

6. Datenerhebungsverfahren

Bei den Datenerhebungsverfahren unterscheiden wir grundsätzlich zwischen Befragung, Beobachtung und Inhaltsanalyse. 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 Subsumierung 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 Analyse 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 qualitativen oder quantitativen 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 Veränderungen bei den Untersuchungsobjekten 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.

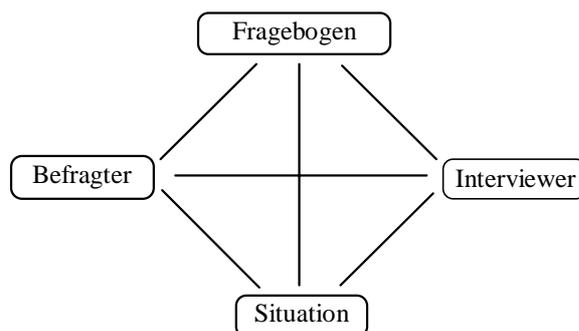
6.1 Befragung

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 soziale Interaktion 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 Definition der Interviewsituation 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):

| Situation | Alte Führungs- schicht | Neue Manager- schicht | Aufstr. Mittel- schicht | Ruhige Mittel- schicht | Aufstr. Arbeiter- schicht | Trad. Arbeiter- schicht | Arme 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 |
| N | 100 (75) | 100 (72) | 100 (293) | 100 (775) | 100 (216) | 100 (299) | 100 (104) |

Ein weiteres bedeutsames Merkmal der Interviewsituation, das zu Veränderungen des Antwortverhalten führen kann, stellt die Anwesenheit Dritter (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.

Neben den bisher diskutierten 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 tiefere 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 Kunstlehre (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 Fremden 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):¹

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?

¹ 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:

- Eigenschaftsfragen (Fragen nach personalen und soziodemographischen Eigenschaften des Befragten, wie z.B. Alter, Bildung, Familienstand, Konfession, usw.)
- Verhaltensfragen (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.)
- Überzeugungsfragen (Fragen nach dem, was der Befragte für wahr oder falsch hält, die sich also auf das „Wissen“ bzw. auf das „Weltbild“ des Befragten 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.)
- Einstellungs- oder Meinungsfragen (Fragen nach der positiven oder negativen Beurteilung eines bestimmten Tatbestands durch den Befragten, 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 Sensibilität 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 der Befragten) und zum anderen die erforderlichen 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 Konsumententscheidung).

Allerdings zeigt gerade die Konsumforschung, daß die Befragten oft nicht in der Lage sind, auf eine direkte Frage nach den Gründen 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 Nescafé 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.)

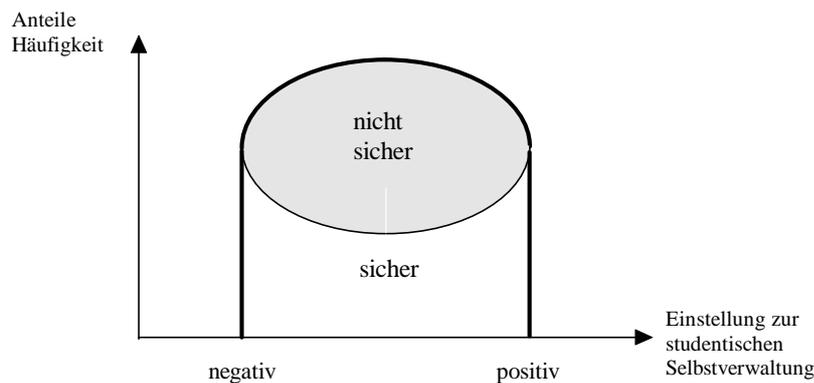
Eines der wichtigsten, aber zugleich auch schwierigsten Probleme bei der Befragung ist die Behandlung von Meinungslosigkeit. Grundsätzlich muß davon ausgegangen werden, daß sich nicht jeder Befragte zu jedem Thema eine Meinung gebildet hat. Die oft geübte Praxis, keine explizite „Weiß nicht“(WN)-Kategorie 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 annähernd 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.

Sensitive Fragen: 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 Trichterung. 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 Aufbau des Fragebogens 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 von 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. Entsprechend der Mean-Shift-Hypothese 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 Antwortverzerrungen 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 Instrumententheorie der Befragung 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 Charakteristika 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 Zustimmungsrates 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 Persönlichkeitsmerkmale, 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 $\xrightarrow{-}$ Ja-Sage-Tendenz $\xrightarrow{+}$ 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 einer 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 Interviewsituation, insbesondere an sozialen Merkmalen des Interviewers; aber auch die Anwesenheit Dritter oder der (erkennbare) Sponsor einer Untersuchung können zur Definition der Be-

fragungssituation 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 und 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 Befragten | ♂ | 3.27 | 3.55 |
| | ♀ | 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 Interviewereffekts, nämlich eine ausreichende kognitive Fähigkeit, um die vom Interviewer ausgehenden sozialen Signale interpretieren zu können, und eine ausgeprägte soziale Anpassungsfähigkeit, bei den mittleren

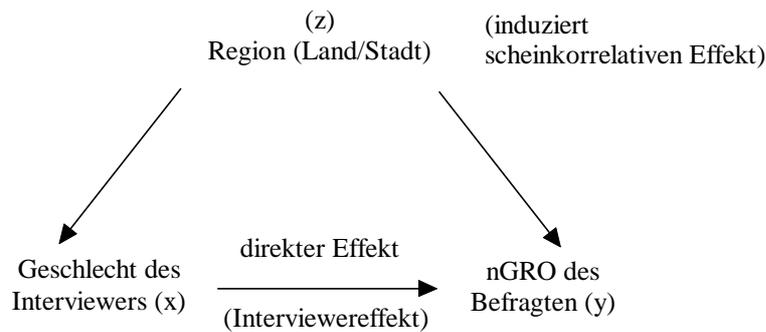
und oberen sozialen Schichten weniger ausgeprägt sein sollten. Es ist eher zu vermuten, daß bei einer geeigneten Thematik eine solche situationspezifische Form der Eigendarstellung auch und gerade bei den gehobenen sozialen Schichten zu beobachten sein sollte. Tatsächlich sind solche Interviewereffekte 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 | | |
|---------|---------|-------|--------|------|
| | | jung | mittel | alt |
| Bildung | niedrig | - 2% | 0% | + 9% |
| | hoch | + 31% | + 30% | - 2% |

Wie ersichtlich ist, zeigt sich ein deutlicher Interviewereffekt bei den Befragten mit hoher 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 Unempfindlichkeit von Beziehungsmustern (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 Kurzsкала 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 Anerkennung in der Interviewsituation gegeben (Interviewereffekt). Welche der beiden Tendenzen überwiegt, kann nur durch eine konkrete Analyse der Kosten- und 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 **Telefoninterview** 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 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 beobachtenden 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 des Beobachtungssystems dar. Auf jeden Fall ist eine Reihe von formalen und inhaltlichen 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 (Verkodung 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 nicht-sprachliche 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 entstammen, wenn also unterstellt werden kann, daß der 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 Textmaterials: 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 subsumiert 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 computergestützten 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 Macht (autoritär bzw. kontrolliert) und Individuum (irrational bzw. rational handelndes), die verwendeten Relationen zwischen den Subjekten, nämlich Macht dominiert das Verhältnis bzw. ausgeglichenes Verhältnis zwischen Macht und Individuum und schließlich die Bewertungskategorien positiv und negativ. 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) herauschälendes 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 Überschriften versehen werden. Anschließend werden Passagen, in denen gleiche oder ähnliche Themen behandelt werden, zusammengestellt. Eine Hauptüberschrift, 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 thematische Vergleich über die verschiedenen Interviews hinweg. 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 soziologischen Konzeptualisierung eine

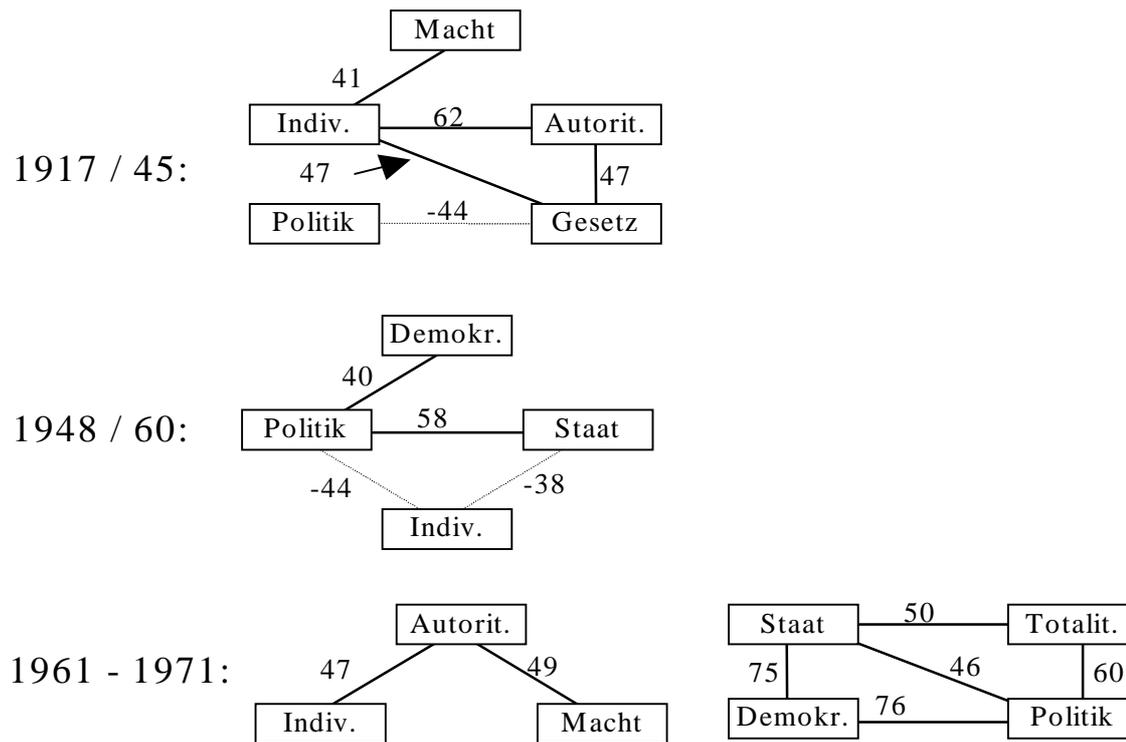
Abwendung von der Terminologie der Interviewten. Die zuvor dem Text entnommenen Begriffe und Überschriften werden nun in soziologische Termini übersetzt, um einen Anschluß der Interpretation an die allgemeine wissenschaftliche Diskussion zu ermöglichen. Das verdichtete Expertenwissen wird quasi im Kontext soziologischer Theorien erneut interpretiert und gleichzeitig theoretisch generalisiert.

Inhaltsanalytische Auswertungstechniken:

Die einfachste Form der Auswertung ist die sog. Frequenzanalyse, in der Textelemente lediglich danach klassifiziert werden, wie häufig die einzelnen 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 Valenzanalysen. 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 syntaktischer Sequenzen von Kategorien dar. Dabei wird in sog. Kontingenzanalysen 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. Weitaus am 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 Bedeutung inhaltsanalytischer Vorgehensweisen in der empirischen Sozialforschung in Zukunft stark zunehmen 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 kontextabhängige 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.

