontrollgruppe vergleicht (vorausgesetzt, die experimentelle Variable ist intervallskaliert und erfüllt die erforderlichen Verteilungsannahmen). Liegt, wie in unserem Fall, der experimentelle Stimulus in mehreren Ausprägungen vor (E<sub>1</sub>, E<sub>2</sub>, K), so muß eine einfaktorielle Varianzanalyse durchgeführt werden (vorausgesetzt, die experimentelle Variable ist intervallskaliert und erfüllt die erforderlichen Verteilungsannahmen). Im Fall des Experiments von Boden u.a. zeigen sich signifikante Langzeiteffekte der Lektüre der beiden Tageszeitungen im Hinblick auf die Einstellung zur Marktwirtschaft (M), der Beurteilung der Gewerkschaften (G) und der Ostpolitik (O), nicht aber für die beiden anderen Skalen S und W. Damit konnte die Verstärkungsthese widerlegt werden, nach der Medienkonsum generell nur zur Bestätigung bereits vorhandener Einstellung führt. Tageszeitungen beeinflussen also ihre Leser durchaus gezielt in Richtung auf die in ihnen vertretene politische Einstellung. Die Leser der Welt wurden konservativer und die Leser der Franfurter Rundschau liberaler.

# Exkurs: Komplexere experimentelle Designs und die mehrfaktorielle Varianzanalyse

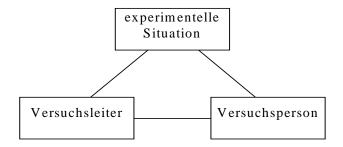
Werden in einem Experiment gleichzeitig mehrere experimentelle Faktoren variiert (in unserem Beispiel also zusätzlich die Studienrichtung der Versuchspersonen), spricht man von

mehrfaktoriellen Designs. Dabei multipliziert sich die Anzahl der Experimental- und Kontrollgruppen entsprechend. Zur Auswertung von mehrfaktoriellen Designs ist eine mehrfaktorielle Varianzanalyse erforderlich. Damit können auch Interaktionseffekte zwischen den Faktoren aufgedeckt werden, insbesondere dann, wenn alle Experimentalgruppen die gleiche Anzahl von Versuchspersonen aufweisen (sog. orthogonale Designs). Ein Spezialfall des mehrfaktoriellen Designs ist das Blockdesign, wo der zusätzlich in das Experiment eingeführte Faktor keine inhaltliche Bedeutung hat, sondern nur dazu dient, die Varianz in der experimentellen Variable zu kontrollieren. Auf diese Weise kann eine Wirkung des experimentellen Stimulus exakter nachgewiesen werden. Werden zusätzlich weitere intervallskalierte unabhängige Variablen (Kovariate) bei den Versuchspersonen kontrolliert, erweitert sich die Varianz- zur Kovarianzanalyse. Können Interaktionseffekte zwischen den Faktoren schon bei der Untersuchungsplanung aufgrund theoretischer Vorüberlegungen ausgeschlossen werden, kann die Anzahl der benötigten Experimentalgruppen deutlich verringert werden (z.B. durch das Design des "lateinischen Quadrates"). Eine Reduzierung der Anzahl der Versuchspersonen ist auch möglich, wenn die gleiche Gruppe von Versuchspersonen nacheinander verschiedenen experimentellen Bedingungen ausgesetzt wird (z.B. bei Cross-Over-Designs). Dabei können Ausstrahlungseffekte (Carry-Over-Effekte) auftreten, die die Validität des Experiments gefährden. Insbesondere für die methodische Erforschung von Design-Effekten im Experiment ist die Vier-Gruppen-Anordnung von Solomon von Bedeutung. In unserer Sprache handelt es sich dabei um eine PEUT-PUT-EUT-UT-Anordnung (vgl. das Schema weiter oben). Durch den Vergleich aller vier Gruppen kann das Auftreten von Pretest-Interaktionseffekten untersucht werden, z.B. gilt:

$$(d_1 - d_2) - (d_3 - d_4) = I_{PE} + I_{PEU}$$
.

## 3.3.3 Forschungsartefakte im Experiment

Wie gesehen, kann die Möglichkeit von Forschungsartefakten im Experiment selbst bei den sorgfältigsten Versuchen zur Kontrolle von Störfaktoren nie ausgeschlossen werden. Forschungsartefakte ergeben sich grundsätzlich aus der Reaktivität der experimentellen Situation (vgl. Bungard u. Lück 1974). Die Versuchspersonen können zusätzlich zu dem experimentellen Stimulus durch den Versuchsleiter und auch durch die experimentelle Situation selbst beeinflußt werden. Wir unterscheiden daher zwischen Versuchsleitereffekten und dem Aufforderungscharakter der experimentellen Situation:



#### Versuchsleitereffekte:

Ist der Versuchsleiter für Fehler verantwortlich, ohne dabei die Reaktionen der Versuchspersonen unmittelbar zu beeinflussen, so spricht man von <u>nichtinteraktionalen Versuchsleitereffekten</u>. Dazu gehört insbesondere die verzerrte Wahrnehmung und Protokollierung von Meßergebnissen, die, wie Studien zeigen, signifikant häufiger in Richtung der jeweils zu bestätigenden experimentellen Hypothese erfolgen. Dabei handelt es sich weniger um absichtliche Fälschungen als um unbewußte Verzerrung der Wahrnehmung der Versuchsleiter.

Bedeutender sind aber die <u>interaktionalen Versuchsleitereffekte</u>, die dadurch zustande kommen, daß die Versuchsleiter neben dem experimentellen Stimulus weitere, nicht beabsichtigte, Signale an die Versuchspersonen senden, die dann neben dem experimentellen Stimulus das Verhalten der Versuchspersonen beeinflussen. Zu diesen Signalen gehören zunächst einmal äußere Merkmale des Versuchsleiters, wie Geschlecht und Alter, die analog zu den Interviewereffekten das Verhalten der Versuchspersonen beeinflussen können (vgl. auch 6.1.2). Im Milgram-Experiment macht man sich sogar die autoritäre Ausstrahlung des Versuchsleiters bewußt zunutze, um die Versuchspersonen entsprechend zu beeinflussen. Was hier als Teil des experimentellen Stimulus eingesetzt wird, kann in seiner Wirkung auch in Situationen, in denen eine neutrales Auftreten des Versuchsleiters erforderlich wäre, nicht ausgeschlossen werden.

die Versuchsleitererwartungseffekte interaktionalen Auch gehören zu den Versuchsleitereffekten. Dabei beeinflussen die Erwartungen des Versuchsleiters in einer oft komplexen, dem Versuchsleiter selbst nicht bewußten Art und Weise das Verhalten der Versuchspersonen im Sinne einer Self-fulfilling-prophecy. Dies geschieht oft durch nichtverbale Signale, wie in dem klassichen Fall des "klugen Hans". Hans war das Pferd des pensionierten Volksschullehrers von Osten, das unter anderem Rechenaufgaben lösen konnte. Hans gab die richtige Antwort durch Hufklopfen an und hatte dabei gelernt, immer dann mit dem Klopfen aufzuhören, wenn der Versuchsleiter (als Aufgabensteller, der die Antwort natürlich kannte) durch unbewußte Entspannungssignale, wie z.B. leichte Kopfbewegung, anzeigte, daß die richtige Zahl erreicht war. Der Berliner Psychologe Pfungst, der (im Gegensatz zu vielen seiner Kollegen) die richtige Erklärung für diese sensationelle Lernleistung des klugen Hans gefunden hatte, war sogar in der Lage, selbst die Rolle des

klugen Hans zu übernehmen und erfolgreich 23 von 25 Rechenaufgaben allein durch die Beobachtung des Versuchsleiters richtig zu beantworten.

Versuchsleitererwartungseffekte sind in vielen Untersuchungen gefunden worden. Meinen etwa Versuchsleiter, sie hätten es mit "besonders klugen" Ratten zu tun, so lernen diese im Vergleich zu "dummen" Ratten besonders schnell, obwohl sie in Wirklichkeit aus der gleichen Population stammen. Die Bedeutung solcher induzierter Erwartungen auch außerhalb der Laborsituation wurde in einem berühmten Experiment von Rosenthal (1970) bestätigt. Rosenthal hat die Ergebnisse seines Experiments in einem Buch mit dem treffenden Titel "Pygmalion im Unterricht. Lehrererwartungen und Intelligenzentwicklung der Schüler" veröffentlicht. Dabei handelt es sich um ein Feldexperiment in einer sechsklassigen Elementarschule in den USA, in der durch die fingierten Ergebnisse eines Intelligenztests bei den Lehrern Erwartungen über die zukünftige intellektuelle Entwicklung der Kinder induziert wurden. Tatsächlich schnitten die Kinder, die entsprechend der induzierten Erwartungen der Lehrer besonders deutliche Lernerfolge aufweisen sollten, bei späteren Tests besser ab. Die Relevanz dieser Ergebnisse liegt auf der Hand: Glauben z.B. Mathematiklehrer, daß Mädchen über eine geringere mathematische Begabung verfügen, so werden sich diese Erwartungen im Sinne einer Self-fulfilling-prophecy verwirklichen.

#### Aufforderungscharakter der Experimentalsituation:

Nicht nur der Versuchsleiter, sondern auch die Merkmale der experimentellen Situation signalisieren den Versuchspersonen bestimmte Erwartungen - mit Orne sprechen wir von dem Aufforderungscharakter (demand-charakteristic) der experimentellen Situation. Die dabei hervorgerufenen Reaktionen der Versuchspersonen können die Effekte des experimentellen Stimulus überlagern und verfälschen.

Werden z.B. in einem Experiment, das die Auswirkungen einer längeren totalen Isolation in einem abgeschlossenen Raum ohne optische und akustische Reize untersuchen soll, zu Beginn des Experiments für die Versuchspersonen gut sichtbar Vorkehrungen für einen medizinischen Notfall getroffen, ist die Panikreaktion der Versuchspersonen vorprogrammiert und nicht etwa dem experimentellen Stimulus zuzuschreiben. Generell gilt, daß jede Versuchsperson Vermutungen über die mutmaßlichen Ziele des Experiments anstellen wird. Dabei geht die erfahrene Versuchsperson zusätzlich davon aus, daß der Versuchsleiter versuchen wird, das wahre Ziel des Experiments zu verheimlichen. Dadurch wird aber die Reaktivität der experimentellen Situation erst recht erhöht, da die Versuchsperson nun zu einer besonders sorgfältigen Beobachtung auch noch so nebensächlich erscheinender Details der experimentellen Situation angeregt wird. Hat die Versuchsperson eine Hypothese über den vermuteten Zweck des Experiments gebildet, so sind in Abhängigkeit von der Motivation der Versuchsperson zwei Reaktionsweisen denkbar: Will die Versuchsperson eine "gute Versuchsperson" sein, wird sie sich bemühen, die (vermutete) Hypothese des Experiments so weit wie möglich zu bestätigen (anderenfalls kann es auch zu systematisch abweichendem Verhalten kommen). So zeigte sich z.B. in einer Untersuchungsreihe im Rahmen der Hawthorne-Experimente, daß die Arbeitsleistung einer Gruppe von Arbeiterinnen unabhängig von der Manipulation der Arbeitsbedingungen (Beleuchtungsstärke und Pausenregelungen als experimenteller Stimulus) ständig zunahm (zunächst hypothesenkonform und dann entgegen den Erwartungen der Versuchsleiter), weil die Arbeiterinnen, wie sich schließlich herausstellte, als "gute Versuchspersonen" besonders motiviert waren, eine herausragende Arbeitsleistung zu erbringen (Hawthorne-Effekt).

# 4. Theoretische Konstrukte und Indikatoren

# 4.1 Konzeptspezifikation

Die in einer Untersuchung verwendeten Begriffe (theoretischen Konstrukte) müssen zunächst definiert werden. Wie in 2.1.4.1 ausführlich begründet wurde, ist eine solche Definition als Nominaldefinition zu verstehen - sie kann im Gegensatz zur Realdefinition nicht wahr oder falsch sein (sie trifft also nicht ein wie auch immer verstandenes "Wesen" des Gegenstandes), sondern nur mehr oder weniger brauchbar. In empirischen Untersuchungen verwendete Begriffe sollten daher präzise definiert sein, konsistent verwendet werden und theoretisch fruchtbar sein. Entscheidend für die Brauchbarkeit eines Begriffes ist die theoretische Fruchtbarkeit, d.h. mit dem Begriff müssen sich Hypothesen bilden lassen, die einen hohen Informationsgehalt haben und sich empirisch bewähren (vgl. die Kriterien für eine "gute" Theorie in 2.1.4.1). Um zu theoretisch fruchtbaren Begriffen zu kommen, ist eine sorgfältige Konzeptspezifikation erforderlich. Dazu bedient man sich des Verfahrens Begriffsexplikation, also der systematischen sprachlichen Präzision von Begriffen. Zunächst wird die Verwendung des Begriffes (z.B. Entfremdung) in den verschiedensten Theoriekontexten untersucht (Literaturstudium: Was verstehen z.B. Hegel, Marx, Weber, Seeman, Luhmann, usw. unter Entfremdung). Auf diese Weise werden unterschiedliche Vorstellungsinhalte zusammengestellt und präzisiert (dimensionale Analyse). In Hinblick auf die Forschungsziele wird dann entschieden, welche Vorstellungsinhalte in die Definition des Begriffes aufgenommen werden sollen (Kriterium der theoretischen Fruchtbarkeit). Im Gegensatz zur Begriffsinterpretation ist es also nicht entscheidend, ob die begriffliche Präzisierung mit der Auffassung eines bestimmten Autors übereinstimmt (z.B. mit der Auffassung von Entfremdung bei Marx). Gefordert ist eine rekonstruktiv-kritische Vorgehensweise, die den zu definierenden Begriff vor dem Hintergrund der vermuteten theoretischen Zusammenhänge (z.B. einer zu entwickelnden Theorie der Entfremdung) präzisiert. Das Ergebnis der Konzeptspezifikation ist also eine Nominaldefinition des zugrundeliegenden theoretischen Konstrukts, die geeignet ist, dieses optimal in eine zu entwickelnde Theorie einzupassen.

# 4.2 Operationalisierung

Für die empirische Forschung genügt es nicht, daß die verwendeten Begriffe explizit definiert sind. Es müssen vielmehr Forschungsoperationen angegeben werden, mit deren Hilfe entscheidbar ist, ob in einer konkreten Situation ein mit dem betreffenden Begriff bezeichnetes Phänomen vorliegt oder nicht (ob z. B. eine bestimmte Person Anzeichen von Entfremdung

zeigt oder nicht). Eine solche Festlegung von Forschungsoperationen nennt man die <u>Operationalisierung</u> eines Begriffs.

Insbesondere für Begriffe ohne direkten empirischen Bezug ist die Operationalisierung ein zweistufiger Prozeß. In diesem Zusammenhang knüpfen wir wieder an die Diskussion der Zweisprachentheorie in 2.1.4.2 an. Für Begriffe der theoretischen Sprache (Begriffe ohne direkten empirischen Bezug, sog. theoretische Konstrukte) muß zunächst festgelegt werden, welche Begriffe der Beobachtungssprache (Begriffe mit direktem empirischen Bezug) als Indikatoren gelten sollen. Dem entspricht die Angabe von Korrespondenzregeln oder die Spezifikation einer Meßtheorie im Sinne des kausalanalytischen Ansatzes (siehe unten).

Wir bezeichnen den Vorgang der Indikatorenauswahl als <u>Operationalisierung im weiteren Sinne</u>. Da die Menge aller möglichen Indikatoren nie genau spezifiziert werden kann, ist die Auswahl von Indikatoren immer selektiv und notwendigerweise unvollständig. Das zentrale Problem besteht also darin, wie die Zuordnung eines Indikators zu einem theoretischen Begriff gerechtfertigt werden kann. Man unterscheidet folgende Lösungen des Korrespondenzproblems (vgl. auch Schnell/Hill/Esser: Methoden der empirischen Sozialforschung, Kap. 4.2):

- Operationalismus
- typologisch-induktiver Ansatz (wird hier nicht weiter behandelt)
- kausalanalytischer Ansatz

Die operationalistische Lösung besteht in der definitorischen Gleichsetzung von theoretischem Konstrukt und Indikator(en). Dies bedeutet, daß man z.B. unter Intelligenz genau das versteht, was die in einem Intelligenztest enthaltenen Aufgaben (Indikatoren) messen. Damit wird die Zweisprachentheorie faktisch wieder aufgegeben. Es gibt im Grunde gar kein unabhängig definiertes theoretisches Konstrukt "Intelligenz", sondern genau so viele empiristische Begriffe der Intelligenz, wie es unterschiedliche Intelligenztests gibt. Damit wird es unmöglich, eine einheitliche Theorie der Intelligenz aufzustellen. Vielmehr gibt es genau so viele, im Sinne von Kuhn untereinander inkommensurable, Aussagen über Intelligenz, wie es verschiedene Untersuchungen über Intelligenz mit unterschiedlichen Intelligenztests gibt. Aus diesem Grunde ist die operationalistische Lösung des Korrespondenzproblems eine Scheinlösung, die den Anforderungen der Sozialwissenschaften nicht gerecht wird.

Der <u>kausalanalytische Ansatz</u> hält an der Zweisprachentheorie fest und betrachtet die Indikatoren als beobachtbare Folgen der latenten Variablen (<u>Wirkungsindikatoren</u>). [Daneben besteht auch die Möglichkeit, Indikatoren als Ursachen von latenten Konstrukten zu konzipieren (Ursachenindikatoren). In einem erweiterten Verständnis des kausalanalytischen Ansatzes sind sowohl Wirkungsindikatoren als auch Ursachenindikatoren zugelassen. Wir beschränken uns im folgenden aber auf die Diskussion von Wirkungsindikatoren.] Es wird also eine explizite Meßtheorie spezifiziert, die die Beziehung zwischen theoretischen Konstrukten und Indikatoren als Kausalbeziehung konkretisiert. Vorteile des kausalanalytischen Ansatzes sind

Kapitel 4.

das Festhalten an der Zweisprachentheorie und die explizite Formulierung einer Meßtheorie (und auch einer Meßfehlertheorie) zusätzlich zur Kerntheorie. Komplementär dazu ergibt sich als (unvermeidlicher) Nachteil (die Welt <u>ist</u> nun einmal kompliziert) ein komplexes mathematisches Modell, zu dessen Lösbarkeit oft stark vereinfachende Annahmen gemacht werden müssen (für eine Formalisierung und genauere Diskussion des kausalanalytischen Ansatzes siehe 4.3).

Auch im kausalanalytischen Ansatz stellt die Auswahl von validen Indikatoren in erster Linie ein theoretisches Problem dar (Meßtheorie). Indikatoren sind genau dann eine valide Operationalisierung eines theoretischen Konstruktes, wenn eine kausale Beziehung zwischen ihnen und dem zugehörigen theoretischen Konstrukt theoretisch einsichtig ist und empirisch bestätigt werden kann. (Nach den Überlegungen in 2.1.4.2 ist klar, daß eine solche empirische Überprüfung immer nur im Verbund von Kern-, Meß- und Meßfehlertheorie stattfinden kann.)

Zur Konkretisierung der Problematik der Indikatorenauswahl gehen wir kurz auf die Messung des Postmaterialismus ein. Die Postmaterialismus-These von R. Inglehart besagt, daß in den westlichen Industriegesellschaften seit dem 2. Weltkrieg eine Entwicklung von materialistischen hin zu postmaterialistischen Einstellungen zu beobachten ist. Zur empirischen Überprüfung dieser These ist die Messung postmaterialistischer Einstellungen erforderlich. Dazu werden die folgenden vier Items (Auswahl von Indikatoren aus dem potentiell unendlichen Indikatorenuniversum) der Postmaterialismusskala verwendet:

A: Aufrechterhaltung von Ruhe und Ordnung

B: Mehr Einfluß der Bürger auf Entscheidungen der Regierung

C: Kampf gegen steigende Preise

D: Schutz des Rechtes auf freie Meinungsäußerung

Die Indikatoren der Postmaterialismusskala stellen also politische Ziele dar, die von den Befragten entsprechend ihrer Wichtigkeit in eine Rangreihe gebracht werden sollen. Dabei sind die Items A und C Indikatoren für eine materialistische Einstellung, während die Items B und D eine postmaterialistische Einstellung anzeigen sollen. Je nachdem, ob der Befragte eher die postmaterialistischen oder die materialistischen Ziele für wichtiger hält, wird ihm ein entsprechender Wert auf der Postmaterialismusskala zugeordnet (eine genauere Diskussion des zugrundeliegenden Meßmodells erfolgt in 5.1).

In dieser Form wird die Postmaterialismusskala üblicherweise eingesetzt. Es existiert aber auch eine andere Version der Postmaterialismusskala, die andere Items verwendet, nämlich:

E: Erhaltung eines hohen Grades von wirtschaftlichem Wachstum

F: Sicherung von starken Verteidigungsstreitkräften

G: Verstärktes Mitspracherecht der Menschen an ihrem Arbeitsplatz und in ihrer Gemeinde

H: Versuche, unsere Städte und ländlichen Gebiete zu verschönern

In dieser zweiten Version sind E und F die materialistischen und G und H die postmaterialistischen Items.

Diese unterschiedlichen Operationalisierungen (im weiteren Sinne) des Postmaterialismus führen zu deutlich verschiedenen empirischen Aussagen (vgl. Hagstotz 1985). (Die folgenden Ergebnisse beziehen sich auf Daten aus dem ALLBUS 1982.) So sind nach Version I 70% der Bundesbürger materialistisch orientiert, nach Version II aber nur 58%. Die Korrelation zwischen den beiden Skalen (hier wird  $\tau_b$  als Assoziationsmaß verwendet; siehe auch 5.2) beträgt nur .27, d.h. die beiden Versionen messen weitgehend unterschiedliche Sachverhalte. Um zwischen den beiden Versionen zu entscheiden, muß auf die genaue Bedeutung des theoretischen Konstruktes Postmaterialismus in der Postmaterialismustheorie Ingleharts eingegangen werden. Zu fragen ist, welche Indikatoren (in welchem länderspezifischen sozialkulturellen Kontext (die Skala wird international vergleichend eingesetzt)) die Inhalte der im Postmaterialismus definierten Werthaltungen am validesten wiedergeben. Neben theoretischinhaltlichen Überlegungen im Hinblick auf die in den einzelnen Items formulierten Sachverhalte (Inhaltsvalidität) ist dabei auch die Überprüfung der theoretisch postulierten Beziehung des Postmaterialismus (P) mit anderen relevanten Variablen der Theorie von Bedeutung (Kriteriumsvalidität, siehe 5.4.1.2). Beschränkt man sich dabei auf die zentralen Variablen Alter (A) und Bildung (B), so ergeben sich für Version I substantiell bedeutsame Korrelationen, die mit den Erwartungen der Theorie im Einklang stehen (r<sub>PA</sub> ist negativ und r<sub>PB</sub> ist positiv) während sich für Version II lediglich eine schwache Beziehung für r<sub>PA</sub> ergibt. Diese Ergebnisse legen, zusammen mit zusätzlichen Überlegungen, die die inhaltliche Bedeutung der einzelnen Items betreffen, den Schluß nahe, daß es sich bei der ersten Version um eine validere Umsetzung des Postmaterialismuskonstruktes aus der Theorie Inglehards handelt.

Unter Operationalisierung im engeren Sinne versteht man die Operationalisierung der einzelnen Indikatoren, also die Angabe von konkreten Meßvorschriften für die Indikatoren. Die Operationalisierung ist also ein zweistufiger Prozeß: Zunächst müssen für ein theoretisches Konstrukt (z. B. die normative Geschlechtsrollenorientierung) Indikatoren gesucht werden. Im Falle einer Likert-Skala zur nGRO (siehe 5.3.1) sind die Items der Likert-Skala die Indikatoren. Zusätzlich müssen diese Indikatoren in konkrete Meßvorschriften umgesetzt werden. Im Falle der Items der Likert-Skala ist etwa über die konkrete Gestaltung des Fragebogens, die Wahl einer 5-stufigen bzw. 6-stufigen Rating-Skala, die genaue Vorgehensweise bei der Befragung, usw. zu entscheiden.

Im Zusammenhang mit jeder Messung müssen die folgenden Gütekriterien beachtet werden (siehe genauer in 5.4):

• **Objektivität:** Die Objektivität gibt an, in welchem Ausmaß eine Messung unabhängig vom Untersucher ist (mangelnde Objektivität kann z. B. durch Interviewereffekte in Ab-

hängigkeit vom Geschlecht des Interviewers bei der Erhebung der nGRO (vgl. 6.1) verursacht werden).

- Reliabilität (Zuverlässigkeit): Die Reliabilität betrifft die Genauigkeit oder Präzision einer Messung. Insbesondere sollten sich bei einer Meßwiederholung genau die gleichen Werte ergeben, wie bei der ersten Messung. Mangelnde Reliabilität wird durch Zufallsfehler bei einer Messung hervorgerufen, z. B. durch Ablesefehler oder mangelnde Konzentration bzw. mangelndes Interesse von Befragten im Interview (vgl. 6.1).
- Validität (Gültigkeit): Die Validität einer Messung hängt davon ab, ob wirklich das gemessen wird, was eigentlich inhaltlich gemessen werden soll. Eine Messung kann präzise im Sinne der Reliabilität sein, ohne jedoch eine Information über den zugrundeliegenden Sachverhalt zu geben (so z. B., wenn ein Befragter sozial erwünschte Antworten gibt oder tendenziell allen Fragen zustimmt (Ja-Sage-Tendenz; siehe auch 6.1)).

## Operationalisierung von theoretischen Konstrukten als zweistufiger Prozeß

(die jeweiligen Gütekriterien sind in eckigen Klammern beigefügt).

L <sub>T</sub> :	$L_T$ :					
Operationalisierung im weiteren Sinne (Lösung des Korrespondenzproblems)		Zuordnung von Indikatoren (Auswahl aus dem Indikatorenuniversum)	[ <b>Validität</b> der Indikatoren]			
L <sub>O</sub> : <u>Indikatoren</u>						
	(z.B. Aufg Intelliger					
Operationalisierung im engeren Sinne		Umsetzen in eine konkrete Meßvorschrift: Wahl des Meßmodells, Gestaltung und Duchführung des Tests	[Objektivität, Reliabilität, Validität]			
Messung:						

# 4.3 Der kausalanalytische Ansatz

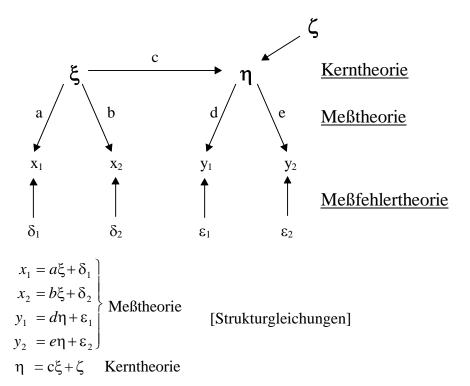
Der kausalanalytische Ansatz bietet die Möglichkeit, die Koeffizienten der Kerntheorie, der Meßtheorie und der Meßfehlertheorie in einem integrierten Kausalmodell zu schätzen - und zwar allein auf der Grundlage der empirisch erhobenen Korrelationen zwischen den Indikato-

ren. Allerdings müssen dazu einige zusätzliche Annahmen (insbesondere über die Meßfehler) gemacht werden, damit das formale Modell lösbar ist. Im folgenden diskutieren wir das einfachste Modell (eine bivariate Beziehung als Kerntheorie; jedes theoretische Konstrukt wird durch zwei Indikatoren erfaßt) in allen Einzelheiten.

In der Praxis verwendet man Modelle mit komplexen multivariaten kausalen Zusammenhängen und einer Vielzahl von Indikatoren, die mit Hilfe von speziellen Analyseprogrammen, wie z. B. LISREL (vgl. Kap. 8), gelöst werden. Ziel unserer Darstellung kann es daher nur sein zu zeigen,

- 1. daß solche Modelle überhaupt gelöst werden können und
- 2. daß mit der zusätzlichen expliziten Berücksichtigung eines Meßmodells (Meßtheorie und Meßfehlertheorie) genauere Aussagen gemacht werden können, die zu inhaltlich bedeutsamen Modifikationen in den Schlußfolgerungen empirischer Untersuchungen führen. Deutlicher ausgedrückt: Untersuchungen ohne die Berücksichtigung von Meßmodellen führen zu verzerrten Ergebnissen und damit auch zu inhaltlich falschen Schlußfolgerungen.

#### Das Kausalmodell im einzelnen:



Zur Lösung des Modells werden die folgenden Annahmen gemacht:

- Standardisierte Variablen: <u>Alle</u> im Modell enthaltenen Variablen sind standardisiert, d.h. es handelt sich um intervallskalierte Variablen mit einem Mittelwert von 0 und einer Standardabweichung von 1.
- Lineare Beziehungen: Alle Beziehungen im Modell sind linear (vgl. die Gleichungen des Modells).

• Unabhängige Meßfehler: Als <u>Meßfehlertheorie</u> spezifizieren wir unabhängige Zufallsfehler, d.h. die Fehlervariablen δ<sub>1</sub>, δ<sub>2</sub>, ε<sub>1</sub>, ε<sub>2</sub> sind <u>untereinander</u> und <u>mit den anderen latenten Variablen</u> unkorreliert. Außerdem sind ξ und ζ unkorreliert. Diese Voraussetzung ist die weitreichenste und zugleich am wenigsten realistische Annahme, die wir zur Lösung des Modells machen müssen (vgl. auch 3.1). Üblicherweise enthalten solche Modelle weitere Indikatoren für die in der Kerntheorie enthaltenen theoretischen Konstrukte und damit auch zusätzliche empirische Informationen (Korrelationen in der zugehörigen Korrelationsmatrix). Ist dies der Fall, können die hier gemachten restriktiven Annahmen schrittweise gelockert werden, so daß eine realitätsnähere Meßfehlertheorie entsteht, die z.B. auch korrelierte Meßfehler zuläßt (vgl. auch die Diskussion der Annahmen der klassischen Testtheorie in 5.4.1).

Grundlage der Berechnung der zu schätzenden fünf Modellparameter a, b, c, d, e ist die empirisch gegebene Korrelationsmatrix der Indikatoren  $x_1$ ,  $x_2$ ,  $y_1$ ,  $y_2$ . Da das Modell mehr empirische Informationen als zu schätzende Parameter enthält, das Modell also überidentifiziert ist (6 empirische Korrelationskoeffizienten bei nur 5 zu schätzenden theoretischen Parametern), bleibt ein Freiheitsgrad zu Testzwecken. Es kann also eine <u>Testgleichung</u> abgeleitet werden, d.h. ein Zusammenhang zwischen den empirisch ermittelten Korrelationen, der erfüllt sein muß, wenn die Voraussetzungen des Modells gelten sollen.

Nach den Regeln der Pfadanalyse (vgl. insbesondere 3.1.1.3) ergeben sich aus den obigen Voraussetzungen die folgenden Gleichungen (Effektzerlegung):

$$r_{x_1x_2} = ab$$
 $r_{y_1y_2} = de$ 
 $r_{x_1y_1} = acd$ 
 $r_{x_1y_2} = ace$ 
 $r_{x_2y_1} = bcd$ 
 $r_{x_2y_2} = bce$ 

Als Testgleichung ergibt sich:

$$r_{x_1y_1}r_{x_2y_2} = r_{x_1y_2}r_{x_2y_1}$$

Ersetzt man in der Testgleichung die Korrelationen durch die zugehörigen Produkte von Pfadkoeffizienten, so erkennt man, daß die Testgleichung tatsächlich aus den Modellvoraussetzungen abgeleitet werden kann:

$$acd \cdot bce = ace \cdot bcd$$
  
 $abc^2 de = abc^2 de$ .

Für die Parameter des Modells ergibt sich daraus:

$$c^{2} = \frac{r_{x_{1}y_{2}} r_{x_{2}y_{1}}}{r_{x_{1}x_{2}} r_{y_{1}y_{2}}} = \frac{r_{x_{1}y_{1}} r_{x_{2}y_{2}}}{r_{x_{1}x_{2}} r_{y_{1}y_{2}}}$$

$$a^{2} = r_{x_{1}x_{2}} \frac{r_{x_{1}y_{2}}}{r_{x_{2}y_{2}}} = r_{x_{1}x_{2}} \frac{r_{x_{1}y_{1}}}{r_{x_{2}y_{1}}}$$

$$b^{2} = r_{x_{1}x_{2}} \frac{r_{x_{2}y_{2}}}{r_{x_{1}y_{2}}} = r_{x_{1}x_{2}} \frac{r_{x_{2}y_{1}}}{r_{x_{1}y_{1}}}$$

$$d^{2} = r_{y_{1}y_{2}} \frac{r_{x_{2}y_{1}}}{r_{x_{2}y_{2}}} = r_{y_{1}y_{2}} \frac{r_{x_{1}y_{1}}}{r_{x_{1}y_{2}}}$$

$$e^{2} = r_{y_{1}y_{2}} \frac{r_{x_{2}y_{2}}}{r_{x_{2}y_{2}}} = r_{y_{1}y_{2}} \frac{r_{x_{1}y_{2}}}{r_{x_{1}y_{2}}}$$

Die Pfadkoeffizienten sind also nur bis auf das Vorzeichen bestimmt. Die Vorzeichen der Pfadkoeffizienten müssen aus theoretischen Überlegungen über die Richtung der Kausalbeziehungen und die Polung der Indikatoren erschlossen werden. Die Polung erfolgt konventionell, ist aber im Rahmen der Überprüfung des Modells fest vorgegeben. Wichtig ist, daß die die Vorzeichen der Pfadkoeffizienten betreffenden theoretischen Überlegungen vor der empirischen Überprüfung des Modells angestellt werden, so daß ein zusätzlicher Modelltest möglich ist, indem die Vorzeichen der empirischen Korrelationen zwischen den Indikatoren (über die Effektzerlegungen) mit den postulierten Vorzeichen der Pfadkoeffizienten verglichen werden.

## Beispiel zum kausalanalytischen Ansatz

Kerntheorie: Soziale Distanz 

→ Ausländerfeindlichkeit

Korrelationsmatrix der Indikatoren aus dem ALLBUS 1980 (6 empirische Informationen):

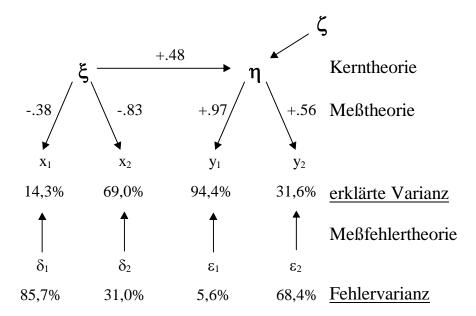
	$\mathbf{r}_{ij}$	$\mathbf{x}_1$	$\mathbf{x}_2$	$y_1$	$y_2$	
Arbeitsplatz	<b>X</b> <sub>1</sub>	1.00				(negative) Indikatoren
Freundeskrei s	<b>x</b> <sub>2</sub>	+.315	1.00			für soziale Distanz $(\xi)$
Ehepartner	<b>y</b> 1	195	370	1.00		(positive) Indikatoren für Ausländerfeind-
politische Tätigkeit	<b>y</b> <sub>2</sub>	094	241	+.563	1.00	lichkeit (η)

Die Indikatoren für soziale Distanz sind jeweils so gepolt, daß ein positiver Wert das Vorliegen von sozialen Beziehungen anzeigt. Die Variablen Arbeitsplatz und Freundeskreis stellen daher negative Indikatoren für soziale Distanz dar. Wir erwarten daher einen negativen Wert für die entsprechenden Faktorladungen a und b. Die Polung für die Indikatoren zur Ausländerfeindlichkeit erfolgte derart, daß ein hoher Wert eine starke Ausländerfeindlichkeit anzeigt. Die Variablen Ehepartner und politische Tätigkeit sind daher positive Indikatoren für Ausländerfeindlichkeit. Dementsprechend erwarten wir positive Werte für die Faktorladungen d und e. Damit können Vorhersagen über die Korrelationen zwischen den Indikatoren abgeleitet werden. Wir erwarten positive Beziehungen zwischen den Indikatoren des gleichen Konstrukts und negative Beziehungen zwischen den Indikatoren verschiedener Konstrukte. Wie ein Blick auf die empirischen Korrelationen aus der ALLBUS-Studie zeigt, bestätigen sich diese theoretischen Voraussagen in unserem Beispiel. Die Überprüfung der Testgleichung ergibt folgendes Ergebnis:

$$r_{x_1y_1}r_{x_2y_2} \stackrel{!}{=} r_{x_1y_2}r_{x_2y_1}$$
?
$$(-.195)(-.241) \stackrel{?}{=} (-.094)(-.370) \quad \underline{\text{Testgleichung}}$$

$$.0469 \approx .0349$$

Für die folgenden Überlegungen gehen wir davon aus, daß die Testgleichung erfüllt ist, d.h. daß die beiden Seiten der Testgleichung nicht signifikant voneinander abweichen. Damit können die Parameter des Modells aus den obigen Gleichungen berechnet werden, wobei zu berücksichtigen ist, daß die beiden Versionen der Parametergleichungen nur dann zu den gleichen Resultaten führen, wenn die Testgleichung exakt erfüllt ist. (In unserem Beispiel wurde das geometrische Mittel aus den beiden Parameterschätzungen gebildet):



Die Kerntheorie wird durch das Modell bestätigt - der Pfadkoeffizient ist mit .48 sogar relativ hoch im Vergleich zu anderen empirischen Untersuchungen (obwohl dies nur einem Anteil

der erklärten Varianz der Ausländerfeindlichkeit von 23% entspricht, während der Rest von 77% auf andere Einflüsse (zusammengefaßt in dem Fehlerterm ζ) zurückgeführt werden muß). Wie an den Faktorladungen (Pfadkoeffizienten) zu erkennen ist, sind die verwendeten Indikatoren von durchaus unterschiedlicher Qualität. Während der Indikator x2 (Freundeskreis) mit einer Faktorladung von -.83 überwiegend von dem zu messenden theoretischen Konstrukt abhängt, gilt dies für x<sub>1</sub> nur zu einem sehr geringen Teil, nämlich zu 14,3%. Genauer ausgedrückt: Die Bestätigung (bzw. Verneinung) der Frage nach einer Bekanntschaft mit Ausländern im Freundeskreis (x2) durch die Befragten läßt sich zu 69% (-.832) auf deren soziale Distanz zu Ausländern zurückführen. Zu 31% hängt die Beantwortung dieses Items von anderen (im Sinne der Meßfehlertheorie zufälligen!) Einflüssen ( $\delta_2$ ) ab. Die Bestätigung (bzw. Verneinung) der Frage nach einer Bekannschaft mit Ausländern am Arbeitsplatz (x<sub>1</sub>) kann dagegen nur zu 14,3% (-.38<sup>2</sup>) auf die soziale Distanz der Befragten zurückgeführt werden - zu 85,7% spielen hier andere (zufällige!) Einflüsse (δ<sub>1</sub>) eine Rolle. Die erklärte Varianz entspricht gerade der Reliabilität der Indikatoren (siehe die Ausführungen zur klassischen Testtheorie in 5.4.1): Die höchste Reliabilität mit .944 hat demnach der Indikator y<sub>1</sub> (Ehepartner). Das zugrundeliegende theoretische Konstrukt Ausländerfeindlichkeit kommt in y<sub>1</sub> in fast reiner Form zum Ausdruck - lediglich ein Varianzanteil von 5,4% ist auf den zugehörigen Zufallsfehler ε<sub>1</sub> zurückzuführen.

Hätte man versucht, die Beziehung zwischen sozialer Distanz und Ausländerfeindlichkeit ohne Spezifikation eines Meßmodells allein aus den Korrelationen der beteiligten Indikatoren zu erschließen, so hätte man je nach Art der verwendeten Indikatoren Werte zwischen .094 und .37 erhalten (vgl. die Korrelationsmatrix). Wie das Modell zeigt, ist die <u>um Meßfehler bereinigte Korrelation</u> zwischen den theoretischen Konstrukten mit .48 deutlich höher. Die empirischen Korrelationen sind also im Vergleich zu der theoretischen Korrelation vermindert, da in ihnen auch die Meßfehler (Zufallsfehler) zum Ausdruck kommen (mangelnde Reliabilität der Indikatoren). Die Korrelationen zwischen den emprischen Indikatoren sind dabei um so höher, je höher die Faktorladungen (theoretische Validität der Indikatoren im Rahmen des Modells; siehe 5.4.1.2) und die zugehörigen Reliabilitäten sind. So ist z.B. die Korrelation zwischen den beiden reliabelsten Indikatoren r<sub>x2y1</sub> mit .37 nur geringfügig kleiner als die zugehörige theoretische Korrelation von .48.

Im Rahmen der klassischen Testtheorie versucht man, die Reliabilität von Skalen, die aus einer Vielzahl von Indikatoren zusammengesetzt sind (z.B. von Likert-Skalen), zu schätzen; eine Möglichkeit ist z.B. die Berechnung von Cronbachs α (vgl. 5.4.1.1). Sind die Reliabilitäten bekannt, so kann mit Hilfe einer mathematischen Formel eine Minderungskorrektur vorgenommen werden, die die durch die Einwirkung von Zufallsfehlern (mangelnde Reliabilität) in den Skalen verursachte Abschwächung der theoretischen Korrelationen zwischen den Skalen wieder ausgleicht (siehe 5.4.1.1). Der kausalanalytische Ansatz ist demgegenüber der fortgeschrittenere und elegantere, da er auf eine Skalenbildung im Sinne von additiven Indizes (wie z.B. bei der Likert-Skalierung; vgl. 5.3.1) verzichtet und die

einzelnen Items explizit im Sinne einer Meß- und Meßfehlertheorie in das Kausalmodell integriert. Eine Minderungskorrektur ist beim kausalanalytischen Ansatz unnötig, da in den berechneten theoretischen Korrelationen die mangelnde Reliabilität der Indikatoren "quasi automatisch" berücksichtigt wird.

## Einfluß unterschiedlicher Reliablitäten

Die Berücksichtigung von Meßmodellen und die damit verbundene Berücksichtigung unterschiedlicher Reliabilitäten kann die inhaltlichen Schlußfolgerungen, die aus einer empirischen Untersuchung gezogen werden müssen, deutlich verändern. Generell gilt, daß die Korrelationen zwischen den theoretischen Konstrukten größer sind als die entsprechenden empirischen Korrelationen zwischen den zugeordneten Indikatoren - und zwar um so mehr, je geringer die Reliabilität der Indikatoren ist. Vergleicht man also verschiedene Gruppen, die sich durch unterschiedliche Reliabilitäten (z.B. durch ein gruppenspezifisches Antwortverhalten bedingt) auszeichnen, so ist dieser Vergleich notwendigerweise verzerrt, solange die erforderlichen Minderungskorrekturen, die in beiden Gruppen unterschiedlich stark ausfallen werden, nicht durchgeführt sind.

# <u>Beispiel:</u> Empirische Untersuchung des Statuszuweisungsprozesses in den USA für Schwarze und Nicht-Schwarze (vgl. Bielby/Hauser/Featherman 1977):

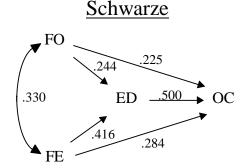
Kerntheorie: Statuszuweisungsmodell (vgl. auch 3.1.1.3)

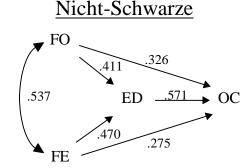
Meßmodell mit 3 Meßzeitpunkten (z. B. fo<sub>1</sub>, fo<sub>2</sub>, fo<sub>3</sub> für FO)

FO = Fathers Occupation ED = Respondents Education

FE = Fathers Education OC = Respondents Occupation

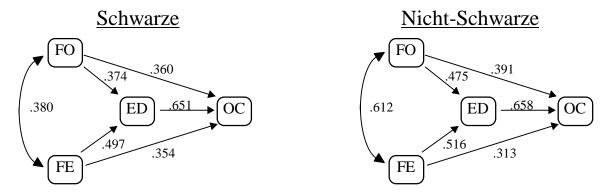
a) <u>unkorrigiertes Modell</u> (nur ein Indikator, z.B.  $FO \equiv fo_1$ )



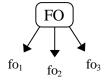


## b) korrigiertes Modell (Modell, das auch Meßfehler berücksichtigt):

Die geringere Reliabilität bei den Angaben der Schwarzen bedingt eine stärkere Minderungskorrektur.



Die Meßmodelle werden nicht explizit mit aufgeführt, z. B.:



Schlußfolgerung: Berücksichtigt man Meßfehler (größere Ungenauigkeit (geringere Reliabilität) in den Antworten der Schwarzen), so ist die Wirkung der Bildung (ED) auf den erreichten Berufsstatus (OC) in beiden Gruppen gleich, d.h. Bildungsinvestitionen sind für Schwarze genauso wirkungsvoll wie für die übrige Bevölkerung in den USA.

[Bemerkung: Genaugenommen ist ein Gruppenvergleich mit standardisierten Variablen nicht so aussagekräftig wie ein entsprechender Vergleich mit unstandardisierten Variablen auf der Grundlage von Kovarianzstrukturen, der auch Unterschiede in den Varianzen zwischen den Gruppen berücksichtigt.]

# 5. Messen

Im vorangehenden Kapitel haben wir einige grundsätzliche Überlegungen zu methodischen Problemen der Operationalisierung, also der Indikatorenwahl und der Spezifizierung einer konkreten Meßvorschrift, angestellt. In diesem Kapitel erfolgt eine tiefergehende und ausführlichere Auseinandersetzung mit der Meßproblematik. Zunächst behandeln wir grundlegende Probleme der Meßtheorie, insbesondere das Repräsentations-, das Eindeutigkeits- und das Bedeutsamkeitsproblem. Dabei unterscheiden wir grundsätzlich zwischen einer Repräsentations- und einer Indexmessung (vgl. Dawes 1964 und Heidenreich 1987) [vgl. 5.1]. Generell gilt, daß das Meßniveau nur im Fall der Repräsentationsmessung eindeutig bestimmt ist, und zwar aufgrund der für einen bestimmten Skalentyp erlaubten Transformationen (Lösung des Eindeutigkeitsproblems). Im Unterschied dazu wird das Skalenniveau bei der Indexmessung konventionell festgelegt. Je nach Meßniveau sind unterschiedliche statistische Modelle der Datenanalyse erlaubt (Lösung des Bedeutsamkeitsproblems) [vgl. 5.2].

Im Anschluß daran werden einige wichtige Skalierungsverfahren eingehender dargestellt, insbesondere die Likert-Skalierung (Indexmessung auf Intervallniveau) und die Guttman-Skalierung (Repräsentationsmessung auf Ordinalniveau) [vgl. 5.3]. Vor dem Hintergrund der klassischen Testtheorie werden dann die gebräuchlichsten Verfahren zur Bestimmung der Reliabilität und Validität einer Skala diskutiert [vgl. 5.4.1]. Abschließend erfolgt eine Kurzcharakterisierung der Grundgedanken der Faktorenanalyse [vgl. 5.4.2] und der mulitdimensionalen Skalierung [vgl. 5.5] in Hinblick auf Möglichkeiten, die Dimensionalität eines theoretischen Konstrukts meßtheoretisch zu bestimmen.

# 5.1 Repräsentations- und Indexmessung

Messen im engeren Sinne heißt die nach bestimmten Regeln vollzogene Zuordnung von Symbolen (im allgemeinen Zahlen) zu beobachteten Merkmalsausprägungen auf der zu messenden Merkmalsdimension. Im Zentrum der Meßproblematik steht dabei die Frage nach dem Informationsgehalt einer Messung, also nach der Art der Regeln, nach der die Zuordnung von Zahlen zu den beobachteten Merkmalsausprägungen vorgenommen wird. Dabei hat der Sozialwissenschaftler die Wahl zwischen dem Einpassen der sozialen Wirklichkeit in ein bestimmtes Meßmodell und der Überprüfung der Passung zwischen der sozialen Wirklichkeit und einem Meßmodell. Nach dem berühmten Diktum des Meßtheoretikers C. Coombs ist es das Problem des Sozialwissenschaftlers, "ob er weiß was er will, oder ob er etwas wissen will". Im ersteren Fall wird das Meßmodell qua Konvention vorausgesetzt (Indexmessung), während im zweiten Fall die Voraussetzungen des Meßmodells überprüft werden (Konsistenztest im Rahmen der Repräsentationsmessung). Diese Möglichkeit der Konsistenzüberprüfung im Rahmen der Repräsentationsmessung entspricht dem folgenden

von Coombs formulierten Prinzip: Keine Eigenschaft der Daten kann als bestehend behauptet werden, solange nicht die Methoden der Datensammlung und Datenanalyse es zulassen, daß andere Eigenschaften zum Vorschein kommen. Nach dieser Vorstellung ist Messen die strukturtreue Abbildung (also ein Homomorphismus) von einem empirischen Relativ in ein numerisches Relativ. Dabei enthält das empirische Relativ die empirischen Beziehungen zwischen den zu messenden Objekten (wie sie im Rahmen des Meßvorgangs erhoben werden), während das numerische Relativ die Meßwerte (also die Zahlen) mit den entsprechenden mathematischen Relationen enthält. Die für die Repräsentationsmessung charakteristische Möglichkeit der Konsistenzüberprüfung bedeutet, daß im Rahmen der Messung darüber entschieden werden kann, ob eine strukturtreue Abbildung auf einem bestimmten Meßniveau überhaupt möglich ist. Eine Repräsentationsmessung kann also scheitern, und zwar in Abhängigkeit von den empirisch erhobenen Eigenschaften des empirischen Relativs. Gelingt aber eine Repräsentationsmessung, so kann der Forscher wegen der vorangegangenen Konsistenzüberprüfung sicher sein, daß in den Meßwerten tatsächlich Eigenschaften des empirischen Relativs zum Ausdruck kommen.

Zur Verdeutlichung betrachten wir im folgenden das Problem der Messung einer Rangordnung, also der Konstruktion einer Ordinalskala, am Beispiel der Messung des Postmaterialismus. Das numerische Relativ besteht in diesem Falle aus der Menge von Rangplätzen (1-4) und der zugehörigen Ordnungsbeziehung (z.B. ist Rangplatz 2 höher als Rangplatz 3). Das empirische Relativ besteht aus einer Menge von Objekten, in unserem Beispiel also den Items A, B, C und D (vgl. 4.2) der Postmaterialismusskala und den Rangbeziehungen zwischen den Objekten (z.B. könnte ein Befragter angeben, daß er Item D als politisches Ziel wichtiger hält als Item C. Die Strukturtreue der Messung bedeutet in diesem Zusammenhang, daß die Zuordnung der Rangplätze zu den Items ordnungserhaltend (homomorph) ist, d.h. in unserem Beispiel, daß auch der dem Item D zugeordnete Rangplatz (Meßwert) höher sein muß, als der dem Item C zugeordnete Rangplatz.

Die übliche Vorgehensweise bei der Messung des Postmaterialismus enthält keine Überprüfung der Strukturtreue der Abbildung ist also keine Repräsentations-, sondern nur Indexmessung. Der Befragte wird nämlich lediglich aufgefordert, die 4 Items A, B, C, D in eine Rangreihe zu bringen, also zu sagen, welches Item für ihn am wichtigsten, am zweitwichtigsten, usw. ist. Aus diesen Antworten wird ein Postmaterialismus-Index berechnet. In unserem Zusammenhang entscheidend ist die Frage, ob in den Antworten des Befragten tatsächlich eine Rangordnung zum Ausdruck kommt, über die der Befragte unabhängig über die Art der Datenerhebung verfügt. Offensichtlich wird nämlich in unserem Beispiel die Rangordnung gerade durch die Frageformulierung erzwungen, da der Befragte explizit aufgefordert wird, eine Rangordnung zu produzieren. Eine Abweichung von diesem durch das Frageformat vorgegebenen Antwortschema ist für den Befragten nicht möglich. Im Sinne unseres oben zitierten Prinzips kann also das Vorliegen einer Rangordnung (Eigenschaft der Daten) nicht als bestehend behauptet werden (zumindest wird sie im

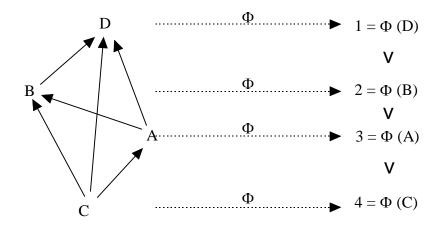
Rahmen des Meßvorgangs nicht empirisch überprüft), da die Methode der Datensammlung kein anderes Resultat zuläßt. Eine solche Messung bezeichnen wir als Indexmessung. Die die Skala charakterisierende Eigenschaft, hier die Messung einer Rangordnung wird vorausgesetzt und durch die Art des Verfahrens sichergestellt. Eine empirische Konsistenzüberprüfung dieser entscheidenden Voraussetzung des Meßmodells findet nicht statt.

In ähnlicher Weise läge eine Indexmessung auf Intervallskalenniveau vor, wenn der Befragte gebeten würde, alle 4 Items der Postmaterialismusskala auf einer von 0 bis 10 reichenden, gleichmäßig untergliederten Rating-Skala mit den Endpunkten 0 = "überhaupt nicht wichtig" und 10 = "sehr wichtig", einzustufen. In diesem Fall ist es sogar möglich, die Abstände zwischen den Wichtigkeitseinstufungen für die einzelnen Items der Postmaterialismusskala zu vergleichen, indem man die Abstände auf der Rating-Skala der Wichtigkeit interpretiert. Allerdings wird auch hier vorausgesetzt, daß der Befragte in der Lage ist, die Items der Postmaterialismusskala auf der Rating-Skala der Wichtigkeit sinnvoll einzuordnen, also die Abstände auf der Rating-Skala (numerisches Relativ) im Sinne gleicher Abstände zwischen den Wichtigkeiten der Items (empirisches Relativ) zu interpretieren. Diese grundlegende Voraussetzung der Messung wird aber im Rahmen der Messung nicht mehr empirisch überprüft. In diesem Sinne wird das Meßniveau rein konventionell durch den Forscher festgelegt, es handelt sich in diesem Falle also um Indexmessung auf Intervallskalenniveau.

Im Gegensatz dazu erlaubt die Repräsentationsmessung, durch sein Antwortverhalten zu dokumentieren, ob er über eine konsistente Rangordnung der Items der Postmaterialimusskala verfügt, d.h., ob eine Messung auf dem vorgegebenen Skalenniveau überhaupt möglich ist. Dazu betrachten wir noch einmal den Versuch, eine Rangordnung der Postmaterialismus-Items empirisch zu erheben. Eine Form der Repräsentationsmessung wäre hier der Paarvergleich, bei dem der Befragte aufgefordert wird, für jedes Paar von Items zu sagen, welches der beiden Items er konkret präferiert. Im Gegensatz zur Rangreihe wird also nicht nach einer umfassenden Rangordnung gefragt, sondern nur nach paarspezifischen Präferenzen (dabei steigt der Aufwand bei der Datenerhebung erheblich an, da sich aus vier Items sechs Paare von Items bilden lassen für die jeweils eine Präferenzeinstufung durch den Befragten erfolgen muß). Entscheidend ist nun, daß die Erhebung einer Rangordnung mit der Methode des Paarvergleichs scheitern kann. Um eine konsistente Rangordnung zu ergeben, müssen die sechs einzelnen Paarvergleiche, nämlich die Transitivitätsbedingungen erfüllen. Wenn der Befragte z.B. A vor B präferiert und B vor C, so muß er also auch A vor C präferieren. Gibt er aber in einem solchen Falle an, daß er C höher einstuft als A, so liegt eine empirische Inkonsistenz vor. Der Befragte ist also in der Lage durch sein Antwortverhalten zu dokumentieren, daß er über keine konsistente Rangordnung der Items verfügt. (Im Falle der Indexmessung einer durch das Frageformat erzwungenen Rangordnung würde in diesem Fall also eine Information erhoben, die keine wirkliche empirische Entsprechung hat.) In diesem Falle wäre die Messung also gescheitert, eine Skalierung der Items auf Ordinalniveau wäre

nicht möglich. Umgekehrt kann gesagt werden, daß bei gelingender Messung die Repräsentationsmessung eine höhere Aussagekraft als die Indexmessung hat, da in diesem Fall eine Konsistenzüberprüfung bestanden wurde. Der Forscher kann also sicher sein, daß der Befragte über eine Rangordnung verfügt und das Antwortmuster nicht einfach ein Artefakt der Frageformulierung ist (wie möglicherweise im Falle der Indexmessung).

Schematische Darstellung einer <u>strukturtreuen (homomorphen) Abbildung ( $\Phi$ )</u> am Beispiel der Messung des Postmaterialismus durch Paarvergleich:



empirisches Relativ

[die empirischen Präferenzurteile (empirische Relation) im Paarvergleich sind durch die sechs Pfeile dargestellt]

## numerisches Relativ

[die Rangfolge der Rangplätze definiert die Ordnungsrelation (mathematische Relation) im numerischen Relativ]

Der Klarheit halber soll noch einmal betont werden, daß der Paarvergleich nur <u>eine</u> Methode der Repräsentationsmessung ist. Daneben existieren <u>andere</u> Methoden der Repräsentationsmessung, wie z.B. die Guttman-Skalierung (Repräsentationsmessung auf Ordinalniveau; vgl. 5.3) oder die hier nicht weiter behandelte additiv verbundene Messung (Conjoint-Messung; vgl. Backhaus u.a. 1993 und Schweikl 1985) (Repräsentationsmessung auf Intervallskalenniveau).

Fazit: Repräsentationsmessungen können auf interne Konsistenz überprüft werden. Damit ist eine strukturtreue Abbildung des empirischen Relativs in das numerische Relativ gesichert. Den Meßwerten und den zwischen ihnen bestehenden numerischen Beziehungen entsprechen also empirische Beziehungen zwischen den zu messenden Objekten. Dagegen wird bei Indexmessungen das Meßniveau qua Konvention festgelegt. Die Qualität des Informationsgehalts einer Indexmessung wird also im Rahmen des Meßvorganges nicht empirisch überprüft, sondern muß im Lichte der externen Gütekriterien Objektivität, Reliabilität und Validität beurteilt werden.

## 5.2 Skalenniveaus

Wie gerade gesehen, stellt eine <u>Skala</u> eine homomorphe Abbildung von einem empirischen Relativ in ein numerisches Relativ dar. Ob eine solche Abbildung möglich ist, ist Gegenstand des <u>Repräsentationsproblems</u>. Zu fragen ist, welche Bedingungen im empirischen Relativ erfüllt sein müssen, damit eine Skala eines bestimmten Typs vorliegt. Die Antwort wird durch <u>Repräsentationstheoreme</u> der unterschiedlichsten Art gegeben. So sichert z.B. im Falle des Paarvergleichs die Transitivität der empirischen Präferenzen, daß eine Rangordnung gebildet werden kann, d.h. daß eine strukturtreue Abbildung in ein numerisches Relativ (Ordnungsrelation) möglich ist. Das zugehörige Repräsentationstheorem lautet in diesem Falle: Ist die in einem empirisichen Relativ gemessene Relation konnex (der Paarvergleich ergibt für jedes Paar eine eindeutige Präferenz des Befragten) und transitiv (wie oben erläutert), so existiert eine homomorphe Abbildung in ein numerisches Relativ mit Ordnungsrelation.

Allerdings gibt es viele Abbildungen in ein numerisches Relativ die die Rangordnungsinformation der empirischen Präferenzen erhalten, also strukturtreu sind. Daraus resultiert das <u>Eindeutigkeitsproblem</u>. Die Frage ist, welche Transformationen im numerischen Relativ zulässig sind, ohne daß die Homomorphie-Eigenschaft (Strukturtreue) der Abbildung verloren geht. Im Falle der Ordinalskala umfaßt die Klasse der erlaubten Transformationen alle monotonen Transformationen (siehe Schaubild weiter unten).

Bei der Auswertung der Meßergebnisse stellt sich das Bedeutsamkeitsproblem, also die Frage, welche statistischen Verfahren bei einem gegebenen Meßniveau angewandt werden können. So ist z.B. auf Ordinalskalenniveau eine Mittelwertbildung nicht sinnvoll, da die Abstände zwischen den einzelnen Rangplätzen nicht metrisch interpretiert werden können. Dagegen enthält der Median eine sinnvolle Information, da er bei jeder erlaubten Transformation (hier monotone Transformation) strukturtreu transformiert wird. Angaben zu einigen erlaubten statistischen Verfahren enthält die folgende Übersicht über die verschiedenen Skalenniveaus:

Die <u>vier wichtigsten Skalentypen</u>: Informationsgehalt, zulässige Transformationen und geeignete statistische Verfahren (nach Heidenreich 1987)

	nicht-metri	sche Skalen	metrisch	ne Skalen
Skalentyp	Nominalskala	Ordinalskala	Intervallskala	Verhältnisskala (Ratioskala)
empirische Operationen	Gleichheit und Ungleichheit	zusätzlich: Rangfolge, z.B. D > B > A > C	zusätzlich: gleiche Intervalle (z.B. D-B ≈ A-C); willkürlich festge- legter Nullpunkt	zusätzlich: gleiche Verhält- nisse $(z.B. \frac{x}{y} \approx \frac{k}{1});$ absoluter Null- punkt
zulässige Trans- formationen	Umbenennung, Permutation	monoton steigende Transformationen	lineare Transfor- mationen: f'(x)=v+u f(x) (wobei u > 0)	Ähnlichkeitstrans- formationen f'(x)=u f(x) (wobei u >0)
Statistische Maß- zahlen (Beispiele)	Häufigkeiten, Modalwert	zusätzlich: Median, Quartile, Prozentrangwerte, Spannweite	zusätzlich: arithmetisches Mittel ( $\overline{x}$ ), Stan- dardabwei- chung (s), Schiefe, Exzeß	zusätzlich: geometrisches Mit- tel, Variabilitätsko- effizient
Zusammenhangs- maße (Beispiele)	Cramers V, Vier- felderkoeffizient (Phi), Prozent- satzdifferenz (d%)	zusätzlich: RangkorrKoeffi- zienten (Tau, Gamma, Rho)	zusätzlich: Produkt-Moment- Korrelation (r), Regressions- koeffizient	
Beispiele	Numerierung von Fußballspielern, Geschlecht (Dichotomie), Konfession, Fami- lienstand	Richtersche Erdbe- bebenskala, Schul- noten, Bildung, Guttman-Skalen	Temperatur nach Celsius, Alter, Einkommen, Bil- dung (?), Likert- Skalen	Länge, Zeit, Temperatur nach Kelvin, Einkommen (?)

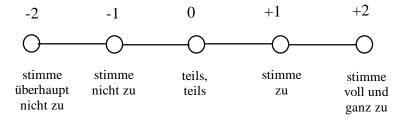
Die Meßniveaus verhalten sich also kumulativ, d.h., der Informationsgehalt nimmt mit steigendem Meßniveau zu. So verfügt z.B. die Intervallskala über die Eigenschaften einer Ordinalskala (Rangordnung) und zusätzlich noch über eine Abstandsinformation. Empirisch getestet wird diese Eigenschaft allerdings nur im Falle der Repräsentationsmessung, bei der Indexmessung wird diese Eigenschaft qua Konvention festgelegt (so z.B., wenn man bei der Likert-Skala voraussetzt, daß der Befragte in der Lage ist, seine Zustimmung oder Ablehnung der vorgelegten Items auf einer Rating-Skala in gleichen Abständen auszudrücken (vgl. 5.3.1)). Über das Meßniveau einer Skala kann also nur im Falle der Repräsentationsmessung eindeutig entschieden werden. Echte Intervallskalen in diesem Sinne sind in den Sozialwissenschaften sehr selten, können aber z.B. mit dem Modell der additiv verbundenen Messung konstruiert werden. Darüberhinaus zeigen empirische Tests der Intervallskalen-Eigenschaft von Rating-Skalen, die in methodischen Untersuchungen durchgeführt wurden, daß die Befragten in der Regel durchaus in der Lage sind, Rating-Skalen konsistent als Intervallskalen zu interpretieren (vgl. z.B. Westermann 1985). Daher besteht eine gewisse Rechtfertigung für die konventionelle Interpretation von Rating-Skalen als Intervallskalen.

# 5.3 Einige gebräuchliche Skalierungsverfahren

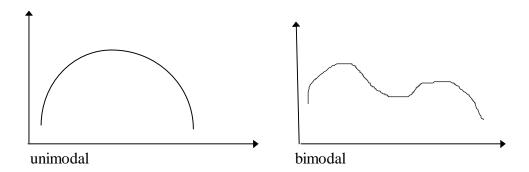
### 5.3.1 Likert-Skala

Die <u>Likert-Skalierung</u> ist die in den Sozialwissenschaften am häufigsten verwandte Skalierungstechnik. Sie ist nicht einfach mit der <u>Methode der summierten Ratings</u> identisch, wie oft irreführend in den Lehrbüchern behauptet wird, sondern umfaßt zusätzlich eine <u>Itemanalyse</u>, die es ermöglicht, Items auszusondern, wenn deren <u>Schwierigkeit</u> und <u>Trennschärfe</u> nicht den Anforderungen an eine Likert-Skala entsprechen.

<u>Rating-Skalen</u> sind Schätz-Skalen, auf denen Befragte ihre Einschätzung (Rating) einer (z.B. in einem Item der Likert-Skala) vorgelegten Aussage ausdrücken können. Man unterscheidet zwischen rein numerischen, graphischen und verbal verankerten Rating-Skalen (vgl. Rohrmann 1978). Vorzuziehen sind Skalen, die sowohl numerisch, als auch verbal chrakterisiert sind, z.B.:



Um bei den Befragten den Eindruck gleicher Abstände zwischen den Antwortkategorien zu verstärken, hat sich die graphische Unterstützung der Skala durch ein Lineal mit gleichen Abständen zwischen den Antwortvorgaben bewährt. Unterschiedliche Meinungen bestehen über die Verwendung einer mittleren Antwortkategorie. Während in psychologischen Tests meistens mit einer geraden Zahl von Antwortkategorien gearbeitet wird (Mittelkategorie entfällt) um die Testperson zu einem profilierten (positiven oder negativen Urteil) zu zwingen, werden in sozialwissenschaftlichen Anwendungen ungerade Skalen vorgezogen, um den Befragten die Möglichkeit einer neutralen Antwort offen zu lassen. Auf diese Weise soll der Gefahr einer Erhebung von Nicht-Einstellungen (geäußerten positiven oder negativen Meinungen, obwohl gar keine profilierte eigene Meinung vorliegt) vorgebeugt werden. In Abhängigkeit von der Thematik sollte aber auch hier auf das Hinzufügen einer expliziten Weiß-Nicht-Kategorie nicht verzichtet werden (vgl. 6.1). Rating-Skalen sollten nicht zu lang sein, in der Regel also nicht mehr als 5 oder 6 Antwortkategorien umfassen. Anderenfalls besteht die Gefahr von bimodalen Verteilungen der Antworten, d.h. der Konzentration auf Extremkategorien, insbesondere dann, wenn die Befragten nicht über eine sehr differenzierte Einstellung (hohes Maß an kognitiver Differenziertheit) verfügen.



Vorgehensweise bei der Konstruktion einer Likert-Skala am Beispiel der Messung der normativen Geschlechtsrollenorientierung (nGRO)

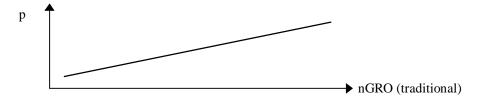
#### 1.Konzeptspezifikation und inhaltliche Analyse des theoretischen Konstrukts

Die normative Geschlechtsrollenorientierung umfaßt alle Vorstellungen einer Person über angemessenes Verhalten von Männern und Frauen in sozialen Situationen, etwa in den Bereichen der beruflichen Tätigkeit, der Kindererziehung, der Arbeitsteilung in Ehe und Partnerschaft, der politischen Betätigung usw.

#### 2.Sammlung von Items

Zu den spezifizierten Bereichen wird nun eine größere Anzahl von Items gesammelt. In den Items werden Aussagen formuliert, von denen angenommen wird, daß sie das zugrundeliegende theoretische Konstrukt messen. Die Zustimmung (bzw. Ablehnung) der Befragten zu diesen Aussagen wird auf einer (meist fünfstufigen) Rating-Skala gemessen. Die Items sollten eine monotone Itemcharakteristik haben, d.h. die Wahrscheinlichkeit der Zustimmung zu einem Item sollte mit zunehmender Ausprägung der theoretischen Variablen (hier willkürlich als traditionale nGRO festgelegt) entweder monoton zunehmen (positives Item) oder monoton abnehmen (negatives Item). Ein positives Item wäre etwa das Item 6: "Eine Frau sollte gegenüber Männern nicht zu strebsam sein" (vgl. Tabelle weiter unten).

Beispiel für eine monotone Itemcharakteristik (positives Item):



Ein negatives (gedrehtes) Item formuliert demgegenüber eine Aussage, bei der die Zustimmung eine emanzipierte nGRO ausdrückt (monoton fallende Itemcharakteristik), etwa das Item 3: "Der Anteil an Zeit und Energie, den jemand seiner Karriere einerseits und der Familie andererseits zukommen läßt, sollte von Interessen und nicht vom Geschlecht der

Person bestimmt sein". Um der Ja-Sage-Tendenz (siehe 6.1) zu begegnen, sollte eine Likert-Skala sowohl positive wie auch negative Items enthalten.

#### 3. Bildung einer vorläufigen Skala (Methode der summierten Ratings)

Die gesammelten Items werden nun einer Eichstichprobe vorgelegt. Anschließend wird die vorläufige Skala gebildet, indem die Antworten der Befragten zunächst numerisch kodiert und dann über alle Items aufsummiert werden. Dabei müssen die Codes für negative Items zunächst gedreht werden, damit die Summe der Ratings in konsistenter Weise interpretiert werden kann.

#### 4. Itemanalyse

[Zur Itemanalyse mit SPSS für Windows (Reliabilitätsanalyse, Faktorenanalyse, t-Test) gibt es Beispieldateien (Acrobat). Diese Dateien können über das WWW bezogen werden]

Der entscheidende Schritt bei der Bildung einer Likert-Skala ist die Itemanalyse. Hier werden alle Items der vorläufigen Skala einer Qualitätskontrolle unterworfen, um ungeeignete Items auszuscheiden. Ungeeignet sind erstens Itmes mit einer zu hohen oder auch zu geringen Schwierigkeit. Die Schwierigkeit eines Items (oder einer Aufgabe bei einem Intelligenztest) entspricht dem Prozentsatz der Befragten, die dem Item zustimmen (oder die Aufgabe lösen können). Extreme Items, die entweder von nahezu allen Befragten abgelehnt werden, oder denen nahezu alle Befragten zustimmen, diskriminieren nicht im Hinblick auf das zu messende theoretische Konstrukt und sind daher vor Bildung der endgültigen Skala auszusondern.

Auch wenn ein Item über eine Schwierigkeit im mittleren Bereich (etwa 35-65%; einige extremere Items im Bereich von 20-80% sollten allerdings auch in der Skala enthalten sein, um zwischen extremeren Einstellungen diskriminieren zu können) verfügt, ist damit noch nicht gesagt, daß das Item auch das zugrundeliegende theoretische Konstrukt mißt. So kann man z.B. fragen, ob die Zustimmung oder Ablehnung zu Item 2: "Die Idee, daß Frauen und Mädchen in der Fußball-Bundesliga spielen, ist einfach lächerlich" etwas über die nGRO der Befragten aussagt. Daher muß zweitens die Trennschärfe der Items überprüft werden. Items, die von Personen mit sehr unterschiedlichen Einstellungen zur nGRO ähnlich beantwortet werden, sind nicht in der Lage, hinsichtlich der zu messenden latenten Variablen zu diskriminieren - das Item besitzt keine Trennschärfe in Hinblick auf das zu messende theoretische Konstrukt. Es gibt zwei Möglichkeiten zur Bestimmung der Trennschärfe (Trennschärfeindizes):

• <u>t-Test</u>: Aus der Eichstichprobe wird in Hinblick auf die vorläufige Skala (Gesamtpunktzahl <u>aller</u> Items) die obere Gruppe (die 25% der Befragten mit der höchsten Punktzahl) und die untere Gruppe (die 25% der Befragten mit der niedrigsten Punktzahl) gebildet. Für <u>jedes</u> Item wird nun mit Hilfe eines t-Tests untersucht, ob sich der Mittelwert für dieses Item in der oberen Gruppe signifikant von dem entsprechenden Mittelwert in der unteren Gruppe

unterscheidet (<u>einseitiger</u> Signifikanztest). Ist dies der Fall, diskriminiert also das Item zwischen der oberen und unteren Gruppe, ist das Item brauchbar und wird in die endgültige Skala aufgenommen. Anderenfalls wird es ausgesondert (vgl. Mayntz u.a. 1978, S. 56f.).

• <u>korrigierte Item- (vorläufige) Gesamtskala-Korrelation</u>: Da man davon ausgehen kann (zumindest gilt dies bei einer theoretisch fundierten inhaltlichen Auswahl der Items), daß die vorläufige Gesamtskala in etwa das zu messende theoretische Konstrukt widerspiegelt, sollte ein trennscharfes Item I (je nach Polung entweder positiv oder negativ) mit der vorläufigen Gesamtskala S - I, die vorher um das zu testende Item bereinigt wurde, korrelieren. Üblicherweise wird gefordert, daß der Trennschärfekoeffizient r<sub>I, S-I</sub> einen absoluten Betrag von mindestens .5 aufweisen sollte, damit das Item I in die endgültige Skala aufgenommen werden kann. Als Beispiel betrachten wir einen Auszug aus der Itemanalyse bei der Bildung einer Likert-Skala zur nGRO (vgl. Krampen 1979).

Beispiel: Schwierigkeitsindizes (p<sub>i</sub>), Trennschärfekoeffizienten (r<sub>i,t-i</sub>) und Faktorladungen ausgewählter Items der nGRO-Skala (nach Krampen 1979)

			Fal	ktor		
Item	$p_i$	$r_{i,t-i}$	I	II	III	IV
1. Es ist für eine Frau wichtiger, den Ehemann bei seiner Karriere zu unterstützen, als selbst Karriere zu machen.	.49	.50	.55	.03	.15	19
2. Die Idee, daß junge Frauen und Mädchen in der Fußball-Bundesliga spielen, ist einfach lächerlich.	.44	.26	.16	.06	.46	19
3.* Der Anteil an Zeit und Energie, den jemand seiner Karriere einerseits und der Familie andererseits zukommen läßt, sollte von Interessen und nicht von dem Geschlecht der Person bestimmt werden.	.62	36	23	46	.08	02
4. Es ist für eine Frau wichtiger, eine gute Figur zu haben und gut gekleidet zu sein, als für einen Mann.	.57	.35	.47	04	.08	18
5. Der alte Ausspruch "die Frau gehört in's Haus und zur Familie" ist im Grunde richtig, und es sollte auch so bleiben.	.50	.56	.72	.16	.12	08
6. Eine Frau sollte gegenüber Männern nicht zu strebsam sein.	.52	.55	.58	.20	.01	04
7. Eine Frau, die in der Öffentlichkeit arbeitet (z.B. eine Verkäuferin), sollte nicht mehr arbeiten, wenn zu sehen ist, daß sie schwanger ist.	.54	.59	.59	.12	.02	18
8. Der Mann solte in einer Familie für alle größeren Entscheidungen verantwortlich und zuständig sein.	.54	.63	.70	.33	.10	.03
9. In einer Gruppe mit weiblichen und männlichen Mitgliedern sollte ein Mann die Führungsposition innehaben.	.49	.61	.70	.23	.15	14
10. Verheiratete Frauen, die Kinder im Schulalter haben, sollten nicht arbeiten, es sei denn, es ist für die Familie finanziell unbedingt notwendig.	.44	.22	.50	46	.41	.11
11. Wenn sich ein gut qualifizierter Mann und eine Frau um eine Arbeitsstelle bewerben, so sollte der Mann die Stelle	.48	.54	.67	02	.32	04

erhalten, da er seine Familie ernähren muß.

Wie ersichtlich, verfügt Item 6 über eine ausreichende Trennschärfe, während Item 2 (vermutlich aus inhaltlichen Gründen) und Item 3 (vermutlich aus Gründen der zu komplizierten Frageformulierung) über eine unzureichende Trennschärfe verfügen.

#### 5. Bilden der endgültigen Skala

Alle Items, die den Anforderungen der Likert-Skalierung in Hinblick auf Schwierigkeit und Trennschärfe nicht entsprechen, werden ausgesondert und die endgültige Skala erneut mit der Methode der summierten Ratings aus den verbleibenden qualitätsgeprüften Items gebildet.

Die Likert-Skala zeichnet sich zwar durch eine Oualitätskontrolle (Schwierigkeitsindex, Trennschärfeindex) aus, die Voraussetzungen des Meßmodells werden aber nicht empirisch überprüft. Es wird zwar qua Konvention davon ausgegangen, daß die Befragten in der Lage sind, Ratingskalen als Intervallskalen zu interpretieren und ihre Zustimmung oder Ablehnung zu den Items gemäß der Konstruktion der Ratingskala in gleichen Abständen auszudrücken. Ein interner Konsistenztest, wie er für eine Repräsentationsmessung erforderlich wäre, findet also nicht statt. Es handelt sich bei der Likert-Skalierung daher um Indexmessung auf Intervallskalenniveau. (Die von einigen Lehrbüchern vertretene "vorsichtige" Strategie, die endgültige Skala lediglich als Ordinalskala zu interpretieren, ist in sich widersprüchlich, da bei der Konstruktion der Skala ausgiebig von der Intervallskalenqualität der Ratings gebrauch gemacht wird, z.B. beim Summieren der Ratings oder bei der Berechnung der Trennschärfekoeffizienten.) Als Indexmessung muß jede Likert-Skala besonders kritisch in Bezug auf die verschiedenen Gütekriterien einer Messung, also im Hinblick auf Objektivität (z.B. eventuelle Interviewereffekte durch das Geschlecht des Interviewers; vgl. 6.1), Reliabilität (z.B. durch Berechnung von Cronbachs-α als Reliabilitätsmaß; vgl. 5.4.1.1) und Validität (z.B. durch die Überprüfung der Kriteriumsvalidität; vgl. 5.4.1.2) geprüft werden.

#### 5.3.2 Guttman-Skala

Zur Bildung einer Guttman-Skala bentötigt man Items, die in Bezug auf das zu messende theoretische Konstrukt (z.B. die Einstellung zur Atomkraft) immer extremer werden. Die Grundidee besteht darin, auf der zu messenden Dimension eine <u>ineinandergreifende Ordnung</u> von Items und Befragten zu bilden. Dazu ist es erforderlich, daß sich die Items so ordnen lassen, daß, falls ein Befragter ein bestimmtes Item dominiert (also der Aussage zustimmt oder die Aufgabe löst), er auch alle Items dominiert, die unterhalb dieses Items angeordnet sind, und falls er ein bestimmtes Item nicht dominiert, er auch alle Items oberhalb dieses Items nicht dominiert. So lassen sich z.B. Rechenaufgaben nach ihrem Schwierigkeitsgrad anordnen; bilden diese Aufgaben eine echte ineinandergreifende Ordnung, dann kann erwartet

werden, daß eine Person, die eine bestimmte Aufgabe löst, auch alle leichteren Aufgaben löst, und eine Person, die eine bestimmte Aufgabe nicht löst, auch alle schwierigeren nicht löst.

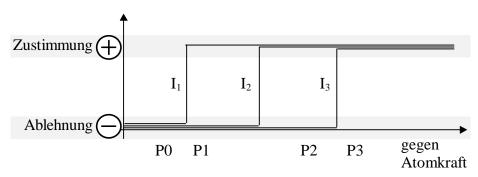
Ist eine solche ineinander greifende Ordnung möglich, so kann jede Person auf der zu messenden Dimension genau zwischen zwei Items plaziert werden, nämlich zwischen dem letzten Item, das sie dominiert und dem ersten Item, das sie nicht dominiert. Zur Konkretisierung des Gesagten betrachten wir die folgenden Items zur Messung der Einstellung zur Atomkraft (vgl. Schnell/Hill/Esser 1992, S. 208):

I<sub>1</sub>: Es sollten keine weiteren Atomkraftwerke geplant werden

I<sub>2</sub>: Es sollten keine weiteren Atomkraftwerke in Betrieb genommen werden

I<sub>3</sub>: Alle Atomkraftwerke sollten sofort abgeschaltet werden.

Die Items werden Befragten vorgelegt, die zu jedem Item entweder zustimmend (+) oder ablehnend (-) Stellung beziehen. Ausgangspunkt des Meßmodells ist das folgende Diagramm, das die für eine Guttman-Skala typischen <u>Itemcharakteristiken</u> der drei Items enthält:



Wie man sieht, handelt es sich bei den Itemcharakteristiken um Treppenfunktionen, die an einem exakt definierten Schwellenwert von Ablehnung in Zustimmung umspringen; es handelt sich um eine spezielle Form einer monton steigenden Itemcharakteristik. Der Guttman-Skala liegt also ein <u>deterministisches Meßmodell</u> zugrunde, denn jede Person, die in Hinblick auf die latente Variable einen niedrigeren Wert als den Schwellenwert des Items aufweist, wird dieses Item mit Sicherheit ablehnen, während eine Person, die oberhalb des Schwellenwertes positioniert ist, mit Sicherheit zustimmen wird. Damit ergibt sich der Skalenwert eines Befragten aus der Nummer desjenigen Items, das er zuletzt akzeptiert. Bezogen auf unser Schema heißt das, daß die Person P<sub>1</sub>, die zwischen I<sub>1</sub> und I<sub>2</sub> positioniert ist, dem ersten Item zustimmt, während sie das zweite und dritte Item ablehnt. Daraus ergeben sich die folgenden erlaubten Antwortmuster:

	$I_1$	$I_2$	$I_3$	Skalenwert
$P_3$	+	+	+	3
$P_2$	+	+	-	2
$P_1$	+	-	-	1
$P_0$	-	-	-	0

Wie man sieht, kann man aus dem Skalenwert auf das zugeordnete Antwortmuster zurückschließen. So entspricht dem Skalenwert 2 das Antwortmuster ++-, d.h. den ersten beiden Items wird zugestimmt, während das dritte Item abgelehnt wird. Insgesamt sind bei n Items genau 2<sup>n</sup> Antwortmuster denkbar. Davon sind aber immer nur n+1 Antwortmuster nach der Logik der Guttman-Skalierung erlaubt. Bei drei Items bedeutet das acht mögliche, aber nur vier erlaubte Antwortmuster.

Die Guttman-Skala sieht also eine <u>interne Konsistenzüberprüfung</u> der Voraussetzungen des Meßmodells vor. Dazu bedient man sich des Verfahrens der <u>Skalogramm-Analyse</u>, etwa in der Form der Cornell-Technik oder der Goodenough-Edwards-Technik. Im folgenden erläutern wir kurz die Goodenough-Edwards-Methode, die einfacher zu handhaben ist, da sie davon ausgeht, daß der Skalenwert einer Person während der Skalogramm-Analyse nicht verändert wird. Ausgangspunkt ist die ursprüngliche Datenmatrix, die die Antworten der Befragten enthält - die Personen bilden die Zeilen und die Items die Spalten der Tabelle. Personen und Items werden nun jeweils so umgeordnet, daß sich eine absteigende Reihenfolge ergibt. Vergleicht man nun für jede Person das vorliegende Antwortmuster mit dem gemäß dem Skalenwert nach der Logik der Guttman-Skala erlaubten Antwortmuster (vgl. die Tabelle oben), so muß jede Nicht-Übereinstimmung als Fehler gezählt werden, da sie ein Abweichen von dem Idealmuster der Guttman-Skala darstellt. Dazu das folgende Beispiel:

ursprüng	liche	Daten-

matrix

	$J_1$	$J_2$	$J_3$	Σ
$P_1$	+	-	+	2
$P_2$	+	+	+	3
<b>P</b> <sub>3</sub>	-	-	-	0
P <sub>4</sub>	+	+	-	2
P <sub>5</sub>	+	-	-	1
$P_6$	+	+	+	3
<b>P</b> <sub>7</sub>	+	-	+	2
P <sub>8</sub>	+	-	-	1
Σ	7	3	4	-

zeilen- und spaltenweise umgeordnete Matrix

	$J_1$	$J_3$	$J_2$	Σ
$P_2$	+	+	+	3
$P_6$	+	+	+	3
$P_1$	+	+	-	2
$P_4$	+	-	$\oplus$	2
<b>P</b> <sub>7</sub>	+	+	-	2
P <sub>5</sub>	+	-	1	1
$P_8$	+	-	-	1
$P_3$	-	-	-	0
$\sum_{i}$	7	4	3	

In unserem Beispiel treten genau 2 Fehler auf. Mit Hilfe der Skalogramm-Analyse ist also eine Überprüfung der Voraussetzungen des Meßmodells der Guttman-Skalierung möglich. Je mehr Fehler auftreten, desto größer ist die Abweichung der tatsächlichen von den erlaubten Antwortmustern in der vorliegenden empirischen Untersuchung. Es handelt sich bei der Guttman-Skalierung daher um Repräsentationsmessung, da eine empirische Überprüfung der Modellannahmen (interner Konsistenztest der Skala) möglich ist. Ein Maß für die interne Konsistenz ist der Reproduzierbarkeitskoeffizient (Rep):

Rep = 
$$1 - \frac{\text{Anzahl der Fehler}}{(\text{Anzahl der Befragten}) \cdot (\text{Anzahl der Items})}$$

In unserem Falle ergibt sich ein Reproduzierbarkeitskoeffizient von

Rep = 
$$1 - \frac{2}{8 \cdot 3} = 1 - \frac{2}{24} = \frac{11}{12} = .917$$

In der Forschungspraxis hat es sich eingebürgert, bis zu 10% Fehler als akzeptabel anzusehen. In unserem Falle liegt der Reproduzierbarkeitskoeffizient mit .917 über dem Schwellenwert von .9. Allerdings muß betont werden, daß es sich bei diesem Schwellenwert um eine reine Konvention handelt. Da die Anzahl der Fehler auch von den Prozentsätzen der Zustimmung für die Items der Guttman-Skala abhängig ist, sollte man in der Forschungspraxis Skalierbarkeitskoeffizienten verwenden, die diese Einflüsse konstant halten (vgl. McIver und Carmines 1981, Kap. 5).

Bei der Guttman-Skalierung können auch negative Items verwendet werden. In unserem Beispiel würde z.B. die Aussage "Die Atomkraft ist die sicherste Energiequelle überhaupt" ein

negatives Item darstellen. Die entsprechende Itemcharakteristik wäre ebenfalls deterministisch, aber monoton fallend. Der Schwellenwert für dieses Item sollte links von Item I<sub>1</sub> in dem obigen Schema liegen. Zur Überprüfung auf Guttman-Skalierbarkeit müßten die Antworten zunächst umcodiert werden, damit sich die obige Aussage als I<sub>0</sub> in das Schema der Skalogramm-Analyse einfügen würde.

Auch bei der Guttman-Skalierung kann es sinnvoll sein, einzelne Items aus der Guttman-Skala auszuschließen, um so den Reproduzierbarkeitskoeffizienten zu erhöhen. Stellt man nämlich während der Skalogramm-Analyse fest, daß sich die Fehler bei einem bestimmten Item häufen, so ist dies ein Zeichen dafür, daß sich das besagte Item nicht in die Guttman-Logik einfügt. Durch Elimination dieses Items kann es durchaus zu einer Verbesserung des Reproduzierbarkeitskoeffizienten kommen, obwohl sich natürlich auch der Nenner des Reproduzierbarkeitskoeffizienten (Anzahl der Items) verringert.

Bei der Guttman-Skalierung handelt es sich um Repräsentationsmessung auf Ordinalskalenniveau. Die Messung kann scheitern, wenn sich in einer empirischen Untersuchung Antwortmuster häufen, die nach der Guttman-Logik verboten sind. Besteht die Skala aber den internen Konsistenztest, so kann der Forscher sicher sein, daß eine ineinandergreifende Ordnung von Personen und Items auf der untersuchten Dimension möglich ist. Da durch das Meßmodell lediglich die Ordnung der Items festgelegt wird, aber nicht die Abstände zwischen ihnen, handelt es sich um eine Messung auf Ordinalskalenniveau.

[Wegen ihrer sehr restriktiven Voraussetzungen wird die Guttman-Skalierung in der Praxis nur sehr selten angewandt. Es existiert aber ein Meßmodell, daß in gewissem Sinn als eine Erweiterung der Guttman-Skalierung angesehen werden kann und daß in der aktuellen meßtheoretischen Diskussion von großer Bedeutung ist, nämlich das <u>Rasch-Modell</u>. Der Rasch-Skalierung liegt eine besondere Form monotoner Itemcharakteristiken zugrunde. Im Gegensatz zum deterministischen Guttman-Modell folgt die Wahrscheinlichkeit der positiven Beantwortung eines Items hier einer logistischen Funktion. Da sich das Rasch-Modell ebenfalls auf dichotome Items bezieht, wird es auch als dichotom-logistisches Modell bezeichnet. Für eine eingehendere Behandlung des Rasch-Modells muß auf die weiterführende Literatur verwiesen werden (vgl. z.B. Andrich 1988).]

# 5.3.3 Semantisches Differential (Polaritätsprofil)

Eine Skalierungstechnik, die ebenfalls auf Rating-Skalen beruht, ist das sogenannte Semantische Differential. Es besteht aus einer Anzahl bipolarer Skalen, deren Endpunkte durch Adjektive gekennzeichnet sind, die die Enden eines semantischen Kontinuums angeben. Die Befragten werden gebeten, ihr Urteil über ein bestimmtes Objekt abzugeben, um so dessen konotative Bedeutung einschätzen zu können. Das resultierende Polaritätsprofil kann z.B. dazu dienen, das Image eines bestimmten Objektes zu erfassen. Die verwendeten

Gegensatzpaare haben dabei meist nur einen metaphorischen Bezug auf das zu beurteilende Objekt. Ein typisches Semantisches Differential könnte etwa die folgenden Skalen umfassen:

		schlecht
		unsozial
		unharmonisch
		nieder
		weich
		leicht
		klein
		nachgiebig
		passiv
		ruhig
		kalt
		stumpf

1,0	0	0
,42	-,19	,18
,49	-,12	,01
,45	,07	,17
-,24	,97	0
-,2	,48	-,02
,09	,21	-,05
-,06	,34	,06
,17	,12	,98
-,15	,03	,26
,12	,09	,26
-,06	,17	,29

В

S

A

Je nach Untersuchungsgegenstand ist es natürlich möglich, die Skalen des Semantischen Differentials objektspezifisch zu ergänzen, z.B. im Falle der Messung eines spezifischen Markenimages. Unterzieht man die Skalen eines Semantischen Differentials einer Faktorenanalyse (vgl. 5.4.2), so erkennt man, daß die unterschiedlichen Gegensatzpaare im wesentlichen drei Dimensionen messen, nämlich Bewertung (B), Stärke (S) und Aktivität (A). Die entsprechenden Faktorladungen für die Gegensatzpaare des Semantischen Differentials sind in der obigen Abbildung zusätzlich aufgeführt. Eine zusätzliche Auswertungsmöglichkeit ist der Vergleich der Polaritätsprofile verschiedener Bewertungsobjekte, etwa bei der Bestimmung der Markenimages in einem bestimmten Produktbereich. Die Polaritätsprofile zweier Marken im semantischen Raum können etwa wie folgt schematisch dargestellt werden:

Gegensatzpaare der Dimension der <b>B</b> ewertung	gut sozial •			schlecht unsozial •
Gegensatzpaare	weich			hart
der Dimension	•		`\	•
der <b>S</b> tärke	•		<i>)</i>	•
	•		· · · · · · · · · · · · · · · · · · ·	•
Gegensatzpaare	passiv			aktiv
der Dimension	•		/	•
der ${f A}$ ktivität	•	,		•
	•	\ /		•

Zur Auswertung wird üblicherweise zunächst die Semantische Distanz zwischen den einzelnen Polaritätsprofilen berechnet. Dazu werden die Differenzen der Ratings für die einzelnen semantischen Gegensatzpaare gebildet, quadriert und über alle Gegensatzpaare aufsummiert (Euklidische Distanz). Zur Bestimmung der Dimensionen, die den semantischen Raum der untersuchten Markenimages aufspannen, kann dann eine multidimensionale Skalierung durchgeführt werden (vgl. 5.4.3).

Da das Semantische Differential Rating-Skalen verwendet, eine interne Konsistenzüberprüfung aber nicht stattfindet, handelt es sich um <u>Indexmessung</u> auf Intervallskalenniveau.

# 5.4 Klassische Testtheorie und dimensionale Analyse

Ausgangspunkt der klassischen Testheorie ist die Annahme, daß das Testergebnis einer Person, bzw. die Erfassung einer spezifischen Einstellungsstärke dieser Person mit einem Testinstrument fehlerbehaftet ist. Jeder empirischen Messung entspricht deswegen ein wahrer Wert, der sich von dem beobachteten durch einen Meßfehler unterscheidet. Der Grundgedanke der klassischen Testtheorie besteht daher in dem Versuch, empirisch gegebene Messungen analytisch von möglichen Meßfehlern zu reinigen (5.4.1). Die Reliabilität einer Messung kann dann als der Anteil der in der Messung enthaltenen Varianz des wahren Wertes bestimmt werden (5.4.1.1). In einem zweiten Schritt wird dann versucht, die Validität der bereinigten Meßwerte zu überprüfen. Dies ist nur mit Hilfe von externen Kriterien möglich, die im Rahmen einer Theorie mit dem zu messenden Konstrukt verbunden sein müssen. Der Ansatz der klassischen Testtheorie ist daher einzubetten in eine umfassende Meßfehlertheorie, die sowohl Zufalls- als auch systematische Fehler berücksichtigt. Gerade dies leistet der kausalanalytische Ansatz, der es ermöglicht, die Beziehungen zwischen theoretischen Konstrukten (Kerntheorie), die Beziehungen zwischen theoretischen

Konstrukten und Indikatoren (Meßtheorie) und die Einflüsse von Zufalls- und systematischen Fehlern (Meßfehlertheorie) in einem Kausalmodell systematisch zu erfassen (5.4.1.2).

Um die <u>Dimensionalität</u> von Skalen zu bestimmen, muß man auf Verfahren der Faktorenanalyse zurückgreifen. Auch die <u>Faktorenanalyse</u> kann als ein Spezialfall des kausalanalytischen Modells verstanden werden (5.4.2). Ein weiteres, im Vergleich zur Faktorenanalyse weniger voraussetzungsvolles Verfahren zur dimensionalen Analyse ist die <u>multidimensionale Skalierung</u> (5.4.3).

## 5.4.1 Reliabilität und Validität

Eine fehlerfreie Messung ist ein Ideal, dem man sich in der Praxis nur mehr oder weniger gut annähern kann. Das Ziel der klassischen Testtheorie (KT) ist es, den Grad dieser Annäherung an eine fehlerfreie Messung zu bestimmen. Die KT geht daher davon aus, daß jede Messung (X) aus einem wahren Wert (T) und einem Meßfehler (E) besteht, formal:

$$X = T + E. (1)$$

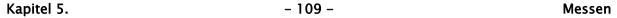
Die Grundidee der KT besteht darin, durch mehrmalige Messung und anschließende Mittelwertbildung Meßfehler herauszurechnen, um so den wahren Wert T (der während der Meßreihe unverändert bleiben soll) zu bestimmen:

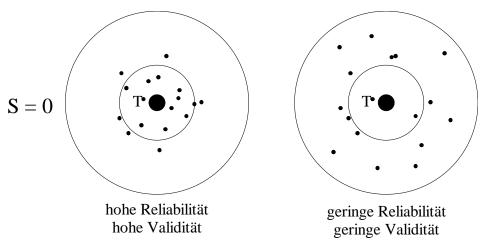
$$X_1 = T + E_1$$
 
$$\vdots \qquad \qquad \Rightarrow \overline{X} = T + \overline{E}$$
 
$$X_n = T + E_n$$

Um etwa die Zeit eines 100-Meter-Laufs zu messen, setzte man (in der guten alten Zeit der Handstoppung) mehrere Zeitnehmer (Paralleltest-Methode) ein und mittelte dann deren Meßergebnisse. Führt man bei einer Längenmessung mit einem Lineal mehrere Messungen hintereinander durch (Test-Retest-Methode), kann man ebenfalls durch Mittelwertbildung Ablesefehler ausgleichen. In beiden Fällen ist diese Vorgehensweise nur sinnvoll, wenn lediglich Zufallsfehler auftreten, die Messung also nicht durch systematische Fehler verzerrt ist. Systematische Fehler (etwa ein verbogenes Lineal oder die verzögerte Reaktion der Zeitnehmer auf den Start schuß) kann durch Mittelwertbildung nicht ausgeglichen werden. Zufallsfehler vermindern die Reliabilität einer Messung und damit natürlich auch die Validität, da eine Messung nur so valide sein kann, wie sie reliabel ist. Systematische Fehler stellen keine Beeinträchtigung der Reliabilität dar - eine Messung kann also in einem formalen Sinne genau, aber in einem inhaltlichen Sinne falsch sein. Erweitert man die Grundgleichung (1) der KT um einen systematischen Fehler S, so erhält man:

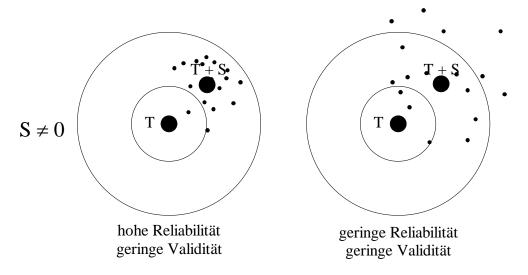
$$X = T + S + E. \tag{2}$$

In Bezug auf mangelnde Reliabilität und Validität kann man dann schematisch folgende Situationen unterscheiden:





In diesem Fall kann man nach der Logik der KT durch mehrmaliges Messen den Meßfehler (Zufallsfehler) durch Mittelwertbildung herausrechnen und damit die Genauigkeit (Reliabilität) der Messung (und wegen S=0 auch die Validität) erhöhen. Wird eine Skala mit mehreren Items konstruiert, die parallele Messungen des zugrundeliegenden theoretischen Konstruktes (T) darstellen, so kann man durch Hinzufügung von Items die Reliabilität (und damit auch die Validität) verbessern (Reliabilitätsverbesserung durch Testverlängerung; s.u.).



In diesem Fall ist eine hohe Validität wegen des systematischen Fehlers grundsätzlich nicht gegeben. Auch hier kann durch Testverlängerung eine hohe Reliabilität erreicht werden, es wird aber lediglich mit zunehmender Genauigkeit der systematisch verzerrte Wert T + S gemessen.

In der KT werden <u>Zufallsfehler</u> vorausgesetzt, d.h. man geht von der Grundgleichung (1) aus. Um systematische Fehler auszuschließen und um E als Zufallsfehler zu spezifizieren, müssen die folgenden Annahmen gemacht werden:

- Der Mittelwert des Meßfehlers ist gleich Null ( $\overline{E} = 0$ ).
- Der Meßfehler korreliert nicht mit dem wahren Wert der Messung ( $r_{TE} = 0$ ). Bezogen auf die Messung der nGRO mit einer Likertskala bedeutet dies, daß die Größe des Meßfehlers

unabhängig von einer eher liberalen oder einer eher traditionalen Einstellung des Befragten ist.

- Die Meßfehler verschiedener Messungen korrelieren nicht ( $r_{E_1E_2} = 0$ ). Diese Voraussetzung ist überaus weitreichend: Zunächst einmal bedeutet sie, daß die Meßfehler zweier Items der gleichen Skala (hier z.B. der nGRO) nicht miteinander korrelieren (unabhängige Messungen). Das gleiche muß aber auch für die Korrelation mit dem Meßfehler einer anderen Skala (z.B. für das Selbstwertgefühl oder für eine Tendenz zur sozialen Erwünschtheit) oder deren Items gelten.
- Der Meßfehler korreliert nicht mit dem wahren Wert einer anderen Messung. Damit sind systematische Einflüsse anderer Variablen auf die Messung ausgeschlossen. So darf es z.B. keinen Einfluß einer Tendenz zur sozialen Erwünschtheit (wahrer Wert einer anderen Messung) auf das Antwortverhalten bei den Items zur nGRO (Meßfehler) geben.

Wie man sieht, sind die Voraussetzungen der KT sehr restriktiv und in der Praxis mehr oder weniger unrealistisch. Im Rahmen einer Erweiterung der KT im kausalanalytischen Ansatz können diese Annahmen jedoch schrittweise gelockert werden (s. u.).

Zunächst halten wir fest, daß die KT Zufallsfehler voraussetzt und daher hervorragend zur Bestimmung der Reliabilität einer Skala geeignet ist. Da systematische Meßfehler aufgrund der Modellannahmen ausgeschlossen sind, kann die Validität einer Messung nur mit Hilfe externer Kriterien (Kriteriumsvalidität) oder durch die Einbettung des Konstruktes in eine komplexere theoretisch abgeleitete Kausalstruktur (Konstruktvalidität) überprüft werden. In diesem Zusammenhang können dann auch die restriktiven Annahmen der KT gelockert werden. So kann z.B. im Rahmen der Meßfehlertheorie im kausalanalytischen Ansatz eine Kausalbeziehung zwischen der Tendenz zur sozialen Erwünschtheit und den Meßfehlern der Items zur nGRO postuliert werden (dies hebt die vierte Annahme der KT wieder auf).

#### 5.4.1.1 Reliabilität

Nach diesem Vorgriff auf neuere Entwicklungen kommen wir auf die Bestimmung der Reliabilität im Rahmen der KT zurück. Die Genauigkeit (Reliabilität) einer Messung ( $R_X$ ) kann im Rahmen der KT spezifiziert werden als der Anteil der Varianz des wahren Wertes ( $s_T^2$ ) an der Gesamtvarianz der Messung ( $s_X^2$ ). Da E nicht mit T korreliert ist, ist die Gesamtvarianz gleich der Summe aus der Varianz des wahren Wertes und der Fehlervarianz ( $s_E^2$ ):

$$R_X = \frac{s_T^2}{s_X^2}$$
  $\left( = \frac{s_T^2}{s_T^2 + s_E^2} = 1 - \frac{s_E^2}{s_X^2} = r_{XT}^2 \right)$ 

Ist die Messung fehlerfrei ( $s_E^2 = 0$ ), so ist  $R_X = 1$ . Nimmt der Anteil der Fehlervarianz an der Gesamtvarianz zu (gleichbedeutend mit einem Sinken des Anteils der Varianz des wahren

Wertes), so sinkt die Reliabilität. Ist in der Messung schließlich nur noch Fehlervarianz enthalten ( $s_T^2 = 0$ ), so ist  $R_X = 0$ .

In der Praxis ist die Varianz des wahren Wertes unbekannt, die Reliabilität einer Messung muß daher auf andere Weise geschätzt werden. Dazu gibt es grundsätzlich zwei verschiedene Vorgehensweisen:

• <u>Test-Retest-Methode</u>: Die Reliabilität wird als zeitliche Stabilität durch die wiederholte Messung mit dem gleichen Meßinstrument bestimmt. Formal gilt:

$$T$$
 $\sqrt{R_X}$ 
 $X_{t_1}$ 
 $X_{t_2}$ 
 $X_{t_2}$ 
 $X_{t_1} = R_{X_{t_1}} = R_{X_{t_2}} = r_{X_{t_1}X_{t_2}}$ 

3 für den Zeitraum der Meßwiederholung die Stab

Dabei muß für den Zeitraum der Meßwiederholung die Stabilität des wahren Wertes vorausgesetzt werden. Bei mangelnder Stabilität des wahren Wertes wird nämlich die Reliabilität unterschätzt, da Schwankungen des wahren Wertes als Ungenauigkeiten der Messung interpretiert werden. Andererseits ist es auch denkbar, daß sich die Befragten um ein konsistentes Antwortverhalten bemühen (Erinnerungseffekte) und so einen systematischen Fehler verursachen, der zu einer Überschätzung der Reliabilität führt. Um diese Fehlerquellen auszuschalten, bedient man sich vorzugsweise der

• Paralleltest-Methode: Hier werden zum gleichen Zeitpunkt zwei vergleichbare Messungen des gleichen theoretischen Konstrukts vorgenommen. Dabei wird vorausgesetzt, daß es sich um exakt parallele Messungen handelt, d.h. die beiden Instrumente (z.B. Items)  $X_1$  und  $X_2$  verfügen über die gleichen Fehlervarianzen ( $s_{E_1}^2 = s_{E_2}^2$ ) und damit über die gleiche Reliabiliät. [Da wir uns bei unseren mathematischen Ableitungen auf standardisierte Variablen beschränken, vernachlässigen wir hier den Unterschied zwischen parallelen,  $\tau$ -äquivalenten und kongenerischen Messungen (vgl. z.B. Wegener 1983).] Dann gilt:

$$\begin{array}{c|c} & & & & \\ \hline \sqrt{R_x} & & & \\ \hline & & & \\ X_1 & & X_2 & & \\ \end{array} \qquad R_X = R_{X_1} = R_{X_2} = r_{X_1 X_2}$$

Die Anwendbarkeit der Paralleltest-Methode ist in dieser Form kaum gegeben, da sie an die Voraussetzung exakt paralleler (d.h. gleich reliabler) Meßinstrumente gebunden ist. Die Reliabilität soll aber gerade erst duch das Verfahren der Parallelmessung bestimmt werden. In der Praxis behilft man sich daher mit der Split-Half-Methode, die z.B. bei Likert-Skalen angewandt werden kann. Dazu splittet man den Gesamttest K in zwei gleichlange Testhälften  $K_1$  und  $K_2$ , wobei die Items des Tests zufällig einer der beiden Testhälften zugeordnet werden (z.B. nach der Odd-Even-Methode). Dadurch entstehen

zwei, bis auf Zufallsschwankungen parallele Testhälften, deren Reliabilität geschätzt werden kann:

$$R_{K_1} = R_{K_2} = r_{K_1K_2}$$

Mit Hilfe der Spearman-Brown-Formel kann dann die <u>Split-Half-Reliabilität</u> des Gesamttests wie folgt berechnet werden:

$$R_{K} = \frac{2 \cdot r_{K_{1}K_{2}}}{1 + r_{K_{1}K_{2}}}$$

Die dabei angewandte Korrekturformel kann im Hinblick auf einen Gesamttest K verallgemeinert werden, der aus N <u>exakt parallelen</u> Items mit einer Einzelreliabilität von  $R_{\rm I}$  besteht. Für die Reliabilität des Gesamttests K gilt dann:

$$R_{K} = \frac{NR_{I}}{1 + (N - 1)R_{I}}$$

dieser verallgemeinerten Spearman-Brown-Formel kann noch einmal der Grundgedanke der KT erläutert werden, nämlich das Herausmitteln von (Zufalls-)Meßfehlern. Je länger der Test ist, um so mehr Einzelmessungen (Items) gehen in die Mittelwertbildung ein und um so reliabler ist der Gesamttest. Diesen Effekt macht man sich bei der Reliabilitätsverbesserung durch Testverlängerung zunutze. Geht man z.B. von (exakt parallelen) Items mit einer Einzelreliabilität von .3 (70% Meßfehleranteil) aus, so hat ein zusammengesetzter Test der Länge N=10 bereits eine Reliabilität von .81 (nur noch 19% Meßfehleranteil).

Da die Split-Half-Reliabilität von dem jeweils konkret vorgenommenen Split der Gesamtskala abhängt, wird sie in der Forschungspraxis kaum verwendet. Stattdessen berechnet man Cronbachs  $\alpha$  als Maß für die sog. interne Konsistenz einer Skala (hat nichts mit dem internen Konsistenztest bei einer Repräsentationsmessung zu tun), das als Mittelwert aller möglichen Split-Half-Koeffizienten interpretiert werden kann. Dabei verwendet man den Mittelwert aller Korrelationen  $\mathbf{r}_{\mathbf{I}_i\mathbf{I}_j}$  der Items eines Tests der Länge N als Maß für die durchschnittliche Reliabilität der einzelnen Items und wendet dann die verallgemeinerte Spearman-Brown-Formel an:

$$\alpha = \frac{N \cdot \overline{r}_{I_i I_j}}{1 + (N - 1) \cdot \overline{r}_{I_i I_j}}$$

 $\alpha$  ist also eine Funktion der Anzahl der Items und der durchschnittlichen Interkorrelation der Items. Werden bei einer Likert-Skala ungeeignete Items (geringe Trennschärfe) bei der Itemanalyse eliminiert (wodurch die durchschnittliche Interkorrelation der verbleibenden Items erhöht wird), ist es im Allgemeinen nicht schwierig, bei einer entsprechenden Länge der Skala den konventionellen Schwellenwert von .8 für Cronbachs  $\alpha$  zu übertreffen. So verfügt z.B. die in 5.3.1 vorgestellte Likert-Skala zur nGRO (bei einer Länge von N = 36)

über eine interne Konsistenz von .92. Allerdings werden in sozialwissenschaftlichen Untersuchungen auch sehr viel kürzere Skalen verwendet, deren Reliabilität dann deutlich niedriger ist.

Im Gegensatz zu der in vielen Lehrbüchern vertretenen Meinung kann die interne Konsistenz (Cronbachs  $\alpha$ ) nicht ohne weiteres als Maß für die <u>Eindimensionalität</u> einer Skala verstanden werden. So konnten z.B. bei der nGRO-Skala trotz der sehr hohen internen Konsistenz mit Hilfe einer Faktorenanalyse vier unabhängige Faktoren isoliert werden (vgl. die Tabelle in 5.3.1). Da  $\alpha$  durch Testverlängerung "quasi beliebig" erhöht werden kann, muß die Dimensionalität einer Skala unabhängig von ihrer internen Konsistenz mit Hilfe einer Faktorenanalyse überprüft werden (vgl. 5.4.2).

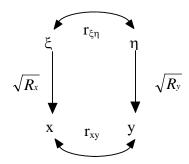
Die Reliabilität einer Messung ist nach der KT um so höher, je geringer der Anteil der Fehlervarianz relativ zu der Varianz des wahren Wertes ist. Die Varianz des wahren Wertes kann aber in Abhängigkeit von der Untersuchungsgesamtheit variieren. Ist die Untersuchungsgesamtheit homogen zusammengesetzt (z.B. bei einer Befragung einer Gruppe feministischer Frauen zur Bestimmung der Reliablität der nGRO Skala), so wird die Reliabilität geringer ausfallen als bei einer heterogenen Stichprobe (z.B. bei der Befragung einer repräsentativen Zufallsstichprobe der Bevölkerung der BRD).

Bei der Überprüfung von Kausalhypothesen sind wir in Kapitel 3 (unrealistischerweise) von fehlerfrei gemessenen Variablen ausgegangen. Ist dies nicht der Fall, müssen Meßfehler bei der Überprüfung von Kausalhypothesen explizit berücksichtigt werden. Dies geschieht im kausalanalytischen Ansatz durch die explizite Spezifikation einer Meß- und einer Meßfehlertheorie. In Übereinstimmung mit unseren Überlegungen zum methodologischen Falsifikationismus (vgl. 2.1.4.2) zeigte die Diskussion des kausalanalytischen Ansatzes in 4.3, daß Kern-, Meß- und Meßfehlertheorie nur im Rahmen eines gemeinsamen Modells überprüft werden können. Geht man bei der Interpretation des Modells von der Gültigkeit der Meß- und Meßfehlertheorie aus, so stellen die Koeffizienten der Kerntheorie um Meßfehler bereinigte Schätzungen der Stärke der Kausalbeziehungen zwischen den theoretischen Konstrukten dar eine Minderungskorrektur aufgrund mangelnder Reliabilität der Messungen wird dabei "quasi automatisch" vorgenommen (vgl. auch das Beispiel in 4.3). Dabei können die in 4.3 gemachten Voraussetzungen als Konkretisierungen der allgemeinen Annahmen der KT verstanden werden (alle Fehler werden als Zufallsfehler vorausgesetzt, d.h. sie haben einen Mittelwert von Null, korrelieren nicht mit dem zugehörigen theoretischen Konstrukt, korrelieren nicht untereinander und korrelieren nicht mit anderen theoretischen Konstrukten).

Möchte man dagegen eine Kausalhypothese in einer Untersuchung überprüfen, in der die entsprechenden Konstrukte  $\xi,\eta$ , .... durch Likert-Skalen x, y, .... nach den Kriterien der KT gebildet wurden, so empfiehlt sich eine <u>Minderungskorrektur</u> der Interskalenkorrelationen nach der Formel:

$$r_{\xi\eta} = \frac{r_{xy}}{\sqrt{R_x}\sqrt{R_y}}$$

Die Korrelation zwischen den wahren Werten wird also berechnet, indem man die durch die in den Skalen enthaltenen Meßfehler (mangelnde Reliabilität) verursachte Verminderung der empirischen Korrelationen rückgängig macht, schematisch im Kausalmodell:



Diese Korrektur fällt umso deutlicher aus, je geringer die Reliabilität der Skalen ist, d.h., je stärker die Messungen durch Meßfehler verunreinigt sind.

#### 5.4.1.2 Validität

Als Theorie, die Meßfehler als Zufallsfehler versteht, ist die KT, wie gesehen, in hervorragender Weise geeignet, Modelle zur Bestimmung der Reliabilität einer Messung bereitzustellen. Da Reliabilität immer auch eine Voraussetzung der Validität ist, kann aus der Reliabilität einer Messung auch ein Maß für deren Validität abgeleitet werden. Versteht man unter der theoretischen Validität die Korrelation zwischen Messung und wahrem Wert  $r_{TX}$ , so gilt (die Annahmen der KT vorausgesetzt):

$$r_{TX} = \sqrt{R_X}$$

(In der Terminologie des Kausalmodells von 4.3 wäre dies gerade die Faktorladung für den zugehörigen Indikator.) Allerdings werden durch die Voraussetzungen der KT systematische Fehler gerade ausgeschlossen, so daß die theoretische Validität nur eine Obergrenze für die tatsächliche Validität darstellt. Diese empirisch zu ermittelnde Validität kann nur im Rahmen einer Theorie bestimmt werden, die das durch die vorliegende Skala gemessene theoretische Konstrukt spezifiziert und mit anderen theoretischen Konstrukten in Beziehung setzt. Auf diese Weise kann die Validität einer Skala entweder korrelativ (Kriteriumsvalidität) oder kausalanalytisch (Konstruktvalidität) empirisch überprüft werden. Wir unterscheiden folgende Formen der Validität:

## • Inhaltsvalidität:

Bei der Beurteilung der Inhaltsvalidität einer Skala muß geprüft werden, ob bei der Auswahl der Indikatoren (z.B. den in den Items einer Likert-Skala formulierten Aussagen) alle inhaltlichen Aspekte berücksichtigt wurden, die bei der Konzeptspezifikation des zu messenden theoretischen Konstruktes zusammengestellt wurden. Ist z.B. in der Konzeptspezifikation der

nGRO unter anderem die Entscheidungsfindung in der Partnerschaft als ein inhaltlicher Aspekt enthalten, so muß die zu entwickelnde Likert-Skala Indikatoren enthalten, die sich auf diesen Aspekt beziehen. Die Überprüfung der Inhaltsvalidität ist also in besonderer Weise geeignet, Aussagen über die Validität der Indikatoren, d.h. über ihre Repräsentativität und Vollständigkeit zu machen. Dabei ist es nicht unbedingt ein Nachteil, daß zur Beurteilung der Inhaltsvalidität keine abstrakten Kriterien und mathematischen Formalismen existieren. Die Inhaltsvalidität kann nur argumentativ im Rahmen eines theoretischen Kontextes beurteilt werden.

#### • Kriteriumsvalidität:

Die Kriteriumsvalidität betrifft den Zusammenhang zwischen der Messung eines theoretischen Konstrukts und einem unabhängig gemessenen externen Kriterium. Offenbar ist eine empirische Überprüfung der Kriteriumsvalidität nur vor dem Hintergrund von theoretischen Annahmen über Art und Stärke dieser Beziehung möglich. Die Korrelation zwischen der Messung und dem externen Kriterium gilt dann als empirisches Maß für die Kriteriumsvalidität. (Natürlich hängt die Höhe der Korrelation auch von der Reliabilität ab, mit der das Kriterium gemessen wurde (vgl. die Ausführungen zur Minderungskorrektur)). Eine Messung ist also im Sinne der Kriteriumsvalidität in dem Maße valide, wie ihre Beziehungen zu anderen Variablen den theoretischen Erwartungen entsprechen. Man unterscheidet zwischen Übereinstimmungsvalidität (speziell: bekannte Gruppen) und Vorhersagevalidität, je nachdem, ob das Kriterium zeitgleich oder zu einem späteren Zeitpunkt gemessen wurde. So validiert z.B. Krampen (1979) die nGRO Skala unter anderem, indem er feststellt, daß Männer eine signifikant traditionalere nGRO haben als Frauen und daß die nGRO hochsignifikant mit Autoritarismus korreliert. Eine Überprüfung der Vorhersagevalidität der nGRO wäre möglich, wenn man bei jungen Frauen, deren nGRO vor einem längeren Zeitraum gemessen wurde, feststellen würde, ob die Frauen mit einer traditionaleren nGRO tatsächlich häufiger die Hausfrauenrolle übernommen haben als Frauen mit einer liberalen nGRO. Dabei ist es in allen Fällen erforderlich, theoretisch zu begründen, warum ein entsprechender empirischer Zusammenhang erwartet wird. In diesem Sinne besteht ein fließender Übergang zur

## • Konstruktvalidität:

Konstruktvalidität liegt dann vor, wenn für ein Konstrukt Aussagen über Zusammenhänge dieses Konstrukts mit anderen Konstrukten theoretisch abgeleitet und empirisch bestätigt werden können. Dies bedeutet die gleichzeitige Überprüfung von Kern-, Meß- und Meßfehlertheorie im Sinne des kausalanalytischen Ansatzes (vgl. 4.3). Die Überprüfung der Konstruktvalidität ist die aussagekräftigste Form der Validitätsüberprüfung, allerdings auch die theoretisch voraussetzungsvollste. Es können nämlich alle Fehlermöglichkeiten auftreten, die bereits im Zusammenhang mit der Diskussion des methodologischen Falsifikationismus (vgl. 2.1.4.2) und des kausal-analytischen Ansatzes angesprochen wurden:

- erstens kann tatsächlich keine Konstruktvalidität vorliegen
- zweitens kann die zur Validierung verwendete Hypothese (Kerntheorie) falsch gewesen sein
- drittens kann das zur Validierung herangezogene zweite Konstrukt selbst nicht reliabel und/oder nicht valide gemessen worden sein (Meß- und Meßfehlertheorie)

So könnte z.B. das kausalanalytische Modell in 4.3 als Beispiel zur Überprüfung der Konstruktvalidität der Messung von Ausländerfeindlichkeit interpretiert werden. Zunächst zeigt sich, daß durch die beiden Indikatoren Ehe und politische Tätigkeit eine reliable Messung der Ausländerfeindlichkeit möglich ist. Zudem steht dieses Konstrukt in einer theoretisch erwarteten Beziehung zu einem anderen Konstrukt (Kerntheorie) nämlich zu der sozialen Distanz, die selbst reliabel und valide durch die beiden Indikatoren Arbeitsplatz und Nachbarschaft gemessen wurde. Der kausalanalytische Ansatz ermöglicht es auch, die Konvergenz (hohe interne Korrelation zwischen den Indikatoren des gleichen Konstrukts) und die Diskriminanz (schwächere Korrelation zwischen den Indikatoren verschiedener Konstrukte) der Indikatoren eines theoretischen Konstrukts zu beurteilen (siehe auch die Ausführungen zur konfirmatorischen Faktorenanalyse in 5.4.2).

## 5.4.2 Faktorenanalyse

[Zur Faktorenanalyse mit SPSS für Windows gibt es eine Beispieldatei (Acrobat). Diese Datei kann über das WWW bezogen werden]

Im Rahmen des Skriptes fehlt der Platz, ein so komplexes Verfahren wie die Faktorenanalyse (FA) auch nur in den Grundzügen darzustellen. Stattdessen illustrieren wir die Vorgehensweise bei der FA anhand einiger Beispiele. Als erstes Beispiel untersuchen wir eine (hypothetische) Korrelationsmatrix, die die Beziehung zwischen den Schulnoten der Fächer Deutsch (D), Englisch (E), Physik (P) und Mathematik (M) enthält:

	D	Е	P	M
D	1.			
Е	.58	1.		
P	.24	.30	1.	
M	.17	.25	.74	1.

Wir gehen von standardisierten Variablen aus  $(z_1=D, z_2=E, z_3=P, z_4=M)$ . Das <u>einfaktorielle Meßmodell</u> würde einen gemeinsamen Begabungsfaktor postulieren, der die empirischen Korrelationen erklären soll. Offenbar reicht aber ein Faktor  $(f_1)$  nicht aus, um die Struktur der Korrelationsmatrix zu rekonstruieren. Zwar korrelieren D und E bzw. P und M hoch miteinander. (<u>Konvergenz</u> der Indikatoren für sprachliche bzw. mathematische Begabung), die Korrelationen zwischen den Indikatoren D und E einerseits und P und M andererseits sind

aber deutlich geringer (<u>Diskriminanz</u> der Indikatoren verschiedener theoretischer Konstrukte [hier Faktoren]). Wir spezifizieren daher ein <u>zweifaktorielles Modell</u>,

$$z_1=a_1f_1+b_1f_2+\epsilon_1$$
 
$$z_2=a_2f_1+b_2f_2+\epsilon_2$$
 [Strukturgleichungen] 
$$z_3=a_3f_1+b_3f_2+\epsilon_3$$
 
$$z_4=a_4f_1+b_4f_2+\epsilon_4$$

das zwei <u>unabhängige (unkorrelierte) Faktoren</u> für die mathematische Begabung (f<sub>1</sub>) und die sprachliche Begabung (f<sub>2</sub>) enthält. Dabei sollten die <u>Faktorladungen</u> von D (a<sub>1</sub>) und E (a<sub>2</sub>) auf dem ersten Faktor (für mathematische Begabung) gering ausfallen, ebenso wie umgekehrt die Faktorladungen von P (b<sub>3</sub>) und M (b<sub>4</sub>) auf dem zweiten Faktor (für sprachliche Begabung). Wie die Lösung der Faktorenanalyse zeigt, ist dies tatsächlich der Fall:

Variabl	le	Faktorla	adungen	Kommunalität	spezifische Varianz
		1 a <sub>i</sub>	2 b <sub>i</sub>	$h_i^2 = a_i^2 + b_i^2$	
D z	<b>z</b> <sub>1</sub>	.10	.80	.65	.35
E z	<b>Z</b> 2	.20	.70	.53	.47
P z	<b>Z</b> 3	.80	.20	.68	.32
M z	<b>Z</b> 4	.90	.10	.82	.18

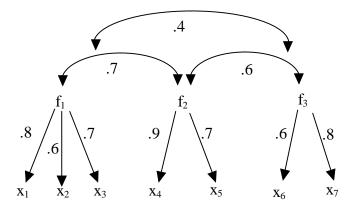
Die Faktorladungen einer Variablen geben ihre Korrelationen mit den (als unabhängig vorausgesetzten!) Faktoren an. Die <u>Kommunalitäten</u> stellen die von den beiden Faktoren gemeinsam erklärte Varianz einer jeden Variable der Faktorenanalyse dar. Aus der Summe der Quadrate der Faktorladungen ergibt sich für jeden Faktor die insgesamt erklärte Varianz und damit seine Erklärkraft im Rahmen der Faktorenanalyse. Die Erklärkraft von Faktor 1 ist demnach 1.50 (= .1² + .2² + .8² + .9²), und die von Faktor 2 1.18 (= .8² + .7² + .2² + .1²). Bei insgesamt vier in die Analyse eingehenden standardisierten Variablen (z<sub>1</sub>, z<sub>2</sub>, z<sub>3</sub>, z<sub>4</sub>) beträgt die insgesamt zu erklärende Varianz 4.0 (= 1² + 1² + 1² + 1²). Damit erklärt der erste Faktor einen Varianzanteil von 37,5% und der zweite einen Varianzanteil von 29,5%. [Im Falle einer <u>Hauptkomponentenanalyse</u> entspricht der <u>Eigenwert</u> eines Faktors seiner Erklärkraft in der unrotierten Lösung.] Aus den angegebenen Faktorladungen können die <u>empirischen</u> Korrelationen der Ausgangsmatrix gemäß der Formel

$$r_{z_i z_j} = a_i a_j + b_i b_j \\$$

rekonstruiert werden. Wie man sieht, ist in diesem konstruierten Beispiel eine exakte Rekonstruktion der Korrelationsmatrix möglich.

Ein etwas komplexeres Modell mit drei <u>korrelierten</u> Faktoren entnehmen wir aus Weede und Jagodzinski (1977). Wir zeichnen das Modell als Pfaddiagramm, das auch die entsprechenden

Korrelationskoeffizienten zwischen den Faktoren enthält (die Faktoren aus der Faktorenanalyse müssen also nicht notwendig unkorreliert sein):



Aus dem Pfadmodell läßt sich die zugrundeliegende Korrelationsmatrix gemäß den Regeln der Pfadanalyse rekonstruieren. Zum Beispiel ist  $r_{12} = .8 \cdot .6 = .48$  und  $r_{16} = .8 \cdot .4 \cdot .6 = .192$ . Insgesamt ergibt sich folgende Korrelationsmatrix:

	$\mathbf{x}_1$	$\mathbf{x}_2$	<b>X</b> 3	X4	X5	X <sub>6</sub>	<b>X</b> 7
<b>X</b> <sub>1</sub>							
X2	0,480						
X3	0,560	0,420					
X4	0,504	0,378	0,441				
X5	0,392	0,294	0,343	0,630			
x <sub>6</sub>	0,192	0,144	0,168	0,324	0,252		
X7	0,256	0,192	0,224	0,432	0,336	0,480	

Berechnet man mit dieser Matrix als Dateneingabe eine <u>explorative</u> FA in SPSS (<u>Hauptachsenmethode</u>, 3-Faktoren-Lösung, oblique (schiefwinklige) Rotation), so erhält man eine Lösung, die die Koeffizienten des Ausgangsmodells sehr genau reproduziert:

	$\mathbf{f}_1$	$f_2$	$f_3$
<b>x</b> <sub>1</sub>	0,80	0,00	0,00
<b>x</b> <sub>2</sub>	0,60	0,00	0,00
Х3	0,70	0,00	0,00
X4	0,09	0,75	0,07
X5	-0,03	0,78	-0,02
X <sub>6</sub>	0,00	-0,05	0,69
X7	0,00	0,08	0,68

	$\mathbf{f}_1$	$f_2$	$f_3$
$f_1$	1		
$f_2$	.68	1	
$f_3$	.39	.57	1

Natürlich würde eine <u>konfirmatorische</u> FA, die die Bedingungen des Ausgangsmodells getreu übernimmt, die exakten Koeffizienten reproduzieren. [Eine solche Analyse könnte z.B. mit LISREL durchgeführt werden (vgl. 8.).] Das Beispiel soll aber gerade zeigen, daß auch bei

einer unbekannten Struktur der Faktoren eine explorative FA in der Lage ist, diese Struktur sehr genau zu reproduzieren.

Wie die explorative FA der Items der Skala zur nGRO von Krampen (1979) zeigt [da es sich um 36 Items handelt, muß eine 36 x 36 Korrelationsmatrix analysiert werden], können in diesem Fall vier unabhängige (orthogonale) Faktoren extrahiert werden (vgl. die Tabelle in 5.3.1). Das Konstrukt der nGRO ist also <u>nicht eindimensional</u>, obwohl die Skala über eine sehr hohe interne Konsistenz verfügt. Inhaltlich bedeutet dies, daß sich unterschiedliche Teilaspekte der nGRO spezifizeren lassen (indem man die Aussagen der Items mit einer besonders hohen Faktorladung auf dem jeweiligen Faktor analysiert), die unabhängig voneinander meßbare <u>Teildimensionen des theoretischen Konstrukts</u> darstellen. (Bevor man allerdings aus einem solchen Ergebnis theoretische Konsequenzen zieht, muß erst seine Robustheit und Stabilität in Nachfolgestudien bestätigt werden, wie dies z.B. für die fundamentalen Faktoren der Bewertung, Stärke und Aktivität beim semantischen Differential geschehen ist.)

Abschließend illustrieren wir die Bedeutung der Faktorenanalyse für die <u>dimensionale</u> <u>Analyse von Skalen</u> an einem Beispiel aus der Forschungspraxis, und zwar der Likert-Skala zur Messung des allgemeinen Selbstwerts aus der Kaufsuchtstudie (vgl. 1.). Die folgende Tabelle zeigt die Items der Skala zusammen mit den <u>Trennschärfekoeffizienten</u> und dem Maß für die interne Konsistenz:

Item		Trennschärfe
$I_1*$	Bin manchmal zu nichts gut	.60
$I_2$	Finde mich ganz in Ordnung	.53
I <sub>3</sub> *	Bin ein niemand	.59
$I_4$	Bin eigentlich mit mir zufrieden	.57
I <sub>5</sub> *	Wünsche, ich wäre nie geboren	.64
$I_6*$	Mangelnde Selbstachtung	.58
$I_7*$	Gefühl der Nutzlosigkeit	.66
$I_8$	Schneide gut ab im Vergleich mit anderen	.48
I <sub>9</sub> *	Verachte mich	.58
$I_{10}$	Bin zufrieden mit mir	.52
Cronb	eachs $\alpha = .86$	

[Für die Berechnung der Trennschärfekoeffizienten und von Cronbachs  $\alpha$  mit Hilfe des SPSS-Programms RELIABILITY und für die folgende Faktorenanalyse mit dem SPSS-Programm FACTOR wurden die positiven Items umgepolt. Im Gegensatz dazu wurde bei der Bildung der Subskalen für positiven (pSW) und negativen (nSW) Selbstwert (siehe unten) die ursprüngliche Polung der Items beibehalten; hohe Werte von pSW (nSW) bedeuten also einen hohen (geringen) Selbstwert.]

Trotz der hohen internen Konsistenz von .86 ist ein interessantes Muster bei den Trennschärfekoeffizienten zu erkennen. Die vier positiven Items, I<sub>2</sub>, I<sub>4</sub>, I<sub>8</sub> und I<sub>10</sub>, verfügen mit Werten zwischen .48 und .57 über leicht geringere Trennschärfen als die sechs negativen Items mit Werten zwischen .58 und .66. Dies deutet auf eine zweidimensionale Struktur hin, die auch durch eine Faktorenanalyse bestätigt wird (Hauptkomponentenmethode, 2-Faktoren-Lösung). Zunächst zeigen wir die <u>unrotierte Lösung</u>:

Item	$f_1$	$f_2$
$I_1^*$	.71	21
$I_2$	.61	.45
$I_3*$	.70	29
${ m I}_4$	.65	.48
$I_5*$	.74	27
$I_6*$	.69	28
$I_7*$	.76	23
${ m I}_8$	.56	.50
$I_9$ *	.69	35
$I_{10}$	.60	.51
Eigenwert	4.52	1.41
erklärte Varianz	(45,2%)	(14,1%)

Wie man sieht, entsprechen die Ladungen auf dem ersten Faktor in etwa den Trennschärfe-koeffizienten und geben demnach die Ladung der Items auf der Gesamtskala an. Der zweite Faktor kann als Methodenfaktor interpretiert werden, da er für positive (negative) Items jeweils positive (negative) Faktorladungen geringerer Größe enthält. Zu fragen bleibt, ob es sich bei der gefundenen zweidimensionalen Struktur nur um ein Methodenartefakt handelt, das durch die unterschiedliche Polung der Items hervorgerufen wird, oder ob sich aus den beiden Faktoren zwei inhaltlich verschiedene Subskalen konstruieren lassen. Dazu zeigen wir zunächst die <u>rotierte Lösung</u> (Varimax-Rotation (beide Faktoren sind unkorreliert); nur Faktorladungen größer als .3):

Item	$\mathbf{f}_1$	$f_2$
I <sub>1</sub> *	.70	
$I_2$		.73
I <sub>3</sub> *	.74	
$I_4$		.77
I <sub>5</sub> * I <sub>6</sub> * I <sub>7</sub> *	.76	
I <sub>6</sub> *	.71	
$I_7*$	.75	
$I_8$		.73
I <sub>9</sub> *	.76	
$I_{10}$		.77

Wir bilden nun die beiden <u>Subskalen</u>

$$nSW = I_1 + I_3 + I_5 + I_6 + I_7 + I_9$$

$$pSW = I_2 + I_4 + I_8 + I_{10},$$

die jeweils den <u>positiven</u>, bzw. <u>negativen</u> Selbstwert messen. Die <u>Unabhängigkeit</u> der beiden Faktoren bedeutet dabei, daß das Fehlen eines positiven Selbstwerts nicht unbedingt mit einem negativen Selbstwertgefühl verbunden sein muß. Um zu überprüfen, ob die beiden Subskalen pSW und nSW unabhängig voneinander valide Konstrukte darstellen, bestimmen wir die <u>Kriteriumsvalidität</u>. Als Kriterien im Rahmen der Kaufsuchtstudie kommen u.a. die Kaufsucht und die materielle Gütergebundenheit in Frage:

	nSW (+)	pSW (+)
Kaufsucht (+)	.37***	12***
mat. Gütergebundenheit (+)	.05 <sup>ns</sup>	.17***

(\*\*\* = signifikant auf dem 0.1% - Niveau)

Wie man sieht, stellen nSW und pSW tatsächlich inhaltlich verschiedene Konstrukte dar. Kaufsucht korreliert deutlich stärker mit dem negativen Selbstwert als mit dem Fehlen eines positiven Selbstwertes, und für die materielle Gütergebundenheit findet man eine signifikante Korrelation lediglich mit dem positiven Selbstwertgefühl. Die zweifaktorielle Lösung ist also kein Methodenartefakt, sondern zeigt eine inhaltlich plausibel zu interpretierende zweidimensionale Struktur des theoretischen Konstruktes Selbstwert an.

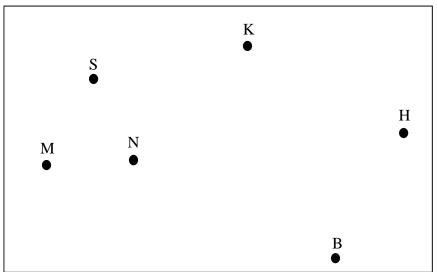
## 5.4.3 Multidimensionale Skalierung

Ziel der multidimensionalen Skalierung (MDS) ist die Darstellung einer Punktekonfiguration in einem Raum möglichst niedriger Dimension derart, daß die <u>Abstände</u> d<sub>ij</sub> der Punkte in der Lösung der MDS den <u>ursprünglichen Distanzen</u> u<sub>ij</sub> (Eingangsinformationen) möglichst genau

entsprechen. Der Grad der Abweichung zwischen u<sub>ij</sub> und d<sub>ij</sub> wird in einem <u>STRESS-Maß</u> ausgedrückt. Durch Verschieben der Punkte in einem Lösungsraum mit definierter Dimensionalität und Metrik (i.a. der euklidischen Metrik) wird versucht, schrittweise (iterativ) die Lösungsdistanzen den Ausgangsdistanzen anzupassen, um so den STRESS der Lösung zu minimieren. [In dem in SPSS enthaltenen ALSCAL-Programm von Young wird zur Minimierung das Verfahren der alternierenden Kleinstquadrateschätzung (ALS) verwendet.] Als Beispiel betrachten wir die folgende Distanzmatrix (vgl. Borg 1981 und Backhaus u.a. 1993):

	В	Н	K	M	N	S
В	-					
Н	105	-				
K	191	142	-			
M	205	250	184	-		
N	153	189	136	61	-	
S	206	216	116	76	64	-

Die Distanzen zwischen den Punkten sind hier als metrische Informationen zu verstehen, d.h. die Distanz zwischen H und M ist mit 250 mehr als viermal so groß wie die Distanz von 61 zwischen M und N (das Argument unterstellt sogar eine Ratioskalierung der Distanzen). Gehen die Distanzen als metrische Information in die MDS ein, spricht man von einer metrischen MDS. Die zweidimensionale Lösung der metrischen MDS der obigen Distanzmatrix hat folgendes Aussehen:



(aus technischen Gründen ist die Darstellung der Lösungen im Skript immer nur approximativ möglich)

Wie man mit einiger Phantasie erkennen kann, wurde unser Beispiel konstruiert, indem die Distanzen zwischen den Städten Berlin (B) Hamburg (H), Köln (K), München (M), Nürnberg (N) und Stuttgart (S) auf einer Landkarte ausgemessen und in die Matrix der Ausgangsdistanzen übertragen wurden. Wie nicht anders zu erwarten, gelingt die perfekte Reproduktion der Landkarte (der STRESS der gezeigten zweidimensionalen Lösung ist Null) mit Hilfe der

metrischen MDS. Allerdings muß die Konfiguration erst um 90° nach links gedreht werden, um die Nord-Süd-Achse in die vertraute Position zu bringen.

Ist die Struktur der Daten komplexer, ist eine perfekte Reproduktion der Ausgangsdistanzen in einem zweidimensionalen Lösungsraum nicht möglich. Man kann allerdings eine Verbesserung der Modellanpassung (Verminderung des STRESS) immer erreichen, indem man die Objekte in einen höherdimensionalen Raum einbettet. Je höher nämlich die Dimension des Lösungsraumes, desto mehr Freiheitsgrade hat man bei der Anordnung der Objekte und desto genauer können die Ausgangsdistanzen u<sub>ij</sub> durch die Distanzen d<sub>ij</sub> im Lösungsraum reproduziert werden. Dem Ziel der Informationsreduktion und übersichtlichen Darstellung entsprechend, bemüht man sich aber in den meisten Fällen um eine zweidimensionale (höchstens dreidimensionale) Darstellung wie im obigen Beispiel der Reproduktion einer Landkarte. Dies ist natürlich nur erlaubt, wenn der STRESS (die Fehlanpassung) der Lösung nicht zu hoch ist. Welche STRESS-Werte in Abhängigkeit von der Dimensionalität des Lösungsraumes und der Zahl der Objekte noch akzeptiert werden können, wird durch Monte-Carlo-Studien ermittelt.

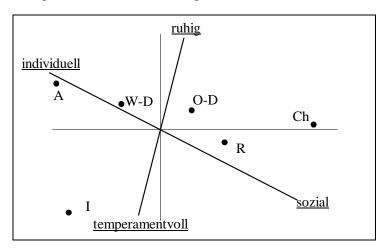
Eine MDS kann aber auch angewandt werden, wenn das Meßniveau der in der Matrix der Ausgangsdistanzen dargestellten Abstände nicht metrisch ist. In der Praxis wird die MDS sogar fast ausschließlich als <u>ordinale MDS</u> angewandt, d.h. es liegt nur eine <u>Rangordnung der Abstände</u> zwischen den Objekten vor. Dazu ein Beispiel einer MDS von Ähnlichkeitsratings zwischen Nationen (nach Steffens 1987):

	Ch	I	O-D	A	W-D	R
Ch: Chinesen	-					
I: Italiener	2.48	-				
O-D: Ostdeutsche	3.38	2.97	-			
A: Amerikaner	2.55	4.59	3.59	-		
W-D: Westdeutsche	2.86	4.55	7.31	6.34	-	
R: Russen	5.24	3.21	6.69	3.31	3.66	-

Die Werte sind als Durchschnittswerte der Ähnlichkeitsratings von Pädagogikstudenten auf einer von 1 bis 10 (maximale Ähnlichkeit) reichenden Ratingskala zu verstehen. Für die folgende Analyse werden diese Informationen nur ordinal interpretiert, d.h. der Wert von 5.24 für die Ähnlichkeit zwischen Russen (R) und Chinesen (Ch) besagt im Vergleich zu dem Wert von 6.34 für die Ähnlichkeit von Westdeutschen (W-D) und Amerikanern (A) nur aus, daß die Distanz zwischen R und Ch in der Lösung der ordinalen MDS größer sein sollte als die zwischen W-D und A. Die ordinale Ausgangsinformation (u<sub>ij</sub>) hat daher folgendes Aussehen:

	Ch	I	O-D	A	W-D	R
Ch: Chinesen	-					
I: Italiener	15	-				
O-D: Ostdeutsche	9	12	-			
A: Amerikaner	14	5	8	-		
W-D: Westdeutsche	13	6	1	3		
R: Russen	4	11	2	10	7	-

Als Lösung der ordinalen MDS ergibt sich:



Der STRESS für diese (zweidimensionale) Lösung beträgt 0.0624, ein Wert, der nach den üblichen Konventionen als eine gute Anpassung der Lösung an die Ausgangsdaten betrachtet werden kann. (Bei der Interpretation des STRESS-Wertes muß berücksichtigt werden, daß sich die von Steffens berichtete Lösung auf insgesamt 10 Nationen bezieht - aus Gründen der Übersichlichkeit werden hier nur 6 Nationen dargestellt.) Grundlage der iterativen Optimierung einer MDS-Lösung ist die Verlustfunktion

$$L = \sum_{i < j} \left( d_{ij} - f(u_{ij}) \right)^2,$$

die die quadrierten Abweichungen zwischen den transformierten Ausgangsdistanzen  $f(u_{ij})$  und den Distanzen im Lösungsraum  $(d_{ij})$  aufsummiert. Entsprechend der Art der MDS ist als Transformation f entweder eine <u>monotone</u> (ordinale MDS) oder <u>affine</u> (metrische MDS) <u>Funktion</u> zugelassen. Durch Standardisierung der Verlustfunktion erhält man als STRESS S:

$$S = \sqrt{\frac{\sum_{i < j} \left(d_{ij} - f(u_{ij})\right)^{2}}{\sum_{i < j} d_{ij}^{2}}}.$$

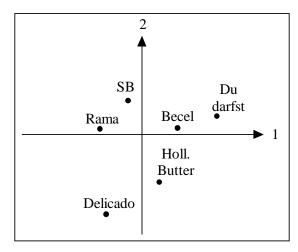
In den gebräuchlichen Programmen zur Berechnung von MDS-Lösungen ist dieses von Kruskall (vgl. z.B. Kruskall und Wish 1977) entwickelte STRESS-Maß auf verschiedene Weisen modifiziert worden, insbesondere, um der Gefahr von <u>entarteten Lösungen</u> zu entgehen. Diese Modifikationen unterscheiden sich vor allem in der Art der Standardisierung der Verlustfunk-

tion. Zur Beurteilung der Güte einer MDS-Lösung wird aber (aus Gründen der Vergleichbarkeit) immer noch auf die obige erste Version eines STRESS-Maßes Bezug genommen.

Zur Interpretation der Metrik des Lösungsraumes können als Interpretationshilfe zusätzliche Skalen in den Lösungsraum hineingelegt werden. In unserem Beispiel wurden die Nationen zusätzlich auf einem semantischen Differential beurteilt, u.a. in Bezug auf die Gegensatzpaare "ruhig - temperamentvoll" und "individuell - sozial". Wie man sieht, kann die zweidimensionale Konfiguration der Nationen durchaus sinnvoll anhand der als Achsen in die zweidimensionale Lösung eingezeichneten Gegensatzpaare interpretiert werden (siehe die oben angegebene zweidimensionale MDS-Lösung). [Auch zur optimalen Einpassung von metrischen Kriterien in den Lösungsraum einer MDS (optimiert wird die durch die MDS-Lösung erklärte Varianz des Kriteriums) existieren mathematische Verfahren.]

Da eine ordinale MDS immer dann durchführbar ist, wenn man über eine ordinale Information über Distanzen zwischen Objekten verfügt, hätte man auch die aus den Profilen der Nationen im semantischen Differential berechnete <u>euklidische Distanz</u> als Ausgangsinformation für eine MDS verwenden können. Eine solche Vorgehensweise empfiehlt sich in der <u>Marktforschung</u> immer dann, wenn als <u>Polaritätsprofile</u> erhobene Images von Objekten (Produkten, Firmen, Marken, usw.) analysiert werden sollen. Die ordinale MDS zeigt dann, durch welche Dimensionen der Bewertungsraum für die Imageobjekte aufgespannt wird.

Das Gewinnen einer metrischen Lösung aus einer ordinalen Ausgangsinformation in der ordinalen MDS ist nur möglich, da sich die ordinale Information in der Ausgangsmatrix bereits auf Abstände bezieht. Dabei besteht die große Stärke der MDS gerade darin, daß bei der Erhebung der Abstände zwar implizit ein Beurteilungsraum der Befragten vorausgesetzt wird, die Erhebung der Abstandsinformation aber ohne ein explizites Wissen über die für den Beurteilungsraum konstitutiven Dimensionen möglich ist (z.B. durch Ähnlichkeitsratings, durch Rangordnen oder durch die Ankerpunktmethode; vgl. Backhaus u.a. 1993). Die Rekonstruktion dieser Dimensionen ist erst das Ergebnis der multidimensionalen Skalierung, nicht aber die Voraussetzung des Verfahrens. Die Dimensionen des Lösungsraums werden als Intervallskalen allein aus den ordinalen Informationen über Abstände rekonstruiert. Abschließend illustrieren wir diese Eigenschaft der MDS noch einmal am Beispiel einer Imageanalyse von Margarinemarken auf der Grundlage von Ähnlichkeitsratings (nach Backhaus u.a. 1993):



Die Lösung ist bereits so rotiert, daß eine anschauliche Interpretation der von den Konsumenten benutzten Beurteilungsdimensionen möglich ist. Die horizontale Dimension kann als "gesunde Ernährung" interpretiert werden und ist besonders durch die beiden fett- und kalorienarmen Marken "Becel" und "Du darfst" charakterisiert. Die vertikale Dimension stellt Margarinemarken (oben) und Buttermarken (unten) gegenüber und kann etwa als "Geschmack" interpretiert werden.

# 6. Datenerhebungsverfahren

Bei den Datenerhebungsverfahren unterscheiden wir grundsätzlich zwischen <u>Befragung</u>, <u>Beobachtung</u> und <u>Inhaltsanalyse</u>. Dabei ist die Entscheidung für ein bestimmtes Erhebungsverfahren unabhängig von der Wahl der Untersuchungsform oder des Auswahlverfahrens. Anders ausgedrückt, Untersuchungsformen, Auswahlverfahren und Erhebungsverfahren können in beliebiger Weise kombiniert werden, um ein für ein spezielles Forschungsproblem maßgeschneidertes Forschungsdesign zu entwickeln.

Dabei ist die Subsummierung der Inhaltsanalyse unter die Erhebungsverfahren im Grunde ungenau, da, wie es der Name schon sagt, eigentlich nicht die Art der Datenerhebung im Vordergrund steht, sondern die inhaltliche <u>Analyse</u> von Texten und Bildern, die z.B. auch aus einer Befragung oder Beobachtung stammen können. Bei der Inhaltsanalyse geht es also zunächst um eine theoriegeleitete und systematisch-interpretative Aufbereitung von Daten, die dann möglicherweise in eine statistische Datenanalyse mit den üblichen Verfahren einmünden kann.

Erhebungsverfahren können nach ihrer eher <u>qualitativen</u> oder <u>quantitativen</u> Ausrichtung und, eng damit verbunden, nach dem Grad ihrer Standardisierung unterschieden werden. Diese Unterscheidung ist wiederum unabhängig von der Art des Erhebungsverfahrens, das heißt sowohl bei der Befragung, wie auch bei Beobachtung und Inhaltsanalyse kann zwischen eher unstandardisiert-qualitativen und standardisiert-quantitativen Verfahren unterschieden werden. Wie bereits eingangs in Kapitel 3 angemerkt, sehen wir in dieser Unterscheidung keine weltanschauliche Frage, die zu einem Bekenntnis für die eine und damit gegen die andere der beiden methodischen Traditionen zwingen würde. Ihre jeweilige Angemessenheit kann nur in Bezug auf das spezielle sozialwissenschaftliche Untersuchungsziel vor dem Hintergrund des jeweiligen theoretischen und methodischen Kenntnisstandes und des empirischen Vorwissens beurteilt werden. Dabei ist sowohl eine zeitliche Abfolge von qualitativ-explorativen und quantitativ-hypothesentestenden Untersuchungen denkbar, wie auch die kombinierte Anwendung in einem umfassenden Forschungsdesign.

Eine weitere zentrale Unterscheidung ist die zwischen reaktiven und nichtreaktiven Erhebungsverfahren. Reaktive Erhebungsverfahren sind solche, bei denen die Erhebungstechnik Untersuchungsobjekten Veränderungen bei den hervorruft, die Gefahr von Forschungsartefakten also immer gegeben ist. Nichtreaktive Erhebungsverfahren sind Verfahren, bei denen sich die Erhebungsperson nicht bewußt ist, daß sie Gegenstand einer Datenerhebung ist, so z.B. die verdeckte Beobachtung oder alle Verfahren der Inhaltsanalyse, sofern sie sich auf prozeßproduzierte Dokumente wie z.B. Zeitungsartikel, Werbeanzeigen, Schulaufsätze, Parteiprogramme usw. beziehen. Dagegen ist die offene Beobachtung ein reaktives Erhebungsverfahren wie auch alle Formen der Befragung, allerdings in unterschiedlichem Ausmaß. Ein besonderes Problem stellt die Reaktivität für die mündliche Befragung (Interview) dar.

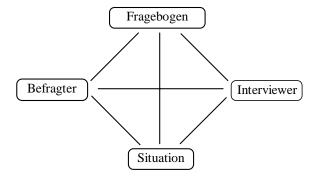
## 6.1 Befragung

Die am häufigsten angewandte Erhebungsmethode ist die Befragung und hier insbesondere das standardisierte Interview, das von R. König (1972) als "Königsweg" der empirischen Sozialforschung betrachtet wird. Grundsätzlich unterscheiden wir zwischen der mündlichen, der schriftlichen und der telefonischen Befragung. Je nach Standardisierungsgrad können folgende Formen der mündlichen Befragung unterschieden werden:

- geringer Standardisierungsgrad: Expertengespräch, narratives Interview, exploratives Interview
- mittlerer Standardisierungsgrad: Leitfadengespräch
- hoher Standardisierungsgrad: standardisiertes Interview

Die Standardisierung eines Interviews bezieht sich einmal auf die Vorgabe festgelegter Fragen mit festgelegten Antwortvorgaben in einer festgelegten Reihenfolge und zum anderen auf die Gleichheit der Interviewsituation, die durch besondere Anforderungen an das Interviewerverhalten sichergestellt werden soll. Der Interviewer wird dabei als neutrale Instanz bzw. als austauschbares Instrument verstanden. Dazu gehört ein seriöser Eindruck und allgemein eine Haltung des freundlichen Gewährenlassens. Insbesondere darf der Interviewer keine eigene Einstellung zum Untersuchungsgegenstand verraten oder Befremden bzw. Zustimmung bei den Antworten äußern.

Dennoch kann nicht übersehen werden, daß das Interview grundsätzlich eine <u>soziale Interaktion</u> zwischen Interviewer und Befragten ist. Schematisch kann die Interviewsituation wie folgt charakterisiert werden:



Die Befragung ist also nie ein einfacher Datenabruf; der Befragte antwortet immer kontextgebunden, d.h. in einer spezifischen sozialen Situation, die einen Einfluß auf sein Antwortverhalten ausübt. Zunächst ist natürlich der Einfluß des Fragebogens hervorzuheben, also der Anordnung und der genauen Formulierung der Fragen, aber auch der Formulierung

und Reihenfolge der Antwortvorgaben. Darüber hinaus reagiert der Befragte auch auf soziale Merkmale des Interviewers, auch dann, wenn dieser sich bemüht, den Regeln für die Befragung entsprechend eine möglichst neutrale Haltung einzunehmen.

Schließlich hat die Interviewsituation einen Einfluß auf das Antwortverhalten des Befragten. Jede soziale Situation wird durch die Beteiligten im Kontext alltäglicher Interpretationsschemata interpretiert. So kann der Befragte den Interviewer als "Spion", als "Vertreter einer Behörde", als "freundlichen Unbekannten", als "jemanden, dem man seine Sorgen anvertraut", als "armen Hund" oder als "jemanden, dem man die Meinung sagt" empfinden und sich im Interview entsprechend verhalten. Dies verdeutlicht die folgende Tabelle aus einer Untersuchung von H. Steinert (1984, S. 37), die die <u>Definition der Interviewsituation</u> durch den Befragten (gemäß der nachträglichen Beurteilung durch den Interviewer) in Abhängigkeit von der Schicht-Selbsteinstufung des Befragten zeigt (Angaben in Prozent):

Definition der Interviewsituation durch den Interviewer nach Schicht-Selbsteinstufung der Befragten (Angaben in Prozent):

	Alte	Neue	Aufstr.	Ruhige	Aufstr.	Trad.	
Situation	Führungs-	Manager-	Mittel-	Mittel-	Arbeiter-	Arbeiter-	Arme
	schicht	schicht	schicht	schicht	schicht	schicht	Leute
"Spion"	5	3	3	5	7	6	15
"Behörde"	27	26	36	41	49	48	50
"freundl. Unbekannter"	32	31	37	30	20	24	20
"Sorgen anvertrauen"	4	7	4	4	6	9	3
"armer Hund"	23	22	15	13	11	6	4
"die Meinung sagen"	10	12	6	7	6	7	8
	100	100	100	100	100	100	100
N	(75)	(72)	(293)	(775)	(216)	(299)	(104)

Ein weiteres bedeutsames Merkmal der Interviewsituation, das zu Veränderungen des Antwortverhalten führen kann, stellt die <u>Anwesenheit Dritter</u> (Familienangehörige, Freunde, ...u.s.w.) dar. Greift ein Dritter aktiv in das Interview ein, sind die Verzerrungen meistens so schwerwiegend, daß das gesamte Interview wertlos wird. Aber auch die bloße Anwesenheit eines Dritten kann zu Verzerrungen, analog zu den Interviewereffekten, führen, da der Befragte bei seinen Antworten mögliche Reaktionen des Dritten in Rechnung stellen wird.

diskutierten Neben den bisher Formen der Einzelbefragung können auch Gruppenbefragungen nach dem Grad ihrer Standardisierung unterschieden werden. Die Gruppendiskussion stellt eine wenig standardisierte Form der Gruppenbefragung dar, bei der die Mitglieder einer Gruppe, vom Interviewer lediglich durch eine thematische Vorgabe und durch Zwischenfragen gelenkt, in freier Form miteinander diskutieren. Der ungezwungene Charakter solcher Gruppendiskussionen ähnelt stark der Atmosphäre alltäglicher informeller Gespräche. Die Gruppendiskussion wird in der Marktforschung häufig eingesetzt, etwa um tiefergehende Einsichten in die Art der Bewertungskriterien zu gewinnen, nach denen Produktkonzepte von Konsumenten beurteilt werden.

## 6.1.1 Fragebogenkonstruktion

Die Konstruktion eines Fragebogens ist eine <u>Kunstlehre</u> (Payne: The Art of Asking Questions) und erfordert in erster Linie Erfahrung und Fingerspitzengefühl. Grundregel ist, daß die Formulierungen möglichst einfach und dem Sprachgebrauch der Befragten angepaßt sein sollten. Die im Fragebogen verwendeten Fragen müssen also nicht unbedingt "druckreif" formuliert sein. Das Interview soll trotz der Standardisierung seinen ungefähren Gesprächscharakter behalten. Der Fragebogen darf also keine zu komplizierten oder wissenschaftlich abstrakten Formulierungen enthalten. Die Fragen sollten ungefähr in der Sprache formuliert sein, die der Befragte in einer Unterhaltung mit einem <u>Fremden</u> benutzen würde. Auch das andere Extrem einer zu starken sprachlichen Anbiederung in Wortwahl und Grammatik muß also vermieden werden. Vor allem müssen die Fragen den Befragten in dem vom Forscher gemeinten Sinn verständlich sein. Fragen sollten daher:

- einfache Worte enthalten
- kurz formuliert sein
- konkret sein
- neutral formuliert sein
- nicht hypothetisch formuliert sein
- nur eine Thematik ansprechen
- keine doppelten Verneinungen enthalten und
- formal balanciert sein.

Eine Frage ist formal balanciert, wenn sie alle negativen und positiven Antwortmöglichkeiten enthält. Eine nicht balancierte Frage kann zu Antwortverzerrungen führen, wie das folgende Beispiel aus dem Jahr 1955 zeigt (vgl. Noelle-Neumann 1963, S.86):<sup>1</sup>

Version I: Finden Sie, daß in einem Betrieb alle Arbeiter in der Gewerkschaft sein sollten?

Version II: Finden Sie, daß in einem Betrieb alle Arbeiter in der Gewerkschaft sein sollten, oder muß man es jedem einzelnen überlassen, ob er in der Gewerkschaft sein will oder nicht?

\_

Auch aktuelle Zahlen aus 1994 bestätigen die beobachteten Antwortverzerrungen (vgl. Noelle-Neumann u. Petersen 1996, S. 132), wobei zusätzlich ein langfristiger Trend gegen eine Zwangsmitgliedschaft zu beobachten ist (Ablehnung in Version II steigt von 70% auf 81%).

	Version I	Version II
ja (alle in der Gewerkschaft)	44%	24%
nein (ist jedem einzelnen überlassen)	20%	70%
unentschieden	36%	6%

Im allgemeinen unterscheidet man zwischen:

- <u>Eigenschaftsfragen</u> (Fragen nach personalen und soziodemographischen Eigenschaften des Befragten, wie z.B. Alter, Bildung, Familienstand, Konfession, usw.)
- <u>Verhaltensfragen</u> (Fragen die sich konkret auf das berichtete Verhalten des Befragten beziehen, wie z.B. Fragen nach der Kirchgangshäufigkeit, der Teilnahme an Demonstrationen, dem Kaufverhalten, dem Umweltverhalten, usw.)
- <u>Überzeugungsfragen</u> (Fragen nach dem, was der Befragte für wahr oder falsch hält, die sich also auf das "Wissen" bzw. auf das "Weltbild" des Befragen beziehen. Es geht aber nicht darum (und kann natürlich in einer Befragung auch gar nicht darum gehen), was objektiv wahr oder falsch ist. Entscheidend ist die Einschätzung des Befragten etwa bezüglich der Frage, ob die Mehrheit der Deutschen seiner Meinung nach Steuerhinterzieher sind oder nicht.)
- <u>Einstellungs- oder Meinungsfragen</u> (Fragen nach der positiven oder negativen Beurteilung eines bestimmten Tatbestands durch den Befragen, wie z.B. die Beurteilung der Items zur normativen Geschlechtsrollenorientierung auf einer Ratingskala mit den Endpunkten "stimme überhaupt nicht zu" und "stimme voll und ganz zu".)

Oft wird kritisch gegen die Befragung eingewandt, daß man durch eine geschickte Fragestellung jede beliebige Antwort provozieren kann. Daß dieser in seiner Pauschalität völlig überzogene Einwand dennoch ein Körnchen Wahrheit enthält, erkennt man an folgendem Beispiel. In einer Umfrage in den USA wurde 1978 die folgende Frage gestellt: "Falls sich in irgendeinem Teil der Welt eine Situation ähnlich der in Vietnam entwickelt, sollten die USA dann Truppen schicken?" Aufgrund des tiefsitzenden Vietnam-Traumas bejahten nur 18% die Frage. Wurde die Frage aber wie folgt ergänzt: " .... Truppen schicken, um eine kommunistische Machtübernahme zu verhindern?", so stieg die Interventionsbereitschaft auf 33% (vgl. Schuman u. Presser 1981, S. 285). Diese Abhängigkeit der prozentualen Befragungsergebnisse von der genauen Frageformulierung öffnet zwar einerseits einer manipulativen Verwendung von Befragungsergebnissen in der Öffentlichkeit Tür und Tor, nämlich dann, wenn (wie üblich!) bei der Präsentation der Ergebnisse auf die genaue Wiedergabe der Frageformulierung verzichtet wird. Andererseits dokumentiert die Sensiblität in Bezug auf die Frageformulierung gerade die Validität der Befragung. Problematisch wäre aus methodischer Sicht allein der umgekehrte Effekt, wenn nämlich trotz unterschiedlicher

Frageformulierungen immer die gleichen Ergebnisse erzielt würden. Dennoch ist bei der Interpretation solcher "rohen" Prozentzahlen immer Vorsicht geboten.

Weit weniger abhängig von der Frageformulierung sind aber im allgemeinen theoretisch fundierte Beziehungen zwischen den Variablen, deren Überprüfung gerade das Hauptziel der empirischen Sozialforschung ist (oder zumindest sein sollte). Bezogen auf unser Beispiel kann man theoretisch gute Gründe für eine positive Beziehung zwischen einer rechts gerichteten politischen Einstellung und der Befürwortung von Militärinterventionen anführen. Die entsprechende empirische Beziehung sollte sich in allen Untersuchungen unabhängig von der konkreten Frageformulierung finden lassen und zwar mit ungefähr gleicher Stärke. Trotz Unterschieden in den Randverteilungen (hier den unterschiedlichen Prozentsätzen für die Militärintervention) sollten also die (bivariaten und multivariaten) Beziehungen zwischen den Variablen weitgehend unverändert bleiben, falls es sich um theoretisch fundierte Beziehungen handelt.

Geschlossene Fragen verlangen von dem Befragten, sich zwischen vorgegebenen Antwortkategorien zu entscheiden, während offene Fragen eine Antwort mit den eigenen Worten des Befragten verlangen. Offene Fragen haben also stärker qualitativen Charakter, da der Befragte in dem ihm eigenen sprachlichen Referenzsystem antworten kann und nicht in ein vorgegebenes Antwortschema hineingezwungen wird. Die vorgängige Formulierung von Antwortvorgaben durch den Sozialforscher setzt ein zutreffendes Vorwissen über die denkbaren Antwortmöglichkeiten voraus. Ein solches System von Antwortvorgaben sollte daher stets in einem Pretest auf Verständlichkeit und Vollständigkeit geprüft werden. Eine gut formulierte geschlossene Frage hat gegenüber der offenen Frage den Vorteil der höheren Standardisierung und der einfacheren Auswertbarkeit. Offene Fragen werden in der Sozialforschung heute nur noch relativ selten eingesetzt, da zum einen der Befragte bei der Formulierung der Antworten oft überfordert ist (unterschiedliches sprachliches Artikulationsvermögen Befragten) die erforderlichen der und zum anderen Auswertungstechniken (Inhaltsanalyse) sehr zeit- und kostenintensiv sind (zur methodischen Problematik der Inhaltsanalyse siehe 6.3). Gerade in Vorstudien sollte aber auf offene Fragen nicht ganz verzichtet werden. Eine weitere Alternative stellen sogenannte Hybridfragen dar, die es dem Befragten ermöglichen, neben den vorformulierten Antwortmöglichkeiten auch eine eigene Antwort zu formulieren (z.B. in einer Kategorie "sonstige Gründe, und zwar ..." bei der Frage nach den Gründen für eine Konsumentscheidung).

Allerdings zeigt gerade die Konsumforschung, daß die Befragten oft nicht in der Lage sind, auf eine direkte <u>Frage nach den Gründen</u> für eine Kaufentscheidung eine zutreffende Antwort zu geben. So berichtet Noelle-Neumann (1963, S. 274), daß bei einer Untersuchung der Gründe für den Kauf eines Radiogeräts das schöne Aussehen des Apparats kaum genannt wurde. Erst eine tiefergehende Analyse zeigte, daß das Design tatsächlich einen ausschlaggebenden Grund für den Kauf eines Radiogeräts darstellte. Solche tiefergehenden Analysen können z.B. mit Hilfe von projektiven Techniken durchgeführt werden (vgl. Green

und Tull 1982, S. 129ff. bzw. Noelle-Neumann u. Petersen 1996, S. 502f.). In einer berühmten Studie in Verbindung mit der Markteinführung von löslichem Kaffee gaben Konsumenten, die den Kauf von Nescafé ablehnten, als Grund den schlechten Geschmack an. Um die Vermutung zu überprüfen, daß sich hinter dieser stereotypen Ablehnung andere Gründe versteckten, wurde der folgende projektive Test entwickelt: In einem Fragebogenexperiment wurde eine Stichprobe zufällig in zwei Gruppen aufgeteilt (Randomisierung; siehe 3.3), wobei jeder Gruppe in ihrer Fragebogenversion eine Hausfrau vorgestellt wurde, die durch eine bestimmte Einkaufsliste charakterisiert war. Diese Einkaufsliste war für beide Gruppen bis auf die experimentelle Variable identisch, und enthielt eine Reihe typischer Lebensmittel, wie sie in jedem Haushalt benötigt werden. Der einzige Unterschied zwischen den beiden Einkaufslisten betraf den Kauf von Kaffee. Der Einkaufszettel für die erste Gruppe enthielt ein Pfund Bohnenkaffee, während in der zweiten Gruppe stattdessen eine Dose Nescafé gekauft wurde. Das Urteil der Befragten in beiden Gruppen fiel entsprechend unterschiedlich aus. Der Hausfrau, die Nescafe einkaufte, wurden wesentlich häufiger negative Eigenschaften, wie "bequem", "faul", "keine gute Hausfrau", usw. zugeschrieben wie der Hausfrau, die Bohnenkaffee bevorzugte. Das Ergebnis dieses projektiven Tests zeigt also die eigentlichen Gründe für die Ablehnung von Nescafé und weist darauf hin, daß das offen verbalisierte Geschmacksmotiv nur einen vorgeschobenen Grund darstellte. (Neuere Studien mit der Einkaufszetteltechnik zeigen übrigens eine deutlich gestiegene Akzeptanz von löslichem Kaffee.)

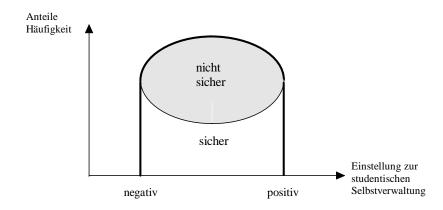
Eines der wichtigsten, aber zugleich auch schwierigsten Probleme bei der Befragung ist die Behandlung von Meinungslosigkeit. Grundsätzlich muß davon ausgegangen werden, daß sich nicht jeder Befragte zu jedem Thema eine Meinung gebildet hat. Die oft geübte Praxis, keine explizite "Weiß nicht"(WN)-Kategorie vorzugeben, um den Befragten dadurch zu zwingen, sich für eine der vorgegebenen inhaltlichen Antwortmöglichkeiten zu entscheiden, gefährdet sowohl die Reliabilität als auch die Validität der Messung. Zum einen besteht nämlich die Gefahr, daß sich der Befragte rein zufällig für eine der inhaltlichen Antwortkategorien entscheidet, was im Extremfall zu einer Reliabilität von annäherend Null führen würde (die Messung enthält nur noch Zufallsfehler). Andererseits sind auch systematische Verzerrungen möglich, etwa wenn nur nach Zustimmung oder Ablehnung gefragt wird und der Befragte in Ermangelung einer eigenen Meinung einfach zustimmt (Ja-Sage-Tendenz bei mangelndem Involvement).

Neben der Vorgabe einer expliziten WN-Kategorie (Quasifilter) besteht auch die Möglichkeit, eine echte Filterfrage vorzuschalten. Dazu ein Beispiel von Schuman und Presser (1981, S.125): In einer Umfrage in den USA aus dem Jahre 1977 wurde gefragt: "Versucht die Regierung der UDSSR, ihre Beziehungen zu den USA zu verbessern?"; eine sicherlich relevante und interessante Meinungsfrage, bei der man allerdings nicht ohne weiteres erwarten kann, daß sich die Befragten (insbesondere in den USA, wo außenpolitische Fragen in der öffentlichen Diskussion nur ein Randthema sind) dazu eine eigene Meinung gebildet

haben. Tatsächlich verneinten 56% der Befragten die vorgeschaltete Frage: "Haben Sie sich zu diesem Thema eine Meinung gebildet?" (echte Filterfrage, d.h. nur diejenigen, die die Filterfrage bejahen, werden um eine inhaltliche Stellungnahme gebeten). Wird auf die vorgeschaltete Filterfrage verzichtet, aber eine explizite WN-Kategorie als Antwortmöglichkeit vorgesehen, wird diese noch von 42% der Gefragten gewählt. Verzichtet man auch auf die explizite Vorgabe einer WN-Kategorie und läßt nur die Wahl zwischen Zustimmung und Ablehnung zu, geben nur 15% spontan an, keine Meinung zu haben, indem sie aus dem Schema der vorgegebenen Antwortkategorien ausbrechen. Eine explizite WN-Kategorie ist also grundsätzlich erforderlich, um Meinungslose herauszufiltern und zu reliablen und validen Daten zu gelangen. Dies gilt um so mehr, als hohe WN-Anteile auch ein Hinweis auf Mängel in der Frageformulierung sein können (z.B. schwer verständliche Formulierungen, doppelte Verneinungen, verschiedene Sachverhalte werden angesprochen, usw.). Daher ist insbesondere in Pretests eine explizite WN-Kategorie unabdingbar.

Man sollte allerdings auch nicht das Kind mit dem Bade ausschütten und durch eine zu scharf formulierte Filterfrage einen großen Teil der Befragten abschrecken, die sich durchaus eine Meinung gebildet haben, durch die vorgeschaltete explizite Filterfrage aber unsicher werden, ob sie kompetent genug sind, sich zu der angesprochenen Thematik äußern zu können. Möchte man etwa die Einstellung zur Atomkraft messen und beginnt mit einer Filterfrage danach, ob die Befragten die öffentliche Diskussion über Sicherheitsrisiken von Leichtwasserreaktoren verfolgt haben, so wird man viele Befragte abschrecken, die sich durchaus eine eigene Meinung zur Atomkraft gebildet haben. Daher sollte eine solche Filterfrage nur dann eingesetzt werden, wenn man wirklich nur an den Befragten interessiert ist, die eine argumentativ fundierte Aussage zu dem angesprochenen Themenbereich machen können (in den meisten Fällen eine Minderheit).

Eine Alternative zu der Strategie, von vorneherein Meinungslose auszufiltern, ist es, im nachhinein zu fragen, wie sicher sich der Befragte seiner Antwort ist. Diese Vorgehensweise wurde von Sample und Warland (1973) bei einer Befragung von Studenten verwandt, in der die Einstellung zur studentischen Selbstverwaltung auf einer Likert-Skala gemessen wurde. Es ergab sich folgende Verteilung der Einstellung zwischen dem negativen und dem positiven Pol der Skala:



Wie man sieht, hat die Gesamtverteilung eine typisch unimodale Form. Die Befragten, die sich ihrer Einstellung nicht sicher waren, gruppieren sich dabei stärker um den Mittelwert der Skala. Betrachtet man dagegen nur die Befragten die sich ihrer Einstellung sicher waren, so erhält man eine bimodale (u-förmige) Verteilung, da sich insbesondere Befragte mit einer extremen Einstellung ihrer Meinung sicher waren. Der Vorteil dieser Vorgehensweise ist es, nicht von vorneherein eine Ja-Nein-Entscheidung der Befragten wie bei einer echten Filterfrage zu verlangen, sondern im nachhinein eine Selbsteinschätzung der Sicherheit der gerade gegebenen Antworten durch den Befragten vornehmen zu lassen. Wie die Ergebnisse von Sample und Warland zeigen, ist die Validität der Messung in den beiden Gruppen sehr verschieden. Verwendet man die Beteiligung an den Wahlen zur studentischen Selbstverwaltung als externes Kriterium (vgl. 5.4.1.2), so zeigt sich nur bei den Befragten, die sich ihrer Antwort sicher waren, eine hohe Validität (die Korrelation zwischen Einstellung und Verhalten beträgt hier .47 im Vergleich zu .06 für die Befragten, die sich ihrer Antwort nicht sicher waren).

Zusammenfassend kann gesagt werden, daß die Meinungslosigkeit durchaus eine erlaubte und inhaltlich sinnvolle Antwortmöglichkeit darstellt, die explizit erhoben werden sollte. Die WN-Kategorie enthält also (im Gegensatz zu Antwortverweigerung) eine echte Information über die Meinung des Befragten - die Bezeichnung Meinungslosigkeit ist im Grunde irreführend und sollte besser durch den in der amerikanischen Literatur üblichen Begriff der Nicht-Einstellung (Non-Attitude) ersetzt werden.

### **Exkurs** zum Problem der Einstellungs-Verhaltens-Konsistenz:

In diesem Zusammenhang ist eine kurze Bemerkung zu der Einstellungs-Verhaltens-Kontroverse in der Sozialpsychologie angebracht. Einstellungen sind relativ dauerhafte, miteinander verknüpfte Überzeugungen in Bezug auf einen Sachverhalt mit einer kognitiven, affektiven und konativen Komponente, die Handlungen gegenüber dem Einstellungsobjekt beschreiben, bewerten und rechtfertigen. Einstellungen sind also hypothetische Konstrukte, denen aber durchaus eine handlungsorientierende Funktion und auch eine konkrete Handlungsrelevanz zukommt. Wie aber (nicht nur) das Beispiel der Beziehung zwischen umweltbewußter Einstellung und umweltgerechtem Verhalten zeigt, kann nicht generell von dem Postulat eine Einstellungs-Verhaltens-Konsistenz ausgegangen werden; die Beziehung zwischen Einstellung und Verhalten ist vielmehr eher indirekt und situationsabhängig. Eine situationsunabhängige Determination des Verhaltens durch Einstellungen würde grundlegenden Einsichten der psychologischen Verhaltenstheorie und der ökonomischen Theorie rationalen Handelns widersprechen (vgl. auch Bierbrauer 1976). Eine konkrete Handlung kann nämlich immer nur vor dem Hintergrund der Interpretation einer spezifischen Situation verstanden werden. So zeigt z.B. die klassische Studie über Vorurteile gegenüber Schwarzen (Minard 1952), daß sich junge weiße Bergarbeiter im Süden der USA in Situationen des Alltagslebens über Tage durchaus einstellungskonform rassistisch verhalten,

unter Tage aber kooperativ mit ihren schwarzen Kollegen zusammenarbeiten - also in einer Situation, die durch eine starke gegenseitige Abhängigkeit, geringe Sichtbarkeit und eine geringe soziale Kontrolle charakterisiert ist. Auch für unser eingangs angesprochenes Beispiel lassen sich aus dem Alltag eines jeden Lesers wohl mühelos Fälle von Diskrepanzen zwischen Umwelteinstellung und Verhalten finden, die auch die Bedeutung der Interpretation der konkreten Situation durch den Handelnden hinweisen. Es ist also nicht erstaunlich, daß in der empirischen Sozialforschung meistens relativ geringe Korrelationen zwischen Einstellung und Verhalten gefunden werden.

Wie aber die oben diskutierte Untersuchung von Sample und Warland zeigt, sind dabei neben dem theoretischen Problem der Situationsabhängigkeit auch grundlegende methodologische Fragen von Bedeutung. Zu fragen ist, ob der Befragte überhaupt über eine strukturierte Einstellung verfügt, die in der Befragungssituation (!) einfach abgerufen werden kann. Ist dies der Fall, so ist, wie die Studie zeigt, durchaus eine substantielle Einstellungs-Verhaltens-Korrelation zu erwarten. Nur in diesem Fall besteht nämlich überhaupt die Chance, eine reliable und valide Messung zu erhalten. Das zugrunde liegende hypothetische Konstrukt muß also mehr sein als nur eine theoretische Konstruktion des Forschers; es muß von dem Befragten im Rahmen seines Weltbildes verstanden und, wenn auch möglicherweise implizit, zur Orientierung in alltäglichen Handlungssituationen verwendet werden. Die WN-Kategorie oder, besser noch, die nachträgliche Frage danach, wie sicher sich der Befragte bei seinen Antworten war, bietet dem Befragten die Möglichkeit anzuzeigen, daß er die Fragestellung in der vorgelegten Form nicht versteht. Damit macht er deutlich, daß Diskrepanzen zwischen seinem lebensweltlich verankerten Wissenshorizont und den abstrakt formulierten, oft aus wissenschaftlichen Theorien abgeleiteten Fragestellungen bestehen.

Natürlich ist es eigentlich die Aufgabe eines Pretestes, solche Diskrepanzen zu erkennen und durch verbesserte Frageformulierungen zu beseitigen. Dies ist aber nur in solchen Fällen möglich, wo die beschriebenen Diskrepanzen auf Schwierigkeiten des sprachlichen Verständnisses zurückzuführen sind. Haben die abzufragenden theoretischen Konstrukte aber in der Alltagswelt des Befragten überhaupt keine Bedeutung, kann auch durch eine noch so geschickt formulierte Frage keine valide Meinungsäußerung des Befragten erhalten werden. Die oft enttäuschenden Ergebnisse von Einstellungsverhaltensstudien, sind also zumindestens teilweise auch auf die leider immer noch übliche Praxis zurückzuführen, auf eine explizite Weiß-Nicht-Kategorie zu verzichten und auf diese Weise Meinungen von Befragten zu erheben, die gar nicht vorhanden sind. Solche Nicht-Einstellungen, die fälschlicherweise als eine valide Messung interpretiert werden, können dann natürlich auch nicht mit dem berichteten Verhalten des Befragten korrelieren.

<u>Sensitive Fragen</u>: In vielen Fällen will der Sozialforscher Dinge von dem Befragten wissen, die dieser nur im Freundeskreis, oder vielleicht nicht einmal dort, äußern würde. Bei solchen sensitiven Themenbereichen besteht die Gefahr eines hohen Anteils von

Antwortverweigerungen bzw. einer geringen Reliabilität und Validität der Antworten. Antworthemmungen werden üblicherweise erzeugt durch Fragen nach Handlungen, die in der Gesellschaft negativ sanktioniert werden, aber auch durch Fragen nach tabuisierten Handlungsbereichen, z.B. zum Intimbereich. Was allerdings zum Intimbereich zählt, ist nicht konstant, sondern unterscheidet sich je nach Kultur und auch nach sozialer Schicht. So ist es bei manchen Personenkreisen schwieriger, Fragen über das Einkommen zu stellen als über das Sexualverhalten (vgl. Scheuch 1973, S. 117). Als noch problematischer erweist sich oft das Thema der körperlichen Hygiene. Gerade in westlichen Gesellschaften sind viele Befragte noch eher geneigt, über abweichendes Sexualverhalten zu sprechen, als über eine Verletzung der Normen der körperlichen Sauberkeit (vgl. Scheuch, 1973, S. 119).

Zur Umgehung von Antworthemmungen bei sensitiven Fragen sind eine Reihe von Techniken entwickelt worden. Grundlegend ist zunächst einmal die Zusicherung absoluter Vertraulichkeit (z.B. auch durch die Randomized-Response-Technik). Eine Methode zur Verringerung von Antworthemmungen ist die Entschärfung oder Verharmlosung der Frage durch geeignete Formulierungen: Statt vom Diebstahl spricht man z.B. davon, daß jemand "etwas weggenommen" hat oder statt von Steuerhinterziehung ist die Rede davon, daß jemand das Finanzamt "über's Ohr haut". Auch der Appell an den Mitläufereffekt hat sich bewährt, etwa wenn man eine Reihe von Fragen zur Steuerhinterziehung mit der Bemerkung einleitet, daß bekanntermaßen eine Mehrheit der Bundesbürger schon einmal in irgendeiner Form eine Steuerhinterziehung begangen habe. Oft erweisen sich auch harte Befragungstechniken als wirksam. Das klassische Beispiel dazu ist der Kinsey-Report: Es wurde sehr schnell gefragt und die Fragen wurden in einem sehr bestimmten Ton vorgetragen. Dabei wurde nicht gefragt, ob der Befragte eine bestimmte Sexualpraktik verwendet, sondern wann er sie zum letzten Mal ausgeübt habe. Charakteristisch in diesem Zusammenhang ist ein Kommentar von Kinsey (1970 (zuerst 1948), S. 62): "Solch ein Schnellfeuer von Fragen bietet eines der wirksamsten Mittel, um Schwindeleien aufzudecken, wie Kriminal- und Polizeibeamte wohl wissen." Gerade dieser letzte Satz charakterisiert treffend die Befragungsatmosphäre, die durch diese Art der Befragung hergestellt wird. Es dürfte zweifelhaft sein, ob eine solche Befragungstechnik heute noch erfolgreich sein würde.

Reihenfolge von Fragen (Halo-Effekt): Eine Frage strahlt auf die nachfolgenden sowohl inhaltlich wie auch emotional aus. Der Inhalt einer gerade beantworteten Frage beschäftigt den Befragten noch, während ihm bereits die nächste Frage gestellt wird. Dadurch wird diese Frage in einem spezifischen Kontext gestellt. Ein Beispiel dafür sind Kontrasteffekte: Wird z.B. in einer Vorfrage gefragt, ob Reis ein typisch deutsches Nahrungsmittel ist, so wird dadurch die Folgefrage danach, ob Kartoffeln ein typisch deutsches Nahrungsmittel sind, in einen speziellen Zusammenhang gerückt. Konkret bedeutet dies, daß die Zustimmung höher ausfallen wird, als dies ohne die Vorfrage der Fall sein würde. Keine Frage wird von dem Befragten isoliert gesehen. Der Ausstrahlungseffekt ist insbesondere auf das Bemühen des

Befragten um Konsistenz bei seinen Antworten zurückzuführen. In der Regel wird der Halo-Effekt einer Frage um so stärker sein, je größer das dadurch induzierte emotionale Engagement des Befragten ist. Die Tatsache des Ausstrahlungseffekts nutzt man gezielt bei der <u>Trichterung</u>. Mit allgemeinen Fragen wird der Befragte zunächst in ein Thema eingeführt. Anschließend geht man mit immer spezielleren Fragen bis in Einzelaspekte der Thematik: Die Fragen über spezielle Einzelaspekte könnte der Befragte vermutlich gar nicht beantworten, wenn nicht die Thematik von Frage zu Frage weiter entfaltet würde. Bei der Trichterung arbeitet man also mit einem gewollten Ausstrahlungseffekt. In anderen Fällen ist dieser Effekt aber absolut unerwünscht, insbesondere bei Meinungsfragen.

Generell gilt, daß der <u>Aufbau des Fragebogens</u> sorgfältig überlegt werden sollte. Die Einleitungsfragen sind besonders wichtig für den Verlauf des Interviews, sie sollten interessant und für den Befragten leicht zu beantworten sein. Auf diese Weise wird der Befragte in zwangloser Weise zum Thema der Befragung hingeführt ("Eisbrecher"-Fragen). Ebenso wichtig ist die übersichtliche Gliederung des Fragebogens nach Fragekomplexen. Dadurch kann die Möglichkeit von Plazierungseffekten (Halo-Effekten) deutlich reduziert werden. Wichtig sind neben der Trichterung von Fragen in den einzelnen Fragekomplexen auch Überleitungsfragen, die quasi als Pufferfragen Ausstrahlungseffekte zwischen verschiedenen Teilen des Fragebogens möglichst verhindern sollen. Schwierige Fragen werden im allgemeinen an das Ende des Fragebogens gestellt, um die Kosten bei einem möglichen Abbruch der Befragung möglichst gering zu halten.

Formulierung der Fragen: Wie bereits mehrfach betont, hat die genaue Wortwahl bei der Frageformulierung einen entscheidenden Einfluß auf die Beantwortung der Frage. So wurde in einem Fragebogenexperiment im Rahmen des ALLBUS 1984 in der ersten Frageversion nach Interessenkonflikten in der Bundesrepublik und in der zweiten Fragebogenversion nach Interessengegensätzen gefragt. Während nur 31% der Befragten einen Interessenkonflikt zwischen arm und reich in der Bundesrepublik erkennen konnten, erhöhte sich die Zustimmungsrate bei der schwächer formulierten Frage nach den Interessengegensätzen erwartungsgemäß auf 40%. Fragt man nach Kontakten zwischen Deutschen und Gastarbeitern (vgl. Reuband 1989), so erhält man bei der Frage nach dem "Zusammenkommen am Arbeitsplatz oder am Wohnort" eine Kontaktrate von 60%. Fragt man, ob der Befragte "persönlich Kontakt" zu Gastarbeitern hat, reduziert sich die Kontaktrate auf 36% und bei der Frage nach einem "persönlichen Kontakt" sogar auf 28%. Fragt man konkret nach "Kontakten im Freundes- oder Bekanntenkreis", sinkt die Kontaktrate weiter auf 15% und fragt man schließlich, "ob der Befragte ab und zu einen gemeinsamen Abend mit Gastarbeitern verbringt", so berichten nur noch 2% der Befragten einen Kontakt mit Gastarbeitern. Wie bereits bemerkt, dokumentiert sich in der Abhängigkeit der Ergebnisse von der konkreten Frageformulierung gerade die Validität der Befragungsmethode. Dies bedeutet aber auf der anderen Seite auch, daß jede Frage bis in die einzelnen Formulierungen hinein genau überdacht sein sollte.

Reihenfolge der Antwortkategorien: Wird dem Befragten eine Liste mit Antwortvorgaben vorgelegt, aus denen er auswählen soll, so zeigt sich bei der schriftlichen Vorlage ein sog. Primacy-Effekt (die ersten Kategorien werden überzufällig häufig ausgewählt) und bei der mündlichen Vorlage (Vorlesen) ein sog. Recency-Effekt (die letzten Vorgaben auf der Liste werden überzufällig häufig ausgewählt). Dabei sind die Unterschiede um so geringer, je eingehender der Befragte sich bereits mit dem Thema beschäftigt hat. Auch durch vorangehende Kontextfragen (Trichterung) können diese Effekte verringert werden. Eine Möglichkeit zur Vermeidung von Reihungseffekten besteht in der Randomisierung der Reihenfolge der Antwortvorgaben, etwa dadurch, daß die Antwortvorgaben in Form eines Kartenspiels vor dem Stellen der Frage durch den Interviewer gemischt werden. Primacy- und Recency-Effekte können mit Hilfe Theorien über die menschliche Informationsverarbeitung erklärt werden, wie sie im Rahmen der kognitiven Psychologie entwickelt wurden (vgl. Sudman u.a. 1996).

Antwortkategorien als Referenzsystem für den Befragten: Die Befragten neigen dazu, aus vorgegebenen Antwortkategorien hypothetische Normen zu konstruieren und sich dann tendenziell am Mittelbereich dieses als gesellschaftlich üblich interpretierten Spektrums zu orientieren. Fragt man z.B. nach der Dauer des Fernsehkonsums, so orientiert sich der Befragte bei seiner Antwort an den vorgegebenen Antwortkategorien. Verwendet man z.B. eine sechsteilige Antwortskala, die von "weniger als eine halbe Stunde" bis zu "mehr als zweieinhalb Stunden" reicht, so wählen nur 16% die Kategorie "mehr als zweieinhalb Stunden". Verändert man dagegen die Antwortvorgaben und fragt mit einer ebenfalls sechsteiligen Skala, die von "weniger als zweieinhalb Stunden" über "zweieinhalb bis drei Stunden" usw. bis zu "mehr als viereinhalb Stunden" reicht, so entscheiden sich 38% für die (insgesamt fünf) Kategorien mit mehr als zweieinhalb Stunden Fernsehkonsum (zusammengefaßt). Der Befragte versucht also, aus den vorgelegten Antwortkategorien in etwa den gesellschaftlich üblichen durchschnittlichen Fernsehkonsum zu erschließen und paßt dann die Angaben über seinen eigenen Fernsehkonsum an diesen Mittelwert an. Antwortalternativen sind also weit davon entfernt, nur eine neutrale Bezugsreferenz zu sein. Sie dienen vielmehr zur Strukturierung des Raumes der Antwortmöglichkeiten durch den Befragten (vgl. Schwarz u.a. 1989). Auch diese Effekte können durch Sätze der kognitiven Psychologie theoretisch abgeleitet werden.

In einem anderen Befragungsexperiment wurden die Befragten gebeten zu sagen, für wie erfolgreich sie sich im Leben halten. Dabei wurde einmal eine von 0 bis 10 reichende Skala mit den Endpunkten "überhaupt nicht erfolgreich" und "sehr erfolgreich" verwendet und in der zweiten Version eine von -5 bis +5 reichende Skala mit den gleichen Endpunkten. Entspre-

chend der <u>Mean-Shift-Hypothese</u> wird die Bedeutung der Skala durch die numerische Bezeichnung der Endpunkte festgelegt. Das bedeutet konkret, daß im ersten Fall (Endpunkt 0) die Kategorie "überhaupt nicht erfolgreich" in der Interpretation des Befragten lediglich das Fehlen von Erfolg anzeigt (unipolare Interpretation), während die gleiche Bezeichnung in der zweiten Version (Endpunkt -5) Mißerfolg bedeutet (bipolare Interpretation). Wie die Ergebnisse des Experiments zeigen, stuften sich die Befragten in der zweiten Version eher als erfolgreich ein (Bestätigung der Mean-Shift-Hypothese).

Abschließend soll noch einmal festgestellt werden, daß ein vorgängiges theoretisches Konzept eine unverzichtbare Voraussetzung für die Gestaltung eines jeden Fragebogens ist. Nur so kann letztlich darüber entschieden werden, welche Informationen überhaupt erhoben werden sollen. Es genügt eben nicht, einfach nur "interessante" Fragen zu stellen und dann darauf zu vertrauen, daß einem bei der Auswertung schon die richtigen Ideen kommen werden. Bei der konkreten Ausgestaltung eines Fragebogens sollte man stets die Möglichkeit in Erwägung ziehen, ob nicht bereits erprobte Erhebungsinstrumente übernommen werden können. In diesem Zusammenhang ist z.B. die von ZUMA entwickelte Standarddemographie zu nennen, die Fragen zu allen üblicherweise zu erhebenden soziodemographischen Variablen enthält, so z.B. Fragen zur Berufstätigkeit und zur Haushaltszusammensetzung. Auch für viele Einstellungen existieren bereits bewährte und testtheoretisch abgesicherte Skalen, die in dem ebenfalls von ZUMA herausgegebenen Skalenhandbuch zusammengefaßt sind. Vor Durchführung einer Befragung sollte unbedingt ein Pretest durchgeführt werden, um den Aufbau des Fragebogens und die Formulierung der einzelnen Fragen zu überprüfen.

# 6.1.2 Antwortverzerrungen (Response Errors) und die Gefahr von Forschungsartefakten

Ein besonderes Problem stellen <u>Antwortverzerrungen</u> im Interview dar, die sich aus der Reaktivität des Verfahrens ergeben und grundsätzlich die Gefahr von Forschungsartefakten bei der Befragung heraufbeschwören. Die Befragung ist als soziale Interaktionssituation grundsätzlich reaktiv, so daß es nicht darum gehen kann, die Reaktivität des Verfahrens gänzlich auszuschalten. Vielmehr muß eine <u>Instrumententheorie der Befragung</u> aus einer allgemeinen Theorie sozialer Interaktion entwickelt und auf die Interviewsituation übertragen werden, die es ermöglicht, die Bedeutung für das Auftreten von Antwortverzerrungen zu klären, um so mögliche Antwortverzerrungen möglichst gering zu halten und in den Auswertungen berücksichtigen zu können.

Zu den bereits behandelten Antwortverzerrungen im Interview gehören:

- die Abgabe einer inhaltlichen Antwort bei Meinungslosigkeit
- die Angabe von "Weiß-nicht" trotz eigener Meinung
- Antwortverweigerung (z.B. bei sensitiven Fragen)

- Ausstrahlungseffekte (Halo-Effekte)
- Frageeffekte (z. B. Primacy- bzw. Recency-Effekte)

Zu den Antwortverzerrungen, die sich allgemein aus grundlegenden Charkteristika der Interviewsituation ergeben, gehören die Zustimmungstendenz und die (kulturelle und situationale) soziale Erwünschtheit.

Zustimmungstendenz unabhängig vom Inhalt der Frage (Ja-Sage-Tendenz): Die Zustimmungstendenz wurde von uns bereits an verschiedenen Stellen erwähnt, z.B. im Zusammenhang mit dem Hinweis auf die Bedeutung negativer Items bei der Konstruktion einer Likert-Skala. Die Zustimmungstendenz erzeugt stabile Antwortmuster und ist daher gewöhnlich mit einer hohen Reliabilität verbunden. Da den Antworten aber keine inhaltliche Bedeutung zukommt, liegt ein systematischer Meßfehler vor - die Validität der Messung ist nicht gegeben. Sehr eingehend wurde die Ja-Sage-Tendenz im Zusammenhang mit der von Adorno u.a. entwickelten Autoritarismus-Skala dokumentiert. (Da die Items der Autoritarismus-Skala durchgehend positiv formuliert sind (vgl. die Ausführungen zur Likert-Skala in 5.3.1), ist die Skala besonders anfällig für die Ja-Sage-Tendenz.) Ein weiteres Beispiel sind Antwortverzerrungen bei unbalancierten Fragen (vgl. z.B. die unbalancierte Frage nach einer obligatorischen Mitgliedschaft in der Gewerkschaft in 6.1.1). Ein anderes Beispiel wird von Schuman und Presser (1981, S. 208) berichtet und stammt aus einer Umfrage in den USA aus dem Jahre 1974. Danach stimmten 57% der Befragten der Aussage zu, daß für die steigende Kriminalität eher die sozialen Bedingungen als der einzelne selbst verantwortlich sind, aber sogar 60% stimmten der umgekehrten Formulierung zu. Werden in einer balancierten Frageversion beide Antwortmöglichkeiten als Alternativen vorgegeben, entscheidet sich eine Mehrheit von 54% für die sozialen Bedingungen. (Berücksichtigt man allerdings den hohen Stellenwert individueller Eigenverantwortung in der amerikanischen Kultur, so ist auch die Möglichkeit nicht ganz von der Hand zu weisen, daß die hohe Zustimmungsrate von 60% für die individuelle Zuschreibung bei der unbalancierten Fragestellung zumindest teilweise auf einen Effekt der kulturellen sozialen Erwünschtheit zurückzuführen ist.)

Manche Autoren versuchen, die Ja-Sage-Tendenz durch Hinweis auf besondere <u>Persönlichkeitsmerkmale</u>, z.B. eine mangelnde Ich-Stärke der Ja-Sager, zu erklären. Zur Begründung wird dabei auf die Beziehung zwischen niedrigem sozialen Status und Zustimmungstendenz verwiesen, wie sie z.B. im Zusammenhang mit der Autoritarismusforschung festgestellt wurde.

(Die Korrelation zwischen niedriger sozialer Schicht und Autoritarismus erscheint daher zumindest teilweise als Methodenartefakt. Sie ist nämlich wenigstens zum Teil auf die Ja-Sage-Tendenz als intervenierende Variable zurückzuführen:

soziale Schicht — Ja-Sage-Tendenz — Autoritarismus).

Theoretisch plausibler erscheint aber eine stärker situationsbezogene Erklärung zu sein, die die Zustimmungstendenz als eine gelernte Anpassungsstrategie in Situationen versteht, die einerseits zwar unklar definiert, andererseits aber für den Betroffenen nicht von besonderer Bedeutung sind. Danach wäre die Ja-Sage-Tendenz eine rationale Strategie in einer Befragungssituation, in der der Befragte (aus seiner Sicht) unklar formulierte und schwer verständliche Fragen beantworten soll, die ihn, wenn überhaupt, nur am Rande interessieren (vgl. Esser 1986). Wir werden auf die unterschiedlichen Erklärungsansätze für Antwortverzerrungen noch einmal systematisch im Zusammenhang mit Instrumententheorie der Befragung zurückkommen.

Kulturelle soziale Erwünschtheit (im Sinne der Orientierung an sozialen Normen für sozial erwünschtes Verhalten): Wie wir aus kontrollierten Untersuchungen aus den USA zur Validität von Interviews wissen, sind die Angaben der Befragten über ihren Zigaretten- bzw. Alkoholkonsum systematisch untertrieben. Wie diese Untersuchungen weiter zeigen, ist die Untertreibung des Alkoholkonsums (in den USA!) bei Protestanten stärker als bei Katholiken. Der umgekehrte Effekt zeigt sich bei Angaben über einen Schwangerschaftsabbruch. Der Effekt der kulturellen sozialen Erwünschtheit ist also immer bei den Gruppen am stärksten, bei denen eine subkulturell besonders ausgeprägte soziale Verhaltensnorm besteht.

Abweichungen zwischen mündlichen und schriftlichen Befragungen können ebenfalls im Sinne eines Effekts der kulturellen sozialen Erwünschtheit gedeutet werden. Wegen der geringeren sozialen Kontrolle sind die Befragten in schriftlichen Befragungen eher bereit, abweichende Einstellungen oder Verhaltensweisen zuzugeben. So gehört es sicher zu den fundamentalsten gesellschaftlichen Normen (vgl. z.B. die Zehn Gebote), daß Kinder ihre Eltern lieben sollten. Vergleicht man nun den Anteil der Befragten die angeben, ihre Mutter sehr (very dearly) zu lieben, so sinkt dieser Anteil in der schriftlichen Befragung im Vergleich zur mündlichen von 37% auf 25% ab, was auf einen geringeren Effekt der kulturellen sozialen Erwünschtheit in der schriftlichen Befragung hinweist (vgl. Esser 1986).

Situationale soziale Erwünschtheit (der Befragte orientiert sich an Charakteristiken der Interviewerstuation, insbesondere an sozialen Merkmalen des Interviewers; aber auch die Anwesenheit Dritter oder der (erkennbare) Sponsor einer Untersuchung können zur Definition der Befragungssituation durch den Befragten beitragen und so sein Antwortverhalten beeinflussen): Die Interviewereffekte stellen die wichtigste Form der Antwortverzerrungen durch situationale soziale Erwünschtheit dar. Im Kern ist das Interview (von Angesicht zu Angesicht) eine soziale Interaktion zwischen Interviewer und Befragtem. Noch bevor das Interview beginnt, entwickelt der Befragte bestimmte Vermutungen hinsichtlich der Einstellungen und Verhaltensweisen des Interviewers, die im wesentlichen von den wahrnehmbaren sozialen Merkmalen des Interviewers, also Geschlecht, Alter, Kleidung, Auftreten, Sprechweise, usw. abhängen. Anhand derartiger Merkmale macht sich der Befragte

ein hypothetisches Bild von der Person des Interviewers und orientiert sein Verhalten daran. Zahlreiche Untersuchungen haben den Einfluß wahrnehmbarer sozialer Merkmale des Interviewers auf das Antwortverhalten des Befragten nachgewiesen:

- Arbeiter stimmen der Forderung nach einem Gesetz gegen spontane Streiks eher zu, wenn der Interviewer der Mittelschicht angehört (59% im Vergleich zu 44%, wenn der Interviewer selbst der Arbeiterschicht angehört) [USA 1942].
- Vorurteile gegenüber Juden ("Juden haben in den USA zuviel Macht") werden gegenüber "jüdisch aussehenden" Interviewern seltener geäußert (6% im Vergleich zu 25%) [USA 1946].
- Die Angaben zur Geschlechtsrollenorientierung sind zumindest teilweise von dem Geschlecht des Interviewers abhängig und zwar geben sich Männer <u>und</u> Frauen eher emanzipierter, wenn sie von einer Frau, und eher traditional, wenn sie von einem Mann interviewt werden. Die folgenden Daten stammen aus dem Schleswig-Holstein-Survey (vgl. 1.). Die nGRO wurde als Likert-Skala gemessen; die zugrundeliegende Ratingskala hatte die Endpunkte 1=traditionaler Pol und 5=emanzipierter Pol:

#### Mittelwerte über die nGRO nach Geschlecht

#### Geschlecht des Interviewers

Geschlecht des Befra	agten

	3	9
$\bigcirc$	3.27	3.55
9	3.48	3.80

In vielen Untersuchungen wurde gefunden, daß solche Interviewereffekte eher bei Personen aus den unteren sozialen Schichten (niedrige Bildung) auftreten. Zur Erklärung wurden ähnliche Argumente vorgebracht, wie im Falle der Ja-Sage-Tendenz. Es ist aber nicht ohne weiteres einsichtig, daß die Voraussetzung für das Eintreten eines Interviereffekts, nämlich eine ausreichende kognitive Fähigkeit, um die vom Interviewer ausgehenden sozialen Signale interpretieren zu können, und eine ausgeprägte soziale Anpassungsfähigkeit, bei den mittleren und oberen sozialen Schichten weniger ausgeprägt sein sollten. Es ist eher zu vermuten, daß bei einer geeigneten Thematik eine solche situationsspezifische Form der Eigendarstellung auch und gerade bei den gehobenen sozialen Schichten zu beobachten sein sollte. Tatsächlich sind solche Interviewereffekte im Hinblick auf den Postmaterialismus gefunden worden, der gerade für junge und hochgebildete Befragte eine starke lebensweltliche Orientierungskraft besitzt. Um den Interviewereffekt zu messen, wurden die Anteile der reinen Postmaterialisten in Abhängigkeit vom Alter des Interviewers untersucht. [Das Alter des Interviewers kann als sichtbares Merkmal von dem Befragten mit dessen Postmaterialismus in Verbindung gebracht

werden (dies setzt allerdings ein entsprechendes Wissen auf Seiten des Befragten über die Beziehung "je jünger, desto postmaterialistischer" voraus).] Ein Interviewereffekt sollte sich dann derart bemerkbar machen, daß der Anteil der Postmaterialisten unter den Befragten bei einem jungen Interviewer höher ist als bei einem alten. In der folgenden Tabelle verwenden wir diese Prozentsatzdifferenz als Maß für den Interviewereffekt:

Interviewereffekt im Hinblick auf das Alter des Interviewers (Prozentsatzdifferenz) in Abhängigkeit von der Bildung und dem Alter des Befragten (vgl. Reuband 1986).

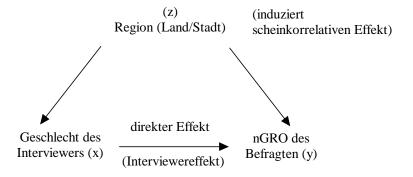
Alter

Bildung

	jung	mittel	alt
niedrig	- 2%	0%	+ 9%
hoch	+ 31%	+ 30%	- 2%

Wie ersichtlich ist, zeigt sich ein deutlicher Interviewereffekt bei den Befragten mit <u>hoher</u> Bildung in den beiden unteren Altersgruppen.

Ein Interviewereffekt liegt vor, wenn das Antwortverhalten des Befragten kausal von einem Merkmal des Interviewers beeinflußt wird. Um das Vorliegen eines Interviewereffekts empirisch zu bestätigen, muß daher eine Beziehung zwischen einem Interviewermerkmal (z.B. Geschlecht) und einem Merkmal des Befragten (z.B. nGRO) gefunden werden. Soll eine kausale Interpretation gerechtfertigt sein, darf diese Beziehung bei Kontrolle relevanter Drittvariablen nicht verschwinden (vgl. 3.1). Tatsächlich haben manche Autoren vermutet, daß viele als Interviewereffekte interpretierte Beziehungen lediglich Scheinkorrelationen sind, die durch regionale Variation der beteiligten Variablen hervorgerufen werden. Sind z.B. die Interviewer in den städtischen Regionen überwiegend weiblich und ist die nGRO der Befragten in diesen Regionen eher emanzipiert (und gilt das entsprechende mit umgekehrten Vorzeichen in ländlichen Regionen), so würde die Variable Region eine Scheinkorrelation induzieren, die keineswegs einen Interviewereffekt, d.h. einen kausalen Einfluß des Interviewermerkmals Geschlecht auf das Antwortverhalten des Befragten (nGRO) anzeigt.



Tatsächlich können, wie die sorgfältige Untersuchung von Hoag und Allerbeck (1981) zeigt, solche scheinkorrelativen Effekte in geringem Umfang nachgewiesen werden. Im

wesentlichen bestätigt die Studie aber die kausale Interpretation der Beziehung als Interviewereffekt (direkter kausaler Effekt).

Allerdings ist das Auftreten von Interviewereffekten eher die Ausnahme als die Regel. Interviewereffekte treten häufiger bei Einstellungs- als bei Verhaltensfragen, bei offenen als bei geschlossenen Fragen und bei komplizierten Fragen auf, die ein Eingreifen durch den Interviewer erfordern. Allgemein konnte gezeigt werden, daß Interviewereffekte um so unwahrscheinlicher sind, je stärker das Interesse des Befragten an der Thematik der Befragung ist. Umgekehrt sind Interviewereffekte um so wahrscheinlicher, je stärker das Interesse des Interviewers an der Thematik der Befragung ist (vgl. Hermann 1983). Gerade das letzte Resultat verdeutlicht noch einmal, wie wichtig eine neutrale Haltung des Interviewers für die Validität der Datenerhebung ist.

Aber selbst das Auftreten starker Interviewereffekte muß nicht unbedingt zu einer grundlegenden Veränderung der Pfadkoeffizienten in einem Kausalmodell führen. So konnte mit Hilfe der Daten des ALLBUS 1980, der eine zusätzliche Interviewerbefragung enthält, gezeigt werden, daß sich Befragte dann verstärkt zu liberalen Erziehungszielen bekannten, wenn diese auch von dem Interviewer präferiert wurden (vgl. Schanz und Schmidt 1984). Dieses Ergebnis ist besonders bemerkenswert, da es sich bei den Erziehungszielen (ebenso wie bei dem Postmaterialismus; siehe oben) um ein nicht direkt beobachtbares Merkmal handelt, und die Befragten daher auf Rückschlüsse aus dem äußeren Erscheinungsbild und dem Verhalten des Interviewers angewiesen waren. Fügt man nun das vom Interviewer präferierte Erziehungsziel als weitere unabhängige (erklärende) Variable in ein Kausalmodell ein, das die Erziehungsziele der Befragten erklären soll, so zeigt sich zwar ein direkter kausaler Effekt (d.h. es handelt sich tatsächlich um einen echten Interviewereffekt und keine Scheinkorrelation); die kausalen Effekte der anderen unabhängigen Variablen im Modell (Merkmale der Befragten) bleiben aber unverändert. Selbst wenn ein Interviewereffekt, d.h. ein systematischer Meßfehler, auftritt, führt dies in der Regel also nicht zu einer Verzerrung der geschätzten kausalen Effekte anderer Variablen. (Dies gilt zumindestens immer dann, wenn das verursachende Interviewermerkmal und die übrigen unabhängigen Variablen des Kausalmodells (Befragtenmerkmale) unkorreliert sind.) Darin zeigt sich noch einmal die weitgehende <u>Unempfindlichkeit von Beziehungsmustern</u> (im Gegensatz zu Randverteilungen; siehe oben in 6.1.1) gegenüber Befragungseffekten.

Eine weitere Möglichkeit zur Kontrolle von Antwortverzerrungen im Interview besteht in der direkten Messung einer Tendenz zu sozial erwünschtem Antwortverhalten oder einer Zustimmungstendenz bei den Befragten. Als Beispiel sei die Kurzskala zur Messung sozialer Erwünschtheit (nach Crowne und Marlowe) erläutert (vgl. das ZUMA-Skalenhandbuch 1983, P 03). Folgende Items mit den Antwortmöglichkeiten "trifft zu" und "trifft nicht zu" wurden den Befragten vorgelegt:

Items	trifft zu	trifft nicht zu
1. Ich sage immer, was ich denke.	+	
2. Ich bin manchmal ärgerlich, wenn ich meinen Willen nicht bekomme.		+
3. Ich bin immer gewillt, einen Fehler, den ich mache, auch zuzugeben.	+	
4. Ich habe gelegentlich mit Absicht etwas gesagt, was die Gefühle des Anderen verletzen könnte.		+

Die Plus-Zeichen zeigen die Antwortvorgaben an, die soziale Erwünschtheit induzieren. Wie man sieht, sind 2 Items negativ gepolt, um der Ja-Sage-Tendenz vorzubeugen. Wird in einer Untersuchung die Tendenz zur sozialen Erwünschtheit explizit gemessen, kann die entsprechende Variable zur Konstruktion und empirischen Überprüfung von Meßfehlertheorien im Rahmen des kausalanalytischen Ansatzes herangezogen werden. So findet Reinecke (1991) in einer Untersuchung eine systematische Untertreibung der Ausländerfeindlichkeit bei Befragten mit einer starken Tendenz zu sozialer Erwünschtheit. Dieses Ergebnis kann als Meßfehlertheorie in ein Kausalmodell zur Erklärung der Ausländerfeindlichkeit integriert werden.

#### Exkurs: Können Befragte lügen? Antwortverzerrungen als rationales Handeln

Allerdings macht die Messung einer Tendenz zur sozialen Erwünschtheit oder einer Zustimmungstendenz nur Sinn, wenn diese Tendenzen als stabile Persönlichkeitsmerkmale angesehen werden können. Gerade dies wird aber in einer Instrumententheorie der Befragung bezweifelt, die die situative Angemessenheit von verschiedenen Antwortstilen aus der Sicht des Befragten auf der Grundlage der Theorie rationalen Handelns erklärt (vgl. Esser 1986). Esser spielt in seinem Aufsatz mit dem provokativen Titel "Können Befragte lügen?" auf die These von Manning an: "The respondent never lies!". Manning argumentiert im Rahmen des interpretativ-interaktionistischen Paradigmas. Danach ist jede Antwort des Befragten, also auch eine im herkömmlichen Sinne ungültige Antwort, wie sie durch die verschiedenen Typen von Antwortverzerrungen repräsentiert wird, als eine aus dessen Sicht sinnvolle und daher auch aus der Interaktionssituation der Befragung heraus zu verstehende Reaktion zu interpretieren. Esser ergänzt diesen Gedanken, indem er die Reaktion des Befragten als Resultat des Abwägens der mit den unterschiedlichen Antwortstrategien verbundenen Kosten und Nutzen erklärt. (Sinnadäquanz und Kausaladäquanz einer Erklärung sozialen Handelns im Sinne Weber; siehe auch 2.2). Danach sind systematische Verzerrungen nicht als Resultat einer stabilen Verhaltensdisposition zu erklären, sondern als rationale Reaktion des Befragten auf situative Bedingungen. Dabei orientieren sich die Befragten je nach wahrgenommener Situation an bestimmten Handlungsroutinen (beschränkte Rationalität im Sinne von Simon). Zufällige Antwortverzerrungen sind danach insbesondere dann zu erwarten, wenn es weder für eine systematische inhaltliche (auf die Frage bezogene), noch für eine systematisch situationsbezogene Reaktion eine Grundlage gibt. Überwiegt in der Befragungssituation eine der beiden angegebenen Tendenzen, kommt es entweder zu validen Antworten (überwiegen der inhaltlichen Tendenz) oder zu systematischen Verzerrungen (überwiegen der situationsbezogenen Tendenz). Der erste Fall liegt z.B. vor, wenn der Befragte stark an dem Thema der Befragung interessiert ist und dazu eine klare Meinung hat, die von ihm mit einer verständlichen Frageformulierung abgefragt wird. Der zweite Fall wäre z.B. bei einem hohen Bedürfnis nach sozialer Anerkennnung in der Interviewsituation gegeben (Interviewereffekt). Welche der beiden Tendenzen überwiegt, kann nur durch eine konkrete Analyse der Kostenund Nutzenkomponenten der Situation aus der Sicht des Befragten ermittelt werden. Die hier kurz skizzierte Instrumententheorie der Befragung von Esser ist besser als der persönlichkeitsbezogene Ansatz in der Lage, daß breit gefächerte Spektrum aller möglichen Antwortverzerrungen aus einem einheitlichen theoretischen Modell heraus zu verstehen.

### 6.1.3 Andere Formen der Befragung

Gegenüber dem Interview bietet die schriftliche Befragung zwei Vorteile: Einmal ist sie im allgemeinen billiger und organisatorisch einfacher durchzuführen. Zum anderen ist die Zusicherung der Anonymität glaubwürdiger. Interviewereinflüsse fallen ganz weg und die Reaktivität des Verfahrens ist generell geringer. Der entscheidende Nachteil ist aber, daß keine Kontrolle der Befragungssituation möglich ist. So ist über die Konzentration, mit der die Fragen beantwortet werden, ebenso wenig eine Aussage möglich wie über die Ernsthaftigkeit der Motivation des Befragten, über das Eingreifen Dritter und schließlich sogar darüber, ob der Befragte den Fragebogen tatsächlich selbst ausgefüllt hat. Weitere Nachteile der schriftlichen Befragung liegen in der hohen Ausfallquote und der Notwendigkeit, Fragen besonders einfach formulieren zu müssen, da Erläuterungen durch den Interviewer nicht möglich sind. Allerdings ist auch bei der schriftlichen Befragung durch die Total-Design-Methode (vgl. Hippler und Seidel 1985) eine deutliche Verbesserung der Rücklaufquoten möglich.

Das <u>Telefoninterview</u> stellt eine Erhebungstechnik dar, die zunehmend öfter eingesetzt wird und z.T. schon das Interview von Angesicht zu Angesicht (face to face) [mündliche Befragung] ersetzt hat (vgl. Frey u.a. 1990). So werden schon heute Befragungen in der Marktforschung überwiegend als Telefonbefragungen durchgeführt. Telefoninterviews sind relativ billig, können schnell durchgeführt werden und erreichen den Befragten direkt in seinem Wohnzimmer. Gerade dieser letzte Punkt wird zunehmend wichtig, da bei mündlichen Befragungen die Anzahl der Antwortverweigerungen deutlich zugenommen hat. Dies liegt daran, daß zum einen viele Befragte nur schwer erreichbar sind und zum anderen viele, insbesondere ältere Befragte, den Kontakt an der Haustür scheuen. Die Stichprobenziehung ist für Telefoninterviews relativ problemlos durch Rückgriff auf das jährlich aktualisierte

Telefonbuch möglich. Da inzwischen fast jeder Haushalt in der Bundesrepublik (zumindest in den alten, aber zunehmend auch in den neuen Bundesländern) über einen eigenen Telefonanschluß verfügt, kommt das Telefonbuch dem statistischen Ideal einer Liste, die alle Mitglieder der Grundgesamtheit enthält, zunehmend näher. (Innerhalb der Haushalte ist natürlich noch eine zusätzliche Selektionsstufe erforderlich (vgl. Kap. 7.).)

Generell gelten für die Fragebogenkonstruktion die gleichen Grundsätze, wie sie für die mündliche Befragung formuliert werden. Durch Reduzierung auf die verbal-akkustische Kommunikation entfallen aber alle visuellen Hilfen wie Karten mit Antwortvorgaben oder Skalendiagramme zur Unterstützung der Fragebeantwortung. Die Kehrseite der Einschränkung der Kommunikationsmöglichkeiten ist natürlich die im Vergleich zur mündlichen Befragung verringerte Reaktivität der Befragungssituation im Falle des Telefoninterviews. Besondere Bedeutung bei einem Telefoninterview kommt der Einleitung zu, die in besonderer Weise geeignet sein muß, Vertrauen zu schaffen und eine hohe Teilnahmemotivation zu erzeugen. Generell sollten gerade in Telefoninterviews kurze und relativ einfache Fragen verwendet werden. Lange Listen mit Antwortvorgaben sind untauglich, da dadurch ein starker Recency-Effekt provoziert würde. Nützlich ist die Verwendung einer stichwortartigen Zusammenfassung der Antwortalternativen am Ende der Frage. Kann auf eine differenzierte Liste von Antwortmöglichkeiten nicht verzichtet werden, empfiehlt sich die Zweistufentechnik, d.h. das Aufspalten der Frage in eine Haupt- und eine differenzierende Zusatzfrage.

Entgegen dem oft geäußerten Vorurteil, daß das Telefoninterview nur für die Abfrage einfacher Sachverhalte in einem kurzen Gespräch geeignet sei, liegen inzwischen Beispiele dafür vor, daß auch komplexere Befragungen mit einem oft relativ langen Fragebogen (Interviewdauer über eine Stunde) erfolgreich als Telefoninterview durchgeführt werden können (vgl. Hormuth und Brückner 1985).

# 6.2 Beobachtung

Die Beobachtung als Erhebungsverfahren zielt auf die Erfassung von konkretem Verhalten von Menschen in einfachen oder komplexen sozialen Situationen. Jedes soziale Handeln ist sowohl durch die Situation und deren subjektive Deutung durch die Handelnden wie auch durch die Intentionen des Handelnden bestimmt. Beobachtung richtet sich also immer auf Verhalten, dem ein subjektiver Sinn und eine objektive soziale Bedeutung zukommt. Deshalb gehört zum Beobachten notwendigerweise die zutreffende Interpretation des subjektiven Sinns und der sozialen Bedeutung einer bestimmten Handlung oder Verhaltenssequenz.

Beobachtungsverfahren lassen sich nach folgenden Gesichtspunkten unterscheiden:

- systematische bzw. unsystematische Beobachtung: relativ unsystematische Beobachtungsverfahren arbeiten mit gering strukturierten Beobachtungskategorien. Dagegen verfügen systematische Beobachtungsverfahren über ein detailliertes und exakt definiertes System von Beobachtungskategorien.
- Beobachtungen in "natürlichen" sozialen Situationen oder in "künstlichen", experimentell erzeugten Situationen: Ob die zu untersuchenden Verhaltensweisen in gleichsam "natürlichen" Situationen oder in hochstandardisierten, experimentell erzeugten Situationen beobachtet werden sollen, hängt in erster Linie von der Forschungsfrage, aber natürlich auch von den technisch-praktischen Möglichkeiten ab, Situationen kontrolliert zu erzeugen, in denen die interessierenden Verhaltensweisen beobachtet werden können.
- Teilnehmende bzw. nichtteilnehmende Beobachtung: Der teilnehmende Beobachter ist selbst Element des sozialen Feldes, in dem er seine Beobachtungen durchführt, während der nichtteilnehmende Beobachter gleichsam von außen her das ihn interessierende Verhalten beobachtet.
- Offene bzw. verdeckte Beobachtung: Haben die Beobachtungspersonen keine Kenntnis von dem Beobachtungsvorgang, so spricht man von einer verdeckten Beobachtung. Im Gegensatz zur offenen Beobachtung stellt sich damit das Problem der Reaktivität des Verfahrens nicht.

Jedes Beobachtungsverfahren muß, damit es überhaupt als wissenschaftlich gelten kann, über ein Minimum an Systematik und Strukturiertheit verfügen. Je expliziter die Theorie über die zu beobachtetenden Vorgänge ist und je besser die Begriffe definiert und in Form von anwendbaren Beobachtungskategorien operationalisiert werden, desto leichter läßt sich ein systematisches Beobachtungsschema aufstellen. In einem solchen Kategoriensystem wird jede auftretende Handlung nach einem festgelegten Schema klassifiziert. Die Entwicklung eines Kategoriensystems, das es ermöglicht, ablaufende soziale Prozesse zu registrieren und in ihrem zeitlichen Ablauf zu protokollieren, ist die Hauptschwierigkeit bei der Konstruktion eines Beobachtungsverfahrens. Zunächst einmal muß eine Liste der für die Untersuchung bedeutsam erscheinenden und theoretisch relevanten Verhaltenseinheiten aufgestellt werden. Diese Verhaltenseinheiten oder Kategorien stellen dann die Variablen Beobachtungssystems dar. Auf jeden Fall ist eine Reihe von formalen und inhaltichen Anforderungen an das Kategoriensystem zu stellen, insbesondere:

- Eindimensionalität der zu messenden Variablen
- Ausschließlichkeit der Kategorien
- Vollständigkeit der Kategorien
- Konkretion der Kategorien (Spezifizierung der zugeordneten beobachtbaren Sachverhalte)

Beobachtungsverfahren sind auf den Beobachter als Meßinstrument zur Entdeckung (Registrierung von Verhaltensweisen) und Protokollierung (Vercodung von Verhaltensweisen gemäß einem Kategorienschema) angewiesen. Von dem Beobachter wird also auf jeden Fall eine Wahrnehmungs-, Selektions- und Reduktionsleistung gefordert. Dabei können eine Reihe von Fehlern auftreten, die die Qualität der Beobachtung beeinflussen. Dazu gehören in erster Linie Beobachterfehler, die bei Aufnahme, Beurteilung und Protokollierung einer bestimmten Verhaltensweise gemäß dem zugrundeliegenden Kategorienschema auftreten können. Bei der offenen Beobachtung muß darüberhinaus auch mit Reaktivitätseffekten gerechnet werden, wenn die Beobachteten unter dem Eindruck der Beobachtungssituation ihr Verhalten verändern.

## 6.3 Inhaltsanalyse

Die Inhaltsanalyse knüpft an das vorwissenschaftliche und alltäglich praktizierte Sprachverstehen an, das es dem Menschen ermöglicht, nicht nur Bedeutungen zu verstehen, die mit Hilfe der Sprache bewußt kommuniziert werden sollen, sondern auch die zugrundeliegenden Absichten, Einstellungen, Situationsdeutungen und stillschweigenden Annahmen, die dem Weltbild des Autors (Textproduzenten) zugrunde liegen. Dieser intuitive Vorgang des Sprachverstehens muß jedoch zum Zweck einer wissenschaftlichen Analyse explizit gemacht, systematisiert und objektiviert werden. Man kann die Inhaltsanalyse daher als eine Forschungstechnik definieren, die sprachliche Eigenschaften eines Textes objektiv und systematisch identifiziert und beschreibt, um daraus Schlußfolgerungen auf nichtsprachliche Eigenschaften von Personen oder gesellschaftlichen Gruppen zu ziehen. Der Text kann dabei in gesprochener (z.B. als Tonbandprotokoll) oder in geschriebener (z.B. als Parteiprogramm) Form vorliegen. In einem weiteren Sinne kann die Inhaltsanalyse auch auf nicht rein sprachliches Material, wie z.B. Filme oder Werbeplakate, angewandt werden.

Das in der Definition genannte Kriterium der Systematik verlangt, daß alle für die Beantwortung einer bestimmten Forschungsfrage relevanten Aspekte des zu analysierenden Textmaterials berücksichtigt werden. Anderenfalls bestünde die Gefahr, daß willkürlich nur die die jeweilige Hypothese bestätigenden Textmerkmale registriert werden. Das systematische Vorgehen ist somit eine Voraussetzung für die Gültigkeit der Ergebnisse. Dabei ist es jedoch nicht unbedingt erforderlich, daß ein solches systematisches Vorgehen zugleich ein quantifizierendes ist. Die qualitative Inhaltsanalyse eignet sich sehr wohl zu explorativen und zu deskriptiven Zwecken. Wie bei anderen Forschungstechniken sind aber auch hier die Ergebnisse einer quantitativen Analyse nicht nur präziser und in der Regel zuverlässiger, sondern auch eine Voraussetzung für die exakte Überprüfung von Hypothesen mittels inhaltsanalytischer Verfahren.

Im Kern besteht die Inhaltsanalyse darin, den Inhalt oder die Bedeutung bestimmter sprachlicher Konfigurationen zu erkennen und entsprechend zu klassifizieren. Dabei stützt sich der Inhaltsanalytiker zunächst einmal auf sein eigenes intuitives Sprachverstehen. Dies ist allerdings nur dann zulässig, wenn Produzent und Analytiker der gleichen Sprachgemeinschaft unterstellt werden daß entstammen, wenn also kann, Inhaltsanalytiker die Bedeutung einer Kommunikation so versteht, wie sie ihr Produzent intendiert hat. In dem Maße, in dem die unterstellte Sprachgemeinschaft nicht vorhanden ist, wird die Gültigkeit der Ergebnisse einer Inhaltsanalyse fragwürdig. Das ist besonders einleuchtend in jenen Fällen, in denen das Textmaterial aus einem fremden Sprachsystem oder auch nur aus dem Sprachsystem einer für den Analytiker fremden Subkultur stammt. Was in solchen Fällen unvermeidlich ist, sollte im Idealfall immer durchgeführt werden, nämlich eine empirisch-semantische Analyse, um zunächst einmal die Bedeutung der sprachlichen Ausdrücke zu ermitteln. Dazu muß natürlich auch der situative Kontext, in dem der Text produziert wurde, berücksichtigt werden.

Das eigentliche Problem besteht darin, aus dem festgestellten Inhalt eines Textes, d.h. aus der Art und Häufigkeit von Sprachelementen bestimmter Bedeutung, auf die zugrundeliegenden latenten Variablen (Absichten, Einstellungen, Werthaltungen, Weltbilder, usw.) zu schließen. In bestimmten Fällen ist es zwar möglich, solche Deutungen durch andere Erhebungsverfahren (z.B. eine Befragung) zu validieren (Verfahren der Triangulation). Das eigentliche Anwendungsfeld der Inhaltsanalyse liegt aber gerade in solchen Bereichen, in denen eine solche direkte Validierung der inhaltsanalytischen Interpretation nicht möglich ist, wie z.B. bei historischem Texten.

#### Vorgehensweise bei der Inhaltsanalyse:

• Forschungsleitende Hypothese: Wie in jedem anderen Forschungsvorhaben in der empirischen Sozialforschung bildet eine theoretisch gut begründete forschungsleitende Hypothese die Grundlage der Untersuchung. Dabei zeichnet sich die Inhaltsanalyse durch eine besonders breite Vielfalt möglicher Forschungsthematiken aus. Besonders hervorzuheben ist die historische Tiefe des Ansatzes, da sich die Inhaltsanalyse auf Materialien anwenden läßt, die unter Umständen bereits vor sehr langer Zeit produziert wurden (sog. prozeßproduzierte Daten). Die Inhaltsanalyse ist daher geradezu dazu prädestiniert, historische Wandlungsprozesse in ihren ökonomischen, politischen und sozialen Facetten zu untersuchen. So überprüft z.B. Weber (1981) seine Theorie politischer Konjunkturzyklen (die eng mit der Kondratieff-These ökonomischer Konjunkturzyklen verbunden ist), indem er Material aus den letzten beiden Jahrhunderten inhaltsanalytisch auswertet. Mit einer ähnlichen Methodik versucht Mohler (1984), den politischen Wertewandel in Deutschland in der Periode von 1917 bis 1971 zu erfassen. In ähnlicher Weise könnte ein Vergleich von Werbematerialien aus den 50er und den 90er Jahren dazu benutzt werden, um einen Wandel in der gesellschaftlichen Stellung der Frauen zu

dokumentieren. An diesen Beispielen wird deutlich, daß sich mit der Inhaltsanalyse Informationen über Einstellungen, Werthaltungen usw. aus lange vergangenen Zeiten analysieren lassen, die für andere Verfahren der Datenerhebung nicht mehr zugänglich sind.

- Bestimmung des relevanten <u>Textmaterials</u>: Anhand der forschungsleitenden Hypothese muß konkretisiert werden, aus welcher Grundgesamtheit die relevanten Texte auszuwählen sind. So wählt Weber die Regierungserklärungen der britischen Regierungen (Thronreden) der Jahre 1795 bis 1972 als Datenbasis, während Mohler den politischen Wertewandel in Deutschland anhand von Abituraufsätzen aus den Jahren 1917 bis 1971 untersucht. In allen Fällen wird dabei aus der Grundgesamtheit aller relevanten Texte eine entsprechende Stichprobe für die eigentliche Inhaltsanalyse gezogen.
- Bestimmung der sprachlichen Einheiten der Analyse und Entwicklung eines inhaltsanalytischen Kategorienschemas: Für die Operationalisierung der in der Hypothese auftretenden Variablen ist zunächst einmal die Art der sprachlichen Einheiten festzulegen, auf die hin die Texte untersucht werden sollen. Dabei kann es sich um Worte, Satzteile, ganze Sätze oder sogar ganze Absätze der vorliegenden Texte handeln. Entscheidend ist dann die genaue inhaltliche Spezifikation der Kategorien, unter die die sprachlichen Einheiten des Textmaterials subsummiert werden sollen. Diese Kategorien bilden das eigentliche Verbindungsstück zwischen den Variablen der Hypothese und den sprachlichen Einheiten des zu analysierenden Textes. Sie stellen gewissermaßen Indikatoren für die einzelnen Merkmalsausprägungen der zu untersuchenden Variablen dar. Das inhaltsanalytische Kategorienschema ist also immer selektiv in Hinsicht auf die forschungsleitende Fragestellung, d.h. es kann sich um keine vollständige Erfassung des vorliegenden Textmaterials hinsichtlich aller darin vorhandenen Inhalte handeln.

Kategorien müssen aufgrund einer gründlichen theoretischen Aufarbeitung des Forschungsproblems festgelegt und definiert werden. Dabei handelt es sich bei den Kategorien im inhaltsanalytischen Sinne zunächst um Oberbegriffe, die mit den problemrelevanten Dimensionen der Forschungsthematik identisch sind, bzw. diese in Teildimensionen untergliedern. Zu jeder Kategorie müssen darüberhinaus Unterkategorien gebildet werden, die angeben, welche Arten von Aussagen je Kategorie unterschieden werden sollen; Unterkategorien entsprechen den Merkmalsausprägungen von Variablen. Wie üblich muß ein solches Kategorienschema einige formale Anforderungen erfüllen; Jede Kategorie darf sich nur auf eine Bedeutungsdimension beziehen, Kategorien müssen einander ausschließen, erschöpfend und voneinander unabhängig sein. In einer einfachen quantitativen Inhaltsanalyse wird das Vorkommen der Kategorien (jeweils indiziert durch das Vorkommen der entsprechenden Unterkategorien) einfach ausgezählt. Dies geschieht in der computergestützen Inhaltsanalyse weitgehend auf der Grundlage inhaltsanalytischer Wörterbücher. So verwendet z.B. Weber in seiner Analyse der britischen Thronreden das im

Lasswell-Value-Dictionary zusammengefaßte Kategorienschema zur Erfassung von Wertsystemen (vgl. auch Weber 1990).

Neben den Auszählungen auf der Grundlage von Wörterbüchern in der computergestützten quantitativen Inhaltsanalyse, die meistens auf eine kontextfreie Einzelwortanalyse hinauslaufen, sind auch stärker qualitativ und interpretativ orientierte Kategorienschemata denkbar. So kategorisiert z.B. Mohler die Abituraufsätze als sprachliche Einheit danach, wie in ihnen das Verhältnis von Individuum und kollektiver Macht dargestellt wird. Das Resultat eines solchen interpretativen Verfahrens auf der Grundlage eines systematisch entwickelten Kategorienschemas ist z.B. die mögliche Vercodung eines Abituraufsatzes durch folgende synthetische Sätze:

Typ 1 (autoritärer Typ): [Autoritäre Macht] ist positiv) und [irrational handelndes Individuum] ist negativ) und ([Macht] dominiert [Individuum]) ist positiv).

Typ 2 (demokratischer Typ): [Kontrollierte Macht] ist positiv) und [rational handelndes Individuum] ist positiv) und (ausgeglichenes Verhältnis zwischen [Macht] und [Individuum]) ist positiv).

In beiden Vercodungen erkennt man einmal die zugrundeliegenden Subjekte, nämlich <u>Macht</u> (autoritär bzw. kontrolliert) und <u>Individuum</u> (irrational bzw. rational handelndes), die verwendeten Relationen zwichen den Subjekten, nämlich Macht <u>dominiert</u> das Verhältnis bzw. <u>ausgeglichenes</u> Verhältnis zwischen Macht und Individuum und schließlich die Bewertungskategorien <u>positiv</u> und <u>negativ</u>. Die Vercodung eines Abituraufsatzes nach dem angegebenen Kategorienschema stellt hohe Anforderungen an die interpretativen Fähigkeiten des Codierers.

Aus methodischer Sicht äußerst problematisch sind hingegen Verfahren der qualitativen Inhaltsanalyse, die ohne ein explizit definiertes Kategoriensystem auskommen wollen. Selbst wenn solche Analysen lediglich explorativen Charakter haben, macht sich der Mangel eines systematisch ausgearbeiteten Referenzsystems störend bemerkbar. Fundierte Aussagen über die Reliabilität und die Validität einer solchen Vorgehensweise sind gewöhnlich nicht zu machen. Als Beispiel diskutieren wir kurz die von Meuser und Nagel (1991) vorgeschlagene Vorgehensweise bei der Inhaltsanalyse von Experteninterviews. Dabei wird davon ausgegangen, daß diese Interviews in transskribierter Form vorliegen. Der erste Schritt ist eine Paraphrasierung der Texte, wobei der Chronologie des Gesprächsverlaufs folgend textgetreu, aber in den Worten des Inhaltsanalytikers zusammengefaßt wird, was der Experte an Meinungen, Urteilen und Beobachtungen geäußert hat. Nach Meuser und Nagel ist die Paraphrasierung eines Textes ein entscheidender Schritt der Inhaltsanalyse, wobei besonders darauf zu achten ist, daß Inhalte nicht durch voreiliges Klassifizieren verzerrt und Informationen nicht durch voreiliges Themenraffen verschenkt werden. "Die Paraphrasierung ist der erste Schritt des Verdichtens des Textmaterials - schon nach wenigen Interviews

entwickelt sich geradezu von selbst ein Muster der Paraphrasierung heraus, das bei den folgenden Texten ohne größere Modifikation zur Anwendung gelangt." (S. 457) Methodisch ist zu diesem Zitat anzumerken, daß ein sich im Prozeß des Paraphrasierens von selbst (?) (wie auch immer) herausschälendes Kategoriensystem zur Paraphrasierung explizit gemacht werden sollte, damit es in einer späteren Phase der Untersuchung kritisch hinterfragt werden kann. Die ganze Problematik dieser unkontrolliert qualitativen Vorgehensweise kommt in der folgenden Aussage von Meuser und Nagel zum Ausdruck: "Die Gültigkeit einer Paraphrase beruht darauf, daß das Expertenwissen so ausgebreitet wird, daß jede, die Interview und Paraphrase miteinander vergleicht, zu dem Schluß gelangt, daß nichts unterschlagen, nichts hinzugefügt und nichts verzerrt wiedergegeben wurde." Das ganze Dilemma des Ansatzes ist, daß er einerseits "nichts unterschlagen, nichts hinzufügen, und nichts verzerrt wiedergeben" will, aber andererseits auf eine Verdichtung und Interpretation des Materials aus ist. Die kritische Beurteilung der Validität der Vorgehensweise bleibt letztlich auf den Vergleich mit dem vollständigen Textmaterial angewiesen.

Der nächste Schritt in der Verdichtung des Materials besteht darin, daß die paraphrasierten Passagen mit <u>Überschriften</u> versehen werden. Anschließend werden Passagen, in denen gleiche oder ähnliche Themen behandelt werden, zusammengestellt. Eine <u>Hauptüberschrift</u>, die den Inhalt sämtlicher subsummierter Passagen abdeckt, wird formuliert. Gegenstand der Auswertung ist in dieser Phase allerdings immer noch das einzelne Interview. Die Verdichtungen, Typisierungen und Abstraktionen, die hier vorgenommen werden, verbleiben in dessen Horizont.

Danach erfolgt der <u>thematische Vergleich über die verschiedenen Interviews hinweg</u>. Dabei werden Überschriften gebildet, die jetzt aber thematisch vergleichbare Textpassagen aus verschiedenen Interviews zusammenfassen. Die Kategorien, die auf diese Weise gebildet werden, sollten kontinuierlich auf ihre Triftigkeit, Vollständigkeit und Gültigkeit hin überprüft werden. Wie dies geschehen soll, wird allerdings nicht gesagt. Auch hier ist die Vorgehensweise wenig systematisch und von dem interpretativen Geschick des Auswerters abhängig.

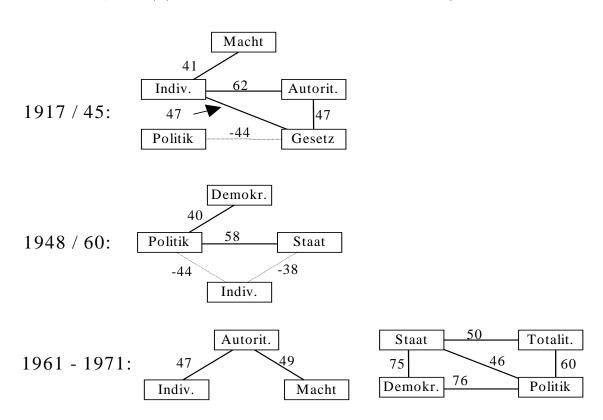
Handelte es sich bisher um eine Verdichtung der Texte, d.h. die gebildeten Kategorien waren an den begrifflichen und metaphorischen Rahmen der von den Experten verwendeten Sprache angelehnt, so folgt im nächsten Schritt der <u>soziologischen Konzeptualisierung</u> eine Abwendung von der Terminologie der Interviewten. Die zuvor dem Text entnommenen Begriffe und Überschriften werden nun in soziologische Termini übersetzt, um einen Anschluß der Interpretation an die allgemeine wissenschaftliche Diskussion zu ermöglichen. Das verdichtete Expertenwissen wird quasi im Kontext soziologischer Theorien erneut interpretiert und gleichzeitig theoretisch generalisiert.

#### Inhalsanalytische Auswertungstechniken:

Die einfachste Form der Auswertung ist die sog. <u>Frequenzanalyse</u>, in der Textelemente lediglich danach klassifiziert werden, wie häufig die einzelen Kategorien des Kategorienschemas in ihnen vorkommen. Wie die Analyse von Weber zeigt, können sich allerdings bereits auf dieser Stufe komplexe Auswertungsprobleme stellen, insbesondere dann, wenn es sich um die Analyse von Zeitreihen handelt. So findet Weber in der Analyse der britischen Thronreden einen zweiundfünfzigjährigen Zyklus, in dem sich bestimmte Thematiken in der politischen Diskussion wiederholen - dies eng gekoppelt mit den entsprechenden ökonomischen Konjunkturzyklen.

Wird nicht nur die Häufigkeit des Auftretens von Kategorien analysiert, sondern gleichzeitig auch die damit verbundene Bewertung (negativ, neutral, positiv) dieser Kategorien, so spricht man von <u>Valenzanalysen</u>. Die Valenzanalyse geht also über eine reine Einzelwortanalyse hinaus, indem sie den Kontext, in dem ein Begriff verwendet wird, mit berücksichtigt.

Eine Erweiterung dieser Vorgehensweise stellt die Analyse <u>syntaktischer Sequenzen</u> von Kategorien dar. Dabei wird in sog. <u>Kontingenzanalysen</u> das Auftreten bestimmter sprachlicher Elemente im Zusammenhang mit anderen Begriffen untersucht. Als Ergebnis einer solchen Kontingenzanalyse zeigen wir die Verbindungen zwischen politischen Kategorien, wie sie sich in den von Mohler analysierten Abituraufsätzen darstellen (vgl. Mohler 1984, S. 294) (als Assoziationsmaß wurde tau b verwendet):



Das zentrale Problem der Inhaltsanalyse ist die Beurteilung der Objektivität, Reliabilität und Validität des Verfahrens. Wie insbesondere im Zusammenhang mit der Vorgehensweise bei

der Auswertung von Experteninterviews deutlich wurde, ist in der explorativ-qualitativen Inhaltsanalyse noch nicht einmal die Objektivität des Verfahrens eindeutig zu beurteilen. Im allgemeinen muß man daher davon ausgehen, daß ein anderer Sozialforscher das gleiche Datenmaterial anders interpretieren würde und zu anderen theoretischen Schlußfolgerungen gelangen würde. Günstiger ist die Situation im Falle einer systematisch-qualitativen Inhaltsanalyse mit einem explizit vorgegebenen Kategorienschema (vgl. z.B. die Vorgehensweise von Mohler). In diesem Fall ist es unabdingbar, die Codierung der gleichen Texte durch verschiedene Codierer vornehmen zu lassen und anschließend eine Intercoderreliabilität zu berechnen. Denn nur bei einer hohen Intercoderreliabilität kann man davon ausgehen, daß die Kategorien des zugrundegelegten Kategorienschemas so eindeutig und trennscharf definiert waren, daß sich für die Coder keine allzu großen Spielräume bei der Interpretation und anschließenden Codierung der Texte ergaben. problematischsten ist die Beurteilung der Validität einer Inhaltsanalyse. Bei der quantitativen Vorgehensweise stellt sich die Frage, ob das entwickelte Kategorienschema valide ist, d.h. ob die in den entsprechenden Wörterbüchern enthaltenen Worte valide Indikatoren für die zugrundeliegenden theoretischen Kategorien darstellen. Darüberhinaus stellt sich natürlich die Frage, ob die verwendeten theoretischen Kategorien in der Lage sind, die relevanten Inhalte der Texte zu erfassen. Wenn möglich, sollte man daher die eher mechanische Vorgehensweise der kategorialen Auszählung in der quantitativen Inhaltsanalyse durch eine stärker systematisch-qualitative Vorgehensweise kreuzvalidieren (wie dies z.B. bei Mohler mit den beiden angegebenen inhaltsanalytischen Strategien geschieht).

Aus dem Gesagten dürfte klar geworden sein, daß die Inhaltsanalyse zur Zeit noch große methodische Probleme aufweist. Dennoch gehen wir davon aus, daß die <u>Bedeutung</u> inhaltsanalytischer Vorgehensweisen in der empirischen Sozialforschung <u>in Zukunft stark</u> <u>zunehmen</u> wird und dies aus zwei Gründen:

- Erstens existiert eine große Fülle sogenannter prozeßproduzierter Daten (z.B. Werbeanzeigen, Zeitungsartikel, Fernsehfilme, historisches Material, usw.), die in Zukunft mit einfachen technischen Mitteln datentechnisch erfaßt werden können (z.B. durch das Scannen von Texten). Auf dieser Grundlage sind dann sozialwissenschaftliche Analysen historischen und aktuellen gesellschaftlichen Wandels möglich, die mit Umfragedaten in dieser Form nicht geleistet werden können.
- Zweitens wird es eine Weiterentwicklung der computergestützten Inhaltsanalyse möglich machen, über einfache kategoriale Häufigkeitsauszählungen hinaus auch stärker kontextab-hängige Analysen durchzuführen. Es sind sogar Entwicklungen im Bereich der künstlichen Intelligenz denkbar, die eine stärker interpretativ orientierte computergestützte Auswertung von Texten möglich erscheinen lässt. Dies würde die Möglichkeiten einer standardisierten, methodisch kontrollierten Inhaltsanalyse enorm erhöhen.

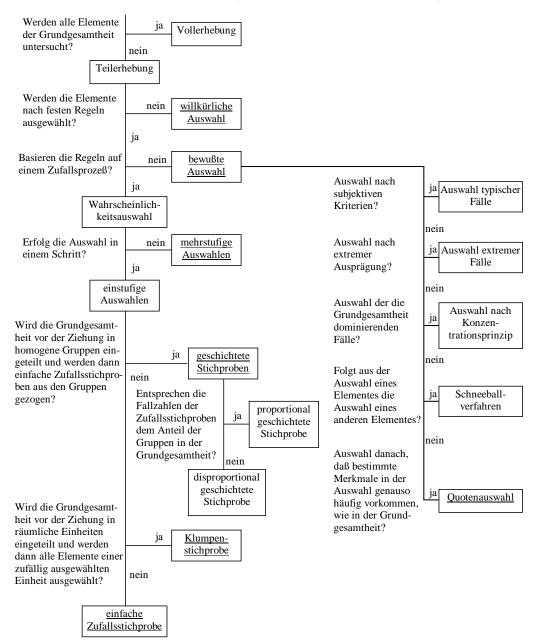
## 7. Auswahlverfahren

## 7.1 Grundgesamtheit und Auswahl

Im Zusammenhang mit der Hypothesenbildung muß auch die Grundgesamtheit (GG), auf die sich die Hypothese bezieht, definiert werden. Im allgemeinen ist eine Vollerhebung aus praktischen Gründen (hoher Zeit- und Kostenaufwand) nicht durchführbar. Es müssen also einige Einheiten der GG für eine Teilerhebung ausgewählt werden. Erfolgt diese Auswahl nach bestimmten Regeln, so spricht man von einer bewußten Auswahl - sonst von einer willkürlichen Auswahl (Auswahl aufs Geratewohl). Willkürliche Auswahlen können in der explorativen Phase einer Untersuchung wertvolle Dienste leisten, für die Überprüfung einer Hypothese sind sie aber wenig aussagekräftig. Bewußte Auswahlen werden im folgenden von Wahrscheinlichkeitsauswahlen unterschieden, bei denen die Auswahlregel auf einem Zufallsprozeß basiert. Bewußte Auswahlen (Auswahlen nach Gutdünken) können sein: Auswahlen extremer Fälle (z.B. Eliteuntersuchungen), Schneeballauswahlen (z.B. Erhebungen egozentrierter Netzwerke) und insbesondere Quotenerhebungen. Bei einer Quotenerhebung wird die Auswahl von Personen durch vorgegebene Quoten in Hinblick auf einige wenige Kontrollmerkmale (wie z.B. Geschlecht, Alter, Bildung, usw.) gesteuert, deren Verteilung in der GG bekannt ist. Die Quotenvorgaben für die einzelnen Interviewer werden so berechnet, daß die Verteilung der Kontrollmerkmale in der Stichprobe gerade der in der GG entspricht. Innerhalb dieser Quotenvorgaben ist der Interviewer frei bei der Auswahl der Befragten. Gerade hierin liegt der Unterschied zur geschichteten Zufallsstichprobe, bei der innerhalb der Quotenvorgaben (Schichten) nach einem Zufallsverfahren ausgewählt wird (siehe 7.3). Quotenauswahlen sind zwar kostengünstig, haben aber einige schwerwiegende Nachteile:

- Quotenauswahlen sind keine Zufallsauswahlen; die Anwendung von statistischen Inferenztechniken (siehe 7.2) ist daher nicht möglich.
- Innerhalb der Quotenvorgaben besteht die Gefahr von erheblichen Verzerrungen durch willkürliche Auswahl der Befragten durch den Interviewer (kooperative Personen, Personen aus dem weiteren Bekanntenkreis des Interviewers, gut erreichbare Personen). Die Ausfallproblematik wird daher nur verdeckt.
- Die Auswahl von Quotierungsmerkmalen wird theoretisch um so problematischer, je indirekter die Beziehungen zwischen den soziodemographischen Hintergrundmerkmalen und den Einstellungen der Befragten ist (Individualisierungsthese).
- Im Grenzfall benötigt man so viele Kontrollmerkmale, daß die Vorteile der Quotenstichprobe wieder verloren gehen und eine einfache Zufallsstichprobe genau so kostengünstig zu realisieren wäre.

### Übersicht über Auswahlverfahren (nach Schnell/Hill/Esser, S. 253)



# 7.2 Zufallsstichproben und Interferenzprobleme

Zufallsstichproben stellen die einzige Gewähr dafür dar, daß aus den Ergebnissen einer Stichprobenerhebung (innerhalb bestimmter statistischer Fehlergrenzen) auf das Vorliegen entsprechender Ergebnisse in der Grundgesamtheit geschlossen werden kann. Ein solcher Repräsentationsschluß ist also nur im Falle von Zufallsstichproben erlaubt, bei denen ein Zufallsprozeß über die Aufnahme eines Elements der GG in die Stichprobe entscheidet. Jedes Element der GG muß demnach eine angebbare (und von Null verschiedene) Auswahlwahrscheinlichkeit haben. Wird aus einer GG eine Stichprobe von n Elementen so daß alle möglichen Stichproben mit n die gezogen, Elementen gleiche Auswahlwahrscheinlichkeit haben, spricht man von einer einfachen Zufallsstichprobe.

Einfache Zufallsstichproben können z.B. durch n unabhängige Ziehungen mit Zurücklegen realisiert werden (Urnenmodell). In der Praxis ist das Ziehen von Zufallsstichproben aber weitaus komplexer (vgl. 7.3).

Eine Zufallsstichprobe kann als konkrete Realisierung eines <u>Zufallsexperiments</u> verstanden werden. Eine <u>Stichprobenstatistik</u> (z.B. der Mittelwert eines Merkmals in der Stichprobe) kann daher als eine <u>Zufallsvariable</u> verstanden werden, deren Verteilung von den entsprechenden <u>Parametern</u> in der GG abhängt. Um wahrscheinlichkeitstheoretische Aussagen über diese Zufallsvariablen machen zu können, müssen bestimmte Annahmen über ihre Verteilung gemacht werden. Ist die <u>Verteilung in der GG</u> bekannt, kann die Verteilung der Stichprobenstatistik exakt abgeleitet werden; anderenfalls können z.B. mit Hilfe des <u>zentralen Grenzwertsatzes</u> approximative Aussagen gemacht werden. Sodann können der <u>Standardfehler</u> der Schätzung berechnet und <u>Konfidenzintervalle</u> bestimmt werden.

Entscheidend für den Test von Hypothesen ist die Formulierung einer Arbeitshypothese H<sub>A</sub> und einer Nullhypothese H<sub>0</sub>. Wenn möglich, testet man jene Behauptung als Nullhypothese, deren irrtümliche Ablehnung die schwerwiegenderen Konsequenzen oder höheren Kosten nach sich zieht (Minimierung des Alpha-Fehlers; Signifikanztest vgl. H<sub>A</sub>). Bei der Hypothesenbildung ist insbesondere auf eine ein- oder zweiseitige Formulierung zu achten (z.B. sollte für den t-Test bei der Trennschärfebestimmung (vgl. 5.3.1) eine einseitge Formulierung gewählt werden). Vor Beginn des Tests muß die Irrtumswahrscheinlichkeit festgelegt werden. Dabei ist die Problematik von Alpha-Fehlern (Fehlern 1. Art) und Beta-Fehlern (Fehlern 2. Art) zu berücksichtigen. Für die genaue Vorgehensweise bei der Durchführung statistischer Tests wird auf die einschlägigen Lehrbücher verwiesen (vgl. z.B. Bortz 1984 und die dort angegebene weiterführende Literatur).

akzeptiert wird (aufgrund der Stichprobenstatistik):

 $H_{O}$ 

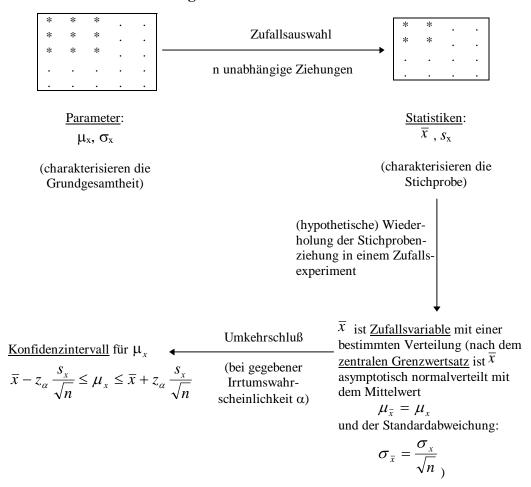
 $H_A$ 

 $\begin{array}{c|c} H_{O} & H_{A} \\ \hline & Fehler zweiter \\ 1-\alpha & Art (\beta-Fehler) \\ \beta & \\ \hline Fehler erster \\ Art (\alpha-Fehler) \\ \alpha & \\ \end{array}$ 

richtig ist (in der GG):

Auswahlverfahren – 162 – Kapitel 7.

#### Logik statistischen Schließens



## 7.3 Praktische Probleme bei der Stichprobenziehung

In der Praxis sind Stichprobenziehungen nach dem Urnenmodell in den meisten Fällen undurchführbar, da keine entsprechenden Listen verfügbar sind. Stattdessen verwendet man mehrstufige Auswahlverfahren, so bei allen repräsentativen Bevölkerungsumfragen wie z.B. dem ALLBUS. Zunächst werden dazu <u>Primäreinheiten</u> definiert, die als Auswahlgesamtheit der ersten Stufe dienen. Die Primäreinheiten sind in den meisten Fällen als Flächen definiert, z.B. Stimmbezirke oder Gitter in einem Stadtplan. Ist dies der Fall, wird die Zufallsauswahl der ersten Stufe als <u>Flächenstichprobe</u> bezeichnet. Auf der zweiten Stufe wird dann in den (auf der ersten Stufe per Zufall ausgewählten) Primäreinheiten eine Zufallstichprobe von <u>Sekundäreinheiten</u> gezogen. Im Falle von Flächenstichproben können die einzelnen räumlichen Einheiten als Cluster von Haushalten verstanden werden. Die Auswahl von Sekundäreinheiten bedeutet dann die Auswahl von Haushalten in den ausgewählten Flächen (<u>Haushaltsstichprobe</u>). Da auch bei diesem Auswahlschritt im allgemeinen auf keine Liste zurückgegriffen werden kann, werden die Haushalte meistens mit Hilfe des Random-Walk-Verfahrens ermittelt, d.h. die Haushaltsadressen werden aufgrund einer Begehungsanweisung von Adressenermittlern erhoben, bevor in einem dritten und letzten Schritt der Befragte

innerhalb eines ausgewählten Haushalts mit Hilfe einer Tabelle von Zufallszahlen ("Schwedenschlüssel") ausgewählt wird (Personenstichprobe).

Eine Flächenstichprobe kann als Spezialfall einer Klumpenstichprobe verstanden werden. Bei einer Klumpenstichprobe werden die Elemente einer GG nicht direkt ausgewählt, sondern stattdessen Klumpen oder Cluster (d.h. zu Gruppen zusammengefaßte Elemente der GG, wie z.B. alle Schüler einer Schulklasse oder alle Arbeiter einer Abteilung). Im Gegensatz zur mehrstufigen Auswahl gehen dann aber alle Elemente der ausgewählten Klumpen in die Stichprobe ein. Mit Hilfe von Klumpenstichproben können daher auch Kontexteffekte untersucht werden (Einfluß der Schulklassse oder der Abteilung). Klumpenstichproben sind im allgemeinen kostengünstiger als einfache Zufallsstichproben; der dadurch erzielten Kostenersparnis steht in vielen Fällen allerdings der Klumpeneffekt gegenüber: Sind nämlich die Klumpen intern homogen, so erhöht sich der Standardfehler der Schätzung, d.h. die Schätzungen aufgrund einer Klumpenstichprobe sind im allgemeinen ungenauer.

Bei einer geschichteten Zufallsstichprobe werden die Elemente der GG zunächst bezüglich eines Merkmals in Gruppen (Schichten) eingeteilt. Sodann wird aus jeder Schicht eine Zufallsstichprobe gezogen. Bei der proportionalen Schichtung werden die Stichprobenumfänge entsprechend dem Anteil der Schichten an der GG gewählt. Auf diese Weise können Zufallschwankungen in Bezug auf das Schichtungsmerkmal ausgeschlossen werden - die resultierende Gesamtstichprobe ist genauer. Sollen kleinere Gruppen (Schichten) besonders berücksichtigt werden, können aber auch disproportional geschichtete Zufallsstichproben gezogen werden (so sind z.B. im ALLBUS 1994 die Ostdeutschen überrepräsentiert).

Ausfallproblematik: Ein äußerst schwerwiegendes Problem der Stichprobenziehung sind Ausfälle. Wir unterscheiden Nicht-Befragbare, Schwer-Erreichbare und die eigentlichen Verweigerer. Die Nicht-Befragbaren stellen nur eine kleine Gruppe dar und verursachen daher in der Regel kaum Verzerrungen (es sei denn, das Thema der Befragung ist z.B. der Gesundheitszustand der Bevölkerung). Schwer-Erreichbare, d.h. Personen, die trotz mehrfacher Kontaktversuche nicht an ihrem Wohnsitz angetroffen wurden, stellen eine zahlenmäßig größere Gruppe dar. Da die Ausfälle darüberhinaus systematisch erfolgen (z.B. hohe berufliche Belastung oder ungewöhnliche Arbeitszeiten), muß auf alle Fälle versucht werden, durch wiederholte Kontaktversuche die Zahl der Schwer-Erreichbaren so gering wie möglich zu halten. Die problematischste Gruppe stellen aber die Verweigerer dar, also Personen, die die Teilnahme an der Befragung explizit verweigern. Gerade hier sind systematische Verzerrungen zu erwarten, obwohl der harte Kern von Verweigerern aus prinzipiellen Gründen eher gering sein dürfte (allerdings mit einer steigenden Tendenz). Die Verweigerungsrate ist insbesondere bei älteren Befragten deutlich höher als im Bevölkerungsdurchschnitt. Entscheidend für die Verweigerung der Teilnahme an einer Untersuchung sind vor allem situationale Bestimmungsgründe (Thema der Untersuchung, Auftreten des Interviewers, generelle Angst vor Kontakten an der Haustür, usw.). Zur Vermeidung von Ausfällen werden in der Befragungspraxis eine Vielzahl von Strategien

angewandt. Dennoch stellt die Ausfallproblematik nach wie vor eine schwerwiegende Bedrohung der Repräsentativität von Zufallsstichproben dar.

Darüber hinaus stellen die komplexen mehrstufigen Auswahlverfahren auch hohe Anforderungen an die Interviewer. Die Realisierung eines Interviews nach den vorgegebenen Regeln des Befragungsinstitutes erfordern oft einen aus Sicht des Interviewers unvertretbar hohen Zeitaufwand, insbesondere wenn man berücksichtigt, daß die Interviewer in der Regel schlecht bezahlt werden. In diesen Fällen ist die Gefahr einer Fälschung von Interviews besonders groß (vgl. Dorroch 1994). Inwieweit hier eine effektive Interviewerkontrolle durch die Befragungsinstitute möglich ist, ist noch weitgehend unklar. Allerdings werden oft noch nicht einmal energische Schritte zur Interviewerkontrolle von Seiten der Befragungsinstitute unternommen, um die Kosten einer Erhebung nicht zu erhöhen. (Hier wäre eine telefonische Befragung mit der Möglichkeit einer zentralen Kontrolle des Erhebungsprozesses eine erwägenswerte kostengünstige Alternative). Langfristig erscheinen hier genauere Kontrollen, verbunden mit einer besseren Bezahlung der Interviewer, unabdingbar. Auch die direkte Ziehung einer Personenstichprobe auf der Grundlage der Daten der Einwohnermeldeämter führt, wie die Erfahrungen mit dem ALLBUS 1994 zeigen, zu einer deutlichen Verbesserung in der Stichprobenqualität - allerdings bei deutlich höheren Kosten als bei der üblichen dreistufigen ADM-Stichprobe.

# 8. Datenaufbereitung und Datenanalyse

(siehe dazu auch 3.1, 3.2, 3.3, 4.3, 5.3 und 5.4)

Eine Einführung in die sozialwissenschaftliche Datenanalyse findet man bei H. Benninghaus in dem gleichnamigen Buch.

Weiterführende analytische Verfahren sind in dem Buch von K. Backhaus u.a. über "Multivariate Analysemethoden" dargestellt.

Einen ersten Einblick in die EDV-unterstützte Datenanalyse mit SPSS für Windows bieten:

- → G. Brosius und F. Brosius: SPSS Base System und Professional Statistics.
- → W.-M. Kähler: SPSS für Windows. Datenanalyse unter Windows.

Daneben gibt es vom Lehrstuhl Beispieldateien zur Datenanalyse zu den folgenden Themen:

- Lineare Regressionsanalyse
- Reliabilitätsanalyse
- Faktorenanalyse
- T-Test
- Tabellenanalyse
- Varianzanalyse

Diese Dateien liegen im Adobe-Acrobat-Format (\*.pdf) vor und sind über das WWW zu beziehen.<sup>1</sup>

Die Windows-Version von SPSS enthält auch das Programmpaket AMOS zur allgemeinen Analyse von Kovarianzstrukturen.<sup>2</sup> Damit können sowohl komplexe Strukturgleichungssysteme wie z.B. nichtrekursive Kausalstrukturen analysiert werden wie auch integrierte Kausalmodelle, die zusätzlich noch Meßmodelle für die jeweiligen theoretischen Konstrukte enthalten (vgl. die Modelle zum kausalanalytischen Ansatz in 4.3). AMOS ist über ein Interface in die SPSS-Programmlogik integriert.

→ Arbuckle, James L.: Amos User's Guide Version 3.6. Chicago 1997.

Genauere Informationen hierzu erteilen gerne die Mitarbeiter des Lehrstuhls. Des weiteren sind Informationen (<u>nicht nur</u>) zu diesen Dateien über das World Wide Web verfügbar. Die Adresse unserer Homepage lautet: <a href="http://wwcont.wiwi.uni-wuppertal.de/kappelhoff/welcome.html">http://wwcont.wiwi.uni-wuppertal.de/kappelhoff/welcome.html</a>.

<sup>&</sup>lt;sup>2</sup> AMOS ersetzt das früher im SPSS-Programmpaket enthaltene LISREL 7.

### Literatur

Albert, Hans: 1968

Traktat über kritische Vernunft, Tübingen: Mohr.

Albert, Hans: 1987

Kritik der reinen Erkenntnislehre. Tübingen: Mohr.

Alemann, Heine von: 1977

Der Forschungsprozeß. Eine Einführung in die Praxis der empirischen Sozialforschung. Stuttgart: Teubner.

Andersson, Gunnar: 1988

Kritik und Wissenschaftsgeschichte. Kuhns, Lakatos' und Feyerabends Kritik des Kritischen Rationalismus.

Tübingen: Mohr.

Andrich, David: 1988

Rasch Models for Measurement. Beverly Hills: Sage.

Arminger, Gerhard: 1976

Anlage und Auswertung von Paneluntersuchungen. S. 134-235 in: K. Holm (Hrsg.): Die Befragung 4.

München: Francke.

Backhaus, K.; B. Erichson; W. Plinke; Chr. Schuchard-Fischer & R. Weiber: 1993

Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Berlin: Springer.

Bartley, William W.: 1987

Flucht ins Engagement. Tübingen: Mohr.

Benninghaus, Hans: 1979

Deskriptive Statistik. Stuttgart: Teubner.

Benninghaus, Hans: 1990

Einführung in die sozialwissenschaftliche Datenanalyse. München: Oldenburg.

Berger, P.L. & T. Luckmann: 1980

Die gesellschaftliche Konstruktion der Wirklichkeit. Frankfurt: Fischer.

Bielby, William T.; Robert M. Hauser & David L. Featherman: 1977

Response Errors of Black and Nonblack Males in Models of the Intergenerational Transmission of

Socioeconomic Status. American Journal of Sociology 82, 1977, S. 1242-1288.

Bierbrauer, G.: 1976

Attitüden: Latente Strukturen oder Interaktonskonzepte. Zeitschrift für Soziologie 5, 1976, S. 4-16.

Blalock, Hubert: 1972

Causal Inferences in Nonexperimental Research. New York: Norton.

Boden, Ulrike; Jürgen Bortz; Paul Braune & Joachim Franke: 1975

Langzeiteffekte zweier Tageszeitungen auf politische Einstellungen der Leser. Kölner Zeitschrift für

Soziologie und Sozialpsychologie 27, 1975, S. 754-780.

Böltken, F.: 1976

Auswahlverfahren. Eine Einführung für Sozialwissenschaftler. Suttgart: Teubner.

Borg, Ingwer: 1981

Anwendungsorientierte multidimensionale Skalierung. Berlin: Springer.

Bortz, Jürgen: 1984

Lehrbuch der empirischen Forschung für Sozialwissenschaftler. Berlin: Springer.

Braun, Michael & Peter Ph. Mohler (Hrsg.): 1994

Blickpunkt Gesellschaft 3, Opladen: Westdeutscher Verlag.

Brosius, Gerhard/Brosius, Felix: 1995

SPSS - Base System und Professional Statistics. Hg. Lammarsch, J. Bonn: International Thomson Publishing.

Brosius, Gerhard: 1988

SPSS/PC+. Basics und Graphics. Einführung und praktische Beispiele. Hamburg: McGraw-Hill.

Brosius, Gerhard: 1989

SPSS/PC+. Advanced Statistics und Tables. Einführung und praktische Beispiele. Hamburg: McGraw-Hill.

Bungard, W. & H.E. Lück: 1974

Forschungsartefakte und nicht-reaktive Meßverfahren. Stuttgart: Teubner.

Chalmers, A.F.: 1986

Wege der Wissenschaft. Einführung in die Wissenschaftstheorie. Berlin: Springer.

Coombs, Clyde H.: 1964

A Theory of Data. New York: Wiley.

Dawes, Robyn M.: 1964

Grundlagen der Einstellungsmessung. Weinheim und Basel: Beltz.

Diekmann, Andreas: 1995

Empirische Sozialforschung - Grundlagen, Methoden, Anwendungen. Hrsg. König, B. Reinbeck: Rowohlt.

Dorroch, Heiner: 1994

Meinungsreport: Wie Umfrageergebnisse entstehen. Göttingen: Steidl.

Dreier, Volker: 1994

Datenanalyse freier Sozialwissenschaftler. München, Oldenburg.

Erbslöh, E.: 1972

Das Interview. Stuttgart: Teubner.

Esser, Hartmut: 1986

Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. Kölner Zeitschrift für Soziologie und Sozialpsychologie 38, 1986, S. 314-336.

Festinger, L.; H.W. Riecken & S. Schachter: 1956

When Prophecy Fails. New York: Harper & Row.

Frey, James A.; Gerhard Kunz & Günther Lüschen: 1990

Telefonumfragen in der Sozialforschung. Methoden, Techniken und Befragungspraxis. Opladen: Westdeutscher Verlag.

Friedrichs, Jürgen: 1973

Lehrbuch empirischer Sozialforschung. Reinbek: Rowohlt.

Glenn, Norval D.: 1977

Cohort Analysis. Beverly Hills: Sage.

Green, Paul E. & Donald S. Tull: 1982

Methoden und Techniken der Marketingforschung. Stuttgart: Pöschel.

Grümer, Karl Wilhelm: 1974 Beobachtung. Stuttgart: Teubner.

Hagstotz, Werner: 1985

Welcher Inglehart-Index ist der richtige? Methodische Anmerkungen zur Messung von Wertorientierungen. ZUMA-Nachrichten 16, 1985, S. 16-38.

Hanefeld, Ute: 1987

Das sozioökonomische Panel. Frankfurt: Campus.

Heidenreich, Klaus: 1987

Grundbegriffe der Meß- und Testtheorie. S. 352-384 in: Erwin Roth (Hrsg.): Sozialwissenschaftliche Methoden. München: Oldenbourg., S. 352-384.

Heise, D.R.: 1970

Causal Inference from Panel Data. Sociological Methodology 1970, S. 3-27.

Hermann, Dieter: 1983

Die Priorität von Einstellungen und Verzerrungen im Interview. Eine Methodenuntersuchung anhand der Daten der Allgemeinen Bevölkerungsumfrage 1980. Zeitschrift für Soziologie 12, 1983, S. 242-252.

Hippler, Hans & Seidel, Kristiane: 1985

Schriftliche Befragung bei allgemeinen Bevölkerungsstichproben. Untersuchungen zur Dillmanschen "Total Design Method". ZUMA-Nachrichten 16: 39-56.

Hoag, Wendy I. & Klaus R. Allerbeck: 1981

Interviewer- und Situationseffekte in Umfragen. Eine log-lineare Analyse. Zeitschrift für Soziologie 10, 1981, S. 413-426.

Hodapp, Volker: 1984

Analyse linearer Kausalmodelle. Bern: Huber.

Hormuth Stefan E. & Erika Brückner: 1985

Telefoninterviews in Sozialforschung und Sozialpsychologie. Kölner Zeitschrift für Soziologie und Sozialpsychologie 37, 1985, S. 526-545.

Hoyningen-Huene, Paul: 1987

Die Wissenschaftsphilosophie Thomas S. Kuhns: Rekonstruktion und Grundlagenprobleme. Braunschweig: Vieweg.

Hunt, Morton: 1991

Die Praxis der Sozialforschung. Reportagen aus dem Alltag der Wissenschaft. Frankfurt.

Kaehler, Wolf-Michael: 1994

SPSS für Windows: Datenanalyse unter Windows. Braunschweig: Vieweg.

Kappelhoff, Peter & Sigrid Schulz: 1991

Determinanten der Berufsverläufe von Frauen. Kiel: Institut für Soziologie.

Karmasin, Fritz & Helene Karmasin: 1977

Einführung in Methoden und Probleme der Umfrageforschung. Köln: Boehlan.

Kinsey, Alfred C.; Wardell B. Pomeroy & Clyde E. Martin: 1970 (1948<sup>1</sup>)

Das sexuelle Verhalten des Mannes. Frankfurt: Fischer.

Kolakowski, Leszek: 1977

Die Suche nach der verlorenen Gewißheit. Stuttgart: Kohlhammer.

König, R.: 1973

Hanbuch der empirischen Sozialforschung, Stuttgart.

König, Rene'(Hrsg.): 1972

Das Interview. Köln: Kiepenheuer & Witsch.

Koolwijk, Jürgen van & Maria Wieken-Mayser (Hrsg.): 1974

Techniken der empirischen Sozialforschung. München: Oldenbourg.

Band IV: Erhebungmethoden: Die Befragung.

Band V: Testen und Messen. Band VIII: Kausalanalyse.

Krampen, Günter: 1979

Eine Skala zur Messung der normativen Geschlechtsrollenorientierung (GRO-

Skala). Zeitschrift für Soziologie 8, 1979, S. 254-266.

Kruskal, Joseph B. & Myron Wish: 1977

Multidimensional Scaling. Beverly Hills: Sage.

Kuhn, Thomas: 1976

Die Struktur wissenschaftlicher Revolutionen. Frankfurt: Suhrkamp.

Lakatos, Imre & Alan Musgrave (Hrsg.): 1974

Kritik und Erkenntnisfortschritt. Braunschweig: Vieweg.

Lakatos, Imre: 1974

Falsifikation und die Methodologie wissenschaftlicher Forschungsprogramme. S. 89-189 in: I. Lakatos & A. Musgrave (Hrsg.): Kritik und Erkenntnisfortschritt. Braunschweig: Vieweg.

Lamnek, Siegfried: 1988

Qualitative Sozialforschung. München: Psychologie Verlags Union.

Band I: Methodologie.

Band II: Methoden und Techniken.

Lisch, Ralf & Jürgen Kriz: 1978

Grundlagen und Modelle der Inhaltsanalyse. Reinbek: Rowohlt.

Lorenz, K.: 1973

Die Rückseite des Spiegels. München: Piper.

Mayntz, Renate; Kurt Holm & Peter Hübner: 1978

Einführung in die Methoden der empirischen Soziologie. Opladen: Westdeutscher Verlag.

McIver, John P. & Edward G. Carmines: 1981 Unidimensional Scaling. Beverly Hills: Sage.

Meinefeld, Werner. 1995

Realität und Konstruktion. Erkenntnistheoretische Grundlagen einer Methodologie der empirischen Sozialforschung. Opladen: Leske & Budrich.

Meulemann, H. & K.-H. Reuband: 1984

Soziale Realität im Interview. Empirische Analysen methodischer Probleme. Frankfurt: Campus.

Meuser, Michael & Ulrike Nagel: 1991

Experteninterviews - vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. S. 441-471 in: D. Garz & K. Kaimer (Hrsg.): Qualitativ-empirische Sozialforschung. Opladen: Westdeutscher Verlag.

Minard, Ralph D.: 1952

Race Relationships in the Pocahontas Coal Field. Journal of Social Issues 8, S. 29-44.

Mohler, Peter Ph. & W. Bandilla (Hrsg.):1992

Blickpunkt Gesellschaft 2. Opladen: Westdeutscher Verlag.

Mohler, Peter Ph.: 1984

Wertewandel in Deutschland: 1917-71 (Abitur 1917-1971). S. 279-308 in: H.-D. Klingemann (Hrsg.): Comupterunterstützte Inhaltsanalyse in der empirischen Sozialforschung. Frankfurt: Campus.

Müller, W.; P. Mohler; B. Erbslöh & M. Wasmer (Hrsg.): 1990 Blickpunkt Gesellschaft. Opladen: Westdeutscher Verlag.

Müller, Walter: 1972

Bildung und Mobilitätsprozeß - Eine Anwendung der Pfadanalyse. Zeitschrift für Soziologie 1, 1972, S. 65-84.

Mummendey, Hans Dieter: 1987

Die Fragebogenmethode. Göttingen: Hogrefe.

Musgrave, Alan: 1993

Alltagswissen, Wissenschaft und Skeptizismus. Tübingen: Mohr.

Noelle-Neumann, Elisabeth & Petersen, Thomas: 1996

Alle, nicht jeder. Einführung in die Methoden der Demoskopie. München dtv.

Noelle-Neumann, Elisabeth: 1963:

Umfragen in der Massengesellschaft. Reinbek: Rowohlt.

Opp, K.D. & Peter Schmidt: 1976

Einführung in die Mehrvariablenanalyse. Reinbek: Rowohlt.

Ostrom, Charles W.: 1990

Time Series Analysis. Regression Techniques. Beverly Hills: Sage.

Popper, Karl R.: 1973

Objektive Erkenntnis. Hamburg: Hoffmann und Campe.

Popper, Karl R.: 1974

Die Normalwissenschaft und ihre Gefahren. S. 51-57 in: I. Lakatos & A. Musgrave (Hrsg.): Kritik und Erkenntnisfortschritt. Braunschweig: Vieweg.

Popper, Karl R.: 1980

Die offene Gesellschaft und ihre Feinde. München: Francke.

Band I: Der Zauber Platons

Band II: Falsche Propheten: Hegel, Marx und die Folgen.

Popper, Karl R.: 1982

Die Logik der Forschung. Tübingen: Mohr.

Prim, R. & H. Tilmann: 1983

Grundlagen einer kritisch-rationalen Sozialwissenschaft. Heidelberg: Quelle & Meyer.

Reinecke, Jost: 1991

Interviewer- und Befragtenverhalten. Theoretische Ansätze und methodische Konzepte. Opladen: Westdeutscher Verlag.

Reuband, Karl-Heinz: 1986

Einflüsse der Interviewsituation auf den Inglehartschen Postmaterialismusindex: die Bedeutung von Interviewermerkmalen für das Antwortverhalten der Befragten. ZA-Information 18, 1986, S. 35-55.

Reuband, Karl-Heinz: 1989

Kontakte zwischen Deutschen und Gastarbeitern: Ein Indikatorenvergleich. ZA-Information 24, 1989, S. 72-83.

Rohrmann, B.: 1978

Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. Zeitschrift für Sozialpsychologie 9, 1978, S. 222-245.

Rosenthal, Robert & Lenore Jacobsen: 1976

Pygmalion im Unterricht. Lehrererwartungen und Intelligenzentwicklung der Schüler. Weinheim: Beltz.

Ross, John & Perry Smith: 1968

Orthodox Empirical Designs. S. 333-389 in: H.M. Blalock und A.B. Blalock (Hrsg.): Methodology in Social Research. London: McGraw Hill

Roth, Erwin (Hrsg.): 1987

Sozialwissenschaftliche Methoden. München: Oldenbourg.

Sample, John & Rex Warland: 1973

Attitude and Prediction of Behavior. Social Forces 51, 1973, S. 292-304.

Schäfer, Lothar: 1988

Karl R. Popper. München: Beck.

Schanz, Volker & Peter Schmidt: 1984

Interviewsituation, Interviewermerkmale und Reaktionen von Befragten im Interview: eine multivariate Analyse. S. 72-113 in: K.U. Mayer & P. Schmidt (Hrsg.): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Frankfurt: Campus.

Scheuch, Erwin K.: 1973

Das Interview in der Sozialforschung; in: R. König 1973, Bd. 2.

Schnell, Rainer; Paul B. Hill & Elke Esser: 1994

Methoden der empirischen Sozialforschung. München: Oldenbourg.

Schumann, H. & S. Presser: 1981

Questions and Answers in Attitude Surveys. New York: Academic Press.

Schwarz, Norbert; Hans-J. Hippler & Elisabeth Noelle-Neumann: 1989

Einflüsse der Reihenfolge von Antwortvorgaben bei geschlossenen Fragen. ZUMA-Nr. 25, 1989, S. 24-38.

Schweikl, Herbert: 1985

Computergestützte Präferenzanalyse mit individuell wichtigen Produktmerkmalen. Berlin: Duncker & Humblot.

Spöhring, Walter: 1989

Qualitative Sozialforschung. Stuttgart: Teubner.

Steffens, Karl: 1987

Skalierung. S. 533-557 in: Erwin Roth (Hrsg.): Sozialwissenschaftliche Methoden. München: Oldenbourg.

Steinert, Heinz: 1984

Das Interview als soziale Situation. S.17-60 in: H. Meulemann & K.-H.Reuband (Hrsg): Soziale Realität im Interview. Frankfurt: Campus.

Sudmann, S. & Bradburn, N. & Schwarz, N.: 1996

Thinking about answers. The application of processes to survey methodology. San Francisco: Jossey-Bass.

Tietzel, Manfred: 1988

Eine Anleitung, empirische Hypothesen unwiderlegbar zu machen, in: K.-D. Freimann und A.E. Ott (Hrsg.): Theorie und Empirie in der Wirtschaftsordnung. Tübingen: Mohr,: 3-23.

Topitsch, Ernst: 1979 (1988<sup>2</sup>)

Erkenntnis und Illusion. Hamburg: Hoffmann & Campe.

Vollmer, Gerhard: 1975

Evolutionäre Erkenntnistheorie. Stuttgart: Hirzel.

Watzlawick, Paul (Hrsg.): 1985

Die erfundene Wirklichkeit. München: Piper.

Watzlawick, Paul: 1989

Wie wirklich ist die Wirklichkeit? München: Piper.

Weber, Max: 1921

Soziologische Grundbegriffe. Tübingen: Mohr.

Weber, Robert Philip: 1981

Society and Economy in the Western World System. Social Forces 59, 1981, S. 1130-1145.

Weber, Robert Philip: 1990

Basic Content Analysis. Beverly Hills: Sage.

Weede, E. & W. Jagodzinski: 1977

Einführung in die konfirmatorische Faktorenanalyse. Zeitschrift für Soziologie 6, 1977, S. 315-333.

Wegener, Bernd: 1983

Wer skaliert? Die Meßfehlertheorie und die Frage nach dem Akteur. In: ZUMA-Handbuch

sozialwissenschaftlicher Skalen (Einleitung).

Westermann, R.: 1985

Empirical Tests of Scale Typs for Individual Ratings. Applied Psychological Measurement 9, 1985, S. 265-

274.

Wiegand, Erich: 1991

Zunahme der Ausländerfeindlichkeit? ZUMA-Nachrichten 31, 1991, S. 7-28.

Zimmermann, E.: 1972

Das Experiment in den Sozialwissenschaften. Stuttgart: Teubner.

ZUMA - Zentrum für Umfragen, Methoden und Analysen: 1983

ZUMA-Handbuch für sozialwissenschaftliche Skalen, Mannheim.

# Übungsaufgaben

## Grundlegende Techniken der Datenanalyse

[Zur linearen Regressionsanalyse und zum t-Test mit SPSS für Windows gibt es Beispieldateien (Acrobat). Diese Dateien können über das WWW bezogen werden]

1. Lineare Regression: Für die beiden Variablen X (normative Geschlechtsrollenorienterung) und Y (politische Partizipation) sind in einer Untersuchung (Stichprobengröße n=5) die folgenden Werte gemessen worden:

	X	Y
1	2	2
2	3	6
3	1	4
4	5	8
5	4	10

Stellen Sie eine Kausalhypothese auf und berechnen Sie die Regression von Y auf X!

- a) Zeichnen Sie zunächst das Streudiagramm und berechnen Sie dann Mittelwert und Standardabweichung für die beiden Variablen.
- b) Bestimmen Sie die Regressionsgerade  $\hat{Y}_i$  =a+b $X_i$  mit Hilfe der Methode der kleinsten Quadrate. Wieviel Prozent der Varianz der abhängigen Variablen Y wird durch die Regression erklärt?
- c) Bestimmen Sie den Produkt-Moment-Korrelationskoeffizienten  $r_{Xy}$  (Pearsons r) und diskutieren Sie die varianzanalytische Interpretation von  $r^2$ .
- d) Standardisieren Sie X und Y und bestimmen Sie die Regressionsgerade erneut!
- e) Welches Ergebnis erhält man für die Regression von X auf Y?

## Stichworte zur Übungsaufgabe

- Variable, Mittelwert, Standardabweichung, Varianz
- lineare Beziehung, Kovarianz, Korrelation
- Vorhersagefehler, Methode der kleinsten Quadrate
- erklärte Varianz, Fehlervarianz
- Regressionskoeffizient
- standardisierte Variable, z-Transformation
- standardisierter Regressionskoeffizient
- Zufallsfehler, Pfadkoeffizient

2. t-Test: In einer Untersuchung soll festgestellt werden, ob sich die Einstellung von Männern und Frauen bezüglich der normativen Geschlechtsrollenorientierung (nGRO) unterscheidet. Es wurden folgende Daten erhoben:

\	24	30	36
$\mapsto$	26	32	38

Skalenwert auf der Likert-Skala zur

nGRO für Männer und Frauen

$S_1$	8	17	8
$S_2$	80	170	80
$S_3$	1	1	1

Besetzungszahlen für die unterschiedlichen Stichprobengrößen  $\, {\rm S}_{1} \, (33), \, {\rm S}_{2} \, (330),$ 

und S<sub>3</sub> (3) (jeweils für Männer und Frauen gleich)

- a) Wird die Hypothese durch die Daten bestätigt? Führen Sie in Abhängigkeit von den jeweiligen Stichprobengrößen einen Signifikanztest durch.
- b) Ändern sich die Schlußfolgerungen, wenn von vornherein postuliert wird, daß Männer eine traditionalere Einstellung zur nGRO (einen niedrigeren Mittelwert) haben als Frauen?

## Stichworte zur Übungsaufgabe

- Logik von Signifikanztests
- Test auf Mittelwertdifferenzen
- einseitige bzw. zweiseitige Fragestellung
- Zentraler Grenzwertsatz
- unabhängige Stichproben bzw. abhängige Stichproben

## II. Der Forschungsprozeß

- Rekapitulieren Sie die in der Vorlesung besprochenen Arbeitsschritte bei der Durchführung eines Forschungsvorhabens. Entwickeln Sie dann für eine der folgenden Problemstellungen einen Forschungsplan:
  - a) Bedingungen des Freizeitverhaltens
  - b) Veränderung im Medienkonsum
  - c) Test der Wirksamkeit einer geplanten Werbemaßnahme
  - d) Umweltproblematik und Konsumentenverhalten
  - e) Vergleichende Untersuchung von Studiendauer und Studienerfolg im Fach WiWi an den Universitäten Köln und Wuppertal
  - f) Freies Thema

Entwickeln Sie zunächst eine Kausalhypothese für die von Ihnen gewählte Problemstellung. Führen Sie dann eine Konzeptspzefikation und Operationalisierung der dabei benutzten Begriffe durch und skizzieren Sie anschließend den von Ihnen ins Auge gefaßten Forschungsplan. Welche Schritte im Forschungsablauf konnten von Ihnen nicht adäquat bearbeitet werden? Benennen Sie möglichst konkret die dabei aufgetretenen Probleme.

### III. Wissenschaftstheorie

### **III.1 Kritischer Rationalismus**

- 1. Was bedeutet die Trennung von Entdeckungs- und Begründungszusammenhang im Kritischen Rationalismus (KR)?
- 2. Erläutern Sie die Stellung des KR zum Induktionsprinzip.
- 3. Welchen Kriterien müssen Nominal- bzw. Realdefinitionen genügen und was folgt daraus für die Brauchbarkeit dieser beiden Definitionsformen?
- 4. Rekapitulieren Sie das Schema mit den verschiedenen Aussageformen und nehmen Sie zu folgender Position Stellung: "Das Ziel der Wissenschaft ist es, zu Aussagen zu gelangen, die Realitätsbezug haben und wahr sind."
- 5. Wie können Theorien nach ihrem Informationsgehalt und ihrem Bewährungsgrad unterschieden werden?
- 6. Geben Sie ein Beispiel für eine deduktiv-nomologische Erklärung.
- 7. Diskutieren Sie den Unterschied zwischen naivem und methodologischem Falsifikationismus.
- 8. Was versteht der KR unter "Wahrheit" und "Objektivität" wissenschaftlicher Erkenntnis?

### III.2 Kritik des Kritischen Rationalismus

- 1. Erläutern Sie den Paradigmabegriff und die Inkommensurabilitätsthese bei Kuhn.
- 2. Stellen Sie den methodologischen und den raffinierten Falsifikationismus gegenüber.

- 3. Was versteht Lakatos unter einem Forschungsprogramm? Kann die neoklassische Tradition innerhalb der Ökonomik als Forschungsprogramm in diesem Sinne verstanden werden?
- 4. Was versteht man unter einer "self-fulfilling-prophecy" bzw. einer "self-destroying-prophecy"? Wird die darin zum Ausdruck kommende Reflexivität gerade sozialwissenschaftlicher Erkenntnis im KR ausreichend berücksichtigt?

### weitere Übungsaufgaben:

- Welche Probleme ergeben sich aus der Theorieabhängigkeit der Erfahrung für die empirische Überprüfung von Theorien? Diskutieren Sie die vom KR vorgeschlagene Lösung. Konkretisieren Sie Ihre Überlegungen an einem Beispiel und verwenden Sie dabei den kausalanalytischen Ansatz. Welche zusätzlichen Schwierigkeiten ergeben sich aus der Theorieabhängigkeit der Erfahrung für den Vergleich konkurrierender Theorien?
- Worin besteht die zentrale Bedeutung des Basissatzproblems für den KR? Diskutieren Sie die von Popper vertretenen Lösung im methodologischen Falsifikationismus an einem Beispiel und nehmen Sie zu den dabei auftretenen Problemen Stellung. Gehen Sie in diesem Zusammenhang auch auf die Kritik Kuhns an dem Falsifikationsgedanken ein. Inwieweit bietet der raffinierte Falsifikationismus von Lakatos hier einen Ausweg?
- Stellen Sie die wissenschaftstheoretische Position von Kuhn dar und konzentrieren Sie sich dabei insbesondere auf die beiden zentralen Begriffe "Paradigma" und "Inkommensurabilität". Welche Implikationen hat diese Position für die Möglichkeit, Theorien zu überprüfen und zwischen konkurrierenden Theorien zu entscheiden? Kontrastieren Sie die Kuhnsche Position mit der des KR und gehen Sie dabei insbesondere auf die Poppersche Sicht vom "Mythos des Rahmens" ein. Nehmen Sie anschließend kritisch zu der gesamten Diskussion Stellung, indem Sie die konträren Argumente gegeneinander abwägen.
- In welchem Sinne kann im KR von der "Wahrheit" wissenschaftlicher Erkenntnis gesprochen werden? Erläutern Sie zunächst die Trennung von Entdeckungs-, Begründungs- und Verwendungszusammenhang, das deduktiv-nomologische Erklärungsmodell und das Basissatzproblem. Diskutieren Sie insbesondere den Unterschied zwischen naivem und methodlogischem Falsifikationismus, und geben Sie dazu ein Beispiel. Was folgt aus der wissenschaftstheoretischen Position des KR konkret in Hinblick auf Möglichkeiten und Grenzen empirischer Forschung? Nehmen Sie kritisch zu der Position des KR Stellung.

# IV. Untersuchungsformen

### IV.1 Qualitative Sozialforschung

- 1. Diskutieren Sie die Auseinandersetzungen um "Verstehen und Erklären" aus Sicht der bekannten Definition von *Soziologie* durch Max Weber.
- 2. Welche Bedeutung hat der gerade in der deutschen geisteswissenschaftlichen Tradition gepflegte Dualismus zwischen einem nomothetischen und einem ideographischen Wissenschaftsverständnis und damit zwischen Natur- und Kulturwissenschaften für das Spannungsverhältnis zwischen quantitativer und qualitativer Sozialforschung?
- 3. Skizzieren Sie das evolutionstheoretische Argument als Stütze einer einheitlichen naturalistischen Wissenschaftsauffassung und geben Sie ein Beispiel für die Erweiterung des deduktiv-nomologischen Erklärungsschemas zu einer genetischen Erklärung eines historischen Ablaufs.
- 4. Kritisieren Sie den "hermeneutischen Positivismus" (Albert und Esser) und zeigen Sie den Hypothesencharakter von Deutungen an einem Beispiel auf. Erläutern Sie in diesem Zusammenhang die in der interpretativen Soziologie (Giddens) hervorgehobene These von der doppelten Hermeneutik in den Sozialwissenschaften.
- 5. Erläutern und kritisieren Sie die von Glaser und Strauss entwickelte Methode der gegenstandsbezogenen Theoriebildung. Inwieweit liegt dieser Kozeptionalisierung des Entdeckungszusammenhangs ein "induktivistisches Selbstmißverständnis" (Kelle) zugrunde?
- 6. Erläutern Sie die Konzepte von "Offenheit" und "Kommunikativität" (Hoffmann-Riem) [oder alternativ von "Gegenstandsangemessenheit", "Vielschichtigkeit der Perspektiven" und "Reflexivität" (Flick)] als Charakteristika qualitativer Sozialforschung.
  - Konkretisieren Sie die sich daraus ergebende Vorgehensweise der qualitativen Sozialforschung an einem Beispiel aus der Marktforschung
- 7. Inwieweit unterliegt der Datenerhebungsprozeß in der qualitativen Sozialforschung systematischen Fehlerquellen? Gehen Sie insbesondere auf die Problematik "Hervorlocken oder Aushandeln" beim narrativen Interview in der Biographieforschung und die Allgegenwärtigkeit von sozialen Einflußfaktoren b ei der Gruppendiskussion ein.
- 8. Inwieweit besteht die Gefahr einer sich selbst erfüllenden Prophezeiung auch bzw. gerade in der qualitativen Sozialforschung?
- 9. Welche Möglichkeiten der Qualitätskontrolle durch Gütekriterien (Objektivität, Reliabilität, Validität,...) bestehen in der qualitativen Sozialforschung?

### IV.2 Querschnittsdesign

- 1. Welche Bedingungen müssen erfüllt sein, um in einer empirischen Untersuchung eine Kausalbeziehung zwischen zwei Variablen nachweisen zu können?
- 2. Beurteilen Sie verschiedene Untersuchungsformen hinsichtlich der Möglichkeit Kausalbeziehungen festzustellen. Geben Sie zur Erläuterung jeweils ein Beispiel.
- Geben Sie eine Typologie aller möglichen Konstellationen bei der Drittvariablenkontrolle für den Fall der Pfadanalyse.
- 4. In einer hypothetischen Untersuchung wurden für die Variablen Alter (A), Bildung (B) und postmaterialistische Einstellung (PM) die folgende Korrelationsmatrix ermittelt:

	A	В	PM
A	*		
В	-0.3	*	
PM	- 0.37	0.475	*

Stellen Sie eine Kausalhypothese auf, bestimmen Sie die Koeffizienten des Kausalmodells und führen Sie eine Effektzerlegung durch.

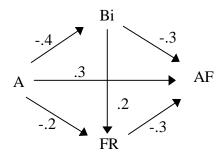
5. Gegeben sei die folgende Korrelationsmatrix (B: Bildung; F: Freundschaftskontakte; AF: Ausländerfeindlichkeit):

	В	F	AF
В	*		
F	0.167	*	
AF	-0.262	-0.370	*

- a. Berechnen und interpretieren Sie zunächst das Kausalmodell mit Ausländerfeindlichkeit als abhängiger Variable.
- b. Kehren Sie die Kausalbeziehung zwischen Ausländerfeindlichkeit und Freundschaftskontakten um, indem Sie nun eine kausale Abhängigkeit der Freundschaftskontakte von der Ausländerfeindlichkeit unterstellen. Berechnen Sie auf der Grundlage der angegebenen Korrelationsmatrix erneut die Pfadkoeffizienten und interpretieren Sie das Ergebnis!

[Zur linearen Regressionsanalyse (Pfadanalyse) mit SPSS für Windows gibt es eine Beispieldatei (Acrobat). Diese Datei kann über das WWW bezogen werden]

6. Gegeben ist das folgende Pfadmodell:



Berechnen Sie die dem Pfadmodell zugrundeliegende Korrelationsmatrix!

- 7. Geben Sie eine Typologie aller möglichen Konstellationen bei der Drittvariablenkontrolle für den Fall der Verfeinerungsanalyse.
- 8. In einer hypothetischen Untersuchung wurden für die Variablen Alter (A), Bildung (B) und postmaterialistische Einstellung (PM) folgende Häufigkeitstabelle ermittelt:

		PM	
A	В	-	+
-	-	24	16
-	+	12	48
+	-	42	18
+	+	20	20

Stellen Sie eine Kausalhypothese auf und überprüfen Sie die Hypothese mit Hilfe des Verfahrens der Verfeinerungsanalyse.

[Zur Tabellenanalyse mit SPSS für Windows gibt es eine Beispieldatei (Acrobat). Diese Datei kann über das WWW bezogen werden]

### weitere Übungsaufgaben:

- Zeigen Sie Möglichkeiten und Grenzen einer Kausalanalyse im Rahmen einer repräsentativen Querschnittserhebung (Survey-Design) auf.
  - a. Geben Sie eine Explikation des Begriffs "Kausalität" und gehen Sie dabei auch auf die wissenschaftstheoretische Grundlegung ein. Unter welchen Bedingungen kann von einer Korrelation zwischen den Variablen x und y auf eine Kausalbeziehung geschlossen werden?
  - b. Wie kann der Einfluß einer dritten Variablen in einem linearen Kausalmodell explizit berücksichtigt werden? Welche Formen von Drittvariableneffekten können unterschieden werden? Inwieweit ist es möglich, mit Hilfe der Drittvariablenkontrolle die Gültigkeit einer Kausalhypothese abzusichern?
  - c. Nehmen Sie an, Sie hätten in einer Umfrage das Alter (A), das Bildungsniveau (B) und das Umweltbewußtsein (U) erhoben und folgende Korrelationsmatrix erhalten:

	A	В	U
A	1.00		
В	30	1.00	
U	25	.53	1.00

Entwickeln Sie eine Kausalhypothese, entwerfen Sie ein Kausalmodell, bestimmen Sie die Kausalkoeffizienten und führen Sie eine Effektzerlegung durch. Interpretieren Sie das Ergebnis und beurteilen Sie die interne Validität der Untersuchung, insbesondere in Hinblick auf nicht berücksichtigte Variablen und andere Störfaktoren.

Stellen Sie die Grundlagen der Kausalanalyse von Querschnittsdaten dar. Gehen Sie dabei insbesondere auf Kausalitätskriterien und die verschiedenen bei der Drittvariablenkontrolle möglichen Konstellationen ein.

Konkretisieren Sie Ihre Überlegungen an dem folgenden Beispiel, das in einer grob vereinfachten Form den tatsächlichen Beziehungen zwischen den behandelten Variablen in einer repräsentativen Querschnittsbefragung in Gesamtdeutschland im Jahre 1992 (ALLBUS 1992) nahekommt:

BI = Bildung (hoch = positiv gepolt)

KB = Kirchenbindung (hoch = positiv gepolt)

nGRO = normative Geschlechtsrollenorientierung (liberal = positiv gepolt)

Es ergeben sich die folgenden Korrelationsmatrizen (idealisierte Daten):

#### Befragte in den alten Bundesländern:

	BI	KB	nGRO
BI	1.00		
KB	20	1.00	
nGRO	+.54	30	1.00

### Befragte in den neuen Bundesländern:

	BI	KB	nGRO
BI	1.00		
KB	.20	1.00	
nGRO	.24	24	1.00

Stellen Sie jeweils eine Kausalhypothese auf und begründen Sie Ihre Entscheidung. (Wählen Sie dabei nGRO als abhängige Variable.) Berechnen Sie die Pfadkoeffizienten der beiden Kausalmodelle und interpretieren Sie die Ergebnisse. Ist ein Test des Kausalmodells möglich?

Erläutern Sie die Vorgehensweise bei der Kausalanalyse von Querschnittsdaten. Gehen Sie dabei insbesondere auf Kausalitätskrterien und die Möglichkeiten der Drittvariablenkontrolle ein. Analysieren Sie als Beispiel die folgende Tabelle der drei dichotomosieten Variablen Alter (-:jung, +:alt), Bildung (-:niedrig, +:hoch) und Karriereorientierung (-:niedrig, +:hoch) (hypothetische Daten aus einer Befragung berufstätiger Frauen):

Alter	Bildung	Karriereor.	Fallzahl
-	-	-	56
-	-	+	14
-	+	-	52
-	+	+	78
+	-	-	117
+	-	+	13
+	+	-	49
+	+	+	21

Stellen Sie eine Kausalhypothese auf (berücksichtigen Sie dabei auch die Möglichkeit eines Interaktionseffektes) und überprüfen Sie die Hypothese. Wie ist das Ergebnis zu interpretieren und wie beurteilen Sie die Gültigkeit Ihrer Schlußfolgerungen?

In einer Untersuchung von Frauen im Alter zwischen 30 und 40 Jahren wurden die folgenden Variablen erhoben:

Familienstand (F): (1) ledig; (v) verheiratet

Kinderzahl (K): (o) ohne Kinder; (m) mindestens 1 Kind

Berufstätigkeit (B): (+) ja; (-) nein

Es ergaben sich die folgenden (hypthetischen) Daten:

		I	3
F	K	-	+
1	0	30	270
1	m	70	130
v	0	90	210
v	m	560	140

Diskutieren Sie mögliche Zusammenhänge zwischen den Variablen vor dem Hintergrund Ihres Alltagswissens und formulieren Sie die entsprechenden Hypothesen.

Untersuchen Sie zunächst alle bivariaten Zusammenhänge. Erläutern Sie dann zunächst abstrakt die Vorgehensweise bei der Drittvariablenkontrolle (Verfeinerungsanalyse).

Welche Konstellationen können dabei auftreten? Führen Sie dann die Verfeinerungsanalyse für mindestens eine bivariate Beziehung durch. Interpretieren Sie zusammenfassend das Resultat Ihrer Analysen, indem Sie die Erwerbsquote der ledigen Frauen ohne Kinder als Basiskategorie verwenden.

### **IV.3 Experiment**

- 1. Welche Formen der Wechselwirkung zwischen Pretest, experimentellem Stimulus und unkontrollierten Ereignissen können im Experiment auftreten? Erläutern Sie die sich daraus ergebenden Probleme anhand eines Feldexperiments, das die Werbewirksamkeit einer Werbekampagne für ein bestimmtes Produkt feststellen soll.
- 2. Welche Störfaktoren sind im Falle des Experiments zu berücksichtigen? Geben Sie für jeden Störfaktor ein Beispiel.
  - Versuchen Sie Ihre Überlegungen auf die anderen Untersuchungsformen zu übertragen. Welche zusätzlichen Probleme bei der Kontrolle von Störfaktoren treten dabei auf?
- 3. Wie sind die interne und externe Validität bei den verschiedenen Untersuchungsformen, insbesondere bei Feld- und Laborexperimenten zu beurteilen?
- 4. Eine Supermarktkette möchte die Wirkung verschiedener Formen der Warendarbietung auf die Absatzmenge von Schokolade prüfen. Als unabhängige Variablen kommen z.B. die Regalhöhe und die Regalbreite in Betracht. Entwerfen Sie eine einfaktorielle und eine zweifaktorielle experimentelle Untersuchungsanordnung.
  - Erscheint Ihnen in diesem Zusammenhang eine Blockanordnung nützlich? Könnten möglicherweise Carry-Over-Effekte auftreten? Welche weiteren Störfaktoren müssen berücksichtigt werden und inwieweit sind sie im Experiment zu kontrollieren?
- 5. Die Elektrizitätswirtschaft plant eine Werbekampagne, um die Akzeptanz der Bürger für die Kernenergie zu verbessern. Die Werbung informiert sachlich über die großen Anstrengungen, die Sicherheitsvorkehrungen im Bundesdeutschen AKW's zu optimieren und weist gleichzeitig auf die mit Kohlekraftwerken verbundenen Umweltrisiken (Treibhauseffekt) hin. In einer Voruntersuchung sollen Sie die Wirksamkeit der Werbemaßnahme testen. Entwerfen Sie eine experimentelle Untersuchungsanordnung und diskutieren Sie ausführlich möglicherweise auftretende Störfaktoren. Untersuchen Sie in diesem Zusammenhang auch die Nützlichkeit von Drei- und Vier-Gruppen-Anordnungen.

Ist es notwendig, einen Pretest durchzuführen? Diskutieren Sie Vor- und Nachteile. Beschreiben Sie ausführlich die Vorgehensweise bei der Durchführung des Experiments bezogen auf die von Ihnen schließlich gewählte Untersuchungsanordnung. Wie beurteilen Sie abschließend die interne und externe Validität des Experiments?

#### weitere Übungsaufgaben:

Diskutieren Sie die Eignung verschiedener Untersuchungsformen zur Überprüfung von Kausalhypothesen am Beispiel der Querschnittsuntersuchung und des Experiments.

- Stellen Sie zunächst den logischen Aufbau der beiden Untersuchungsdesigns dar und diskutieren Sie dann vergleichend Stärken und Schwächen der beiden Designs vor dem Hintergrund eines konkreten, von Ihnen zu konstruierenden Anwendungsfalls. Wie ist die interne Validität der beiden Untersuchungsformen im Vergleich zu beurteilen?
- Stellen Sie die beiden experimentellen Untersuchungsanordnungen PEUT-PUT und EUT-UT jeweils an einem Beispiel dar und erörtern Sie, unter welchen Bedingungen auf einen Pretest verzichtet werden kann.
- Worin besteht der Unterschied zwischen einer Panel-Anordnung und einem echten Experiment bei der Untersuchung von Verhaltensänderungen? Erläutern Sie Ihre Überlegungen an einem Beispiel. Welches Verfahren würden Sie wählen, wenn Sie eine Kausalhypothese zur Verursachung von Verhaltensänderungen überprüfen sollten?
- Eine Umweltorganisation möchte das Umweltbewußtsein von Hausfrauen mit Hilfe einer gezielten Informations- und Aufklärungsaktion verbessern. Es ist geplant, Veranstaltungen durchzuführen, in denen u.a. ein Film vorgeführt werden soll, der über mögliche Umweltgefährdungen im Haushalt (Verwendung umweltbelastender Produkte, anfallender Hausmüll usw.) informiert und auf umweltschonende Alternativen hinweist. Zu diesem Zweck stehen zwei Filme zur Auswahl: ein eher sachlich informierender und ein stärker an Emotionen appellierender Film. In einer Voruntersuchung soll experimentell geprüft werden, welcher der beiden Filme eher geeignet ist, das Umweltbewußtsein zu verändern.
  - a. Diskutieren Sie mögliche experimentelle Untersuchungsanordnungen in Hinblick auf eventuell auftretende Störfaktoren. Welche Anordnung erscheint Ihnen auch unter Kostengesichtspunkten am geeignetesten?
  - b. Skizzieren Sie kurz die Vorgehensweise bei der Durchführung des Experiments. Gehen Sie dabei insbesondere auf die Bestimmung der Experimental- und (eventuell) Kontrollgruppe ein.
  - c. Wie ist die externe Validität des Experiments zu beurteilen?

### IV.4 Komplexe Untersuchungsformen (folgt im Sommersemester)

### V. Theoretische Konstrukte und Indikatoren

- 1. Erläutern Sie die Vorgehensweise bei der Explikation eines Begriffes an einem Beispiel (z.B. soziale Schicht, Intelligenz, habitualisiertes Kaufverhalten, Kaufsucht).
- 2. Welche Lösung schlägt der Operationalismus für das Problem der Zuordnung von Indikatoren zu theoretischen Konstrukten vor? Kritisieren Sie diese Lösung.
- 3. Wie versucht der kausalanalytische Ansatz mit dem gleichen Problem fertig zu werden? Worin bestehen die Vorteile dieser Lösung und welche neuen Probleme handelt man sich dadurch ein?

- 4. Suchen Sie zwei Indikatoren, mit denen Sie das theoretische Konstrukt "Kaufsucht" (η) messen wollen und stellen Sie eine einfache Kausalhypothese in der Form "ξ wirkt auf η"auf. Suchen Sie dazu ein geeignetes theoretisches Konstrukt ξ von dem Sie annehmen, daß es kausal auf die Kaufsucht wirkt und bestimmen Sie ebenfalls zwei Indikatoren für das theoretische Konstrukt.
- 5. In einer Befragung von Frauen wurden als Indikatoren für politische Partizipation die Variablen politisches Interesse (PI) und Mitgliedschaft in einer politischen Partei (PM) erhoben; die nGRO wurde mit den Statements "Frauen sollten im Beruf nicht zu strebsam sein" (B) und "In der Familie sollten die meisten Entscheidungen vom Mann getroffen werden" (E) erfaßt.
  - a. Formulieren Sie eine Kausalhypothese und formalisieren Sie Ihre Hyptohese gemäß dem kausalanalytischen Ansatz als Kausaldiagramm und als Gleichungssystem.
  - b. Welche Voraussetzungen müssen für die Lösbarkeit des Gleichungssystems gemacht werden?
  - c. Berechnen Sie die Koeffizienten des Modells auf der Grundlage der folgenden Korrelationsmatrix (nächste Seite):

	PI	PM	В	Е
PI	*			
PM	.24	*		
В	18	12	*	
Е	24	16	.48	*

- d. Wie ist die Reliabilität der Indikatoren für die politische Partizipation und für die nGRO zu beurteilen?
- e. Was kann über die Validität der Messung der beiden Konstrukte gesagt werden? Welche Korrelation besteht zwischen den beiden latenten Konstrukten?
- f. Erläutern Sie dazu auch den Begriff der Minderungskorrektur.

## weitere Übungsaufgaben:

- In der empirischen Wirtschafts- und Sozialforschung hat man es häufig, wenn nicht überwiegend, mit Begriffen ohne direkten empirischen Bezug zu tun. In solchen Fällen ist es notwendig, die theoretischen Konstrukte durch Indikatoren zu operationalisieren.
  - a. Diskutieren Sie verschiedene Möglichkeiten, die Beziehung zwischen theoretischen Konstrukten und Indikatoren zu konzeptualisieren auch im Lichte der wissenschaftstheoretischen Überlegungen.
  - b. Konkretesieren Sie das unter a. Gesagte für das theoretische Konstrukt "Umweltbewußtsein". Geben Sie zunächst eine Explikation des Begriffs (im Sinne von Einstellungen zur Umweltproblematik und verschieden von umweltbezogenem Wissen und

umweltbezogenem Handeln). Modellieren Sie die Beziehung zwischen Umweltbewußtsein und umweltbewußtem Handeln (jeweils durch mindestens zwei Indikatoren Ihrer Wahl gemessen) als Kausaldiagramm und erläutern Sie, welche Teile des Modells als Kern- und Meßtheorie fungieren.

- c. Welche Annahmen über mögliche Meßfehler müssen gemacht werden, damit die Kausalkoeffizienten des Modells identifiziert werden können?
- Entwickeln Sie in Anknüpfung an die Überlegungen zum methodologischen Falsifikationismus die im kausalanalytsichen Ansatz gefundene Lösung des Korrespondenzproblems und stellen Sie die formalen Eigenschaften des zugehörigen Kausalmodells dar. Konkretisieren Sie Ihre Überlegungen anhand des folgenden (hypothetischen) Beispiels:

In einer Untersuchung zum Konsumentenverhalten soll der Zusammenhang zwischen Kaufsucht und bestimmten Persönlichkeitseigenschaften überprüft werden. Kaufsucht wurde durch die beiden Indikatoren Verlangen (V) (Wenn ich durch ein Kaufhaus gehe, fühle ich ein starkes Verlangen, etwas zu kaufen) und Impuls (I) (Manchmal sehe ich etwas und fühle den unwiderstehlichen Impuls, es zu kaufen) erhoben. Als Persönlichkeitsmerkmal wurde u.a. das Selbstwertgefühl durch die Indikatoren Zufriedenheit (Z) (Eigentlich bin ich mit mir ganz zufrieden) und Achtung (A) (Ich wollte, ich könnte mehr Achtung vor mir haben) erhoben. Die Zustimmung zu diesen Items wurde jeweils auf einer 5-stufigen Ratingskala mit den Endpunkten 1 (trifft überhaupt nicht zu) und 5 (trifft voll und ganz zu) erfaßt. Es ergab sich folgende Korrelationsmatrix:

	V	I	Z	A
V	*			
I	.48	*		
Z	20	15	*	
A	.24	.18	30	*

Stellen Sie zunächst eine Kausalhypothese auf und spezifizieren Sie das zugehörige Kausalmodell. Berechnen Sie dann die Koeffizienten des Modells und interpretieren Sie das Ergebnis. Was kann über die Qualität der einzelnen Indikatoren ausgesagt werden? Vergleichen Sie auch die Korrelation zwischen den theoretischen Konstrukten mit den entsprechenden empirischen Beziehungen zwischen den zugehörigen Indikatoren.

In einer repräsentativen Bevölkerungsbefragung sollte u.a. die Beziehung zwischen sozialem Status und liberalen Erziehungszielen untersucht werden. Für den sozialen Status wurden als Indiaktoren Bildung (X<sub>1</sub>) und Einkommen (X<sub>2</sub>) erhoben; in Hinblick auf den Erziehungsstil wurde u.a. die Wichtigkeit folgender Erziehungsziele erfragt: Selbständigkeit (Y<sub>1</sub>) und Pünktlichkeit (Y<sub>2</sub>). Es ergab sich folgende Korrelationsmatrix (hypothetische Daten):

	$X_{\scriptscriptstyle 1}$	$X_2$	$\mathbf{Y}_{\scriptscriptstyle 1}$	$Y_2$
$X_1$	*			
$X_2$	.24	*		
$\mathbf{Y}_{1}$	.24	.16	*	
$\mathbf{Y}_2$	18	12	48	*

Stellen Sie eine Kausalhypothese auf, entwickeln Sie das zugehörige Kausalmodell und bestimmen Sie die Kausalkoeffizienten des Modells. Welche Voraussetzungen müssen dabei gemacht werden? Interpretieren Sie das Ergebnis in Hinblick auf die in Kern- und Meßtheorie formulierten Hypothesen. Wie ist die Reliabilität der Indikatoren zu beurteilen? Erläutern Sie in diesem Zusammenhang auch den Begriff der Minderungskorrektur.

### VI. Messen

### VI.1 Grundlagen

- 1. Worin besteht der Unterschied zwischen einer Repräsentations- und einer Indexmessung? Konkretisieren Sie den Unterschied am Beispiel der Messung von Markenpräferenzen.
- 2. Erläutern Sie die Unterschiede zwischen den verschiedenen Skalentypen. Was versteht man in diesem Zusammenhang unter dem Repräsentations-, dem Eindeutigkeits- und dem Bedeutsamkeitsproblem?
- 3. Nehmen Sie an, für vier verschiedene empirische Objekte seien die Werte 3, 7, 18 und 19 gemessen worden. Nennen Sie für jedes Skalenniveau vier transformierte Meßwerte, die in Bezug auf das angenommene Skalenniveau aus einer erlaubten Transformation hervorgehen.
- 4. Erläutern Sie das Bedeutsamkeitsproblem, in dem sie den Zusammenhang zwischen dem Skalenniveau einer Messung und erlaubten Verfahren zur Bestimmung der Beziehungen zwischen zwei Variablen (Korrelationskoeffizienten) darstellen.

## VI.2 Gebräuchliche Skalierungsverfahren

- 1. Stellen Sie die Vorgehensweise bei der Konstruktion einer Likert-Skala am Beispiel der Messung von Kaufsucht dar. Welches Meßniveau wird erreicht? Liegt Repräsentationsoder Indexmessung vor?
- 2. Versuchen Sie die Einstellung gegenüber alten Leuten als Guttman-Skala zu konzipieren. Welche Annahmen müssen dabei gemacht werden und welche Möglichkeiten zur Überprüfung des Meßmodells gibt es?

#### 3. Semantisches Differential:

- a. Stellen Sie die Grundidee des Semantischen Differentials dar und geben Sie dazu ein Beispiel aus der Marktforschungspraxis. Welche Auswertungsmöglichkeiten ergeben sich aus der Struktur der erhaltenen Daten (Datenquader)?
- b. Erläutern Sie die Vorgehensweise bei der Konstruktion eines Semantischen Differentials.
- c. Wie geht man bei der individuellen bzw. segmentspezifischen Konstruktion von Imageräumen auf der Grundlage eines Semantischen Differentials vor?
- d. Handelt es sich bei dem Semantischen Differential um Repräsentations- oder Indexmessung? Welches Meßniveau wird erreicht?

#### VI.3 Klassische Testtheorie

- 1. Auf welchem Grundgedanken baut die Klassische Testtheorie auf und welche formalen Annahmen müssen im einzelnen gemacht werden?
- Diskutieren Sie verschiedene Möglichkeiten zur Bestimmung der Reliabilität einer Skala.
- 3. Erläutern Sie die Strategie der Reliabilitätsverbesserung durch Testverlängerung. Wieviel Items mit einer Reliabilität von 0,2 werden für einen Test mit einer Reliabilität von 0,9 benötigt?
- 4. Welche Möglichkeiten gibt es, die Validität einer Skala zu beurteilen?
- 5. In einer Untersuchung wurde das theoretische Konstrukt "Ausländerfeindlichkeit" durch die folgenden vier Indikatoren erfaßt:
  - Y<sub>1</sub>: Wenn Arbeitsplätze knapp werden, sollte man die Gastarbeiter wieder in ihre Heimat zurückschicken.
  - Y<sub>2</sub>: Man sollte Gastarbeitern jede politische Tätigkeit in Deutschland untersagen.
  - Y<sub>3</sub>: Gastarbeiter sollten sich ihre Ehepartner unter ihren eigenen Landsleuten wählen.
  - Y<sub>4</sub>: Gastarbeiter sollten ihren Lebensstil ein bißchen besser an den der Deutschen anpassen.

Es ergab sich folgende Korrelationsmatrix:

	$\mathbf{Y}_1$	$\mathbf{Y}_2$	$\mathbf{Y}_3$	$Y_4$
Yı	*	.56	.56	.48
$Y_2$	.56	*	.49	.42
$Y_3$	.56	.49	*	.42
Y <sub>4</sub>	.48	.42	.42	*

- a. Bilden Sie gemäß der Logik der Likert-Skala ein Index für Ausländerfeindlichkeit aus den Variablen Y<sub>1</sub>-Y<sub>4</sub> und berechnen Sie die Reliabilität mit Hilfe der Spearman-Brown-Formel.
- b. Berechnen Sie die Testgleichungen für das einfaktorielle Meßmodell mit den vier Variablen Y<sub>1</sub>-Y<sub>4</sub>.
- c. Berechnen Sie die Koeffizienten des Modells aus den angegebenen Korrelationen.
- d. Interpretieren Sie Ihre Ergebnisse.

### VI.4 Faktorenanalyse und multidimensionale Skalierung

- 1. Stellen Sie Zielsetzung und Anwendungsvoraussetzungen der Faktorenanalyse grundsätzlich und an einem Beispiel dar. Diskutieren Sie die Grundlagen des Modells und geben Sie eine kausalanalytische Interpretation. Was ist der Unterschied zwischen einer exploratorischen und einer konfirmatorischen Faktorenanalyse?
- 2. Erläutern Sie das Verfahren der Faktorenextraktion in der Hauptkomponentenanalyse. Geben Sie eine varianzanalytische Interpretation der Eigenwertzerlegung. Wie kann die Anzahl der zu extrahierenden Faktoren bestimmt werden?
- 3. Wie können die Faktoren inhaltlich interpretiert werden? Welche Bedeutung haben in diesem Zusammenhang die rotierten Faktorladungen? Wie können Reliabilität und Validität der erhaltenen Faktoren beurteilt werden?
- 4. Diskutieren Sie das Problem der Kommunalitätenschätzung und erläutern Sie in diesem Zusammenhang auch die Hauptachsenmethode.
- 5. Erläutern Sie den Grundgedanken der Multidimensionalen Skalierung. Geben Sie auch ein Anwendungsbeispiel.
- 6. Stellen Sie die Vorgehensweise bei der Konstruktion einer dimensionalen Repräsentation in der MDS in den Grundzügen dar. Was ist der Unterschied zwischen einer ordinalen und einer metrischen MDS?
- 7. Wie kann die Dimensionalität des Lösungsraumes beurteilt werden? Handelt es sich bei der MDS um Repräsentations- oder Indexmessung? Welches Meßniveau wird erreicht?
- 8. Wie kann eine MDS-Lösung interpretiert werden? Gehen Sie in diesem Zusammenhang auch auf das Vektor- und Idealpunkt-Modell ein.

## weitere Übungsaufgaben:

Erläutern Sie den Unterschied zwischen einer Repräsentations- und einer Indexmessung und geben Sie dazu ein Beispiel. Welche unterschiedlichen Vorstellungen von Messen liegen den beiden Verfahren zugrunde? Worin besteht die Überlegenheit der Repräsentationsmessung, und warum wird in der Praxis dennoch überwiegend die Indexmessung verwandt?

- Unter Messen wird gemäß einer verbreiteten Definition die Zuordnung von Zahlen zu Objekten gemäß festgelegter Regeln verstanden. Zeigen Sie die Unvollständigkeit dieser Definition auf, indem Sie das Repräsentationsproblem diskutieren und geben Sie ein Beispiel für eine Repräsentationsmessung. Was versteht man unter dem Eindeutigkeitsproblem und wie ergeben sich daraus die unterschiedlichen Meßniveaus? Erläutern Sie auch das Bedeutsamkeitsproblem, indem Sie zeigen, daß die Mittelwertbildung kein zulässiges statistisches Verfahren für Ordinalskalen darstellt; geben Sie auch dazu ein Beispiel.
- Erläutern Sie die grundlegenden Annahmen der Klassischen Testtheorie (KT) sowohl inhaltlich wie auch formal; geben Sie dazu ein Beispiel. Welche Möglichkeiten zur Bestimmung der Reliabilität von Skalen ergeben sich daraus? Stellen Sie mindestens zwei unterschiedliche Vorgehensweisen zur Reliabilitätsbestimmung jeweils an einem Beispiel dar. Kann im Rahmen der KT auch die Validität von Skalen überprüft werden?
- Erläutern Sie die verschiedenen Gütekriterien zur Beurteilung einer Messung. Verdeutlichen Sie Ihre Argumentation, indem Sie am Beispiel einer hypothetischen Likert-Skala zur normativen Geschlechtsrollenorientierung zeigen, welche Möglichkeiten zur Überprüfung der einzelnen Gütetkriterien bestehen. Schildern Sie dabei möglichst konkret die methodische Vorgehensweise bei der Überprüfung der Gütekriterien und zeigen Sie die Schwachstellen der einzelnen Verfahren auf.
- Diskutieren Sie unterschiedliche Modelle der Skalenbildung am Beispiel der Messung der Einstellung zum Schwangerschaftsabbruch. Skizzieren Sie an diesem Beispiel die Vorgehensweise bei der Konstruktion einer Likert-Skala und einer Guttman-Skala. Worin sehen Sie die Vor- und Nachteile der beiden Skalierungsmethdoen im Vergleich? Erläutern Sie dazu auch auf abstrakter Ebene den grundlegenden Unterschied im Verständnis von Messen zwischen den beiden Verfahren.

## VII. Befragung

- 1. Stellen Sie die Stärken und Schwächen unterschiedlicher Befragungsformen vergleichend gegenüber und geben Sie für jede Befragungsform ein Forschungsproblem an, bei dem der Einsatz der entsprechenden Befragungsmethode angezeigt ist.
- 2. Diskutieren Sie unterschiedliche Formen von Antwortverzerrungen im Interview und geben Sie für die verschiedenen Möglichkeiten jeweils ein Beispiel.
- 3. Welche Möglichkeiten zur Erklärung von Antwortverzerrungen im Sinne einer Intsrumtentheorie sehen Sie? Welche Empfehlungen ergeben sich daraus für die Vermeidung von Artefakten in Befragungssituationen?
- 4. Beurteilen Sie die Anfälligkeit unterschiedlicher Befragungsformen für Antwortverzerrungen.

(wird im Sommersemester erweitert)

# VIII. Auswahlverfahren

- 1. Worin besteht der Unterschied zwischen einer geschichteten Zufallsstichprobe und einer Quotenauswahl? Geben Sie dazu ein Beispiel.
- 2. Stellen Sie vergleichend Vor- und Nachteile von Zufallsstichprobe und Quotenauswahl gegenüber.