

Antwortverzerrungen im Interview - wie läßt sich die Güte der Daten verbessern?

Cannell, Charles F.

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Cannell, C. F. (1984). Antwortverzerrungen im Interview - wie läßt sich die Güte der Daten verbessern? *ZUMA Nachrichten*, 8(15), 3-17. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-210414>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Antwortverzerrungen im Interview – Wie läßt sich die Güte der Daten verbessern?

Forschung und Weiterentwicklung von Techniken im Bereich der Datenerhebung haben nicht Schritt gehalten mit der in anderen Bereichen der Umfrageforschung. Für die Erstellung von Stichproben gibt es sowohl ein solides theoretisches Fundament wie eine Reihe von erprobten Prozeduren für die Ziehung von Zufallsstichproben samt Methoden zur Abschätzung von Stichprobenfehlern. Im Bereich der Datenanalyse ist eine beeindruckende Vielfalt von Techniken entwickelt worden, die komplexe Auswertungen erlauben. Mit der Hilfe moderner Computerprogramme können selbst große und kompliziert aufgebaute Datensätze mit wenig Mühe und großer Schnelligkeit ausgewertet werden. Die Techniken der Datenerhebung bleiben dahinter deutlich zurück. Dabei ist gerade diese Phase im Forschungsprozeß besonders wichtig. Bezogen auf den Gesamtfehler, mit dem Umfragedaten behaftet sind, ist der größte Teil schon diesem Stadium zuzurechnen.

Der Grund, warum den Techniken der Datenerhebung wenig Beachtung geschenkt worden ist, liegt vermutlich in ihrer scheinbaren Einfachheit. Einfache Fragen zu stellen und darauf Antworten zu bekommen, ist ein natürlicher Vorgang in der alltäglichen Kommunikation. Wir alle sind Experten im Fragenstellen und Fragenbeantworten. Während man sehr früh in der Geschichte der Meinungsforschung den Problemen von Frageformulierung und einem möglichen Einfluß des Interviewers große Beachtung geschenkt hatte, wurden sie in der Folgezeit vernachlässigt. Erst in jüngster Zeit werden wieder einschlägige Forschungsergebnisse vorgelegt.

Der Hauptteil der Umfrageforschung in den frühen Jahren war auf Einstellungen und Meinungen konzentriert. Da die Genauigkeit der Angaben bei diesen Phänomenen sehr schwierig zu bewerten ist, waren die Forscher geneigt, die Antworten schlicht als gültig zu akzeptieren. Die vermeintliche Einfachheit des Frage- und Antwort-Spiels hat sowohl die Datenanalytiker wie die Erhebungsspezialisten geblendet, sie die spezifischen Fehlerquellen nicht sehen lassen.

Als Umfragen mehr und mehr dazu benutzt wurden, soziale Probleme zu untersuchen und insbesondere dann auch Daten über individuelles Verhalten zu

ZUMA

sammeln, wurden die Probleme der Verzerrungen bei Antworten deutlicher. Im allgemeinen wurden diese Probleme durch die Forscher in der Phase der Datenanalyse erkannt, und da die analysierenden Forscher Spezialisten in bestimmten inhaltlichen Teilgebieten sind, entdeckten sie die Schwierigkeiten auch immer nur als besondere Schwierigkeiten bei besonderen Themen. Daraus resultierte die Tendenz, Techniken zu entwickeln, die bestimmte Meßinstrumente für diese einschlägigen Variablen verbesserten, statt in allgemeinerer Art die Frage der Gültigkeit von Antworten anzugehen. Obwohl diese Bemühungen ganz ohne Zweifel partiell zu erheblichen Verbesserungen geführt haben, haben sie das Gültigkeitsproblem nicht grundlegend in Angriff nehmen können. Die eigentliche Fehlerquelle für Antwortverzerrungen liegt nicht bei bestimmten Variablen, sondern eher bei den generelleren Problemen der Kommunikation im Interview.

Gültigkeitsprobleme bei Umfragedaten stellen sich in vielfältiger Weise. Nicht überraschend sind Schwierigkeiten bei Angaben zu Einkommen und Ersparnissen, zum Wahlverhalten, zum Gebrauch von Alkohol und Drogen oder zu strafbaren Handlungen. Aber auch bei weniger heiklen Fragen weisen die Antworten ähnliche Verzerrungen auf. So lehrt die Erfahrung, daß auch Angaben über das Datum und den Preis von Käufen für den Haushalt, Angaben zu Dauer und Ziel von Reisen, Angaben über den Fernsehkonsum, die Lektüre von Zeitungen oder auch Arztbesuche beträchtlichen Verzerrungen unterliegen. Um diese entmutigende Liste von Problemen fortzusetzen, ist auch davon auszugehen, daß die Zuordnung von Interviewern zu Befragten einen erheblichen Einfluß hat. Sowohl die Antworten zu faktischen Verhaltensfragen wie zu Einstellungen und Meinungen hängen in signifikanter Weise davon ab, welcher spezifische Interviewer die Fragen gestellt hat. Die systematische Analyse dieser Fehler zeigt schnell, daß sie nicht zufällig entstehen, sondern bestimmten Mustern folgen. Diese Muster wiederum stehen in spezifischem Zusammenhang mit der Art der Frage, die gestellt wird.

Im folgenden sollen einige dieser Verzerrungsmuster anhand einer Studie dokumentiert werden, die vor einigen Jahren in Zusammenarbeit mit dem US Gesundheitsdienst und dem Statistischen Amt der USA durchgeführt wurde (CANNELL, FISHER & BAKKER, 1965). Die Studie basierte auf etwa 2000 Krankenhaus-Akten. Diese Akten gaben jeweils Auskunft über das Datum der Aufnahme, die Länge der Aufenthaltsdauer und die Diagnose. Damit war eine objektive Ver-

ZUMA

gleichsbasis für eine anschließende Befragung der entlassenen Patienten gegeben. Hierfür wurde der Standard-Fragebogen aus der Nationalen Gesundheitsumfrage verwendet. Die Befragungen wurden von den gleichen Interviewern durchgeführt, die sonst bei den Erhebungen des Statistischen Amtes tätig waren. Die folgenden Tabellen geben die Antworten auf die folgende einfache Frage wieder:

"Während des letzten Jahres - also vom (Datum) bis (Datum) - waren da Sie oder irgendein Familienmitglied für mindestens eine Nacht in einem Krankenhaus?"

Insgesamt 12% der Krankenhausaufenthalte, die in den Akten dokumentiert waren, wurden nicht angegeben. Dieser Prozentsatz reduzierte sich auf 9%, wenn der frühere Krankenhauspatient selbst die Auskunft gegeben hatte, und erhöhte sich auf 17%, wenn die Auskunft über ein anderes Familienmitglied eingeholt worden war.

Tabelle 1: Anteil der nicht angegebenen Krankenhausaufenthalte nach Zeitabstand zwischen Krankenhausaufenthalt und Interview (enthält auch Angaben durch andere Familienmitglieder)

Zahl der Wochen zwischen Krankenhausausschlassung und Interview	Zahl der Krankenhausaufenthalte (nach Krankenhausunterlagen)	% nicht im Interview angegebener Aufenthalte
1-10	114	3
11-20	426	6
21-30	459	9
31-40	339	11
41-50	364	16
51-53	131	42

Anhand dieser und weiterer Tabellen sollen insbesondere drei Faktoren, die zu Verzerrungen bei den Antworten führen, im einzelnen illustriert werden.

1. Zeitabstand. Tabelle 1 zeigt, daß eine korrekte Angabe von Krankenhausaufenthalten über die Zeit abnimmt: Je länger der Aufenthalt zurückliegt, desto wahrscheinlicher wird es, daß er nicht mehr angegeben wird. Dieses Ergebnis allein ist nicht sonderlich überraschend. Überraschender ist eher die Schnelligkeit, mit der die Wahrscheinlichkeit abnimmt, daß das Ereignis korrekt berichtet wird. Dies illustrieren auch Ergebnisse einer ande-

ren Studie über ambulante Behandlungen. Wurden nach nur einer Woche bereits 15% der Arztbesuche nicht mehr angegeben, so waren es bei einem Abstand von zwei Wochen schon 30%.

2. Bedeutsamkeit oder Wichtigkeit des Ereignisses. Der Einfluß dieses Faktors wird durch Tabelle 2 gut illustriert. Wir benutzen dazu die Länge des Krankenhausaufenthaltes als Indikator für die Bedeutsamkeit dieses Ereignisses: Je länger der Aufenthalt gedauert hat, desto wahrscheinlicher ist es, daß das Ereignis für den Befragten relevant war. Generell gilt die Aussage, daß Ereignisse, die für das Individuum wichtig gewesen sind, vollständiger und im Detail genauer berichtet werden als solche, die als weniger wichtig empfunden werden.

Tabelle 2: Nicht angegebene Krankenhausaufenthalte nach Dauer des Krankenhausaufenthaltes (enthält auch Angaben durch andere Familienmitglieder)

Dauer des Krankenhausaufenthaltes (in Tagen)	Zahl der Fälle nach den Krankenhausunterlagen	% nicht im Interview angegebener Aufenthalte
1	150	26
2- 4	646	14
5- 7	456	10
8-14	352	10
15-21	111	6
22-30	58	2
31 and over	46	8

In der zugeschriebenen Bedeutsamkeit drückt sich wahrscheinlich die psychische Gegenwärtigkeit vergangener Ereignisse mit ihrem möglichen Einfluß auf das jetzige Leben aus. Forschungsergebnisse in diesem Bereich beruhen darauf, daß man beobachtbare Charakteristika von Ereignissen als Indikatoren heranzieht, von denen man auf die Wichtigkeit des Ereignisses für das Individuum schließt. Neben der hier benutzten Länge des Krankenhausaufenthaltes wäre es etwa auch denkbar, danach zu unterscheiden, ob während des Krankenhausaufenthaltes eine Operation erfolgt ist.

3. Soziale Wünschbarkeit. Der dritte Faktor - hier wäre es besser, von einem ganzen Bündel von Faktoren zu sprechen - liegt in der psychischen Ver-

ZUMA

arbeitung des fraglichen Ereignisses beim Befragten und in seiner Wahrnehmung, wie andere dieses Ereignis beurteilen.

Tabelle 3: Nicht angegebene Krankenhausaufenthalte nach Bedrohlichkeit der Diagnose (enthält Angaben durch andere Familienmitglieder)

Einschätzung der Diagnose	Zahl der Fälle nach Krankenhausunterlagen	% der nicht angegebenen Fälle
sehr bedrohlich	235	21
unter Umständen bedrohlich	421	14
nicht bedrohlich	1164	10

Fragen können als 'heikel' empfunden werden oder sogar eine Bedrohung darstellen. Den Ergebnissen, die in der Tabelle 3 dargestellt sind, liegt ein Indikator zugrunde, mit dem das Ausmaß einer solchen Bedrohung gemessen wurde. Ereignisse werden tendenziell eher so dargestellt, daß sie als sozial erwünscht erscheinen. Wenn nun ein Ereignis als peinlich, als prinzipiell die Intimsphäre berührend oder als Abweichung vom eigenen Selbstbild empfunden wird, ist es sehr wahrscheinlich, daß es entweder gar nicht berichtet wird, oder daß gewisse Modifikationen in der Darstellung erfolgen. Um das Ausmaß der Verzerrungen durch diesen Faktor der sozialen Wünschbarkeit bei Angaben über Krankenhausaufenthalte zu bestimmen, wurde eine 'Bedrohungsskala' für die einzelnen Diagnosen entwickelt, bei der zwischen drei unterschiedlichen Stufen von Bedrohung oder Peinlichkeit unterschieden wurde. Alle Krankheiten, die in der Einschätzung der Experten als sehr peinlich oder bedrohend einzuordnen sind, wurden zur Stufe 1 zusammengefaßt. In Stufe 3 wurden die Krankheitsbilder zusammengefaßt, die weder als sonderlich bedrohend noch als peinlich betrachtet wurden. In der mittleren Kategorie befinden sich die übrigen Krankheitsdiagnosen, die zum Teil eine Bedrohung darstellen, oder zumindest für einen Teil der Personen als bedrohlich erscheinen, ohne daß dies generell plausibel zu unterstellen ist. Der Vergleich der Anteile nicht angegebener Krankenhausaufenthalte in Abhängigkeit von dieser dreistufigen Skala zeigt, daß hier ein sehr wichtiger Einflußfaktor liegt. Wie Tabelle 3 zeigt, werden Krankenhausaufenthalte unter als bedrohlich oder peinlich erscheinenden Randbedingungen doppelt so oft nicht berichtet wie Krankenhausaufenthalte bei weniger heiklen Diagnosen.

Tabelle 4: Vergleich der Angaben zur Diagnose nach Krankenhausunterlagen und im Interview (enthält auch Angaben durch andere Familienmitglieder)

Krankheitsgruppe*	Zahl der Fälle nach den Interviews	Prozentuale Abweichung der Interviewangaben auf der Basis der Krankenhausdaten
Gutartige und nicht spezifizierte Gewächse	87	+51
Infektiöse und parasitäre Krankheiten	23	+45
Magen- und Zwölffingerdarmgeschwüre	36	+12
Krankheiten der Gallenblase	46	+10
...		
Andere Krankheiten des Verdauungssystems	72	-37
Brusterkrankungen und Störungen im Genitalbereich (Frauen)	52	-44
Erkrankungen des Nervensystems und der Sinnesorgane	47	-45
Psychische Störungen	8	-67

*Die Gruppenbildung folgt der International Classification of Diseases, 1955 Revision

Die Tabelle 4 illustriert diesen Effekt noch klarer. Der Befragte mag die Tatsache des Krankenhausaufenthaltes noch angeben, aber die Angabe über den Grund wird häufig modifiziert in der Weise, daß eine weniger bedrohliche oder als weniger peinlich empfundene Krankheit angegeben wird.

Die drei hier diskutierten Faktoren, die einen Einfluß auf die Genauigkeit der Verhaltensangaben ausüben - Zeitabstand, Wichtigkeit und psychische Bedeutung - haben sich über eine ganze Reihe von Studien (auch zu anderen Gegenstandsbereichen) hinweg immer wieder als bedeutsam erwiesen. Fehler und Verzerrungen in Umfragedaten sind gründlich und umfassend dokumentiert. Unglückseligerweise ist sehr viel weniger bekannt, wie diese Verzerrungen im einzelnen zustandekommen, und noch weniger ist bekannt darüber, wie sich diese Verzerrungen ganz ausschalten oder zumindest reduzieren lassen. Die aktuelle Herausforderung an Methodologen besteht darin, Hypothesen, Modelle und Theorien zu entwickeln und zu testen, um die Genauigkeit von Umfragedaten zu verbessern.

Die oben dargestellten Daten belegen zweifelsfrei, daß auf seiten des Befragten nicht die Information produziert wurde, die gewünscht war. Es scheint so, als ob die Aufgabe, über vergangenes Verhalten zu berichten, schwieriger zu bewältigen ist, als man zunächst geneigt war anzunehmen. Wir wollen diese Schwierigkeiten im einzelnen etwas näher beschreiben, und zwar

ZUMA

mit Hilfe eines Modells, das schematisch den Ablauf von der Fragestellung bis zur Hervorbringung einer Antwort darstellt.

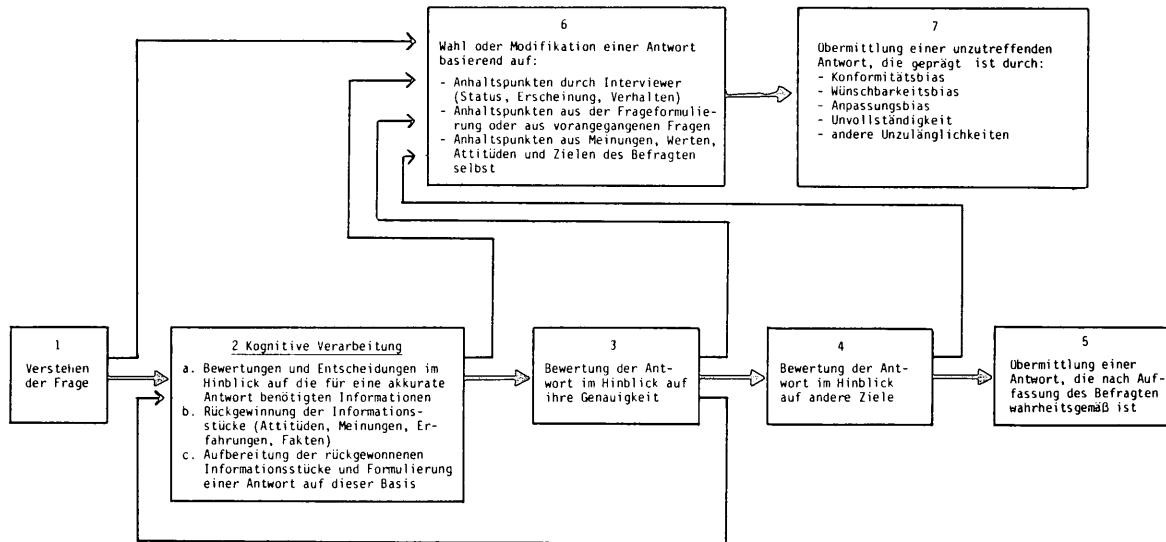
Das in Abbildung 1 dargestellte schematische Modell geht zurück auf MARQUIS und ist in der Zwischenzeit mehrfach verändert worden. Die jetzige Fassung findet sich bei CANNELL, MILLER & OKSEBERG (1981). Bei unserer Diskussion wollen wir zunächst den Weg betrachten, den ein pflichtbewußter Befragter durchläuft, nämlich die Stationen 1 bis 5 in unserer Abbildung. Dabei bleiben mögliche Abweichungen in die Stationen 6 und 7, die unerwünscht sind, zunächst außer Betracht. Der Beantwortungsprozeß beginnt mit dem Verstehen der Frage (Kästchen 1). Dabei sind alle die Punkte wichtig, die üblicherweise in den Methodenlehrbüchern diskutiert werden: Einfachheit/Komplexität der Satzstruktur, Benutzung von Fremdwörtern, Unzweideutigkeit der Begriffe. Darüber hinaus müssen wir jedoch auch das allgemeinere Problem der Frageninterpretation betrachten, also die Art und Weise, in der sich der Befragte einen Bezugsrahmen oder einen Orientierungspunkt verschafft, auf den hin er über die Frage nachdenkt. Dabei mag es durchaus mehrere solcher Bezugsrahmen geben, die plausibel erscheinen. Der Befragte muß dann diese unterschiedlichen Möglichkeiten durchdenken und eine Art inneren, zusätzlichen Fragebogen beantworten, um zu entscheiden, auf welchen Bezugsrahmen er sich beziehen soll. Ein Beispiel soll dies verdeutlichen; die Frage mag lauten:

"Wie oft haben Sie während des letzten Monats mit Ihrem Arzt über Ihre Gesundheit gesprochen?"

Der Befragte muß nun für sich entscheiden, ob er telefonische Kontakte mit hinzurechnen soll, ob Besuche etwa bei einem Chiropraktiker mitgezählt werden sollen, ob Impfungen Teil von 'Gesundheit' sind, und schließlich, ob der 'letzte' Monat sich auf die letzten vier Wochen oder einen Kalendermonat bezieht. Es ist dabei nicht entscheidend, ob der Befragte sich diese verschiedenen Möglichkeiten explizit bewußt macht. In jedem Fall ist seine Erinnerungsleistung abhängig davon, in welcher Art und Weise er die Frage versteht.

Der zweite Schritt (Kästchen 2) ist die Stufe der Informationsverarbeitung. Hier muß der Befragte entscheiden, welche Information im einzelnen benötigt wird, um die Frage genau zu beantworten, und welche Anhaltspunkte hierfür

Abbildung 1: Modell des Frage-Antwort-Prozesses



ZUMA

nützlich sein können. In diesem Schritt wird das Gedächtnis nach relevanter Information durchforscht und das gefundene Material strukturiert, um eine kohärente Antwort zu geben.

Im dritten Schritt (Kästchen 3) muß der Befragte entscheiden, ob die so vorformulierte Antwort den Intentionen der Frage gerecht wird. Wenn die zunächst gefundene Antwort als inadäquat empfunden wird, erfolgt ein nochmaliger Suchprozeß, d.h. die zweite Station wird noch einmal durchlaufen.

Im vierten Schritt wird eine zweite Art von Bewertung vorgenommen. Der Befragte überlegt hier, ob die soziale Bedeutung der Antwort verträglich ist mit persönlichen Zielen, die er ganz unabhängig von dieser Umfrage verfolgt. Auch im Prinzip kooperationsbereite Befragte mögen in Konflikt geraten zwischen dem Ziel, eine möglichst genaue und umfassende Antwort zu geben, und dem Ziel, sich in gewünschter Weise darzustellen. Wenn die vorläufig bereitgestellte Antwort als nicht schädlich bewertet wird, wird sie auch tatsächlich gegeben, und der Prozeß endet mit einer - zumindest aus der Sicht des Befragten - akkuraten Antwort (Kästchen 5).

Diese kurze Beschreibung des Frage-Antwort-Prozesses - wenn er in 'idealer Weise' verläuft - macht die Anforderungen deutlich, die an den Befragten gestellt werden, und illustriert die mögliche Komplexität beim Bearbeiten scheinbar einfacher Fragen. Die einzelnen Stufen dieses Prozesses zu durchlaufen, kann sehr schwierig sein, insbesondere dann, wenn die Frage einen größeren Aufwand zu ihrer genauen Beantwortung verlangt, oder wenn der Befragte die wahrheitsgemäße Antwort als peinlich oder in anderer Weise als persönlich schädlich empfindet. Ganz ohne Zweifel gibt es Befragte, die sich nicht ohne Einschränkung auf diesen Prozeß einlassen. Sie werden genaue Antworten nur solange zu geben versuchen, wie es nicht besonders aufwendig für sie ist, oder solange sie dadurch nicht in Verlegenheit gebracht werden.

Unsere schematische Darstellung in Abbildung 1 macht weiterhin deutlich, welche Abweichungen von diesem 'idealen' Verlauf erfolgen können. Während der 'ideale' Befragte nur die Stufen 1 bis 4 durchläuft und dann eine - zumindest aus seiner Sicht - angemessene Antwort gibt (Stufe 5), mag es bei jedem der ersten vier Schritte Abweichungen geben und ein Übergang zu ande-

ZUMA

rem Antwortverhalten erfolgen, was dann eine zumindest teilweise unzutreffende Antwort zur Folge hat (Stufe 7).

Diejenigen Befragten, die die Frage nicht verstanden haben oder die nicht ausreichend geübt oder motiviert sind, um durch den Prozeß des Erinnerns und Antwort-Formulierens hindurchzugehen, unterliegen in besonderer Weise den durch die Situation gegebenen Umständen und reagieren mit ihrer Antwort stärker auf die Situation als auf die eigentliche Frage. Aber auch Befragte, die zunächst die Stufe 3 adäquat durchlaufen, d.h. daß sie zunächst für sich eine Antwort formulieren, die sie selbst als wahrheitsgemäß empfinden, werden stets diese mögliche Antwort unter anderen Aspekten bewerten, also insbesondere abwägen, ob diese mögliche Antwort in Konflikt steht mit anderen Zielen. Wenn sich daraus eine 'Bedrohung' ergibt, mag der Befragte vom idealen Weg abweichen und die mögliche Antwort modifizieren oder sogar durch eine andere vollständig ersetzen.

Es scheint, daß kognitive und motivationale Schwierigkeiten bei der Beantwortung von Fragen weitaus häufiger sind und sehr viel ernstere Konsequenzen haben, als man üblicherweise annimmt. Fragen mögen die kognitiven Fähigkeiten des Befragten überfordern, oder sie mögen unrealistische Annahmen darüber machen, wie gut der Befragte in der Lage ist, vergangene Ereignisse zu erinnern und sich alle Informationen in vollständiger Weise verfügbar zu machen, um eine umfassende Antwort zu geben. Schließlich mögen die sozialen Implikationen, eine Antwort zu geben, die wahrheitsgemäß die Überzeugungen oder das vergangene Verhalten des Befragten reflektiert, dazu führen, daß diese Information unterdrückt wird oder in eine stärker den wahrgenommenen sozialen Normen entsprechenden Form verändert wird.

Die einzelnen Mechanismen der Veränderung der Antwort sind in Kästchen 7 von Abbildung 1 näher dargestellt. Ihre Effekte führen zu gewichtigen Verzerrungen in den Umfragedaten, d.h. die angegebene Information weicht von der 'Wahrheit' in bestimmter Weise ab.

Die wahrscheinlich häufigste Verzerrung besteht darin, daß Informationen nicht gegeben werden. Dies kann einerseits durch Probleme beim Prozeß des Erinnerns (Stufe 2) bedingt sein, also durch tatsächliche Gedächtnislücken; mag aber auch bedingt sein durch Nachlässigkeit und mangelnde Bereitschaft,

sehr gründlich nachzudenken, um die geforderte Information bereitzustellen. Obwohl also die Aufgabe, bestimmte Informationen umfassend und wahrheitsgemäß zu berichten, durchaus sowohl Sorgfalt wie einige Anstrengungen auf seiten des Befragten voraussetzt, sind diese Ansprüche wiederum auch nicht so hoch, als daß sie unüberwindbar für den Befragten wären. Geht man Verzerrungen, wie wir sie oben beispielhaft diskutiert haben, weiter nach, so zeigt sich in den meisten Fällen, daß die Befragten nicht ausreichend motiviert wurden, ihren Teil der Aufgabe zu erfüllen. Um etwas mehr über die Gründe für die mangelnde Kooperationsbereitschaft herauszufinden, haben wir bei rund 400 Personen eine Nachbefragung zum Nationalen Gesundheitssurvey der USA durchgeführt, und zwar genau einen Tage später. Obwohl nach der Intervieweranweisung die Befragten bei der ursprünglichen Erhebung darüber zu informieren waren, daß die Umfrage für den US Gesundheitsdienst vom Statistischen Amt durchgeführt würde, wußten fast die Hälfte der Befragten (45%) nicht, welche Behörde die Umfrage durchführte bzw. für welche Behörde diese Umfrage bestimmt war. Nur 11% konnten korrekt das Statistische Amt als durchführende Behörde identifizieren. Unter denjenigen, die überhaupt eine Angabe machten, waren die Vorstellungen sehr vage. Das gleiche gilt für die Frage nach Sinn und Ziel dieser Untersuchung. Mehr als die Hälfte der Befragten sagten, daß sie hierüber nichts wußten. Und auch die übrigen äußerten höchstens sehr allgemeine Vermutungen, wie etwa, daß die Umfrage 'statistischen Zwecken' dienen würde. Weiterhin wurde in der Studie untersucht, wie die Befragten die Erwartungen des Interviewers wahrgenommen haben. Ein größerer Teil des Interviews in der Gesundheitsumfrage richtete sich auf Fragen, die sehr spezifisch nach Krankheiten, nach versäumten Arbeitstagen, nach eingenommenen Arzneien und chronischen Behinderungen fragten. Andere Fragen zielten auf Dinge wie die Zahl der Arztbesuche und der Krankenhausaufenthalte. Der Fragebogen war entwickelt worden, um sehr detaillierte und vollständige Angaben zu bekommen. Jedoch gaben nur etwa die Hälfte der Befragten an, daß der Interviewer umfassende und detaillierte Informationen wollte. Nach Eindruck der anderen ging es eher um ungefähre Angaben. Drei Viertel der Befragten waren der Auffassung, daß der Interviewer einen vollständigen Bericht über alle Ereignisse wollte, während beim Rest die Meinung herrschte, daß nur die wichtigeren Dinge berichtet werden sollten.

Soweit diese konkreten Beispiele. Sie illustrieren, daß sich die gegenwärtigen Interviewtechniken nicht primär auf gesicherte Forschungsergebnisse

oder theoretische Einsichten stützen, sondern viel eher auf pragmatische Erfahrungswerte. Über die Jahre hinweg hat sich eine ganze Reihe von Regeln für die Durchführung von Interviews entwickelt, die ziemlich ungeteilten Konsens findet. Diese Prinzipien gehen alle von der Annahme aus, daß die Kommunikation im Interview nur ein besonderer Fall von interpersonaler Interaktion ist. Diese Prinzipien basieren auf punktueller Erfahrung und allgemeiner Einschätzung und sind nur in seltenen Fällen von empirisch geprüften Annahmen abgeleitet. Die vielen Belege für Antwortverzerrungen illustrieren jedoch nachdrücklich, daß diese Techniken sehr häufig nicht dazu in der Lage sind, gültige Daten zu produzieren und daß alle diese wohl etablierten Regeln einer systematischen Überprüfung bedürfen.

Eines dieser wohl etablierten Prinzipien, das in Frage gestellt werden sollte, ist zum Beispiel die Forderung, daß es wichtig ist, 'Rapport' herzustellen. Dieses Prinzip besagt, daß zunächst eine persönliche Beziehung mit dem Befragten hergestellt werden muß, daß so am wirkungsvollsten eine Motivation, das Interview in der gewünschten Form mitzumachen, herbeizuführen ist. Diesem althergebrachten Grundsatz mag man die These entgegenstellen, daß es weitaus besser ist, eine 'aufgabenorientierte' Situation im Interview herzustellen. Wie schon Herbert HYMAN und seine Mitarbeiter vor rund 30 Jahren sinngemäß formuliert haben, ist ein Interview in einer Umfrage kein Kaffeekränzchen. Rapport mag sogar unerwünschte Konsequenzen haben. Wenn es der Befragte darauf anlegt, die persönliche Wertschätzung aufrechtzuerhalten und auszubauen, mag ihn das dazu führen, Angaben über Verhalten, was als weniger günstig wahrgenommen werden könnte, zu unterdrücken. Rapport kann so selbst zum verzerrenden Faktor werden. Es gibt gute Gründe zu vermuten, daß ein Interview, das streng an der Erfüllung der Aufgaben orientiert ist, eher die Voraussetzung dafür schafft, daß der Befragte konstruktiv mitarbeitet und dadurch Verzerrungen vermieden werden. Allerdings bedarf es weiterer systematischer Forschungen, um zu bestimmen, wie idealerweise die soziale Interaktion im Interview gestaltet werden sollte, d.h. auf welche Weise die umfassende Mitarbeit des Befragten sicherzustellen ist und möglichst viele verzerrende Einflüsse ausgeschlossen werden können.

Ein anderes Prinzip aus dem allgemein akzeptierten Erfahrungsschatz zur Datenerhebung ist, daß der Interviewer einen möglichst großen Spielraum haben

sollte, um die Techniken auf die Bedürfnisse des jeweiligen Befragten abzustimmen. Folglich werden Interviewer während ihrer Schulung nur allgemein mit Formen akzeptabler Verhaltensweisen vertraut gemacht, und es bleibt ihnen überlassen, diese nach eigener Einschätzung der Notwendigkeit zu variieren. So werden während der Schulung Formen der Nachfrage allgemein diskutiert, aber es bleibt offen, wie oft und in genau in welcher Form nachgefragt werden soll. Es wird gelehrt, welche Formen der Rückkopplung an den Befragten generell akzeptierbar sind, aber es bleibt für den Einzelfall jeweils offen, zu welchem Zeitpunkt und in welchem Umfang dies geschehen sollte.

Es gibt eigentlich nur eine einzige allgemein akzeptierte Regel für die Standardisierung des Interviewprozesses, nämlich die, daß die Fragen genau so vorgelesen werden sollen, wie sie im Fragebogen stehen. Für alle anderen Verhaltensweisen werden lediglich generelle Prinzipien vermittelt, wobei dem Interviewer im Einzelfall ein großer Spielraum bleibt. In den Naturwissenschaften baut jede quantitative Analyse darauf auf, daß die Meßtechniken voll standardisiert sind. Die Techniken müssen im Detail festgelegt sein, nicht nur als generelle Prinzipien formuliert werden, die mit individueller Flexibilität im Einzelfall angewendet werden. Wir wissen, daß Interviewer sich durchaus darin unterscheiden, wie und wann sie diese Techniken anwenden, und daß diese Variationen sowohl die Vergleichbarkeit zwischen Interviewern wie die Gültigkeit der Antworten beeinflussen.

Fast niemand bestreitet im Prinzip den wissenschaftlichen Wert, Meßtechniken zu standardisieren, aber es gibt erhebliche Meinungsunterschiede darüber, wie weit man gehen kann, das Interviewerverhalten im einzelnen festzulegen. Wir glauben jedoch, daß es möglich ist, mehr als bisher die Verhaltensweise des Interviewers während der Befragung zu standardisieren und auf diese Weise die Gültigkeit und Verlässlichkeit der erhobenen Daten zu verbessern.

Ein anderes Argument, das einer weitergehenden Standardisierung von Interviewtechniken entgegensteht, ist, daß spezifische kulturelle Einflüsse so groß sind, daß die Erfahrungen, die in einem Land gesammelt werden, in der Regel nicht auf ein zweites übertragen werden können. Mit diesem Argument werden relativ häufig neue Einsichten ohne weitere Prüfung zurückgewiesen.

Wenn dieses Argument in vollem Umfang gilt, dann ist es allerdings um die Möglichkeit international vergleichender Forschung generell schlecht bestellt. Zumindest wäre es unmöglich, eine allgemein gültige Methodologie der Umfrageforschung zu entwickeln. Wir glauben daher, daß es notwendig ist, stärker als bisher Forschungsergebnisse in anderen Kontexten zu replizieren, um auf diese Weise dazu beizutragen, eine Umfragemethodologie zu entwickeln, die jenseits von nationalen oder sprachlichen Grenzen gilt.

Unsere eigenen Forschungen sind ein Versuch, Methoden zu entwickeln, mit der die Gültigkeit der Antworten verbessert werden kann und mit denen striktere Anforderungen an den Meßprozeß erfüllt werden können. Dies kann am ehesten durch eine Verbesserung der Techniken zur Fragebogenkonstruktion erreicht werden. Dabei sind für eine erfolgreiche Durchführung des Interviews die folgenden Probleme zu lösen:

1. Der Befragte muß im einzelnen darüber unterrichtet werden, was von ihm ganz allgemein erwartet wird, um seine Aufgabe angemessen wahrzunehmen.
2. Der Befragte muß weitgehende Hilfestellung bekommen, um einzelne Fragen möglichst effizient beantworten zu können.
3. Der Befragte muß motiviert und ermutigt werden, so genau wie möglich nachzudenken und die Informationen konsistent zu organisieren sowie selbst möglicherweise peinliche Fakten preiszugeben.
4. Generell müssen Prozeduren entwickelt werden, um die Techniken zu standardisieren, so daß die Vergleichbarkeit zwischen einzelnen Interviewern erhöht wird. Dieses kann u.a. dadurch erreicht werden, daß in den Fragebogen selbst genauere und ausführlichere Anleitungen und Hinweise für den Interviewer aufgenommen werden.

Ein genauerer Überblick über Experimente dieser Art findet sich in CANNELL, MILLER & OKSENBERG (1981) sowie CANNELL, OKSENBERG & CONVERSE (1979). Die dort näher dargestellten Experimente basieren auf unserer Suche nach theoretischen Erklärungen in Psychologie und Soziologie, die die oben benannten Phänomene aufklären können. Besonders relevant sind Theorien und Hypothesen über Interaktion und Motivation. Wir haben Experimente mit 'aufgabenorientierten' Interviews angestellt, in denen wir verschiedene lerntheoretische Elemente benutzt haben. In einzelnen Studien haben wir versucht, die Anwei-

sungen für die Interviewer im Fragebogen so detailliert zu halten, daß damit der Ablauf des Interviews fast vollständig festgeschrieben war. Die Resultate belegen nachdrücklich, daß Angaben über tatsächliches Verhalten durch diese Maßnahmen positiv beeinflußt werden können. Sie belegen auch, daß eine stärkere Standardisierung des Interviewerverhaltens möglich ist und dies durch Spezifikationen im Fragebogen erfolgen kann.

Alle diese Studien jedoch sind noch nicht mehr als erste Experimente. Obwohl eine Reihe von signifikanten Ergebnissen erzielt werden konnten, liegt noch ein weiter Weg vor uns, bevor wir in der Lage sind, unsere Umfragedaten weitgehend fehlerfrei zu erheben. Noch viele andere Theorien und Hypothesen müssen überprüft, operationalisiert und auf die Situation im Umfrageinterview hin getestet werden. Besonders wichtig ist es, daß solche Experimente von anderen Forschern repliziert werden; insbesondere auch in anderen nationalen Kontexten, um den Beitrag derartiger Ergebnisse für eine generelle Methodologie der Umfrageforschung zu bestimmen.

Dieser Bericht ist die schriftliche Fassung eines Vortrags, den Charles F. Cannell (Institute for Social Research der University of Michigan, Ann Arbor) während seines Gastaufenthaltes bei ZUMA in den Monaten Mai und Juni 1984 gehalten hat. Die deutsche Bearbeitung besorgte Manfred Küchler.

Literatur

- CANNELL, C.F., FISHER, G. & BAKKER, Th. Vital and Health Statistics. Washington D.C.: U.S. Public Health Service, Juli 1965, Serie 2, Nr. 6.
- CANNELL, C.F., OKSENBERG, L. & CONVERSE, J.M. (eds.). Experiments in Interviewing Techniques: Field Experiments in Health Reporting, 1971-77. Research Report Series, Ann Arbor, Mi.: Survey Research Center, Institute for Social Research, 1979.
- CANNELL, C.F., MILLER, P.V. & OKSENBERG, L. Research on Interviewing Techniques. In: S. LEINHARDT (ed.): Sociological Methodology. San Francisco: Jossey Bass, 1981.